Taller 8

Métodos Computacionales para Políticas Públicas - URosario

Entrega: viernes 12-abr-2019 11:59 PM

Francisco Monsalve

francisco.monsalve@urosario.edu.co

Instrucciones:

- Guarde una copia de este *Jupyter Notebook* en su computador, idealmente en una carpeta destinada al material del curso.
- Modifique el nombre del archivo del *notebook*, agregando al final un guión inferior y su nombre y apellido, separados estos últimos por otro guión inferior. Por ejemplo, mi *notebook* se llamaría: mcpp_taller8_santiago_matallana
- Marque el *notebook* con su nombre y e-mail en el bloque verde arriba. Reemplace el texto "[Su nombre acá]" con su nombre y apellido. Similar para su e-mail.
- Desarrolle la totalidad del taller sobre este *notebook*, insertando las celdas que sea necesario debajo de cada pregunta. Haga buen uso de las celdas para código y de las celdas tipo *markdown* según el caso.
- Recuerde salvar periódicamente sus avances.
- Cuando termine el taller:
  1. Descárguelo en PDF. Si tiene algún problema con la conversión, descárguelo en HTML.
  2. Suba todos los archivos a su repositorio en GitHub, en una carpeta destinada exclusivamente para este taller, antes de la fecha y hora límites.

---

1. [1 punto]

Usando expresiones regulares extraiga en una lista todos los números presentes en el siguiente objeto de Python:

ob1 = "JEFF BEZOS, the founder of Amazon, has reached a divorce settlement with his wife, MacKenzie. Mr Bezos will keep all the shares in the Washington Post and Blue Origin, a space-exploration firm, as well as 75% of the couple's Amazon stock. Mrs Bezos will retain a 4% stake in the tech giant, worth nearly $36bn, which is likely to make her the third-richest woman alive when the divorce is finalised."

In [1]:

```
ob1 = "JEFF BEZOS, the founder of Amazon, has reached a divorce settlement with his wife, MacKenzie. Mr Bezos will keep all the shares in the Washington Post and Blue Origin, a space-exploration firm, as well as 75% of the couple's Amazon stock. Mrs Bezos will retain a 4% stake in the tech giant, worth nearly $36bn, which is likely to make her the third-richest woman alive when the divorce is finalised."
```

ch is likely to make her the third-richest woman alive when the divorce is finalised.

In [2]:

```
import re
```

In [3]:

```
nums_1 = re.findall("[0-9]+", ob1)
nums_1
```

Out[3]:

['75', '4', '36']

2. [1 punto]

Usando expresiones regulares ahora extraiga de *ob1* sólo los números que correspondan a porcentajes.

In [4]:

```
nums_2 = re.findall("([0-9]+)%", ob1)
nums_2
#debido a que sólo se extraen los números, no tengo en cuenta el símbolo %
```

Out[4]:

['75', '4']

3. [2 puntos]

Usando expresiones regulares, escriba una función de Python que reciba una fecha en formato **Marzo 7, 2019** y retorne la fecha en formato **2019-07-03**

In [5]:

```
months={"Enero":"01","Febrero":"02","Marzo":"03","Abril":"04","Mayo":"05","Junio":"06","Julio":"07","Agosto":"
08","Septiembre":"09","Octubre":"10","Noviembre":"11","Diciembre":"12"}
def date_1(dates):
    year = re.findall("([0-9][0-9][0-9][0-9])", dates)
    año = year[0]
    month = re.findall("^([\w]+).", dates)
    mes = months[""".join(month)]
    day = re.findall("([0-9]+),", dates)
    día = "0"+ day[0]
    fecha = año+"-"+día+"-"+mes
    return fecha
```

In [6]:

```
date_1("Marzo 7, 2019")
```

Out[6]:

'2019-07-03'

In [7]:

```
date_1("Abril 5, 2020")
```

Out[7]:

'2020-05-04'

In [8]:

```
date_1("Diciembre 12, 1996")
```

Out[8]:

'1996-012-12'

4. [3 puntos]

ob2 es un string que reune una lista de clases en una universidad. Use expresiones regulares para extraer los códigos de cada una de las clases. Ejemplo: El código de la clase **COMPSCI 143 (Spring 2012): Machine Learning** es 143.

ob2 = "COMPSCI 270 (Spring 2019): Introduction to Artificial Intelligence. COMPSCI 590.2 (Fall 2018): Computational Microeconomics: Game Theory, Social Choice, and Mechanism Design. COMPSCI 223 (Spring 2018): Computational Microeconomics. COMPSCI 570 (Fall 2017): Artificial Intelligence. COMPSCI 590.3 (Fall 2017): Ethics and AI. COMPSCI 590.2 (Spring 2017): Computation, Information, and Learning in Market Design. COMPSCI 590.4 (Spring 2016): Computational Microeconomics: Game Theory, Social Choice, and Mechanism Design. COMPSCI 290.4/590.4 (Spring 2015): Crowdsourcing Societal Tradeoffs. COMPSCI 570 (Fall 2014): Artificial Intelligence. COMPSCI 590.4 (Spring 2014): Computational Microeconomics: Game Theory, Social Choice, and Mechanism Design. COMPSCI 590.1 (Fall 2012): Linear and Integer Programming. COMPSCI 173 (Spring 2012): Computational Microeconomics. COMPSCI 296.1 (Fall 2011): Computational Microeconomics: Game Theory, Social Choice, and Mechanism Design. COMPSCI 296.1 (Fall 2010): Linear and Integer Programming. COMPSCI 173 (Spring 2010): Computational Microeconomics. COMPSCI 196.1/296.1 (Fall 2009): Computational Microeconomics: Game Theory, Social Choice, and Mechanism Design. COMPSCI 170 (Spring 2009): Introduction to Artificial Intelligence. COMPSCI 270 (Fall 2008): Artificial Intelligence. COMPSCI 196/296.2 (Spring 2008): Linear and Integer Programming. COMPSCI 196.2 (Fall 2007): Introduction to Computational Economics. COMPSCI 296.3 (Spring 2007): Topics in Computational Economics. COMPSCI 296.2 (Fall 2006): Computational Game Theory and Mechanism Design."

In [9]:

```
ob2 = "COMPSCI 270 (Spring 2019): Introduction to Artificial Intelligence. COMPSCI 590.2 (Fall 2018): Com
putational Microeconomics: Game Theory, Social Choice, and Mechanism Design. COMPSCI 223 (Spring 20
18): Computational Microeconomics. COMPSCI 570 (Fall 2017): Artificial Intelligence. COMPSCI 590.3 (Fall
2017): Ethics and AI. COMPSCI 590.2 (Spring 2017): Computation, Information, and Learning in Market Des
ign. COMPSCI 590.4 (Spring 2016): Computational Microeconomics: Game Theory, Social Choice, and Mec
hanism Design. COMPSCI 290.4/590.4 (Spring 2015): Crowdsourcing Societal Tradeoffs. COMPSCI 570 (Fa
ll 2014): Artificial Intelligence. COMPSCI 590.4 (Spring 2014): Computational Microeconomics: Game Theory
, Social Choice, and Mechanism Design. COMPSCI 590.1 (Fall 2012): Linear and Integer Programming. CO
MPSCI 173 (Spring 2012): Computational Microeconomics. COMPSCI 296.1 (Fall 2011): Computational Micr
oeconomics: Game Theory, Social Choice, and Mechanism Design. COMPSCI 296.1 (Fall 2010): Linear and
Integer Programming. COMPSCI 173 (Spring 2010): Computational Microeconomics. COMPSCI 196.1/296.1
(Fall 2009): Computational Microeconomics: Game Theory, Social Choice, and Mechanism Design. COMPS
CI 170 (Spring 2009): Introduction to Artificial Intelligence. COMPSCI 270 (Fall 2008): Artificial Intelligence. C
OMPSCI 196/296.2 (Spring 2008): Linear and Integer Programming. COMPSCI 196.2 (Fall 2007): Introducti
on to Computational Economics. COMPSCI 296.3 (Spring 2007): Topics in Computational Economics. COM
PSCI 296.2 (Fall 2006): Computational Game Theory and Mechanism Design."
```

In [10]:

```
ob2
```

Out[10]:

'COMPSCI 270 (Spring 2019): Introduction to Artificial Intelligence. COMPSCI 590.2 (Fall 2018): Computational Microeconomics: Game Theory, Social Choice, and Mechanism Design. COMPSCI 223 (Spring 2018): Computational Microeconomics. COMPSCI 570 (Fall 2017): Artificial Intelligence. COMPSCI 590.3 (Fall 2017): Ethics and AI. COMPSCI 590.2 (Spring 2017): Computation, Information, and Learning in Market Design. COMPSCI 590.4 (Spring 2016): Computational Microeconomics: Game Theory, Social Choice, and Mechanism Design. COMPSCI 290.4/590.4 (Spring 2015): Crowdsourcing Societal Tradeoffs. COMPSCI 570 (Fall 2014): Artificial Intelligence. COMPSCI 590.4 (Spring 2014): Computational Microeconomics: Game Theory, Social Choice, and Mechanism Design. COMPSCI 590.1 (Fall 2012): Linear and Integer Programming. COMPSCI 173 (Spring 2012): Computational Microeconomics. COMPSCI 296.1 (Fall 2011): Computational Microeconomics: Game Theory, Social Choice, and Mechanism Design. COMPSCI 296.1 (Fall 2010): Linear and Integer Programming. COMPSCI 173 (Spring 2010): Computational Microeconomics. COMPSCI 196.1/296.1 (Fall 2009): Computational Microeconomics: Game Theory, Social Choice, and Mechanism Design. COMPSCI 170 (Spring 2009): Introduction to Artificial Intelligence. COMPSCI 270 (Fall 2008): Artificial Intelligence. COMPSCI 196/296.2 (Spring 2008): Linear and Integer Programming. COMPSCI 196.2 (Fall 2007): Introduction to Computational Economics. COMPSCI 296.3 (Spring 2007): Topics in Computational Economics. COMPSCI 296.2 (Fall 2006): Computational Game Theory and Mechanism Design.'

In [11]:

```
code_class1 = re.findall("[COMPSCI] (\d+\.\d+)", ob2)
code_class2 = re.findall("[COMPSCI] ([0-9]+)", ob2)
code_class3 = re.findall("/(\d+\.\d+)", ob2)
print(code_class1,code_class2, code_class3)
```

['590.2', '590.3', '590.2', '590.4', '290.4', '590.4', '590.1', '296.1', '296.1', '196.1', '196.2', '296.3', '296.2'] ['270', '590', '223', '570', '590', '590', '590', '290', '570', '590', '590', '173', '296', '296', '173', '196', '170', '270', '196', '196', '296', '296'] ['590.4', '296.1', '296.2']

In [12]:

```
len(code_class1)
```

Out[12]:

13

In [13]:

```
len(code_class2)
```

Out[13]:

22

In [14]:

```
len(code_class3)
```

Out[14]:

3

In [15]:

```
code_class = code_class1 + code_class2 + code_class3
print(code_class)
```

['590.2', '590.3', '590.2', '590.4', '290.4', '590.4', '590.1', '296.1', '296.1', '196.1', '196.2', '296.3', '296.2', '270', '590', '223', '570', '590', '590', '590', '290', '570', '590', '590', '173', '296', '296', '173', '196', '170', '270', '196', '196', '296', '296', '590.4', '296.1', '296.2']

In [16]:

```
len(code_class)
```

Out[16]:

38

In [33]:

```
#Otra manera de sacar los códigos
code_class_other=re.findall('COMPSCI (\S+)', ob2)
len(code_class_other)
```

Out[33]:

22

In [34]:

```
print(code_class_other)
```

['270', '590.2', '223', '570', '590.3', '590.2', '590.4', '290.4/590.4', '570', '590.4', '590.1', '173', '296.1', '296.1', '173', '196.1/296.1', '170', '270', '196/296.2', '196.2', '296.3', '296.2']

5. [5 puntos]

ob3 es un string que reune una lista de publicaciones. Use expresiones regulares para extraer todos los *Journals* en los cuales el autor ha publicado. Ejemplo: El paper **Bail, CA. "The configuration of symbolic boundaries against immigrants in Europe." American Sociological Review 73.1 (January 1, 2008): 37-59. Full Text** fue publicado en el Journal *American Sociological Review*

ob3 = "Bail, CA, Argyle, LP, Brown, TW, Bumpus, JP, Chen, H, Hunzaker, MBF, Lee, J, Mann, M, Merhout, F, and Volfovsky, A. "Exposure to opposing views on social media can increase political polarization." Proceedings of the National Academy of Sciences of the United States of America 115.37 (September 2018): 9216-9221. Full Text Open Access Copy. Bail, CA, Merhout, F, and Ding, P. "Using Internet search data to examine the relationship between anti-Muslim and pro-ISIS sentiment in U.S. counties." Science Advances 4.6 (June 6, 2018): eaao5948-null. Full Text Open Access Copy. Bail, CA, Brown, TW, and Mann, M. "Channeling Hearts and Minds: Advocacy Organizations, Cognitive-Emotional Currents, and Public Conversation." American Sociological Review 82.6 (December 1, 2017): 1188-1213. Full Text. Bail, CA. "Taming Big Data: Using App Technology to Study Organizational Behavior on Social Media." Sociological Methods and Research 46.2 (March 1, 2017): 189-217. Full Text. McDonnell, TE, Bail, CA, and Tavory, I. "A Theory of Resonance." Sociological Theory 35.1 (March 1, 2017): 1-14. Full Text. Bail, CA. "Combining natural language processing and network analysis to examine how advocacy organizations stimulate conversation on social media." Proceedings of the National Academy of Sciences of the United States of America 113.42 (October 2016): 11823-11828. Full Text. Bail, CA. "Emotional Feedback and the Viral Spread of Social Media Messages About Autism Spectrum Disorders." American journal of public health 106.7 (July 2016): 1173-1180. Full Text. Bail, CA. "The public life of secrets: Deception, disclosure, and discursive framing in the policy process." Sociological Theory 33.2 (January 1, 2015): 97-124. Full Text. Bail, CA. "The cultural environment: Measuring culture with big data." Theory and Society 43.3 (January 1, 2014): 465-524. Full Text."

In [17]:

```
ob3 ='ob3 = "Bail, CA, Argyle, LP, Brown, TW, Bumpus, JP, Chen, H, Hunzaker, MBF, Lee, J, Mann, M, Merhout, F, and Volfovsky, A. "Exposure to opposing views on social media can increase political polarization." Proceedings of the National Academy of Sciences of the United States of America 115.37 (September 2018): 9216-9221. Full Text Open Access Copy. Bail, CA, Merhout, F, and Ding, P. "Using Internet search data to examine the relationship between anti-Muslim and pro-ISIS sentiment in U.S. counties." Science Advances 4.6 (June 6, 2018): eaao5948-null. Full Text Open Access Copy. Bail, CA, Brown, TW, and Mann, M. "Channel
```

ing Hearts and Minds: Advocacy Organizations, Cognitive-Emotional Currents, and Public Conversation." American Sociological Review 82.6 (December 1, 2017): 1188-1213. Full Text. Bail, CA. "Taming Big Data: Using App Technology to Study Organizational Behavior on Social Media." Sociological Methods and Research 46.2 (March 1, 2017): 189-217. Full Text. McDonnell, TE, Bail, CA, and Tavory, I. "A Theory of Resonance." Sociological Theory 35.1 (March 1, 2017): 1-14. Full Text. Bail, CA. "Combining natural language processing and network analysis to examine how advocacy organizations stimulate conversation on social media." Proceedings of the National Academy of Sciences of the United States of America 113.42 (October 2016): 11823-11828. Full Text. Bail, CA. "Emotional Feedback and the Viral Spread of Social Media Messages About Autism Spectrum Disorders." American journal of public health 106.7 (July 2016): 1173-1180. Full Text. Bail, CA. "The public life of secrets: Deception, disclosure, and discursive framing in the policy process." Sociological Theory 33.2 (January 1, 2015): 97-124. Full Text. Bail, CA. "The cultural environment: Measuring culture with big data." Theory and Society 43.3 (January 1, 2014): 465-524. Full Text."'

In [18]:

```
journal = re.findall('." ([^0-9]+)', ob3)
print(journal)
```

['Proceedings of the National Academy of Sciences of the United States of America ', 'Science Advances ', 'American Sociological Review ', 'Sociological Methods and Research ', 'Sociological Theory ', 'Proceedings of the National Academy of Sciences of the United States of America ', 'American journal of public health ', 'Sociological Theory ', 'Theory and Society ']

6. [10 puntos]

Vamos a hacer "scraping" a esta página: https://archive.ics.uci.edu/ml/datasets.php (https://archive.ics.uci.edu/ml/datasets.php), que contiene un listado de 468 bases de datos que hacen parte del repositorio de la Universidad de California, Irvine.

Su tarea consiste en crear un "Pandas dataframe" que contenga 468 filas (una por base de datos) y las siguientes columnas:

- Nombre de la base de datos
- Link a la base de datos
- Tipo de datos
- Tipo de tarea a resolver (default task)
- Tipo de las variables
- Número de observaciones
- Número de variables
- Año
- Descripción de la base (Pista: Utilice la opción list view: https://archive.ics.uci.edu/ml/datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list (https://archive.ics.uci.edu/ml/datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list))

Diviértase.

In [19]:

```
import requests
from bs4 import BeautifulSoup
```

In [163]:

```
import pandas as pd
```

```
import numpy as np
```

In [20]:

```
#Obtener contenido html de la página
html = requests.get("https://archive.ics.uci.edu/ml/datasets.php").text
```

In [21]:

```
html
```

Out[21]:

'<!DOCTYPE HTML PUBLIC \\"-//W3C//DTD HTML 4.01 Transitional//EN\\">\n<html>\n<head>\n<title>UCI Machine Learning Repository: Data Sets</title>\n\n<!-- Stylesheet link -->\n<link rel="stylesheet" type="text/css" href="assets/ml.css" />\n\n<script language="JavaScript" type="text/javascript">\n<!--\nfunction checkform (form)\n{\n // see http://www.thesitewizard.com/archive/validation.shtml\n // for an explanation of this script and how to use it on your\n // own website\n\n // ** START **\n if (form.q.value == "")\n {\n   alert("Please enter search terms.");\n   form.q.focus();\n   return false ;\n }\n\n if (getCheckedValue(form.sitesearch) == "ics.uci.edu" && form.q.value.indexOf("site:archive.ics.uci.edu/ml") == -1)\n {\n   form.q.value = form.q.value + " site:archive.ics.uci.edu/ml";\n }\n\n // ** END **\n return true ;\n}\n\n// return the value of the radio button that is checked\n// return an empty string if none are checked, or\n// there are no radio buttons\nfunction getCheckedValue(radioObj) {\n\tif(!radioObj)\n\t\treturn "";\n\tvar radioLength = radioObj.length;\n\tif(radioLength == undefined)\n\t\tif(radioObj.checked)\n\t\t\treturn radioObj.value;\n\t\telse\n\t\t\treturn "";\n\tfor(var i = 0; i < radioLength; i++) {\n\t\tif(radioObj[i].checked) {\n\t\t\treturn radioObj[i].value;\n\t\t}\n\t}\n\treturn "";\n}\n\n\n\n\n\n//-->\n</script>\n\n</head>\n\n<body>\n\n\n<!-- SITE HEADER (INCLUDES LOGO AND SEARCH BOX) -->\n\n<!-- SITE HEADER (INCLUDES LOGO AND SEARCH BOX) -->\n\n<table width=100% bgcolor="#003366">\n<tr>\n\t<td>\n\t\t<a href="index.html" alt="Home"><img src="assets/logo.gif" border=0></img></a>
          <a href="http://cml.ics.uci.edu"><font color="FFDD33">Center for Machine Learning and Intelligent Systems</font></a>\n\t</td>\n\t<td width=100% valign=top align="right">\n\t\t\n\t\t<a href="about.html">About</a> \n\t\t<a href="citation_policy.html">Citation Policy</a> \n\t\t<a href="donation_policy.html">Donate a Data Set</a> \n\t\t<a href="contact.html">Contact</a>\n\t\t\n\n\t\t
\n\t\t
\n\t\t<!-- Search Google -->\n\n\t\t<FORM method=GET action=http://www.google.com/custom onsubmit="return checkform(this);">\n\t\t<INPUT TYPE=text name=q size=30 maxlength=255 value="">\n\t\t<INPUT type=submit name=sa VALUE="Search">\n\t\t<INPUT type=hidden name=cof VALUE="AH:center;LH:130;L:http://archive.ics.uci.edu/assets/logo.gif;LW:384;AWFID:869c0b2eaa8d518e;">\n\t\t<input type=hidden name=domains value="ics.uci.edu">\n\t\t
\n\t\t<input type=radio name=sitesearch value="ics.uci.edu" checked> <font size="1">Repository</font>\n\t\t<input type=radio name=sitesearch value=""> <font size="1">Web</font>\n\t\t   \n\t\t<A HREF=http://www.google.com/search><IMG SRC=http://www.google.com/logos/Logo_25blk.gif border=0 ALT=Google align=middle height=27></A>\n\t\t
\n\t\t</FORM>\n\t\t<!-- Search Google -->\n\n\n\t\t<a href="datasets.php"><font size="3" color="#FFDD33">View ALL Data Sets</font></a>\n\t\t
\n\t</td>\n</tr>\n</table>\n\n
\n\n<table cellpadding=3>\n\t<tr>\n\t\t<td valign="top">\n\n\t\t\t<table><tr><td> Browse Through: </td></tr></table>\n\n\n\n<table border=2 cellpadding=2 width=150>\n\t\t<tr> <td bgcolor="#003366"> Default Task\t \n\t\t</td>\n\t</tr>\n\t<tr>\n\t\t<td valign=top> <a href=\'datasets.php?format=&task=cla&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table\'>Classification</a> <font color=red>(350)</font>
<a href=\'datasets.php?format=&task=reg&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table\'>Regression</a> <font color=red>(96)</font>
<a href=\'datasets.php?format=&task=clu&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table\'>Clustering</a> <font color=red>(84)</font>
<a href=\'datasets.php?format=&task=other&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table\'>Other</a> <font color=red>(55)</font>\t \n\t\t</td>\n\t</tr>\n\t<tr> <td bgcolor="#003366"> Attribute Type \t \n\t\t</td>\n\t</tr>\n\t<tr>\n\t\t<td valign=top> <a href=\'datasets.php?format=&task=&att=cat&area=&numAtt=&numIns=&type=&sort=nameUp&view=table\'>Categorical</a> <font color=red>(38)</font>
<a href=\'datasets.php?format=&task=&att=num&area=&numAtt=&numIns=&type=&sort=nameUp&view=table\'>Numerical</a> <font color=red>(307)</font>
<a href=\'datasets.php?format=&task=&att=mix&area=&numAtt=&numIns=&type=&sort=nameUp&view=table\'>Mixed</a> <font color=red>(55)</font>\t \n\t\t</td>\n\t</tr>\n\t<tr> <td bgcolor="#003366"> Data Type \t \n\t\t</td>\n\t</tr>\n\t<tr>\n\t\t<td valign=top> <a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=mvar&sort=nameUp&view=table\'>Multivariate</a> <font color=red>(357)</font>
<a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=uvar&sort=nameUp&view=table\'>Univariate</a> <font color=red>(23)</font>
<a href=\'datasets.

php?format=&task=&att=&area=&numAtt=&numIns=&type=seq&sort=nameUp&view=table\'>Sequential</a> <font color=red>(47)</font>
<a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=ts&sort=nameUp&view=table\'>Time-Series</a> <font color=red>(91)</font>
<a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=text&sort=nameUp&view=table\'>Text</a> <font color=red>(53)</font>
<a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=dt&sort=nameUp&view=table\'>Domain-Theory</a> <font color=red>(23)</font>
<a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=other&sort=nameUp&view=table\'>Other</a> <font color=red>(21)</font>
\t \n\t\t</td>\n\t</tr>\n\t<tr><td bgcolor="#003366"> Area\t \n\t\t</td>\n\t</tr>\n\t<tr>\n\t\t<td valign=top> <a href=\'datasets.php?format=&task=&att=&area=life&numAtt=&numIns=&type=&sort=nameUp&view=table\'>Life Sciences</a> <font color=red>(107)</font>
<a href=\'datasets.php?format=&task=&att=&area=phys&numAtt=&numIns=&type=&sort=nameUp&view=table\'>Physical Sciences</a> <font color=red>(49)</font>
<a href=\'datasets.php?format=&task=&att=&area=comp&numAtt=&numIns=&type=&sort=nameUp&view=table\'>CS / Engineering</a> <font color=red>(170)</font>
<a href=\'datasets.php?format=&task=&att=&area=soc&numAtt=&numIns=&type=&sort=nameUp&view=table\'>Social Sciences</a> <font color=red>(26)</font>
<a href=\'datasets.php?format=&task=&att=&area=bus&numAtt=&numIns=&type=&sort=nameUp&view=table\'>Business</a> <font color=red>(29)</font>
<a href=\'datasets.php?format=&task=&att=&area=game&numAtt=&numIns=&type=&sort=nameUp&view=table\'>Game</a> <font color=red>(10)</font>
<a href=\'datasets.php?format=&task=&att=&area=other&numAtt=&numIns=&type=&sort=nameUp&view=table\'>Other</a> <font color=red>(73)</font>\t \n\t\t</td>\n\t</tr>\n\t<tr><td bgcolor="#003366"> # Attributes\t \n\t\t</td>\n\t</tr>\n\t<tr>\n\t\t<td valign=top> <a href=\'datasets.php?format=&task=&att=&area=&numAtt=less10&numIns=&type=&sort=nameUp&view=table\'>Less than 10</a> <font color=red>(113)</font>
<a href=\'datasets.php?format=&task=&att=&area=&numAtt=10to100&numIns=&type=&sort=nameUp&view=table\'>10 to 100</a> <font color=red>(210)</font>
<a href=\'datasets.php?format=&task=&att=&area=&numAtt=greater100&numIns=&type=&sort=nameUp&view=table\'>Greater than 100</a> <font color=red>(84)</font>\t \n\t\t</td>\n\t</tr>\n\t<tr><td bgcolor="#003366"> # Instances </td>\n\t\t</tr>\n\n\t\t<tr>\n\t\t\t<td valign=top> \n\t\t<a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=less100&type=&sort=nameUp&view=table\'>Less than 100</a> <font color=red>(27)</font>
<a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=100to1000&type=&sort=nameUp&view=table\'>100 to 1000</a> <font color=red>(162)</font>
<a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=greater1000&type=&sort=nameUp&view=table\'>Greater than 1000</a> <font color=red>(246)</font>\t \n\t\t</td>\n\t</tr>\n\t<tr><td bgcolor="#003366"> Format Type\t \n\t\t</td>\n\t</tr>\n\t<tr>\n\t\t<td valign=top> <a href=\'datasets.php?format=mat&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table\'>Matrix</a> <font color=red>(324)</font>
<a href=\'datasets.php?format=nonmat&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table\'>Non-Matrix</a> <font color=red>(145)</font>\t \n\t\t</td>\n\t</tr></table>\n\t</td>\n\n\t<td valign=top>\n\n<table width=100%>\n\t\t<tr>\n\t\t\t<td> 469 Data Sets </td>\n\t\t\t<td align="right"> <font color=gray>Table View</font> <a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list\'>List View</a> </td>\n\t\t</tr>\n\t</table><table border = 1 cellpadding = 5><tr bgcolor="#003366">\n\t\t<td class="normal, whitetext"> <a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameDown&view=table\'>Name</a> </td>\n\t\t<!-- <td> Abstract </td> -->\n\t\t<td> <a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=typeUp&view=table\'>Data Types</a> </td>\n\t\t<td> <a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=taskUp&view=table\'>Default Task</a> </td>\n\t\t<td> <a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=attTypeUp&view=table\'>Attribute Types</a> </td>\n\t\t<td> <a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=instUp&view=table\'># Instances</a> </td>\n\t\t<td> <a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=attUp&view=table\'># Attributes</a> </td>\n\t\t<td> <a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=dateUp&view=table\'>Year</a> </td>\n\n\t\t<!-- <td> Area </td> -->\n\n\n\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Abalone"><img src="assets/MLimages/SmallLarge1.jpg" border=1 /></a> </td><td> <a href="datasets/Abalone">Abalone</a> </td></tr></table></td>\n\t\t\t<!-- <td> Predict the age of abalone from physical measurements </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer, Real </td>\n\t\t\t<td> 4177 </td>\n\t\t\t<td> 8 </td>\n\t\t\t<td> 1995 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Adult"><img src="assets/MLimages/SmallLarge2.jpg" border=1 /></a> </td><td> <a href="datasets/Adult">Adult</a> </td></tr></table></td>\n\t\t\t<!-- <td> Predict whether income exceeds $50K/yr based on census data. Also known as "Census Income" dataset. </td> -

->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<td> 48842 </td>\n\t\t\t<td> 14 </td>\n\t\t\t<td> 1996 </td>\n\t\t\t<!-- <td> Social </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Annealing"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Annealing">Annealing</a> </td></tr></table></td>\n\t\t\t<!-- <td> Steel annealing data </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer, Real </td>\n\t\t\t<td> 798 </td>\n\t\t\t<td> 38 </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Anonymous+Microsoft+Web+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Anonymous+Microsoft+Web+Data">Anonymous Microsoft Web Data</a> </td></tr></table></td>\n\t\t\t<!-- <td> Log of anonymous users of www.microsoft.com; predict areas of the web site a user visited based on data on other areas the user visited. </td> -->\n\t\t\t<td>   </td>\n\t\t\t<td> Recommender-Systems </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 37711 </td>\n\t\t\t<td> 294 </td>\n\t\t\t<td> 1998 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Arrhythmia"><img src="assets/MLimages/SmallLarge5.jpg" border=1 /></a> </td><td> <a href="datasets/Arrhythmia">Arrhythmia</a> </td></tr></table></td>\n\t\t\t<!-- <td> Distinguish between the presence and absence of cardiac arrhythmia and classify it in one of the 16 groups. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer, Real </td>\n\t\t\t<td> 452 </td>\n\t\t\t<td> 279 </td>\n\t\t\t<td> 1998 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Artificial+Characters"><img src="assets/MLimages/SmallLarge6.jpg" border=1 /></a> </td><td> <a href="datasets/Artificial+Characters">Artificial Characters</a> </td></tr></table></td>\n\t\t\t<!-- <td> Dataset artificially generated by using first order theory which describes structure of ten capital letters of English alphabet </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer, Real </td>\n\t\t\t<td> 6000 </td>\n\t\t\t<td> 7 </td>\n\t\t\t<td> 1992 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Audiology+%28Original%29"><img src="assets/MLimages/SmallLarge7.jpg" border=1 /></a> </td><td> <a href="datasets/Audiology+%28Original%29">Audiology (Original)</a> </td></tr></table></td>\n\t\t\t<!-- <td> Nominal audiology dataset from Baylor </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 226 </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 1987 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Audiology+%28Standardized%29"><img src="assets/MLimages/SmallLarge7.jpg" border=1 /></a> </td><td> <a href="datasets/Audiology+%28Standardized%29">Audiology (Standardized)</a> </td></tr></table></td>\n\t\t\t<!-- <td> Standardized version of the original audiology database </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 226 </td>\n\t\t\t<td> 69 </td>\n\t\t\t<td> 1992 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Auto+MPG"><img src="assets/MLimages/SmallLarge9.jpg" border=1 /></a> </td><td> <a href="datasets/Auto+MPG">Auto MPG</a> </td></tr></table></td>\n\t\t\t<!-- <td> Revised from CMU StatLib library, data concerns city-cycle fuel consumption </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Categorical, Real </td>\n\t\t\t<td> 398 </td>\n\t\t\t<td> 8 </td>\n\t\t\t<td> 1993 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Automobile"><img src="assets/MLimages/SmallLarge10.jpg" border=1 /></a> </td><td> <a href="datasets/Automobile">Automobile</a> </td></tr></table></td>\n\t\t\t<!-- <td> From 1985 Ward\'s Automotive Yearbook </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Categorical, Integer, Real </td>\n\t\t\t<td> 205 </td>\n\t\t\t<td> 26 </td>\n\t\t\t<td> 1987 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href

="datasets/Badges"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p cla
ss="normal"><a href="datasets/Badges">Badges</a> </td></tr></table></td>\n\t\t\t<!-- <td><p class
="normal">Badges labeled with a "+" or "-" as a function of a person\'s name </td> -->\n\t\t\t<td><p cl
ass="normal">Univariate, Text </td>\n\t\t\t<td> Classification </td>\n\t\t\t
<td>   </td>\n\t\t\t<td> 294 </td>\n\t\t\t<td><p class="n
ormal">1 </td>\n\t\t\t<td> 1994 </td>\n\t\t\t<!-- <td>
Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Balanc
e+Scale"><img src="assets/MLimages/SmallLarge12.jpg" border=1 /></a> </td><td> <
b><a href="datasets/Balance+Scale">Balance Scale</a> </td></tr></table></td>\n\t\t\t<!-- <td><p class
="normal">Balance scale weight & distance database </td> -->\n\t\t\t<td> Multivari
ate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Cat
egorical </td>\n\t\t\t<td> 625 </td>\n\t\t\t<td> 4
; </td>\n\t\t\t<td> 1994 </td>\n\t\t\t<!-- <td> Social </p
></td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Balloons"><img src="assets/MLimages/Sm
allLarge13.jpg" border=1 /></a> </td><td> <a href="datasets/Balloons">Balloons</
a> </td></tr></table></td>\n\t\t\t<!-- <td> Data previously used in cognitive psycholog
y experiment; 4 data sets represent different conditions of an experiment </td> -->\n\t\t\t<td><p class
="normal">Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td><p
class="normal">Categorical </td>\n\t\t\t<td> 16 </td>\n\t\t\t<td><p class
="normal">4 </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> So
cial </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Breast+C
ancer"><img src="assets/MLimages/SmallLarge14.jpg" border=1 /></a> </td><td>
<a href="datasets/Breast+Cancer">Breast Cancer</a> </td></tr></table></td>\n\t\t\t<!-- <td><p class="
normal">Breast Cancer Data (Restricted Access) </td> -->\n\t\t\t<td> Multivariate&
nbsp; </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categor
ical </td>\n\t\t\t<td> 286 </td>\n\t\t\t<td> 9
</td>\n\t\t\t<td> 1988 </td>\n\t\t\t<!-- <td> Life </td> -
->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Breast+Cancer+Wisconsin+%28Original%29"><im
g src="assets/MLimages/SmallLarge14.jpg" border=1 /></a> </td><td> <a href="da
tasets/Breast+Cancer+Wisconsin+%28Original%29">Breast Cancer Wisconsin (Original)</a> </td></tr>
</table></td>\n\t\t\t<!-- <td> Original Wisconsin Breast Cancer Database </td> -->\
n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification
</td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 699 </td>\n\t\
t\t<td> 10 </td>\n\t\t\t<td> 1992 </td>\n\t\t\t<!-- <td><
p class="normal">Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="
datasets/Breast+Cancer+Wisconsin+%28Prognostic%29"><img src="assets/MLimages/SmallLarge14.jpg" bord
er=1 /></a> </td><td> <a href="datasets/Breast+Cancer+Wisconsin+%28Prognosti
c%29">Breast Cancer Wisconsin (Prognostic)</a> </td></tr></table></td>\n\t\t\t<!-- <td><p class="norm
al">Prognostic Wisconsin Breast Cancer Database </td> -->\n\t\t\t<td> Multivariate
  </td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td><p class="no
rmal">Real </td>\n\t\t\t<td> 198 </td>\n\t\t\t<td> 34&
nbsp; </td>\n\t\t\t<td> 1995 </td>\n\t\t\t<!-- <td> Life <
/p></td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Breast+Cancer+Wisconsin+%28Diagnost
ic%29"><img src="assets/MLimages/SmallLarge14.jpg" border=1 /></a> </td><td>
<a href="datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29">Breast Cancer Wisconsin (Diagnostic)</a></
b> </td></tr></table></td>\n\t\t\t<!-- <td> Diagnostic Wisconsin Breast Cancer Database&
nbsp; </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classi
fication </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 569&nb
sp; </td>\n\t\t\t<td> 32 </td>\n\t\t\t<td> 1995 </td
>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><
tr><td><a href="datasets/Pittsburgh+Bridges"><img src="assets/MLimages/SmallLarge18.jpg" border=1 /></a>
 </td><td> <a href="datasets/Pittsburgh+Bridges">Pittsburgh Bridges</a>
</td></tr></table></td>\n\t\t\t<!-- <td> Bridges database that has original and numeric-discreti
zed datasets </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td><p class="
normal">Classification </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<
td> 108 </td>\n\t\t\t<td> 13 </td>\n\t\t\t<td><p class="
normal">1990 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td
><table><tr><td><a href="datasets/Car+Evaluation"><img src="assets/MLimages/SmallLarge19.jpg" border=1
/></a> </td><td> <a href="datasets/Car+Evaluation">Car Evaluation</a> </
td></tr></table></td>\n\t\t\t<!-- <td> Derived from simple hierarchical decision model, this dat
abase may be useful for testing constructive induction and structure discovery methods. </td> -->\n\t\
t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td

>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 1728 </td>\n\t\t\t<td> 6 </td>\n\t\t\t<td> 1997 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Census+Income"><img src="assets/MLimages/SmallLarge2.jpg" border=1 /></a> </td><td> <a href="datasets/Census+Income">Census Income</a> </td></tr></table></td>\n\t\t\t<!-- <td> Predict whether income exceeds $50K/yr based on census data. Also known as "Adult" dataset. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<td> 48842 </td>\n\t\t\t<td> 14 </td>\n\t\t\t<td> 1996 </td>\n\t\t\t<!-- <td> Social </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Chess+%28King-Rook+vs.+King-Knight%29"><img src="assets/MLimages/SmallLarge24.jpg" border=1 /></a> </td><td> <a href="datasets/Chess+%28King-Rook+vs.+King-Knight%29">Chess (King-Rook vs. King-Knight)</a> </td></tr></table></td>\n\t\t\t<!-- <td> Knight Pin Chess End-Game Database Creator </td> -->\n\t\t\t<td> Multivariate, Data-Generator </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 22 </td>\n\t\t\t<td> 1988 </td>\n\t\t\t<!-- <td> Game </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Chess+%28King-Rook+vs.+King-Pawn%29"><img src="assets/MLimages/SmallLarge24.jpg" border=1 /></a> </td><td> <a href="datasets/Chess+%28King-Rook+vs.+King-Pawn%29">Chess (King-Rook vs. King-Pawn)</a> </td></tr></table></td>\n\t\t\t<!-- <td> King+Rook versus King+Pawn on a7 (usually abbreviated KRKPA7). </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 3196 </td>\n\t\t\t<td> 36 </td>\n\t\t\t<td> 1989 </td>\n\t\t\t<!-- <td> Game </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Chess+%28King-Rook+vs.+King%29"><img src="assets/MLimages/SmallLarge24.jpg" border=1 /></a> </td><td> <a href="datasets/Chess+%28King-Rook+vs.+King%29">Chess (King-Rook vs. King)</a> </td></tr></table></td>\n\t\t\t<!-- <td> Chess Endgame Database for White King and Rook against Black King (KRK). </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<td> 28056 </td>\n\t\t\t<td> 6 </td>\n\t\t\t<td> 1994 </td>\n\t\t\t<!-- <td> Game </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Chess+%28Domain+Theories%29"><img src="assets/MLimages/SmallLarge24.jpg" border=1 /></a> </td><td> <a href="datasets/Chess+%28Domain+Theories%29">Chess (Domain Theories)</a> </td></tr></table></td>\n\t\t\t<!-- <td> 6 different domain theories for generating legal moves of chess </td> -->\n\t\t\t<td> Domain-Theory </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Game </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Bach+Chorales"><img src="assets/MLimages/SmallLarge25.jpg" border=1 /></a> </td><td> <a href="datasets/Bach+Chorales">Bach Chorales</a> </td></tr></table></td>\n\t\t\t<!-- <td> Time-series data based on chorales; challenge is to learn generative grammar; data in Lisp </td> -->\n\t\t\t<td> Univariate, Time-Series </td>\n\t\t\t<td>   </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<td> 100 </td>\n\t\t\t<td> 6 </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Connect-4"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Connect-4">Connect-4</a> </td></tr></table></td>\n\t\t\t<!-- <td> Contains connect-4 positions </td> -->\n\t\t\t<td> Multivariate, Spatial </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 67557 </td>\n\t\t\t<td> 42 </td>\n\t\t\t<td> 1995 </td>\n\t\t\t<!-- <td> Game </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Credit+Approval"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Credit+Approval">Credit Approval</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data concerns credit card applications; good mix of attributes </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer, Real </td>\n\t\t\t<td> 690 </td>\n\t\t\t<td> 15 </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Financial </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Japanese+Credit

</td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Japanese+Credit
+Screening"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="no
rmal"><a href="datasets/Japanese+Credit+Screening">Japanese Credit Screening</a> </td></tr></
table></td>\n\t\t\t<!-- <td> Includes domain theory (generated by talking to Japanese domain
experts); data in Lisp </td> -->\n\t\t\t<td> Multivariate, Domain-Theory <
/td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Real, I
nteger </td>\n\t\t\t<td> 125 </td>\n\t\t\t<td> </
p></td>\n\t\t\t<td> 1992 </td>\n\t\t\t<!-- <td> Financial </p
></td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Computer+Hardware"><img src="assets/M
Limages/SmallLarge29.jpg" border=1 /></a> </td><td> <a href="datasets/Compute
r+Hardware">Computer Hardware</a> </td></tr></table></td>\n\t\t\t<!-- <td> Relativ
e CPU Performance Data, described in terms of its cycle time, memory size, etc. </td> -->\n\t\t\t<td><
p class="normal">Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<t
d> Integer </td>\n\t\t\t<td> 209 </td>\n\t\t\t<td><p cla
ss="normal">9 </td>\n\t\t\t<td> 1987 </td>\n\t\t\t<!-- <td><p class="nor
mal">Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="dataset
ts/Contraceptive+Method+Choice"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>
</td><td> <a href="datasets/Contraceptive+Method+Choice">Contraceptive Method Choi
ce</a> </td></tr></table></td>\n\t\t\t<!-- <td> Dataset is a subset of the 1987 Nationa
l Indonesia Contraceptive Prevalence Survey. </td> -->\n\t\t\t<td> Multivariate&nbs
p; </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical
, Integer </td>\n\t\t\t<td> 1473 </td>\n\t\t\t<td> 9&nbs
p; </td>\n\t\t\t<td> 1997 </td>\n\t\t\t<!-- <td> Life
</td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Covertype"><img src="assets/MLimages/Sm
allLarge31.jpg" border=1 /></a> </td><td> <a href="datasets/Covertype">Covertyp
e</a> </td></tr></table></td>\n\t\t\t<!-- <td> Forest CoverType dataset </t
d> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification&nbs
p; </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<td> 5810
12 </td>\n\t\t\t<td> 54 </td>\n\t\t\t<td> 1998 </
p></td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><t
able><tr><td><a href="datasets/Cylinder+Bands"><img src="assets/MLimages/SmallLargedefault.jpg" border=1
/></a> </td><td> <a href="datasets/Cylinder+Bands">Cylinder Bands</a>
</td></tr></table></td>\n\t\t\t<!-- <td> Used in decision tree induction for mitigating process d
elays known as "cylinder bands" in rotogravure printing </td> -->\n\t\t\t<td> Multiva
riate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> C
ategorical, Integer, Real </td>\n\t\t\t<td> 512 </td>\n\t\t\t<td><p class="
normal">39 </td>\n\t\t\t<td> 1995 </td>\n\t\t\t<!-- <td><p class="normal"
>Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Dermatology"><img src
="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="dat
asets/Dermatology">Dermatology</a> </td></tr></table></td>\n\t\t\t<!-- <td> Aim for
this dataset is to determine the type of Eryhemato-Squamous Disease. </td> -->\n\t\t\t<td><p class="
normal">Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td><p cl
ass="normal">Categorical, Integer </td>\n\t\t\t<td> 366 </td>\n\t\t\t<td>
33 </td>\n\t\t\t<td> 1998 </td>\n\t\t\t<!-- <td><p clas
s="normal">Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datas
ets/Diabetes"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="n
ormal"><a href="datasets/Diabetes">Diabetes</a> </td></tr></table></td>\n\t\t\t<!-- <td><p class="n
ormal">This diabetes dataset is from AIM \'94 </td> -->\n\t\t\t<td> Multivariate, Tim
e-Series </td>\n\t\t\t<td>   </td>\n\t\t\t<td> Categorical
, Integer </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 20 </
p></td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Life </td> -->\
n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/DGP2+-+The+Second+Data+Generation+Program">
<img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a
href="datasets/DGP2+-+The+Second+Data+Generation+Program">DGP2 - The Second Data Generation Progr
am</a> </td></tr></table></td>\n\t\t\t<!-- <td> Generates application domains based
on specific parameters, number of features, and proportion of positive to negative examples </td> --
>\n\t\t\t<td> Data-Generator </td>\n\t\t\t<td>   </td>\n\
t\t\t<td> Real </td>\n\t\t\t<td>   </td>\n\t\t\t<td><p clas
s="normal"> </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Ot
her </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Documen
t+Understanding"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class
="normal"><a href="datasets/Document+Understanding">Document Understanding</a> </td></tr><
/table></td>\n\t\t\t<!-- <td> Five concepts, expressed as predicates, to be learned
</td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classificat
</td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 1000
</td>\n\t\t\t<td> 10 </td>\n\t\t\t<td> 1990 </td>\n\t\t\t<
!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr
><td><a href="datasets/Drug+Consumption+(Quantified)"><img src="assets/MLimages/SmallLargedefault.jpg"
border=1 /></a> </td><td> <a href="datasets/Drug+Consumption+(Quantified)">Dr
ug Consumption (Quantified)</a> </td></tr></table></td>\n\t\t\t<!-- <td> Fifteen
personality measurements including neuroticism, extraversion, openness to experience, agreeableness, and
conscientiousness, plus level of education, age, gender, country of residence and ethnicity for 1885 respon
dents. Participants were also asked about their use of 18 legal and illegal drugs (alcohol, amphetamines, a
mphetamine substitutes, benzodiazepines, cannabis, chocolate, cocaine, caffeine, crack, ecstasy, heroin, ket
amine, legal highs, LSD, methadone, mushrooms, nicotine and volatile substance abuse) and one fictitious d
rug (Semeron) which was introduced to identify over-claimers. For each drug they selected from seven answer
s: never used, used over a decade ago, used in last decade, used in last year, used in last month, used in
last week, used in last day. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t
\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer
</td>\n\t\t\t<td> 1885 </td>\n\t\t\t<td> 32 </td>\n\t\t
\t<td> 2016 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</
tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Drug+Consumption+(Quantified)"><im
g src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a
href="datasets/Drug+Consumption+(Quantified)">Drug Consumption (Quantified)</a> </td></tr></
table></td>\n\t\t\t<!-- <td> Fifteen personality measurements including neuroticism, extrav
ersion, openness to experience, agreeableness, and conscientiousness, plus level of education, age, gender,
country of residence and ethnicity for 1885 respondents. Participants were also asked about their use of 1
8 legal and illegal drugs (alcohol, amphetamines, amphetamine substitutes, benzodiazepines, cannabis, choco
late, cocaine, caffeine, crack, ecstasy, heroin, ketamine, legal highs, LSD, methadone, mushrooms, nicotine
and volatile substance abuse) and one fictitious drug (Semeron) which was introduced to identify over-claim
ers. For each drug they selected from seven answers: never used, used over a decade ago, used in last decad
e, used in last year, used in last month, used in last week, used in last day. </td> -->\n\t\t\t<td><
p class="normal">Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t
\t<td> Categorical, Integer </td>\n\t\t\t<td> 1885 </td>\n\t\t
\t<td> 32 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td><p
class="normal">Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="d
atasets/Dry+Bean+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><
td> <a href="datasets/Dry+Bean+Dataset">Dry Bean Dataset</a> </td></tr></table>
</td>\n\t\t\t<!-- <td> Images of 13,611 grains of 7 different registered dry beans were taken
with a high-resolution camera. Later, a total of 16 features; 12 dimensions and 4 shape forms, were obtain
ed from the grains. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td><p
class="normal">Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<
td> 13611 </td>\n\t\t\t<td> 16 </td>\n\t\t\t<td><p class="
normal">2020 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bg
color="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Dry+Bean+Dataset"><img src="assets/MLimages/
SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Dry+Bean+Da
taset">Dry Bean Dataset</a> </td></tr></table></td>\n\t\t\t<!-- <td> Images of 13,
611 grains of 7 different registered dry beans were taken with a high-resolution camera. Later, a total of
16 features; 12 dimensions and 4 shape forms, were obtained from the grains. </td> -->\n\t\t\t<td><
p class="normal">Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t
\t<td> Integer, Real </td>\n\t\t\t<td> 13611 </td>\n\t\t
\t<td> 16 </td>\n\t\t\t<td> 2020 </td>\n\t\t\t<!-- <td><p
class="normal">Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="d
atasets/Dutch+Energy+Companies+Gas+Consumption"><img src="assets/MLimages/SmallLargedefault.jpg" bord
er=1 /></a> </td><td> <a href="datasets/Dutch+Energy+Companies+Gas+Consumption
">Dutch Energy Companies Gas Consumption</a> </td></tr></table></td>\n\t\t\t<!-- <td><p class="norm
al">Gas consumption data of Dutch energy companies. </td> -->\n\t\t\t<td> Multivar
iate, Time-Series </td>\n\t\t\t<td>   </td>\n\t\t\t<td> Re
al </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n
\t\t\t<td>   </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t
\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Dutch+Energy+Companies+Gas+Consu
mption"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="norma
l"><a href="datasets/Dutch+Energy+Companies+Gas+Consumption">Dutch Energy Companies Gas Consumption
</a> </td></tr></table></td>\n\t\t\t<!-- <td> Gas consumption data of Dutch energy
companies. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<
td>   </td>\n\t\t\t<td> Real </td>\n\t\t\t<td><p class="no
rmal"> </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </t
d>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><
table><tr><td><a href="datasets/Electrical+Grid+Stability+Simulated+Data"><img src="assets/MLimages/Sma
llLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Electrical+Gri
d+Stability+Simulated+Data">Electrical Grid Stability Simulated Data</a> </td></tr></table></td>
\n\t\t\t<!-- <td> The data is generated for the 4-node star system where the electricity pro
ducer is in the center. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td
> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td><
p class="normal">10000 </td>\n\t\t\t<td> 14 </td>\n\t\t\t<td><p class="norm
al">2015 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor
="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Electrical+Grid+Stability+Simulated+Data"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a
href="datasets/Electrical+Grid+Stability+Simulated+Data">Electrical Grid Stability Simulated Data</a></b
> </td></tr></table></td>\n\t\t\t<!-- <td> The data is generated for the 4-node star sy
stem where the electricity producer is in the center. </td> -->\n\t\t\t<td> Multiva
riate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> R
eal </td>\n\t\t\t<td> 10000 </td>\n\t\t\t<td> 14
</td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Life </td>
-->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Electricity+Load+Diagrams
+20112014"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="norm
al"><a href="datasets/Electricity+Load+Diagrams+20112014">Electricity Load Diagrams 20112014</a>
</td></tr></table></td>\n\t\t\t<!-- <td> Electricity consumption data of 370 points/cl
ients. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td><p
class="normal"> </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> &
nbsp; </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t
\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table
><tr><td><a href="datasets/Electricity+Load+Diagrams+20112014"><img src="assets/MLimages/SmallLarged
efault.jpg" border=1 /></a> </td><td> <a href="datasets/Electricity+Load+Diagr
ams+20112014">Electricity Load Diagrams 20112014</a> </td></tr></table></td>\n\t\t\t<!-- <td><p
class="normal">Electricity consumption data of 370 points/clients. </td> -->\n\t\t\t<td><p class="no
rmal">Multivariate, Time-Series </td>\n\t\t\t<td>   </td>\n\t\t\t<td><p cl
ass="normal">Real </td>\n\t\t\t<td>   </td>\n\t\t\t<td> &
nbsp; </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Life </
td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Energy+Efficiency"><
img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a
href="datasets/Energy+Efficiency">Energy Efficiency</a> </td></tr></table></td>\n\t\t\t<!-- <td><p
class="normal">This study looked into the assessment of the heating load and cooling load requirements of
buildings (that is, energy efficiency) as a function of building parameters. </td> -->\n\t\t\t<td><
p class="normal">Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td
> Real </td>\n\t\t\t<td> 768 </td>\n\t\t\t<td><p class="no
rmal">8 </td>\n\t\t\t<td> 2012 </td>\n\t\t\t<!-- <td> Phy
sical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Ener
gy+Efficiency"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class
="normal"><a href="datasets/Energy+Efficiency">Energy Efficiency</a> </td></tr></table></td>
\n\t\t\t<!-- <td> This study looked into the assessment of the heating load and cooling load
requirements of buildings (that is, energy efficiency) as a function of building parameters. </td>
-->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression
</td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 768 </td>\n\t\t
\t<td> 8 </td>\n\t\t\t<td> 2012 </td>\n\t\t\t<!-- <td><p
class="normal">Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="d
atasets/Facebook+Live+Sellers+in+Thailand"><img src="assets/MLimages/SmallLargedefault.jpg" border=1
/></a> </td><td> <a href="datasets/Facebook+Live+Sellers+in+Thailand">Facebo
ok Live Sellers in Thailand</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dat
aset was collected from the Facebook pages of 10 Thai fashion and cosmetics retail sellers selling products
via live broadcasts. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td>
Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t
\t<td> 1048 </td>\n\t\t\t<td> 7 </td>\n\t\t\t<td><p class="
normal">2016 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bg
color="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Facebook+Live+Sellers+in+Thailand"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a
href="datasets/Facebook+Live+Sellers+in+Thailand">Facebook Live Sellers in Thailand</a> </td></tr
></table></td>\n\t\t\t<!-- <td> This dataset was collected from the Facebook pages of 10 T
hai fashion and cosmetics retail sellers selling products via live broadcasts. </td> -->\n\t\t\t<td>
Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t
\t<td> Integer, Real </td>\n\t\t\t<td> 1048 </td>\n\t\t
\t<td> 7 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td><p
class="normal">Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="d
atasets/Facebook+Metrics"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td>
<td> <a href="datasets/Facebook+Metrics">Facebook Metrics</a> </td></tr></
table></td>\n\t\t\t<!-- <td> This dataset contains metrics on a variety of posts from a Thai
cosmetics brand's Facebook page. </td> -->\n\t\t\t<td> Multivariate </td>
\n\t\t\t<td> Regression </td>\n\t\t\t<td> Integer </td>\n\t\t
\t<td> 500 </td>\n\t\t\t<td> 19 </td>\n\t\t\t<td><p class="
normal">2016 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bg
color="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Facebook+Metrics"><img src="assets/MLimages/
SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Facebook+Me
trics">Facebook Metrics</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset
contains metrics on a variety of posts from a Thai cosmetics brand's Facebook page. </td> -->\n\t\t\t
<td> Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t
\t<td> Integer </td>\n\t\t\t<td> 500 </td>\n\t\t\t<td><p
class="normal">19 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td><p class="
normal">Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datas
ets/Facebook+Page+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><
td> <a href="datasets/Facebook+Page+Data">Facebook Page Data</a> </td></tr></
table></td>\n\t\t\t<!-- <td> Facebook data has been collected from the Facebook pages of
various businesses. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td><p
class="normal">Regression </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td><p
class="normal">500 </td>\n\t\t\t<td> 19 </td>\n\t\t\t<td>
2016 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">
\n\t\t\t<td><table><tr><td><a href="datasets/Facebook+Page+Data"><img src="assets/MLimages/SmallLarged
efault.jpg" border=1 /></a> </td><td> <a href="datasets/Facebook+Page+Data">Fa
cebook Page Data</a> </td></tr></table></td>\n\t\t\t<!-- <td> Facebook data has
been collected from the Facebook pages of various businesses. </td> -->\n\t\t\t<td><p class="normal
">Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td><p class="
normal">Integer </td>\n\t\t\t<td> 500 </td>\n\t\t\t<td>
19 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Life
</td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Facebook+Polit
ical+Ads"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="norm
al"><a href="datasets/Facebook+Political+Ads">Facebook Political Ads</a> </td></tr></table>
</td>\n\t\t\t<!-- <td> Facebook data has been collected from the Facebook pages of various
businesses. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td><p class="
normal">Regression </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td><p class="
normal">500 </td>\n\t\t\t<td> 19 </td>\n\t\t\t<td> 2016&
nbsp; </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t
\t<td><table><tr><td><a href="datasets/Facebook+Political+Ads"><img src="assets/MLimages/SmallLarged
efault.jpg" border=1 /></a> </td><td> <a href="datasets/Facebook+Political+Ads
">Facebook Political Ads</a> </td></tr></table></td>\n\t\t\t<!-- <td> Facebook da
ta has been collected from the Facebook pages of various businesses. </td> -->\n\t\t\t<td><p class="
normal">Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td><p cl
ass="normal">Integer </td>\n\t\t\t<td> 500 </td>\n\t\t\t<td><p class="no
rmal">19 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Li
fe </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Facebo
ok+Post+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class
="normal"><a href="datasets/Facebook+Post+Data">Facebook Post Data</a> </td></tr></table></td>
\n\t\t\t<!-- <td> Facebook data has been collected from the Facebook pages of various
businesses. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td><p class="
normal">Regression </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td><p class="
normal">500 </td>\n\t\t\t<td> 19 </td>\n\t\t\t<td> 2016&
nbsp; </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t
\t<td><table><tr><td><a href="datasets/Facebook+Post+Data"><img src="assets/MLimages/SmallLargedefa
ult.jpg" border=1 /></a> </td><td> <a href="datasets/Facebook+Post+Data">Face
book Post Data</a> </td></tr></table></td>\n\t\t\t<!-- <td> Facebook data has been
collected from the Facebook pages of various businesses. </td> -->\n\t\t\t<td> Mult
ivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td>
Integer </td>\n\t\t\t<td> 500 </td>\n\t\t\t<td> 19 </
p></td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Life </td
> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Facebook+Total+Interac
tions"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="norma
l"><a href="datasets/Facebook+Total+Interactions">Facebook Total Interactions</a> </td></tr>
</table></td>\n\t\t\t<!-- <td> Facebook data has been collected from the Facebook pages of
various businesses. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td><p
class="normal">Regression </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td><p
class="normal">500 </td>\n\t\t\t<td> 19 </td>\n\t\t\t<td>
2016 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">
\n\t\t\t<td><table><tr><td><a href="datasets/Facebook+Total+Interactions"><img src="assets/MLimages/Sma
llLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Facebook+Total
+Interactions">Facebook Total Interactions</a> </td></tr></table></td>\n\t\t\t<!-- <td><p class="
normal">Facebook data has been collected from the Facebook pages of various businesses. </td> -->\n
\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression </td>
\n\t\t\t<td> Integer </td>\n\t\t\t<td> 500 </td>\n\t\t\t<td
> 19 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td><p cl
ass="normal">Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="dat
asets/Facebook+Video+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td>
<td> <a href="datasets/Facebook+Video+Data">Facebook Video Data</a> </td></tr>
</table></td>\n\t\t\t<!-- <td> Facebook data has been collected from the Facebook pages of
various businesses. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t

</table></td>\n\t\t\t<!-- <td> Five concepts, expressed as predicates, to be learned </td> -->\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 1994 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/EBL+Domain+Theories"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/EBL+Domain+Theories">EBL Domain Theories</a> </td></tr></table></td>\n\t\t\t<!-- <td> Assorted small-scale domain theories </td> -->\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Echocardiogram"><img src="assets/MLimages/SmallLarge38.jpg" border=1 /></a> </td><td> <a href="datasets/Echocardiogram">Echocardiogram</a> </td></tr></table></td>\n\t\t\t<!-- <td> Data for classifying if patients will survive for at least one year after a heart attack </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer, Real </td>\n\t\t\t<td> 132 </td>\n\t\t\t<td> 12 </td>\n\t\t\t<td> 1989 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Ecoli"><img src="assets/MLimages/SmallLarge120.jpg" border=1 /></a> </td><td> <a href="datasets/Ecoli">Ecoli</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data contains protein localization sites </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 336 </td>\n\t\t\t<td> 8 </td>\n\t\t\t<td> 1996 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Flags"><img src="assets/MLimages/SmallLarge40.jpg" border=1 /></a> </td><td> <a href="datasets/Flags">Flags</a> </td></tr></table></td>\n\t\t\t<!-- <td> From Collins Gem Guide to Flags, 1986 </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<td> 194 </td>\n\t\t\t<td> 30 </td>\n\t\t\t<td> 1990 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Function+Finding"><img src="assets/MLimages/SmallLarge41.jpg" border=1 /></a> </td><td> <a href="datasets/Function+Finding">Function Finding</a> </td></tr></table></td>\n\t\t\t<!-- <td> Cases collected mostly from investigations in physical science; intention is to evaluate function-finding algorithms </td> -->\n\t\t\t<td>   </td>\n\t\t\t<td> Function-Learning </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 352 </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 1990 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Glass+Identification"><img src="assets/MLimages/SmallLarge42.jpg" border=1 /></a> </td><td> <a href="datasets/Glass+Identification">Glass Identification</a> </td></tr></table></td>\n\t\t\t<!-- <td> From USA Forensic Science Service; 6 types of glass; defined in terms of their oxide content (i.e. Na, Fe, K, etc) </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 214 </td>\n\t\t\t<td> 10 </td>\n\t\t\t<td> 1987 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Haberman%27s+Survival"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Haberman%27s+Survival">Haberman\'s Survival</a> </td></tr></table></td>\n\t\t\t<!-- <td> Dataset contains cases from study conducted on the survival of patients who had undergone surgery for breast cancer </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 306 </td>\n\t\t\t<td> 3 </td>\n\t\t\t<td> 1999 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Hayes-Roth"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Hayes-Roth">Hayes-Roth</a> </td></tr></table></td>\n\t\t\t<!-- <td> Topic: human subjects study </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 160 </td>\n\t\t\t<td> 5 </td>\n\t\t\t<td> 1989 </td>\n\t\t\t<!-- <td> Social </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Heart+Disease"><img src="assets/MLimages/SmallLarge45.jpg" border=1 /></a> </td><td> <a href="datasets/Heart+Dis

ease">Heart Disease</a> </td></tr></table></td>\n\t\t\t<!-- <td> 4 databases: Clevel and, Hungary, Switzerland, and the VA Long Beach </td> -->\n\t\t\t<td> Multivariat e </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Cate gorical, Integer, Real </td>\n\t\t\t<td> 303 </td>\n\t\t\t<td> 75 </td>\n\t\t\t<td> 1988 </td>\n\t\t\t<!-- <td> Lif e </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Hepatitis"><img src="assets/MLimages/SmallLarge46.jpg" border=1 /></a> </td><td> <a href=" datasets/Hepatitis">Hepatitis</a> </td></tr></table></td>\n\t\t\t<!-- <td> From G.Gon g: CMU; Mostly Boolean or numeric-valued attribute types; Includes cost data (donated by Peter Turney) </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classificati on </td>\n\t\t\t<td> Categorical, Integer, Real </td>\n\t\t\t<td> 155 </td>\n\t\t\t<td> 19 </td>\n\t\t\t<td> 198 8 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr> <td><a href="datasets/Horse+Colic"><img src="assets/MLimages/SmallLarge47.jpg" border=1 /></a> </t d><td> <a href="datasets/Horse+Colic">Horse Colic</a> </td></tr></table></td>\ n\t\t\t<!-- <td> Well documented attributes; 368 instances with 28 attributes (continuous, discr ete, and nominal); 30% missing values </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integ er, Real </td>\n\t\t\t<td> 368 </td>\n\t\t\t<td> 27&nbs p; </td>\n\t\t\t<td> 1989 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/ICU"><img src="assets/ MLimages/SmallLarge49.jpg" border=1 /></a> </td><td> <a href="datasets/ICU">I CU</a> </td></tr></table></td>\n\t\t\t<!-- <td> Data set prepared for the use of partici pants for the 1994 AAAI Spring Symposium on Artificial Intelligence in Medicine. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td>   </td>\n\t\t\t<td> Real </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Life&n bsp; </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Image+Segmentation"><img src="a ssets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datas ets/Image+Segmentation">Image Segmentation</a> </td></tr></table></td>\n\t\t\t<!-- <td> Image data described by high-level numeric-valued attributes, 7 classes </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 2310 </td>\n\t\t\t<td> 19 </td>\n\t\t\t<td> 1990 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/In ternet+Advertisements"><img src="assets/MLimages/SmallLarge51.jpg" border=1 /></a> </td><td><p cla ss="normal"><a href="datasets/Internet+Advertisements">Internet Advertisements</a> </td></tr></t able></td>\n\t\t\t<!-- <td> This dataset represents a set of possible advertisements on Intern et pages. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer, Real </td>\n\t\t \t<td> 3279 </td>\n\t\t\t<td> 1558 </td>\n\t\t\t<td> 1998 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr> <tr>\n\t\t\t<td><table><tr><td><a href="datasets/Ionosphere"><img src="assets/MLimages/SmallLarge52.jpg" b order=1 /></a> </td><td> <a href="datasets/Ionosphere">Ionosphere</a> </td></tr></table></td>\n\t\t\t<!-- <td> Classification of radar returns from the ionosphere&nb sp; </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classific ation </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 3 51 </td>\n\t\t\t<td> 34 </td>\n\t\t\t<td> 1989 </ p></td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<t d><table><tr><td><a href="datasets/Iris"><img src="assets/MLimages/SmallLarge53.jpg" border=1 /></a>&nbs p;</td><td> <a href="datasets/Iris">Iris</a> </td></tr></table></td>\n\t\t\t<!-- <td> Famous database; from Fisher, 1936 </td> -->\n\t\t\t<td> Multi variate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 150 </td>\n\t\t\t<td> 4 </ p></td>\n\t\t\t<td> 1988 </td>\n\t\t\t<!-- <td> Life </td > -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/ISOLET"><img src="assets/MLimages/SmallLar gedefault.jpg" border=1 /></a> </td><td> <a href="datasets/ISOLET">ISOLET</a> </td></tr></table></td>\n\t\t\t<!-- <td> Goal: Predict which letter-name was spoken-- a simple classification task. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 7797 </td>\n\t\t\t<td> 617 </td>\n\t\t\t<td> 1994 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bg

color="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Kinship"><img src="assets/MLimages/SmallLarg
e55.jpg" border=1 /></a> </td><td> <a href="datasets/Kinship">Kinship</a></p
></td></tr></table></td>\n\t\t\t<!-- <td> Relational dataset </td> -->\n\t\t\t<td><p cl
ass="normal">Relational </td>\n\t\t\t<td> Relational-Learning </td>\n\t\t\
t<td> Categorical </td>\n\t\t\t<td> 104 </td>\n\t\t\t<td
> 12 </td>\n\t\t\t<td> 1990 </td>\n\t\t\t<!-- <td><p cla
ss="normal">Social </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Labor+Relati
ons"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <
b><a href="datasets/Labor+Relations">Labor Relations</a> </td></tr></table></td>\n\t\t\t<!-- <td><p cla
ss="normal">From Collective Bargaining Review </td> -->\n\t\t\t<td> Multivariate&
nbsp; </td>\n\t\t\t<td>   </td>\n\t\t\t<td> Categorical, Integer,
Real </td>\n\t\t\t<td> 57 </td>\n\t\t\t<td> 16 </p
></td>\n\t\t\t<td> 1988 </td>\n\t\t\t<!-- <td> Social </t
d> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/LED+Display+Domain"><i
mg src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a hr
ef="datasets/LED+Display+Domain">LED Display Domain</a> </td></tr></table></td>\n\t\t\t<!-- <td><p
class="normal">From Classification and Regression Trees book; We provide here 2 C programs for generating
sample databases </td> -->\n\t\t\t<td> Multivariate, Data-Generator </td
>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical </p
></td>\n\t\t\t<td>   </td>\n\t\t\t<td> 7 </td>\n\t\t\t<td><
p class="normal">1988 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</t
r><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Lenses"><img src="assets/MLimages/SmallLargedefault.jpg"
border=1 /></a> </td><td> <a href="datasets/Lenses">Lenses</a> </td></t
r></table></td>\n\t\t\t<!-- <td> Database for fitting contact lenses </td> -->\n\t\t\t<td
> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\
t\t<td> Categorical </td>\n\t\t\t<td> 24 </td>\n\t\t\t<td
> 4 </td>\n\t\t\t<td> 1990 </td>\n\t\t\t<!-- <td><p clas
s="normal">Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="data
sets/Letter+Recognition"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td>
<a href="datasets/Letter+Recognition">Letter Recognition</a> </td></tr></table
></td>\n\t\t\t<!-- <td> Database of character image features; try to identify the letter </p
></td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification&
nbsp; </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 20000
</td>\n\t\t\t<td> 16 </td>\n\t\t\t<td> 1991 </td>\n\
t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="d
atasets/Liver+Disorders"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td>
<a href="datasets/Liver+Disorders">Liver Disorders</a> </td></tr></table></td>\
n\t\t\t<!-- <td> BUPA Medical Research Ltd. database donated by Richard S. Forsyth <
/p></td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td>   <
/td>\n\t\t\t<td> Categorical, Integer, Real </td>\n\t\t\t<td> 345&n
bsp; </td>\n\t\t\t<td> 7 </td>\n\t\t\t<td> 1990 </td
>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><
tr><td><a href="datasets/Logic+Theorist"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>
 </td><td> <a href="datasets/Logic+Theorist">Logic Theorist</a> </td></tr
></table></td>\n\t\t\t<!-- <td> All code for Logic Theorist </td> -->\n\t\t\t<td><p clas
s="normal">Domain-Theory </td>\n\t\t\t<td>   </td>\n\t\t\t<td><p class="
normal"> </td>\n\t\t\t<td>   </td>\n\t\t\t<td> </p
></td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Computer </td
> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Lung+Cancer"><img src="assets/MLimages/Sm
allLarge62.jpg" border=1 /></a> </td><td> <a href="datasets/Lung+Cancer">Lung
Cancer</a> </td></tr></table></td>\n\t\t\t<!-- <td> Lung cancer data; no attribute defi
nitions </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td><p class="normal
">Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td><p class="norma
l">32 </td>\n\t\t\t<td> 56 </td>\n\t\t\t<td> 1992
</td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td>
<table><tr><td><a href="datasets/Lymphography"><img src="assets/MLimages/SmallLargedefault.jpg" border=
1 /></a> </td><td> <a href="datasets/Lymphography">Lymphography</a>
</td></tr></table></td>\n\t\t\t<!-- <td> This lymphography domain was obtained from the Univ
ersity Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia. (Restricted access) </td> -->\n\t\t
\t<td> Multivariate </td>\n\t\t\t<td> Classification </td
>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 148 </td>\n\t
\t\t<td> 18 </td>\n\t\t\t<td> 1988 </td>\n\t\t\t<!-- <td>

Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Mechanical+Analysis"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Mechanical+Analysis">Mechanical Analysis</a> </td></tr></table></td>\n\t\t\t<!-- <td> Fault diagnosis problem of electromechanical devices; also PUMPS DATA SET is newer version with domain theory and results </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer, Real </td>\n\t\t\t<td> 209 </td>\n\t\t\t<td> 8 </td>\n\t\t\t<td> 1990 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Meta-data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Meta-data">Meta-data</a> </td></tr></table></td>\n\t\t\t<!-- <td> Meta-Data was used in order to give advice about which classification method is appropriate for a particular dataset (taken from results of Statlog project). </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer, Real </td>\n\t\t\t<td> 528 </td>\n\t\t\t<td> 22 </td>\n\t\t\t<td> 1996 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Mobile+Robots"><img src="assets/MLimages/SmallLarge66.jpg" border=1 /></a> </td><td> <a href="datasets/Mobile+Robots">Mobile Robots</a> </td></tr></table></td>\n\t\t\t<!-- <td> Learning concepts from sensor data of a mobile robot; set of data sets </td> -->\n\t\t\t<td> Domain-Theory </td>\n\t\t\t<td>   </td>\n\t\t\t<td> Categorical, Integer, Real </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 1995 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29"><img src="assets/MLimages/SmallLarge67.jpg" border=1 /></a> </td><td> <a href="datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29">Molecular Biology (Promoter Gene Sequences)</a> </td></tr></table></td>\n\t\t\t<!-- <td> E. Coli promoter gene sequences (DNA) with partial domain theory </td> -->\n\t\t\t<td> Sequential, Domain-Theory </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 106 </td>\n\t\t\t<td> 58 </td>\n\t\t\t<td> 1990 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%29"><img src="assets/MLimages/SmallLarge67.jpg" border=1 /></a> </td><td> <a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular Biology (Protein Secondary Structure)</a> </td></tr></table></td>\n\t\t\t<!-- <td> From CMU connectionist bench repository; Classifies secondary structure of certain globular proteins </td> -->\n\t\t\t<td> Sequential </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 128 </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29"><img src="assets/MLimages/SmallLarge67.jpg" border=1 /></a> </td><td> <a href="datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29">Molecular Biology (Splice-junction Gene Sequences)</a> </td></tr></table></td>\n\t\t\t<!-- <td> Primate splice-junction gene sequences (DNA) with associated imperfect domain theory </td> -->\n\t\t\t<td> Sequential, Domain-Theory </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 3190 </td>\n\t\t\t<td> 61 </td>\n\t\t\t<td> 1992 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/MONK%27s+Problems"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/MONK%27s+Problems">MONK\'s Problems</a> </td></tr></table></td>\n\t\t\t<!-- <td> A set of three artificial domains over the same attribute space; Used to test a wide range of induction algorithms </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 432 </td>\n\t\t\t<td> 7 </td>\n\t\t\t<td> 1992 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Moral+Reasoner"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Moral+Reasoner">Moral Reasoner</a> </td></tr></table></td>\n\t\t\t<!-- <td> Horn-clause model that qualitatively simulates moral reasoning; Theory includes negated literals </td> -->\n\t\t\t<td> Domain-Theory </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 202 </td>\n\t\t\t<td>   </td>\n\t\t\t<td><p

class="normal">1994 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Multiple+Features"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Multiple+Features">Multiple Features</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset consists of features of handwritten numerals (`0\'--`9\') extracted from a collection of Dutch utility maps </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 2000 </td>\n\t\t\t<td> 649 </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Mushroom"><img src="assets/MLimages/SmallLarge73.jpg" border=1 /></a> </td><td> <a href="datasets/Mushroom">Mushroom</a> </td></tr></table></td>\n\t\t\t<!-- <td> From Audobon Society Field Guide; mushrooms described in terms of physical characteristics; classification: poisonous or edible </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 8124 </td>\n\t\t\t<td> 22 </td>\n\t\t\t<td> 1987 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Musk+%28Version+1%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Musk+%28Version+1%29">Musk (Version 1)</a> </td></tr></table></td>\n\t\t\t<!-- <td> The goal is to learn to predict whether new molecules will be musks or non-musks </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 476 </td>\n\t\t\t<td> 168 </td>\n\t\t\t<td> 1994 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Musk+%28Version+2%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Musk+%28Version+2%29">Musk (Version 2)</a> </td></tr></table></td>\n\t\t\t<!-- <td> The goal is to learn to predict whether new molecules will be musks or non-musks </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 6598 </td>\n\t\t\t<td> 168 </td>\n\t\t\t<td> 1994 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Nursery"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Nursery">Nursery</a> </td></tr></table></td>\n\t\t\t<!-- <td> Nursery Database was derived from a hierarchical decision model originally developed to rank applications for nursery schools. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 12960 </td>\n\t\t\t<td> 8 </td>\n\t\t\t<td> 1997 </td>\n\t\t\t<!-- <td> Social </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Othello+Domain+Theory"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Othello+Domain+Theory">Othello Domain Theory</a> </td></tr></table></td>\n\t\t\t<!-- <td> Used in research to generate features for an inductive learning system </td> -->\n\t\t\t<td> Domain-Theory </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 1991 </td>\n\t\t\t<!-- <td> Game </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Page+Blocks+Classification"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Page+Blocks+Classification">Page Blocks Classification</a> </td></tr></table></td>\n\t\t\t<!-- <td> The problem consists of classifying all the blocks of the page layout of a document that has been detected by a segmentation process. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 5473 </td>\n\t\t\t<td> 10 </td>\n\t\t\t<td> 1995 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Optical+Recognition+of+Handwritten+Digits"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Optical+Recognition+of+Handwritten+Digits">Optical Recognition of Handwritten Digits</a> </td></tr></table></td>\n\t\t\t<!-- <td> Two versions of this database available; see folder </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 5620 </td>\n\t\t\t<td> 64 </td>\n\t\t\t<td> 1998 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits"><img src="assets/M

Limages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits">Pen-Based Recognition of Handwritten Digits</a> </td></tr></table></td>\n\t\t\t<!-- <td> Digit database of 250 samples from 44 writers </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 10992 </td>\n\t\t\t<td> 16 </td>\n\t\t\t<td> 1998 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Post-Operative+Patient"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Post-Operative+Patient">Post-Operative Patient</a> </td></tr></table></td>\n\t\t\t<!-- <td> Dataset of patient features </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<td> 90 </td>\n\t\t\t<td> 8 </td>\n\t\t\t<td> 1993 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Primary+Tumor"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Primary+Tumor">Primary Tumor</a> </td></tr></table></td>\n\t\t\t<!-- <td> From Ljubljana Oncology Institute </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 339 </td>\n\t\t\t<td> 17 </td>\n\t\t\t<td> 1988 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Prodigy"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Prodigy">Prodigy</a> </td></tr></table></td>\n\t\t\t<!-- <td> Assorted domains like blocksworld, eightpuzzle, and schedworld. </td> -->\n\t\t\t<td> Domain-Theory </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Qualitative+Structure+Activity+Relationships"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Qualitative+Structure+Activity+Relationships">Qualitative Structure Activity Relationships</a> </td></tr></table></td>\n\t\t\t<!-- <td> Two sets of datasets are given: pyrimidines and triazines </td> -->\n\t\t\t<td> Domain-Theory </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Quadruped+Mammals"><img src="assets/MLimages/SmallLarge86.jpg" border=1 /></a> </td><td> <a href="datasets/Quadruped+Mammals">Quadruped Mammals</a> </td></tr></table></td>\n\t\t\t<!-- <td> The file animals.c is a data generator of structured instances representing quadruped animals </td> -->\n\t\t\t<td> Multivariate, Data-Generator </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 72 </td>\n\t\t\t<td> 1992 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Servo"><img src="assets/MLimages/SmallLarge87.jpg" border=1 /></a> </td><td> <a href="datasets/Servo">Servo</a> </td></tr></table></td>\n\t\t\t<!-- <td> Data was from a simulation of a servo system </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<td> 167 </td>\n\t\t\t<td> 4 </td>\n\t\t\t<td> 1993 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Shuttle+Landing+Control"><img src="assets/MLimages/SmallLarge92.jpg" border=1 /></a> </td><td> <a href="datasets/Shuttle+Landing+Control">Shuttle Landing Control</a> </td></tr></table></td>\n\t\t\t<!-- <td> Tiny database; all nominal values </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 15 </td>\n\t\t\t<td> 6 </td>\n\t\t\t<td> 1988 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Solar+Flare"><img src="assets/MLimages/SmallLarge89.jpg" border=1 /></a> </td><td> <a href="datasets/Solar+Flare">Solar Flare</a> </td></tr></table></td>\n\t\t\t<!-- <td> Each class attribute counts the number of solar flares of a certain class that occur in a 24 hour period </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 1389 </td>\n\t\t\t<td> 10 </td>\n\t\t\t<td> 1989 </td>\n\t\t\t<!-

- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Soybean+%28Large%29"><img src="assets/MLimages/SmallLarge90.jpg" border=1 /></a> </td><td> <a href="datasets/Soybean+%28Large%29">Soybean (Large)</a> </td></tr></table></td>\n\t\t\t<!-- <td> Michalski\'s famous soybean disease database </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 307 </td>\n\t\t\t<td> 35 </td>\n\t\t\t<td> 1988 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Soybean+%28Small%29"><img src="assets/MLimages/SmallLarge90.jpg" border=1 /></a> </td><td> <a href="datasets/Soybean+%28Small%29">Soybean (Small)</a> </td></tr></table></td>\n\t\t\t<!-- <td> Michalski\'s famous soybean disease database </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 47 </td>\n\t\t\t<td> 35 </td>\n\t\t\t<td> 1987 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Challenger+USA+Space+Shuttle+O-Ring"><img src="assets/MLimages/SmallLarge92.jpg" border=1 /></a> </td><td> <a href="datasets/Challenger+USA+Space+Shuttle+O-Ring">Challenger USA Space Shuttle O-Ring</a> </td></tr></table></td>\n\t\t\t<!-- <td> Task: predict the number of O-rings that experience thermal distress on a flight at 31 degrees F given data on the previous 23 shuttle flights </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 23 </td>\n\t\t\t<td> 4 </td>\n\t\t\t<td> 1993 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Low+Resolution+Spectrometer"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Low+Resolution+Spectrometer">Low Resolution Spectrometer</a> </td></tr></table></td>\n\t\t\t<!-- <td> From IRAS data -- NASA Ames Research Center </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 531 </td>\n\t\t\t<td> 102 </td>\n\t\t\t<td> 1988 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Spambase"><img src="assets/MLimages/SmallLarge94.jpg" border=1 /></a> </td><td> <a href="datasets/Spambase">Spambase</a> </td></tr></table></td>\n\t\t\t<!-- <td> Classifying Email as Spam or Non-Spam </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 4601 </td>\n\t\t\t<td> 57 </td>\n\t\t\t<td> 1999 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/SPECT+Heart"><img src="assets/MLimages/SmallLarge45.jpg" border=1 /></a> </td><td> <a href="datasets/SPECT+Heart">SPECT Heart</a> </td></tr></table></td>\n\t\t\t<!-- <td> Data on cardiac Single Proton Emission Computed Tomography (SPECT) images. Each patient classified into two categories: normal and abnormal. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 267 </td>\n\t\t\t<td> 22 </td>\n\t\t\t<td> 2001 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/SPECTF+Heart"><img src="assets/MLimages/SmallLarge45.jpg" border=1 /></a> </td><td> <a href="datasets/SPECTF+Heart">SPECTF Heart</a> </td></tr></table></td>\n\t\t\t<!-- <td> Data on cardiac Single Proton Emission Computed Tomography (SPECT) images. Each patient classified into two categories: normal and abnormal. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 267 </td>\n\t\t\t<td> 44 </td>\n\t\t\t<td> 2001 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Sponge"><img src="assets/MLimages/SmallLarge97.jpg" border=1 /></a> </td><td> <a href="datasets/Sponge">Sponge</a> </td></tr></table></td>\n\t\t\t<!-- <td> Data on sponges; Attributes in spanish </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Clustering </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<td> 76 </td>\n\t\t\t<td> 45 </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Statlog+Project"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Statlog+Project">Statlog Project</a> </td></tr></table></td>\n\t\t\t<!-- <td> Various Databases: Vehicle silhouttes, Lan-

dsat Sattelite, Shuttle, Australian Credit Approval, Heart Disease, Image Segmentation, German Credit </td> -->\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 1992 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Student+Loan+Relational"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Student+Loan+Relational">Student Loan Relational</a> </td></tr></table></td>\n\t\t\t<!-- <td> Student Loan Relational Domain </td> -->\n\t\t\t<td> Domain-Theory </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 1000 </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 1993 </td>\n\t\t\t<!-- <td> Social </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Teaching+Assistant+Evaluation"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Teaching+Assistant+Evaluation">Teaching Assistant Evaluation</a> </td></tr></table></td>\n\t\t\t<!-- <td> The data consist of evaluations of teaching performance; scores are "low", "medium", or "high" </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<td> 151 </td>\n\t\t\t<td> 5 </td>\n\t\t\t<td> 1997 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Tic-Tac-Toe+Endgame"><img src="assets/MLimages/SmallLarge101.jpg" border=1 /></a> </td><td> <a href="datasets/Tic-Tac-Toe+Endgame">Tic-Tac-Toe Endgame</a> </td></tr></table></td>\n\t\t\t<!-- <td> Binary classification task on possible configurations of tic-tac-toe game </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 958 </td>\n\t\t\t<td> 9 </td>\n\t\t\t<td> 1991 </td>\n\t\t\t<!-- <td> Game </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Thyroid+Disease"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Thyroid+Disease">Thyroid Disease</a> </td></tr></table></td>\n\t\t\t<!-- <td> 10 separate databases from Garavan Institute </td> -->\n\t\t\t<td> Multivariate, Domain-Theory </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Real </td>\n\t\t\t<td> 7200 </td>\n\t\t\t<td> 21 </td>\n\t\t\t<td> 1987 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Trains"><img src="assets/MLimages/SmallLarge103.jpg" border=1 /></a> </td><td> <a href="datasets/Trains">Trains</a> </td></tr></table></td>\n\t\t\t<!-- <td> 2 data formats (structured, one-instance-per-line) </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 10 </td>\n\t\t\t<td> 32 </td>\n\t\t\t<td> 1994 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/University"><img src="assets/MLimages/SmallLarge104.jpg" border=1 /></a> </td><td> <a href="datasets/University">University</a> </td></tr></table></td>\n\t\t\t<!-- <td> Data in original (LISP-readable) form </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<td> 285 </td>\n\t\t\t<td> 17 </td>\n\t\t\t<td> 1988 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Congressional+Voting+Records"><img src="assets/MLimages/SmallLarge105.jpg" border=1 /></a> </td><td> <a href="datasets/Congressional+Voting+Records">Congressional Voting Records</a> </td></tr></table></td>\n\t\t\t<!-- <td> 1984 United Stated Congressional Voting Records; Classify as Republican or Democrat </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 435 </td>\n\t\t\t<td> 16 </td>\n\t\t\t<td> 1987 </td>\n\t\t\t<!-- <td> Social </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Water+Treatment+Plant"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Water+Treatment+Plant">Water Treatment Plant</a> </td></tr></table></td>\n\t\t\t<!-- <td> Multiple classes predict plant state </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Clustering </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 527 </td>\n\t\t\t<td> 38 </td>\n\t\t\t<td> 1993 </td>\n\t\t\t<!-- <td> Physical </td>

> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Waveform+Database+Generator+%28Version+1%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Waveform+Database+Generator+%28Version+1%29">Waveform Database Generator (Version 1)</a> </td></tr></table></td>\n\t\t\t<!-- <td> CART book\'s waveform domains </td> -->\n\t\t\t<td> Multivariate, Data-Generator </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 5000 </td>\n\t\t\t<td> 21 </td>\n\t\t\t<td> 1988 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Waveform+Database+Generator+%28Version+2%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Waveform+Database+Generator+%28Version+2%29">Waveform Database Generator (Version 2)</a> </td></tr></table></td>\n\t\t\t<!-- <td> CART book\'s waveform domains </td> -->\n\t\t\t<td> Multivariate, Data-Generator </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 5000 </td>\n\t\t\t<td> 40 </td>\n\t\t\t<td> 1988 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Wine"><img src="assets/MLimages/SmallLarge109.jpg" border=1 /></a> </td><td> <a href="datasets/Wine">Wine</a> </td></tr></table></td>\n\t\t\t<!-- <td> Using chemical analysis determine the origin of wines </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 178 </td>\n\t\t\t<td> 13 </td>\n\t\t\t<td> 1991 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Yeast"><img src="assets/MLimages/SmallLarge110.jpg" border=1 /></a> </td><td> <a href="datasets/Yeast">Yeast</a> </td></tr></table></td>\n\t\t\t<!-- <td> Predicting the Cellular Localization Sites of Proteins </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 1484 </td>\n\t\t\t<td> 8 </td>\n\t\t\t<td> 1996 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Zoo"><img src="assets/MLimages/SmallLarge111.jpg" border=1 /></a> </td><td> <a href="datasets/Zoo">Zoo</a> </td></tr></table></td>\n\t\t\t<!-- <td> Artificial, 7 classes of animals </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<td> 101 </td>\n\t\t\t<td> 17 </td>\n\t\t\t<td> 1990 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Undocumented"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Undocumented">Undocumented</a> </td></tr></table></td>\n\t\t\t<!-- <td> Various datasets without documentation (feel free to explore!) </td> -->\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Twenty+Newsgroups"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Twenty+Newsgroups">Twenty Newsgroups</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data set consists of 20000 messages taken from 20 newsgroups. </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 20000 </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 1999 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Australian+Sign+Language+signs"><img src="assets/MLimages/SmallLarge114.jpg" border=1 /></a> </td><td> <a href="datasets/Australian+Sign+Language+signs">Australian Sign Language signs</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data consists of sample of Auslan (Australian Sign Language) signs. Examples of 95 signs were collected from five signers with a total of 6650 sign samples. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Real </td>\n\t\t\t<td> 6650 </td>\n\t\t\t<td> 15 </td>\n\t\t\t<td> 1999 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Australian+Sign+Language+signs+%28High+Quality%29"><img src="assets/MLimages/SmallLarge114.jpg" border=1 /></a> </td><td> <a href="datasets/Australian+Sign+Language+signs+%28High+Quality%29">Australian Sign Language signs (High Quality)</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data consists of sample of Auslan (Australian Sign Language) signs. 27 examples of each of 95

Auslan signs were captured from a native signer using high-quality position trackers </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 2565 </td>\n\t\t\t<td> 22 </td>\n\t\t\t<td> 2002 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/US+Census+Data+%281990%29"><img src="assets/MLimages/SmallLarge2.jpg" border=1 /></a> </td><td> <a href="datasets/US+Census+Data+%281990%29">US Census Data (1990)</a> </td></tr></table></td>\n\t\t\t<!-- <td> The USCensus1990raw data set contains a one percent sample of the Public Use Microdata Samples (PUMS) person records drawn from the full 1990 census sample. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Clustering </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 2458285 </td>\n\t\t\t<td> 68 </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Social </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Census-Income+%28KDD%29"><img src="assets/MLimages/SmallLarge2.jpg" border=1 /></a> </td><td> <a href="datasets/Census-Income+%28KDD%29">Census-Income (KDD)</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data set contains weighted census data extracted from the 1994 and 1995 current population surveys conducted by the U.S. Census Bureau. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<td> 299285 </td>\n\t\t\t<td> 40 </td>\n\t\t\t<td> 2000 </td>\n\t\t\t<!-- <td> Social </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Coil+1999+Competition+Data"><img src="assets/MLimages/SmallLarge118.jpg" border=1 /></a> </td><td> <a href="datasets/Coil+1999+Competition+Data">Coil 1999 Competition Data</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data set is from the 1999 Computational Intelligence and Learning (COIL) competition. The data contains measurements of river chemical concentrations and algae densities. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td>   </td>\n\t\t\t<td> Categorical, Real </td>\n\t\t\t<td> 340 </td>\n\t\t\t<td> 17 </td>\n\t\t\t<td> 1999 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Corel+Image+Features"><img src="assets/MLimages/SmallLarge119.jpg" border=1 /></a> </td><td> <a href="datasets/Corel+Image+Features">Corel Image Features</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset contains image features extracted from a Corel image collection. Four sets of features are available based on the color histogram, color histogram layout, color moments, and co-occurence </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td>   </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 68040 </td>\n\t\t\t<td> 89 </td>\n\t\t\t<td> 1999 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/E.+Coli+Genes"><img src="assets/MLimages/SmallLarge120.jpg" border=1 /></a> </td><td> <a href="datasets/E.+Coli+Genes">E. Coli Genes</a> </td></tr></table></td>\n\t\t\t<!-- <td> Data giving characteristics of each ORF (potential gene) in the E. coli genome. Sequence, homology (similarity to other genes) and structural information, and function (if known) are provided. </td> -->\n\t\t\t<td> Relational </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2001 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/EEG+Database"><img src="assets/MLimages/SmallLarge121.jpg" border=1 /></a> </td><td> <a href="datasets/EEG+Database">EEG Database</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data arises from a large study to examine EEG correlates of genetic predisposition to alcoholism. It contains measurements from 64 electrodes placed on the scalp sampled at 256 Hz </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td>   </td>\n\t\t\t<td> Categorical, Integer, Real </td>\n\t\t\t<td> 122 </td>\n\t\t\t<td> 4 </td>\n\t\t\t<td> 1999 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/El+Nino"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/El+Nino">El Nino</a> </td></tr></table></td>\n\t\t\t<!-- <td> The data set contains oceanographic and surface meteorological readings taken from a series of buoys positioned throughout the equatorial Pacific. </td> -->\n\t\t\t<td> Spatio-temporal </td>\n\t\t\t<td>   </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 178080 </td>\n\t\t\t<td> 12 </td>\n\t\t\t<td> 1999 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Entree+Chicago+Recom

mendation+Data"><img src="assets/MLimages/SmallLarge123.jpg" border=1 /></a> </td><td> <a href="datasets/Entree+Chicago+Recommendation+Data">Entree Chicago Recommendation Data</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data contains a record of user interactions with the Entree Chicago restaurant recommendation system. </td> -->\n\t\t\t<td> Transactional, Sequential </td>\n\t\t\t<td> Recommender-Systems </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 50672 </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2000 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/CMU+Face+Images"><img src="assets/MLimages/SmallLarge124.jpg" border=1 /></a> </td><td> <a href="datasets/CMU+Face+Images">CMU Face Images</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data consists of 640 black and white face images of people taken with varying pose (straight, left, right, up), expression (neutral, happy, sad, angry), eyes (wearing sunglasses or not), and size </td> -->\n\t\t\t<td> Image </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 640 </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 1999 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Insurance+Company+Benchmark+%28COIL+2000%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Insurance+Company+Benchmark+%28COIL+2000%29">Insurance Company Benchmark (COIL 2000)</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data set used in the CoIL 2000 Challenge contains information on customers of an insurance company. The data consists of 86 variables and includes product usage data and socio-demographic data </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression, Description </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<td> 9000 </td>\n\t\t\t<td> 86 </td>\n\t\t\t<td> 2000 </td>\n\t\t\t<!-- <td> Social </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Internet+Usage+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Internet+Usage+Data">Internet Usage Data</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data contains general demographic information on internet users in 1997. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td>   </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<td> 10104 </td>\n\t\t\t<td> 72 </td>\n\t\t\t<td> 1999 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/IPUMS+Census+Database"><img src="assets/MLimages/SmallLarge2.jpg" border=1 /></a> </td><td> <a href="datasets/IPUMS+Census+Database">IPUMS Census Database</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data set contains unweighted PUMS census data from the Los Angeles and Long Beach areas for the years 1970, 1980, and 1990. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td>   </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<td> 256932 </td>\n\t\t\t<td> 61 </td>\n\t\t\t<td> 1999 </td>\n\t\t\t<!-- <td> Social </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Japanese+Vowels"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Japanese+Vowels">Japanese Vowels</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset records 640 time series of 12 LPC cepstrum coefficients taken from nine male speakers. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 640 </td>\n\t\t\t<td> 12 </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/KDD+Cup+1998+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/KDD+Cup+1998+Data">KDD Cup 1998 Data</a> </td></tr></table></td>\n\t\t\t<!-- <td> This is the data set used for The Second International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-98 </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<td> 191779 </td>\n\t\t\t<td> 481 </td>\n\t\t\t<td> 1998 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/KDD+Cup+1999+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/KDD+Cup+1999+Data">KDD Cup 1999 Data</a> </td></tr></table></td>\n\t\t\t<!-- <td> This is the data set used for The Third International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-99 </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical,

Integer </td>\n\t\t\t<td> 4000000 </td>\n\t\t\t<td> 42 </td>\n\t\t\t<td> 1999 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/M.+Tuberculosis+Genes"><img src="assets/MLimages/SmallLarge131.jpg" border=1 /></a> </td><td> <a href="datasets/M.+Tuberculosis+Genes">M. Tuberculosis Genes</a> </td></tr></table></td>\n\t\t\t<!-- <td> Data giving characteristics of each ORF (potential gene) in the M. tuberculosis bacterium. Sequence, homology (similarity to other genes) and structural information, and function (if known) are provided </td> -->\n\t\t\t<td> Relational </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2001 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Movie"><img src="assets/MLimages/SmallLarge132.jpg" border=1 /></a> </td><td> <a href="datasets/Movie">Movie</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data set contains a list of over 10000 films including many older, odd, and cult films. There is information on actors, casts, directors, producers, studios, etc. </td> -->\n\t\t\t<td> Multivariate, Relational </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 10000 </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 1999 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/MSNBC.com+Anonymous+Web+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/MSNBC.com+Anonymous+Web+Data">MSNBC.com Anonymous Web Data</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data describes the page visits of users who visited msnbc.com on September 28, 1999. Visits are recorded at the level of URL category (see description) and are recorded in time order. </td> -->\n\t\t\t<td> Sequential </td>\n\t\t\t<td>   </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 989818 </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/NSF+Research+Award+Abstracts+1990-2003"><img src="assets/MLimages/SmallLarge134.jpg" border=1 /></a> </td><td> <a href="datasets/NSF+Research+Award+Abstracts+1990-2003">NSF Research Award Abstracts 1990-2003</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data set consists of (a) 129,000 abstracts describing NSF awards for basic research, (b) bag-of-word data files extracted from the abstracts, (c) a list of words used for indexing the bag-of-word </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 129000 </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2003 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Pioneer-1+Mobile+Robot+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Pioneer-1+Mobile+Robot+Data">Pioneer-1 Mobile Robot Data</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset contains time series sensor readings of the Pioneer-1 mobile robot. The data is broken into "experiences" in which the robot takes action for some period of time and experiences a control </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td>   </td>\n\t\t\t<td> Categorical, Real </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 1999 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Pseudo+Periodic+Synthetic+Time+Series"><img src="assets/MLimages/SmallLarge136.jpg" border=1 /></a> </td><td> <a href="datasets/Pseudo+Periodic+Synthetic+Time+Series">Pseudo Periodic Synthetic Time Series</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data set is designed for testing indexing schemes in time series databases. The data appears highly periodic, but never exactly repeats itself. </td> -->\n\t\t\t<td> Univariate, Time-Series </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 100000 </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 1999 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Reuters-21578+Text+Categorization+Collection"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Reuters-21578+Text+Categorization+Collection">Reuters-21578 Text Categorization Collection</a> </td></tr></table></td>\n\t\t\t<!-- <td> This is a collection of documents that appeared on Reuters newswire in 1987. The documents were assembled and indexed with categories. </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 21578 </td>\n\t\t\t<td> 5 </td>\n\t\t\t<td> 1997 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Robot+Execution+Failures"><img src="assets/MLimages/SmallLarge138.j

pg" border=1 /></a> </td><td> <a href="datasets/Robot+Execution+Failures">Rob
ot Execution Failures</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset contain
s force and torque measurements on a robot after failure detection. Each failure is characterized by 15 force/tor
que samples collected at regular time intervals </td> -->\n\t\t\t<td> Multivariate, Ti
me-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td><p class="norm
al">Integer </td>\n\t\t\t<td> 463 </td>\n\t\t\t<td> 90&
nbsp; </td>\n\t\t\t<td> 1999 </td>\n\t\t\t<!-- <td> Physical&n
bsp; </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Synthetic+Control+Chart+Time+Ser
ies"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <b
><a href="datasets/Synthetic+Control+Chart+Time+Series">Synthetic Control Chart Time Series</a> </
td></tr></table></td>\n\t\t\t<!-- <td> This data consists of synthetically generated control char
ts. </td> -->\n\t\t\t<td> Time-Series </td>\n\t\t\t<td> C
lassification, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td><p class="n
ormal">600 </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 1999&n
bsp; </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t
\t<td><table><tr><td><a href="datasets/Syskill+and+Webert+Web+Page+Ratings"><img src="assets/MLimages
/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Syskill+and+
Webert+Web+Page+Ratings">Syskill and Webert Web Page Ratings</a> </td></tr></table></td>\n\t\t\t<
!-- <td> This database contains HTML source of web pages plus the ratings of a single user
on these web pages. Web pages are on four seperate subjects (Bands- recording artists; Goats; Sheep; and Bi
oMedical) </td> -->\n\t\t\t<td> Multivariate, Text </td>\n\t\t\t<td><p class
="normal">Classification </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td><p
class="normal">332 </td>\n\t\t\t<td> 5 </td>\n\t\t\t<td><p class="normal
">1998 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td>
<table><tr><td><a href="datasets/UNIX+User+Data"><img src="assets/MLimages/SmallLarge141.jpg" border=
1 /></a> </td><td> <a href="datasets/UNIX+User+Data">UNIX User Data</a><
/p></td></tr></table></td>\n\t\t\t<!-- <td> This file contains 9 sets of sanitized user data draw
n from the command histories of 8 UNIX computer users at Purdue over the course of up to 2 years.
</td> -->\n\t\t\t<td> Text, Sequential </td>\n\t\t\t<td>
</td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td><p c
lass="normal"> </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td>
Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Vol
canoes+on+Venus+-+JARtool+experiment"><img src="assets/MLimages/SmallLarge142.jpg" border=1 /></a>&
nbsp;</td><td> <a href="datasets/Volcanoes+on+Venus+-+JARtool+experiment">Volcan
oes on Venus - JARtool experiment</a> </td></tr></table></td>\n\t\t\t<!-- <td> The J
ARtool project was a pioneering effort to develop an automatic system for cataloging small volcanoes in the larg
e set of Venus images returned by the Magellan spacecraft. </td> -->\n\t\t\t<td> I
mage </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td>
  </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t
\t\t<td>   </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t<
/tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Statlog+%28Australian+Credit+Approval%29"><img src="as
sets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="dataset
s/Statlog+%28Australian+Credit+Approval%29">Statlog (Australian Credit Approval)</a> </td></tr></tab
le></td>\n\t\t\t<!-- <td> This file concerns credit card applications. This database exists elsew
here in the repository (Credit Screening Database) in a slightly different form </td> -->\n\t\t\t<td><p cl
ass="normal">Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td
> Categorical, Integer, Real </td>\n\t\t\t<td> 690 </td
>\n\t\t\t<td> 14 </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td><
p class="normal">Financial </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a h
ref="datasets/Statlog+%28German+Credit+Data%29"><img src="assets/MLimages/SmallLargedefault.jpg" bord
er=1 /></a> </td><td> <a href="datasets/Statlog+%28German+Credit+Data%29">S
tatlog (German Credit Data)</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset
classifies people described by a set of attributes as good or bad credit risks. Comes in two formats (one all num
eric). Also comes with a cost matrix </td> -->\n\t\t\t<td> Multivariate </td
>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer&
nbsp; </td>\n\t\t\t<td> 1000 </td>\n\t\t\t<td> 20 </
td>\n\t\t\t<td> 1994 </td>\n\t\t\t<!-- <td> Financial </t
d> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Statlog+%28Heart%29"><img src="assets/MLi
mages/SmallLarge45.jpg" border=1 /></a> </td><td> <a href="datasets/Statlog+%
28Heart%29">Statlog (Heart)</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset
is a heart disease database similar to a database already present in the repository (Heart Disease databases)
but in a slightly different form </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\

t<td> Classification </td>\n\t\t\t<td> Categorical, Real </td>\n\t\t\t<td> 270 </td>\n\t\t\t<td> 13 </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Statlog+%28Landsat+Satellite%29"><img src="assets/MLimages/SmallLarge146.jpg" border=1 /></a> </td><td> <a href="datasets/Statlog+%28Landsat+Satellite%29">Statlog (Landsat Satellite)</a> </td></tr></table></td>\n\t\t\t<!-- <td> Multi-spectral values of pixels in 3x3 neighbourhoods in a satellite image, and the classification associated with the central pixel in each neighbourhood </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 6435 </td>\n\t\t\t<td> 36 </td>\n\t\t\t<td> 1993 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Statlog+%28Image+Segmentation%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Statlog+%28Image+Segmentation%29">Statlog (Image Segmentation)</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset is an image segmentation database similar to a database already present in the repository (Image segmentation database) but in a slightly different form. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 2310 </td>\n\t\t\t<td> 19 </td>\n\t\t\t<td> 1990 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Statlog+%28Shuttle%29"><img src="assets/MLimages/SmallLarge92.jpg" border=1 /></a> </td><td> <a href="datasets/Statlog+%28Shuttle%29">Statlog (Shuttle)</a> </td></tr></table></td>\n\t\t\t<!-- <td> The shuttle dataset contains 9 attributes all of which are numerical. Approximately 80% of the data belongs to class 1 </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 58000 </td>\n\t\t\t<td> 9 </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Statlog+%28Vehicle+Silhouettes%29"><img src="assets/MLimages/SmallLarge149.jpg" border=1 /></a> </td><td> <a href="datasets/Statlog+%28Vehicle+Silhouettes%29">Statlog (Vehicle Silhouettes)</a> </td></tr></table></td>\n\t\t\t<!-- <td> 3D objects within a 2D image by application of an ensemble of shape feature extractors to the 2D silhouettes of the objects. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 946 </td>\n\t\t\t<td> 18 </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Connectionist+Bench+%28Nettalk+Corpus%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Connectionist+Bench+%28Nettalk+Corpus%29">Connectionist Bench (Nettalk Corpus)</a> </td></tr></table></td>\n\t\t\t<!-- <td> The file "nettalk.data" contains a list of 20,008 English words, along with a phonetic transcription for each word. The task is to train a network to produce the proper phonemes </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td>   </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 20008 </td>\n\t\t\t<td> 4 </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29">Connectionist Bench (Sonar, Mines vs. Rocks)</a> </td></tr></table></td>\n\t\t\t<!-- <td> The task is to train a network to discriminate between sonar signals bounced off a metal cylinder and those bounced off a roughly cylindrical rock. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 208 </td>\n\t\t\t<td> 60 </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29">Connectionist Bench (Vowel Recognition - Deterding Data)</a> </td></tr></table></td>\n\t\t\t<!-- <td> Speaker independent recognition of the eleven steady state vowels of British English using a specified training set of lpc derived log area ratios. </td> -->\n\t\t\t<td>   </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 528 </td>\n\t\t\t<td> 10 </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Econ

omic+Sanctions"><img src="assets/MLimages/SmallLarge153.jpg" border=1 /></a> </td><td> <a href="datasets/Economic+Sanctions">Economic Sanctions</a> </td></tr></table></td>\n\t\t\t<!-- <td> Domain Theory on Economic Sanctions; Undocumented </td> -->\n\t\t\t<td> Domain-Theory </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Financial </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Protein+Data"><img src="assets/MLimages/SmallLarge154.jpg" border=1 /></a> </td><td> <a href="datasets/Protein+Data">Protein Data</a> </td></tr></table></td>\n\t\t\t<!-- <td> Undocumented </td> -->\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Cloud"><img src="assets/MLimages/SmallLarge155.jpg" border=1 /></a> </td><td> <a href="datasets/Cloud">Cloud</a> </td></tr></table></td>\n\t\t\t<!-- <td> Little Documentation </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td>   </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 1024 </td>\n\t\t\t<td> 10 </td>\n\t\t\t<td> 1989 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/CalIt2+Building+People+Counts"><img src="assets/MLimages/SmallLarge156.jpg" border=1 /></a> </td><td> <a href="datasets/CalIt2+Building+People+Counts">CalIt2 Building People Counts</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data comes from the main door of the CalIt2 building at UCI. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td>   </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<td> 10080 </td>\n\t\t\t<td> 4 </td>\n\t\t\t<td> 2006 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Dodgers+Loop+Sensor"><img src="assets/MLimages/SmallLarge157.jpg" border=1 /></a> </td><td> <a href="datasets/Dodgers+Loop+Sensor">Dodgers Loop Sensor</a> </td></tr></table></td>\n\t\t\t<!-- <td> Loop sensor data was collected for the Glendale on ramp for the 101 North freeway in Los Angeles </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td>   </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<td> 50400 </td>\n\t\t\t<td> 3 </td>\n\t\t\t<td> 2006 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Poker+Hand"><img src="assets/MLimages/SmallLarge158.jpg" border=1 /></a> </td><td> <a href="datasets/Poker+Hand">Poker Hand</a> </td></tr></table></td>\n\t\t\t<!-- <td> Purpose is to predict poker hands </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<td> 1025010 </td>\n\t\t\t<td> 11 </td>\n\t\t\t<td> 2007 </td>\n\t\t\t<!-- <td> Game </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/MAGIC+Gamma+Telescope"><img src="assets/MLimages/SmallLarge159.jpg" border=1 /></a> </td><td> <a href="datasets/MAGIC+Gamma+Telescope">MAGIC Gamma Telescope</a> </td></tr></table></td>\n\t\t\t<!-- <td> Data are MC generated to simulate registration of high energy gamma particles in an atmospheric Cherenkov telescope </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 19020 </td>\n\t\t\t<td> 11 </td>\n\t\t\t<td> 2007 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/UJI+Pen+Characters"><img src="assets/MLimages/SmallLarge160.jpg" border=1 /></a> </td><td> <a href="datasets/UJI+Pen+Characters">UJI Pen Characters</a> </td></tr></table></td>\n\t\t\t<!-- <td> Data consists of written characters in a UNIPEN-like format </td> -->\n\t\t\t<td> Multivariate, Sequential </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 1364 </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2007 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Mammographic+Mass"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Mammographic+Mass">Mammographic Mass</a> </td></tr></table></td>\n\t\t\t<!-- <td> Discrimination of benign and malignant mammographic masses based on BI-RADS attributes and the patient\'s age. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 961 </td>\n\t\t\t<td> 6 </td>\n\t\t\t<td><p

p class="normal">2007 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr b
gcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Forest+Fires"><img src="assets/MLimages/Sm
allLarge162.jpg" border=1 /></a> </td><td> <a href="datasets/Forest+Fires">Fore
st Fires</a> </td></tr></table></td>\n\t\t\t<!-- <td> This is a difficult regression task,
where the aim is to predict the burned area of forest fires, in the northeast region of Portugal, by using meteorol
ogical and other data (see details at: http://www.dsi.uminho.pt/~pcortez/forestfires). </td> -->\n\t\t\t<td
> Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t
\t<td> Real </td>\n\t\t\t<td> 517 </td>\n\t\t\t<td><p cl
ass="normal">13 </td>\n\t\t\t<td> 2008 </td>\n\t\t\t<!-- <td><p class="no
rmal">Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Reuters+Transcri
bed+Subset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="no
rmal"><a href="datasets/Reuters+Transcribed+Subset">Reuters Transcribed Subset</a> </td></tr>
</table></td>\n\t\t\t<!-- <td> This dataset is created by reading out 200 files from the 10 large
st Reuters \r\nclasses and using an Automatic Speech Recognition system to create \r\ncorresponding transcrip
tions. </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td> Classi
fication </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 200 </
p></td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2008 </td>\n\t\t\t<
!-- <td> Business </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr>
<td><a href="datasets/Bag+of+Words"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&n
bsp;</td><td> <a href="datasets/Bag+of+Words">Bag of Words</a> </td></tr></t
able></td>\n\t\t\t<!-- <td> This data set contains five text collections in the form of bags-of-wo
rds. </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td> Clusteri
ng </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 8000000
  </td>\n\t\t\t<td> 100000 </td>\n\t\t\t<td> 2008
</td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a
href="datasets/Concrete+Compressive+Strength"><img src="assets/MLimages/SmallLarge165.jpg" border=1 />
</a> </td><td> <a href="datasets/Concrete+Compressive+Strength">Concrete Co
mpressive Strength</a> </td></tr></table></td>\n\t\t\t<!-- <td> Concrete is the most i
mportant material in civil engineering. The concrete compressive strength is a highly nonlinear function of age a
nd ingredients.   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td><p class
="normal">Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td><p class="n
ormal">1030 </td>\n\t\t\t<td> 9 </td>\n\t\t\t<td> 2007
  </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF"
>\n\t\t\t<td><table><tr><td><a href="datasets/Hill-Valley"><img src="assets/MLimages/SmallLarge166.jpg" bord
er=1 /></a> </td><td> <a href="datasets/Hill-Valley">Hill-Valley</a> </td></
tr></table></td>\n\t\t\t<!-- <td> Each record represents 100 points on a two-dimensional grap
h. When plotted in order (from 1 through 100) as the Y co-ordinate, the points will create either a Hill (a &#65533;bump
&#65533; in the terrain) or a Valley (a &#65533;dip&#65533; in the terrain). </td> -->\n\t\t\t<td> Sequenti
al </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real
  </td>\n\t\t\t<td> 606 </td>\n\t\t\t<td> 101
</td>\n\t\t\t<td> 2008 </td>\n\t\t\t<!-- <td> Other </td
> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Arcene"><img src="assets/MLimages/SmallLarg
e167.jpg" border=1 /></a> </td><td> <a href="datasets/Arcene">Arcene</a></
p></td></tr></table></td>\n\t\t\t<!-- <td> ARCENE\'s task is to distinguish cancer versus norm
al patterns from mass-spectrometric data. This is a two-class classification problem with continuous input variab
les. This dataset is one of 5 datasets of the NIPS 2003 feature selection challenge. </td> -->\n\t\t\t<td
> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\
t\t<td> Real </td>\n\t\t\t<td> 900 </td>\n\t\t\t<td><p cl
ass="normal">10000 </td>\n\t\t\t<td> 2008 </td>\n\t\t\t<!-- <td><p class
="normal">Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="dataset
s/Dexter"><img src="assets/MLimages/SmallLarge168.jpg" border=1 /></a> </td><td>
<a href="datasets/Dexter">Dexter</a> </td></tr></table></td>\n\t\t\t<!-- <td> DEX
TER is a text classification problem in a bag-of-word representation. This is a two-class classification problem w
ith sparse continuous input variables. This dataset is one of five datasets of the NIPS 2003 feature selection ch
allenge.\r\n </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td><p class="no
rmal">Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td><p class="n
ormal">2600 </td>\n\t\t\t<td> 20000 </td>\n\t\t\t<td>
2008 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table
><tr><td><a href="datasets/Dorothea"><img src="assets/MLimages/SmallLarge169.jpg" border=1 /></a>
</td><td> <a href="datasets/Dorothea">Dorothea</a> </td></tr></table></td>\n\t\
t\t<!-- <td> DOROTHEA is a drug discovery dataset. Chemical compounds represented by st
ructural molecular features must be classified as active (binding to thrombin) or inactive. This is one of 5 datase

ructural molecular features must be classified as active (binding to thrombin) or inactive. This is one of 5 datase ts of the NIPS 2003 feature selection challenge. </td> -->\n\t\t\t<td> Multivariate&n bsp; </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer& nbsp; </td>\n\t\t\t<td> 1950 </td>\n\t\t\t<td> 100000 < /p></td>\n\t\t\t<td> 2008 </td>\n\t\t\t<!-- <td> Life </td > -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Gisette"><img src="assets/ MLimages/SmallLarge170.jpg" border=1 /></a> </td><td> <a href="datasets/Gisett e">Gisette</a> </td></tr></table></td>\n\t\t\t<!-- <td> GISETTE is a handwritten digit recognition problem. The problem is to separate the highly confusible digits \'4\' and \'9\'. This dataset is one of five datasets of the NIPS 2003 feature selection challenge.\r\n </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 13500 </td>\n\t\t\t<td> 5000 </td>\n\t\t\t<td> 2008 </td>\n\t\t\t<!-- <td> Com puter </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Madelon"><img src="assets /MLimages/SmallLarge171.jpg" border=1 /></a> </td><td> <a href="datasets/Mad elon">Madelon</a> </td></tr></table></td>\n\t\t\t<!-- <td> MADELON is an artificial d ataset, which was part of the NIPS 2003 feature selection challenge. This is a two-class classification problem w ith continuous input variables. The difficulty is that the problem is multivariate and highly non-linear.   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification& nbsp; </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 4400 </ td>\n\t\t\t<td> 500 </td>\n\t\t\t<td> 2008 </td>\n\t\t\t< !-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td ><a href="datasets/Ozone+Level+Detection"><img src="assets/MLimages/SmallLarge172.jpg" border=1 /></a> </td><td> <a href="datasets/Ozone+Level+Detection">Ozone Level Detection</a> </td></tr></table></td>\n\t\t\t<!-- <td> Two ground ozone level data sets are include d in this collection. One is the eight hour peak set (eighthr.data), the other is the one hour peak set (onehr.data). Those data were collected from 1998 to 2004 at the Houston, Galveston and Brazoria area. </td> - ->\n\t\t\t<td> Multivariate, Sequential, Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 2536 </td>\n\t\t\t<td> 73 </td>\n\t\t\t<td> 2008&nbs p; </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><t d><a href="datasets/Abscisic+Acid+Signaling+Network"><img src="assets/MLimages/SmallLargedefault.jpg" bo rder=1 /></a> </td><td> <a href="datasets/Abscisic+Acid+Signaling+Network">Abs cisic Acid Signaling Network</a> </td></tr></table></td>\n\t\t\t<!-- <td> The objective is to determine the set of boolean rules that describe the interactions of the nodes within this plant signaling net work. The dataset includes 300 separate boolean pseudodynamic simulations using an asynchronous update scheme.   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Causal-Discovery </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 300 </td>\n\t\t\t<td> 43 </td>\n\t\t\t<td> 2 008 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF"> \n\t\t\t<td><table><tr><td><a href="datasets/Parkinsons"><img src="assets/MLimages/SmallLarge174.jpg" bord er=1 /></a> </td><td> <a href="datasets/Parkinsons">Parkinsons</a> </td ></tr></table></td>\n\t\t\t<!-- <td> Oxford Parkinson\'s Disease Detection Dataset </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification&n bsp; </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 197 </t d>\n\t\t\t<td> 23 </td>\n\t\t\t<td> 2008 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Char acter+Trajectories"><img src="assets/MLimages/SmallLarge175.jpg" border=1 /></a> </td><td> <a href="datasets/Character+Trajectories">Character Trajectories</a> </td></tr></table> </td>\n\t\t\t<!-- <td> Multiple, labelled samples of pen tip trajectories recorded whilst writing in dividual characters. All samples are from the same writer, for the purposes of primitive extraction. Only charact ers with a single pen-down segment were considered. </td> -->\n\t\t\t<td> Time-S eries </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 2858 </td>\n\t\t\t<td> 3 </td>\n\t\t\t<td> 2008 </td>\n\t\t\t<!-- <td> Compute r </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Blood+Tran sfusion+Service+Center"><img src="assets/MLimages/SmallLarge176.jpg" border=1 /></a> </td><td> <a href="datasets/Blood+Transfusion+Service+Center">Blood Transfusion Service Center</ a> </td></tr></table></td>\n\t\t\t<!-- <td> Data taken from the Blood Transfusion Ser vice Center in Hsin-Chu City in Taiwan -- this is a classification problem.   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 748 </td>\n\t\t\t<td><p class="nor

mal">5 </td>\n\t\t\t<td> 2008 </td>\n\t\t\t<!-- <td> Business </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/UJI+Pen+Characters+%28Version+2%29"><img src="assets/MLimages/SmallLarge160.jpg" border=1 /></a> </td><td> <a href="datasets/UJI+Pen+Characters+%28Version+2%29">UJI Pen Characters (Version 2)</a> </td></tr></table></td>\n\t\t\t<!-- <td> A pen-based database with more than 11k isolated handwritten characters </td> -->\n\t\t\t<td> Multivariate, Sequential </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 11640 </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2009 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Semeion+Handwritten+Digit"><img src="assets/MLimages/SmallLarge178.jpg" border=1 /></a> </td><td> <a href="datasets/Semeion+Handwritten+Digit">Semeion Handwritten Digit</a> </td></tr></table></td>\n\t\t\t<!-- <td> 1593 handwritten digits from around 80 persons were scanned, stretched in a rectangular box 16x16 in a gray scale of 256 values. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 1593 </td>\n\t\t\t<td> 256 </td>\n\t\t\t<td> 2008 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/SECOM"><img src="assets/MLimages/SmallLarge179.jpg" border=1 /></a> </td><td> <a href="datasets/SECOM">SECOM</a> </td></tr></table></td>\n\t\t\t<!-- <td> Data from a semi-conductor manufacturing process </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Causal-Discovery </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 1567 </td>\n\t\t\t<td> 591 </td>\n\t\t\t<td> 2008 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Plants"><img src="assets/MLimages/SmallLarge180.jpg" border=1 /></a> </td><td> <a href="datasets/Plants">Plants</a> </td></tr></table></td>\n\t\t\t<!-- <td> Data has been extracted from the USDA plants database. It contains all plants (species and genera) in the database and the states of USA and Canada where they occur. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Clustering </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 22632 </td>\n\t\t\t<td> 70 </td>\n\t\t\t<td> 2008 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Libras+Movement"><img src="assets/MLimages/SmallLarge181.jpg" border=1 /></a> </td><td> <a href="datasets/Libras+Movement">Libras Movement</a> </td></tr></table></td>\n\t\t\t<!-- <td> The data set contains 15 classes of 24 instances each. Each class references to a hand movement type in LIBRAS (Portuguese\r\nname \'L&#65533;ngua BRAsileira de Sinais\', oficial brazilian signal language). </td> -->\n\t\t\t<td> Multivariate, Sequential </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 360 </td>\n\t\t\t<td> 91 </td>\n\t\t\t<td> 2009 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Concrete+Slump+Test"><img src="assets/MLimages/SmallLarge165.jpg" border=1 /></a> </td><td> <a href="datasets/Concrete+Slump+Test">Concrete Slump Test</a> </td></tr></table></td>\n\t\t\t<!-- <td> Concrete is a highly complex material. The slump flow of concrete is not only determined by the water content, but that is also influenced by other concrete ingredients. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 103 </td>\n\t\t\t<td> 10 </td>\n\t\t\t<td> 2009 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Communities+and+Crime"><img src="assets/MLimages/SmallLarge183.jpg" border=1 /></a> </td><td> <a href="datasets/Communities+and+Crime">Communities and Crime</a> </td></tr></table></td>\n\t\t\t<!-- <td> Communities within the United States. The data combines socio-economic data from the 1990 US Census, law enforcement data from the 1990 US LEMAS survey, and crime data from the 1995 FBI UCR. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 1994 </td>\n\t\t\t<td> 128 </td>\n\t\t\t<td> 2009 </td>\n\t\t\t<!-- <td> Social </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Acute+Inflammations"><img src="assets/MLimages/SmallLarge184.jpg" border=1 /></a> </td><td> <a href="datasets/Acute+Inflammations">Acute Inflammations</a> </td></tr></table></td>\n\t\t\t<!-- <td> The data was created by a medical expert as a data set to test the expert system, \r\nwhich will perform the presumptive diagnosis of two diseases of the urinary system.\r\n </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td><p class="normal"

>Classification </td>\n\t\t\t<td> Categorical, Integer </td>\n\t\t\t<td> 120 </td>\n\t\t\t<td> 6 </td>\n\t\t\t<td> 2009 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t</tr><tr>\n\t\t<td><table><tr><td><a href="datasets/Wine+Quality"><img src="assets/MLimages/SmallLarge186.jpg" border=1 /></a> </td><td> <a href="datasets/Wine+Quality">Wine Quality</a> </td></tr></table></td>\n\t\t<!-- <td> Two datasets are included, related to red and white vinho verde wine samples, from the north of Portugal. The goal is to model wine quality based on physicochemical tests (see [Cortez et al., 2009], http://www3.dsi.uminho.pt/pcortez/wine/). </td> -->\n\t\t<td> Multivariate </td>\n\t\t<td> Classification, Regression </td>\n\t\t<td> Real </td>\n\t\t<td> 4898 </td>\n\t\t<td> 12 </td>\n\t\t<td> 2009 </td>\n\t\t<!-- <td> Business </td> -->\n\t\t</tr><tr bgcolor="DDEEFF">\n\t\t<td><table><tr><td><a href="datasets/URL+Reputation"><img src="assets/MLimages/SmallLarge187.jpg" border=1 /></a> </td><td> <a href="datasets/URL+Reputation">URL Reputation</a> </td></tr></table></td>\n\t\t<!-- <td> Anonymized 120-day subset of the ICML-09 URL data containing 2.4 million examples and 3.2 million features. </td> -->\n\t\t<td> Multivariate, Time-Series </td>\n\t\t<td> Classification </td>\n\t\t<td> Integer, Real </td>\n\t\t<td> 2396130 </td>\n\t\t<td> 3231961 </td>\n\t\t<td> 2009 </td>\n\t\t<!-- <td> Computer </td> -->\n\t\t</tr><tr>\n\t\t<td><table><tr><td><a href="datasets/p53+Mutants"><img src="assets/MLimages/SmallLarge188.jpg" border=1 /></a> </td><td> <a href="datasets/p53+Mutants">p53 Mutants</a> </td></tr></table></td>\n\t\t<!-- <td> The goal is to model mutant p53 transcriptional activity (active vs inactive) based on data extracted from biophysical simulations.\r\n </td> -->\n\t\t<td> Multivariate </td>\n\t\t<td> Classification </td>\n\t\t<td> Real </td>\n\t\t<td> 16772 </td>\n\t\t<td> 5409 </td>\n\t\t<td> 2010 </td>\n\t\t<!-- <td> Life </td> -->\n\t\t</tr><tr bgcolor="DDEEFF">\n\t\t<td><table><tr><td><a href="datasets/Parkinsons+Telemonitoring"><img src="assets/MLimages/SmallLarge174.jpg" border=1 /></a> </td><td> <a href="datasets/Parkinsons+Telemonitoring">Parkinsons Telemonitoring</a> </td></tr></table></td>\n\t\t<!-- <td> Oxford Parkinson\'s Disease Telemonitoring Dataset </td> -->\n\t\t<td> Multivariate </td>\n\t\t<td> Regression </td>\n\t\t<td> Integer, Real </td>\n\t\t<td> 5875 </td>\n\t\t<td> 26 </td>\n\t\t<td> 2009 </td>\n\t\t<!-- <td> Life </td> -->\n\t\t</tr><tr>\n\t\t<td><table><tr><td><a href="datasets/Demospongiae"><img src="assets/MLimages/SmallLarge190.jpg" border=1 /></a> </td><td> <a href="datasets/Demospongiae">Demospongiae</a> </td></tr></table></td>\n\t\t<!-- <td> Marine sponges of the Demospongiae class classification domain. </td> -->\n\t\t<td> Multivariate </td>\n\t\t<td> Classification </td>\n\t\t<td> Integer </td>\n\t\t<td> 503 </td>\n\t\t<td>   </td>\n\t\t<td> 2010 </td>\n\t\t<!-- <td> Life </td> -->\n\t\t</tr><tr bgcolor="DDEEFF">\n\t\t<td><table><tr><td><a href="datasets/Opinosis+Opinion+%26frasl%3B+Review"><img src="assets/MLimages/SmallLarge191.jpg" border=1 /></a> </td><td> <a href="datasets/Opinosis+Opinion+%26frasl%3B+Review">Opinosis Opinion ⁄ Review</a> </td></tr></table></td>\n\t\t<!-- <td> This dataset contains sentences extracted from user reviews on a given topic. Example topics are “performance of Toyota Camry” and “sound quality of ipod nano”.   </td> -->\n\t\t<td> Text </td>\n\t\t<td>   </td>\n\t\t<td>   </td>\n\t\t<td> 51 </td>\n\t\t<td>   </td>\n\t\t<td> 2010 </td>\n\t\t<!-- <td> Computer </td> -->\n\t\t</tr><tr>\n\t\t<td><table><tr><td><a href="datasets/Breast+Tissue"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Breast+Tissue">Breast Tissue</a> </td></tr></table></td>\n\t\t<!-- <td> Dataset with electrical impedance measurements of freshly excised tissue samples from the breast. </td> -->\n\t\t<td> Multivariate </td>\n\t\t<td> Classification </td>\n\t\t<td> Real </td>\n\t\t<td> 106 </td>\n\t\t<td> 10 </td>\n\t\t<td> 2010 </td>\n\t\t<!-- <td> Life </td> -->\n\t\t</tr><tr bgcolor="DDEEFF">\n\t\t<td><table><tr><td><a href="datasets/Cardiotocography"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Cardiotocography">Cardiotocography</a> </td></tr></table></td>\n\t\t<!-- <td> The dataset consists of measurements of fetal heart rate (FHR) and uterine contraction (UC) features on cardiotocograms classified by expert obstetricians. </td> -->\n\t\t<td> Multivariate </td>\n\t\t<td> Classification </td>\n\t\t<td> Re

al </td>\n\t\t\t<td> 2126 </td>\n\t\t\t<td> 23 </td>\n\t\t\t<td> 2010 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Wall-Following+Robot+Navigation+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Wall-Following+Robot+Navigation+Data">Wall-Following Robot Navigation Data</a> </td></tr></table></td>\n\t\t\t<!-- <td> The data were collected as the SCITOS G5 robot navigates through the room following the wall in a clockwise direction, for 4 rounds, using 24 ultrasound sensors arranged circularly around its \'waist\'. </td> -->\n\t\t\t<td> Multivariate, Sequential </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 5456 </td>\n\t\t\t<td> 24 </td>\n\t\t\t<td> 2010 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Spoken+Arabic+Digit"><img src="assets/MLimages/SmallLarge195.jpg" border=1 /></a> </td><td> <a href="datasets/Spoken+Arabic+Digit">Spoken Arabic Digit</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset contains timeseries of mel-frequency cepstrum coefficients (MFCCs) corresponding to spoken Arabic digits. Includes data from 44 male and 44 female native Arabic speakers. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 8800 </td>\n\t\t\t<td> 13 </td>\n\t\t\t<td> 2010 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Localization+Data+for+Person+Activity"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Localization+Data+for+Person+Activity">Localization Data for Person Activity</a> </td></tr></table></td>\n\t\t\t<!-- <td> Data contains recordings of five people performing different activities. Each person wore four sensors (tags) while performing the same scenario five times.   </td> -->\n\t\t\t<td> Univariate, Sequential, Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 164860 </td>\n\t\t\t<td> 8 </td>\n\t\t\t<td> 2010 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/AutoUniv"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/AutoUniv">AutoUniv</a> </td></tr></table></td>\n\t\t\t<!-- <td> AutoUniv is an advanced data generator for classifications tasks. The aim is to reflect the nuances and heterogeneity of real data. Data can be generated in .csv, ARFF or C4.5 formats. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical, Integer, Real </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2010 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Steel+Plates+Faults"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Steel+Plates+Faults">Steel Plates Faults</a> </td></tr></table></td>\n\t\t\t<!-- <td> A dataset of steel plates’ faults, classified into 7 different types. \r\nThe goal was to train machine learning for automatic pattern recognition.\r\n </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 1941 </td>\n\t\t\t<td> 27 </td>\n\t\t\t<td> 2010 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/MiniBooNE+particle+identification"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/MiniBooNE+particle+identification">MiniBooNE particle identification</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset is taken from the MiniBooNE experiment and is used to distinguish electron neutrinos (signal) from muon neutrinos (background). </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 130065 </td>\n\t\t\t<td> 50 </td>\n\t\t\t<td> 2010 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/YearPredictionMSD"><img src="assets/MLimages/SmallLarge203.jpg" border=1 /></a> </td><td> <a href="datasets/YearPredictionMSD">YearPredictionMSD</a> </td></tr></table></td>\n\t\t\t<!-- <td> Prediction of the release year of a song from audio features. Songs are mostly western, commercial tracks ranging from 1922 to 2011, with a peak in the year 2000s. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 515345 </td>\n\t\t\t<td> 90 </td>\n\t\t\t<td> 2011 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/PEMS-SF"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/PEMS-SF"

>PEMS-SF</a> </td></tr></table></td>\n\t\t\t<!-- <td> 15 months worth of daily data (440 daily records) that describes the occupancy rate, between 0 and 1, of different car lanes of the San Francisco bay area freeways across time. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 440 </td>\n\t\t\t<td> 138672 </td>\n\t\t\t<td> 2011 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/OpinRank+Review+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/OpinRank+Review+Dataset">OpinRank Review Dataset</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data set contains user reviews of cars and and hotels collected from Tripadvisor (~259,000 \r\nreviews) and Edmunds (~42,230 reviews).   </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2011 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Relative+location+of+CT+slices+on+axial+axis"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Relative+location+of+CT+slices+on+axial+axis">Relative location of CT slices on axial axis</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset consists of 384 features extracted from CT images. The class variable is numeric and denotes the relative location of the CT slice on the axial axis of the human body. </td> -->\n\t\t\t<td> Domain-Theory </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 53500 </td>\n\t\t\t<td> 386 </td>\n\t\t\t<td> 2011 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Online+Handwritten+Assamese+Characters+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Online+Handwritten+Assamese+Characters+Dataset">Online Handwritten Assamese Characters Dataset</a> </td></tr></table></td>\n\t\t\t<!-- <td> This is a dataset of 8235 online handwritten assamese characters. The “online” process involves capturing of data as text is written on a digitizing tablet with an electronic pen. </td> -->\n\t\t\t<td> Multivariate, Sequential </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 8235 </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2011 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/PubChem+Bioassay+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/PubChem+Bioassay+Data">PubChem Bioassay Data</a> </td></tr></table></td>\n\t\t\t<!-- <td> These highly imbalanced bioassay datasets are from the differing types of screening that can be performed using HTS technology. 21 datasets were created from 12 bioassays. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2011 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Record+Linkage+Comparison+Patterns"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Record+Linkage+Comparison+Patterns">Record Linkage Comparison Patterns</a> </td></tr></table></td>\n\t\t\t<!-- <td> Element-wise comparison of records with personal data from a record linkage setting. The task is to decide from a comparison pattern whether the underlying records belong to one person. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 5749132 </td>\n\t\t\t<td> 12 </td>\n\t\t\t<td> 2011 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Communities+and+Crime+Unnormalized"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Communities+and+Crime+Unnormalized">Communities and Crime Unnormalized</a> </td></tr></table></td>\n\t\t\t<!-- <td> Communities in the US. Data combines socio-economic data from the \'90 Census, law enforcement data from the 1990 Law Enforcement Management and Admin Stats survey, and crime data from the 1995 FBI UCR </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 2215 </td>\n\t\t\t<td> 147 </td>\n\t\t\t<td> 2011 </td>\n\t\t\t<!-- <td> Social </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Vertebral+Column"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Vertebral+Column">Vertebral Column</a> </td></tr></table></td>\n\t\t\t<!-- <td> Data set containing values for six biomechanical features used to classify orthopaedic patients into 3 classes (normal, disk hernia or spondilol

ysthesis) or 2 classes (normal or abnormal). </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 310 </td>\n\t\t\t<td> 6 </td>\n\t\t\t<td> 2011 </td>\n\t\t\t<!-- <td>   </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/EMG+Physical+Action+Data+Set"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/EMG+Physical+Action+Data+Set">EMG Physical Action Data Set</a> </td></tr></table></td>\n\t\t\t<!-- <td> The Physical Action Data Set includes 10 normal and 10 aggressive physical actions that measure the human activity. The data have been collected by 4 subjects using the Delsys EMG wireless apparatus. </td> -->\n\t\t\t<td> Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 10000 </td>\n\t\t\t<td> 8 </td>\n\t\t\t<td> 2011 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Vicon+Physical+Action+Data+Set"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Vicon+Physical+Action+Data+Set">Vicon Physical Action Data Set</a> </td></tr></table></td>\n\t\t\t<!-- <td> The Physical Action Data Set includes 10 normal and 10 aggressive physical actions that measure the human activity. The data have been collected by 10 subjects using the Vicon 3D tracker. </td> -->\n\t\t\t<td> Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 3000 </td>\n\t\t\t<td> 27 </td>\n\t\t\t<td> 2011 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Amazon+Commerce+reviews+set"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Amazon+Commerce+reviews+set">Amazon Commerce reviews set</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset is used for authorship identification in online Writeprint which is a new research field of pattern recognition.   </td> -->\n\t\t\t<td> Multivariate, Text, Domain-Theory </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 1500 </td>\n\t\t\t<td> 10000 </td>\n\t\t\t<td> 2011 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Amazon+Access+Samples"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Amazon+Access+Samples">Amazon Access Samples</a> </td></tr></table></td>\n\t\t\t<!-- <td> Amazon\'s InfoSec is getting smarter about the way Access data is leveraged. This is an anonymized sample of access provisioned within the company. </td> -->\n\t\t\t<td> Time-Series, Domain-Theory </td>\n\t\t\t<td> Regression, Clustering, Causal-Discovery </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 30000 </td>\n\t\t\t<td> 20000 </td>\n\t\t\t<td> 2011 </td>\n\t\t\t<!-- <td> Business </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Reuter_50_50"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Reuter_50_50">Reuter_50_50</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset is used for authorship identification in online Writeprint which is a new research field of pattern recognition.   </td> -->\n\t\t\t<td> Multivariate, Text, Domain-Theory </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 2500 </td>\n\t\t\t<td> 10000 </td>\n\t\t\t<td> 2011 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Farm+Ads"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Farm+Ads">Farm Ads</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data was collected from text ads found on twelve websites that deal with various farm animal related topics. The binary labels are based on whether or not the content owner approves of the ad. </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 4143 </td>\n\t\t\t<td> 54877 </td>\n\t\t\t<td> 2011 </td>\n\t\t\t<!-- <td> Business </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/DBWorld+e-mails"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/DBWorld+e-mails">DBWorld e-mails</a> </td></tr></table></td>\n\t\t\t<!-- <td> It contains 64 e-mails which I have manually collected from DBWorld mailing list. They are classified in: \'announces of conferences\' and \'everything else\'. </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 64 </td>\n\t\t\t<td> 4702 </td>\n\t\t\t<td> 2011 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a h

ref="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29"><img src="assets/MLimages/SmallLarge default.jpg" border=1 /></a> </td><td> <a href="datasets/KEGG+Metabolic+Relatio n+Network+%28Directed%29">KEGG Metabolic Relation Network (Directed)</a> </td></tr></table></td >\n\t\t\t<!-- <td> KEGG Metabolic pathways modeled as directed relation network. Variety of graphical features presented. </td> -->\n\t\t\t<td> Multivariate, Univariate, Text&nb sp; </td>\n\t\t\t<td> Classification, Regression, Clustering </td>\n\t\t\t<td><p cl ass="normal">Integer, Real </td>\n\t\t\t<td> 53414 </td>\n\t\t\t<td><p cl ass="normal">24 </td>\n\t\t\t<td> 2011 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/KE GG+Metabolic+Reaction+Network+%28Undirected%29"><img src="assets/MLimages/SmallLargedefault.jpg" b order=1 /></a> </td><td> <a href="datasets/KEGG+Metabolic+Reaction+Network+ %28Undirected%29">KEGG Metabolic Reaction Network (Undirected)</a> </td></tr></table></td>\n\t\t\ t<!-- <td> KEGG Metabolic pathways modeled as un-directed reaction network. Variety of gra phical features presented. </td> -->\n\t\t\t<td> Multivariate, Univariate, Text </td>\n\t\t\t<td> Classification, Regression, Clustering </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 65554 </td>\n\t\t\t<td> 29 </td>\n\t\t\t<td> 2011 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Bank+Marketing"><img s rc="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="d atasets/Bank+Marketing">Bank Marketing</a> </td></tr></table></td>\n\t\t\t<!-- <td> The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The class ification goal is to predict if the client will subscribe a term deposit (variable y). </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 45211 </td>\n\t\t\t<td><p cla ss="normal">17 </td>\n\t\t\t<td> 2012 </td>\n\t\t\t<!-- <td> Business </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="dataset s/YouTube+Comedy+Slam+Preference+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </a> </td><td> <a href="datasets/YouTube+Comedy+Slam+Preference+Data">Yo uTube Comedy Slam Preference Data</a> </td></tr></table></td>\n\t\t\t<!-- <td> Thi s dataset provides user vote data on which video from a pair of videos is funnier collected on YouTube Comedy Slam. The task is to automatically predict this preference based on video metadata. </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td >   </td>\n\t\t\t<td> 1138562 </td>\n\t\t\t<td> 3 </td>\n\t\t\t<td> 2012 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Gas+Sensor+Array+ Drift+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Gas+Sensor+Array+Drift+Dataset">Gas Sensor Array Drift Dataset</a> < /td></tr></table></td>\n\t\t\t<!-- <td> This archive contains 13910 measurements from 16 che mical sensors utilized in simulations for drift compensation in a discrimination task of 6 gases at various levels o f concentrations. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td><p clas s="normal">Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 13910 </td>\n\t\t\t<td> 128 </td>\n\t\t\t<td> 2012 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor=" DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/ILPD+%28Indian+Liver+Patient+Dataset%29"><img src ="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="dat asets/ILPD+%28Indian+Liver+Patient+Dataset%29">ILPD (Indian Liver Patient Dataset)</a> </td></tr> </table></td>\n\t\t\t<!-- <td> This data set contains 10 variables that are age, gender, total Bil irubin, direct Bilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT and Alkphos. </td> -->\n\t\t\t<td > Multivariate </td>\n\t\t\t<td> Classification </td>\n\ t\t\t<td> Integer, Real </td>\n\t\t\t<td> 583 </td>\n\t\t\t<td> 10 </td>\n\t\t\t<td> 2012 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/OPPORTUNI TY+Activity+Recognition"><img src="assets/MLimages/SmallLarge226.jpg" border=1 /></a> </td><td> <a href="datasets/OPPORTUNITY+Activity+Recognition">OPPORTUNITY Activity Recogniti on</a> </td></tr></table></td>\n\t\t\t<!-- <td> The OPPORTUNITY Dataset for Huma n Activity Recognition from Wearable, Object, and Ambient Sensors is a dataset devised to benchmark human activity recognition algorithms (classification, automatic data segmentation, sensor fusion, feature extraction, et c). </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td><p clas s="normal">Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 2551 </td>\n\t\t\t<td> 242 </td>\n\t\t\t<td> 2012 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="D DEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Nomao"><img src="assets/MLimages/SmallLargedefault.j

pg" border=1 /></a> </td><td> <a href="datasets/Nomao">Nomao</a> </td></tr></table></td>\n\t\t\t<!-- <td> Nomao collects data about places (name, phone, localization...) from many sources.\r\nDeduplication consists in detecting what data refer to the same place.\r\nInstances in the dataset compare 2 spots. </td> -->\n\t\t\t<td> Univariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 34465 </td>\n\t\t\t<td> 120 </td>\n\t\t\t<td> 2012 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/SMS+Spam+Collection"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/SMS+Spam+Collection">SMS Spam Collection</a> </td></tr></table></td>\n\t\t\t<!-- <td> The SMS Spam Collection is a public set of SMS labeled messages that have been collected for mobile phone spam research. </td> -->\n\t\t\t<td> Multivariate, Text, Domain-Theory </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 5574 </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2012 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Skin+Segmentation"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Skin+Segmentation">Skin Segmentation</a> </td></tr></table></td>\n\t\t\t<!-- <td> The Skin Segmentation dataset is constructed over B, G, R color space. Skin and Nonskin dataset is generated using skin textures from face images of diversity of age, gender, and race people. </td> -->\n\t\t\t<td> Univariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 245057 </td>\n\t\t\t<td> 4 </td>\n\t\t\t<td> 2012 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Planning+Relax"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Planning+Relax">Planning Relax</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset concerns with the classification of two mental stages from recorded EEG signals: Planning (during imagination of motor act) and Relax state.   </td> -->\n\t\t\t<td> Univariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 182 </td>\n\t\t\t<td> 13 </td>\n\t\t\t<td> 2012 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/PAMAP2+Physical+Activity+Monitoring"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/PAMAP2+Physical+Activity+Monitoring">PAMAP2 Physical Activity Monitoring</a> </td></tr></table></td>\n\t\t\t<!-- <td> The PAMAP2 Physical Activity Monitoring dataset contains data of 18 different physical activities, performed by 9 subjects wearing 3 inertial measurement units and a heart rate monitor. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 3850505 </td>\n\t\t\t<td> 52 </td>\n\t\t\t<td> 2012 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Restaurant+%26+consumer+data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Restaurant+%26+consumer+data">Restaurant & consumer data</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset was obtained from a recommender system prototype. The task was to generate a top-n list of restaurants according to the consumer preferences.   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 138 </td>\n\t\t\t<td> 47 </td>\n\t\t\t<td> 2012 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/CNAE-9"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/CNAE-9">CNAE-9</a> </td></tr></table></td>\n\t\t\t<!-- <td> This is a data set containing 1080 documents of free text business descriptions of Brazilian companies categorized into a\r\nsubset of 9 categories </td> -->\n\t\t\t<td> Multivariate, Text </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 1080 </td>\n\t\t\t<td> 857 </td>\n\t\t\t<td> 2012 </td>\n\t\t\t<!-- <td> Business </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Individual+household+electric+power+consumption"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Individual+household+electric+power+consumption">Individual household electric power consumption</a> </td></tr></table></td>\n\t\t\t<!-- <td> Measurements of electric power consumption in one household with a one-minute sampling rate over a period of almost 4 years. Different electrical quantities and some sub-metering values are available. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>

values are available. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Regression, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 2075259 </td>\n\t\t\t<td> 9 </td>\n\t\t\t<td> 2012 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/seeds"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/seeds">seeds</a> </td></tr></table></td>\n\t\t\t<!-- <td> Measurements of geometrical properties of kernels belonging to three different varieties of wheat. A soft X-ray technique and GRAINS package were used to construct all seven, real-valued attributes. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 210 </td>\n\t\t\t<td> 7 </td>\n\t\t\t<td> 2012 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Northix"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Northix">Northix</a> </td></tr></table></td>\n\t\t\t<!-- <td> Northix is designed to be a schema matching benchmark problem for data integration of two entity relationship databases.   </td> -->\n\t\t\t<td> Multivariate, Univariate, Text </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 115 </td>\n\t\t\t<td> 200 </td>\n\t\t\t<td> 2012 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/QtyT40I10D100K"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/QtyT40I10D100K">QtyT40I10D100K</a> </td></tr></table></td>\n\t\t\t<!-- <td> Since there is no numerical sequential data stream available in standard data sets, this data set is generated from the original T40I10D100K data set </td> -->\n\t\t\t<td> Sequential </td>\n\t\t\t<td>   </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 3960456 </td>\n\t\t\t<td> 4 </td>\n\t\t\t<td> 2012 </td>\n\t\t\t<!-- <td>   </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Legal+Case+Reports"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Legal+Case+Reports">Legal Case Reports</a> </td></tr></table></td>\n\t\t\t<!-- <td> A textual corpus of 4000 legal cases for automatic summarization and citation analysis. For each document we collect catchphrases, citations sentences, citation catchphrases and citation classes. </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2012 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Human+Activity+Recognition+Using+Smartphones"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Human+Activity+Recognition+Using+Smartphones">Human Activity Recognition Using Smartphones</a> </td></tr></table></td>\n\t\t\t<!-- <td> Human Activity Recognition database built from the recordings of 30 subjects performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 10299 </td>\n\t\t\t<td> 561 </td>\n\t\t\t<td> 2012 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/One-hundred+plant+species+leaves+data+set"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/One-hundred+plant+species+leaves+data+set">One-hundred plant species leaves data set</a> </td></tr></table></td>\n\t\t\t<!-- <td> Sixteen samples of leaf each of one-hundred plant species. For each sample, a shape descriptor, fine scale margin and texture histogram are given. </td> -->\n\t\t\t<td>   </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 1600 </td>\n\t\t\t<td> 64 </td>\n\t\t\t<td> 2012 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Energy+efficiency"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Energy+efficiency">Energy efficiency</a> </td></tr></table></td>\n\t\t\t<!-- <td> This study looked into assessing the heating load and cooling load requirements of buildings (that is, energy efficiency) as a function of building parameters. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 768 </td>\n\t\t\t<td> 8 </td>\n\t\t\t<td> 2012 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Yacht+Hydrodynamics"><i
mg src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>   </td><td> <a hr

mg src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a hr
ef="datasets/Yacht+Hydrodynamics">Yacht Hydrodynamics</a> </td></tr></table></td>\n\t\t\t<!-- <td><
p class="normal">Delft data set, used to predict the hydodynamic performance of sailing yachts from dimension
s and velocity. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td><p class=
"normal">Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td><p class="no
rmal">308 </td>\n\t\t\t<td> 7 </td>\n\t\t\t<td> 2013&n
bsp; </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n
\t\t\t<td><table><tr><td><a href="datasets/Fertility"><img src="assets/MLimages/SmallLargedefault.jpg" border
=1 /></a> </td><td> <a href="datasets/Fertility">Fertility</a> </td></tr></ta
ble></td>\n\t\t\t<!-- <td> 100 volunteers provide a semen sample analyzed according to the
WHO 2010 criteria. Sperm concentration are related to socio-demographic data, environmental factors, health
status, and life habits </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td><p
class="normal">Classification, Regression </td>\n\t\t\t<td> Real </td>\n\
t\t\t<td> 100 </td>\n\t\t\t<td> 10 </td>\n\t\t\t<td><p cl
ass="normal">2013 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\
t\t<td><table><tr><td><a href="datasets/Daphnet+Freezing+of+Gait"><img src="assets/MLimages/SmallLarged
efault.jpg" border=1 /></a> </td><td> <a href="datasets/Daphnet+Freezing+of+Gai
t">Daphnet Freezing of Gait</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset
contains the annotated readings of 3 acceleration sensors at the hip and leg of Parkinson\'s disease patients th
at experience freezing of gait (FoG) during walking tasks.\r\n </td> -->\n\t\t\t<td>
Multivariate, Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td>
Real </td>\n\t\t\t<td> 237 </td>\n\t\t\t<td><p class="
normal">9 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td>
Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/3D+Road
+Network+%28North+Jutland%2C+Denmark%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=
1 /></a> </td><td> <a href="datasets/3D+Road+Network+%28North+Jutland%2C+
Denmark%29">3D Road Network (North Jutland, Denmark)</a> </td></tr></table></td>\n\t\t\t<!-- <td><
p class="normal">3D road network with highly accurate elevation information (+-20cm) from Denmark used in e
co-routing and fuel/Co2-estimation routing algorithms. </td> -->\n\t\t\t<td> Sequen
tial, Text </td>\n\t\t\t<td> Regression, Clustering </td>\n\t\t\t<td><p clas
s="normal">Real </td>\n\t\t\t<td> 434874 </td>\n\t\t\t<td><p class="nor
mal">4 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Co
mputer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/ISTANBUL+STOCK+EXC
HANGE"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="norma
l"><a href="datasets/ISTANBUL+STOCK+EXCHANGE">ISTANBUL STOCK EXCHANGE</a> </td><
/tr></table></td>\n\t\t\t<!-- <td> Data sets includes returns of Istanbul Stock Exchange with se
ven other international index; SP, DAX, FTSE, NIKKEI, BOVESPA, MSCE_EU, MSCI_EM from Jun 5, 2009 to F
eb 22, 2011. </td> -->\n\t\t\t<td> Multivariate, Univariate, Time-Series <
/td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Real&
nbsp; </td>\n\t\t\t<td> 536 </td>\n\t\t\t<td> 8 </td
>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Business </td>
-->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Buzz+in+social+media+"><im
g src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href
="datasets/Buzz+in+social+media+">Buzz in social media </a> </td></tr></table></td>\n\t\t\t<!-- <td><p
class="normal">This data-set contains examples of buzz events from two different social networks: Twitter, and
Tom\'s Hardware, a forum network focusing on new technology with more conservative dynamics. </
td> -->\n\t\t\t<td> Time-Series, Multivariate </td>\n\t\t\t<td> Reg
ression, Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td><p c
lass="normal">140000 </td>\n\t\t\t<td> 77 </td>\n\t\t\t<td><p class="nor
mal">2013 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t
<td><table><tr><td><a href="datasets/First-order+theorem+proving"><img src="assets/MLimages/SmallLarged
efault.jpg" border=1 /></a> </td><td> <a href="datasets/First-order+theorem+provi
ng">First-order theorem proving</a> </td></tr></table></td>\n\t\t\t<!-- <td> Given a t
heorem, predict which of five heuristics will give the fastest proof when used by a first-order prover. A sixth pred
iction declines to attempt a proof, should the theorem be too difficult. </td> -->\n\t\t\t<td><p class="no
rmal">Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td><p clas
s="normal">Real </td>\n\t\t\t<td> 6118 </td>\n\t\t\t<td><p class="normal
">51 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Comp
uter </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Wearabl
e+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29"><img src="assets/M
Limages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Wea
rable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29">Wearable Comp

uting: Classification of Body Postures and Movements (PUC-Rio)</a> </td></tr></table></td>\n\t\t\t<!-- <td> A dataset with 5 classes (sitting-down, standing-up, standing, walking, and sitting) collected on 8 hours of activities of 4 healthy subjects. We also established a baseline performance index. </td> -->\n\t\t\t<td> Sequential </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 165632 </td>\n\t\t\t<td> 18 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Gas+sensor+arrays+in+open+sampling+settings"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Gas+sensor+arrays+in+open+sampling+settings">Gas sensor arrays in open sampling settings</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset contains 18000 time-series recordings from a chemical detection platform at six different locations in a wind tunnel facility in response to ten high-priority chemical gaseous substances </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 18000 </td>\n\t\t\t<td> 1950000 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Climate+Model+Simulation+Crashes"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Climate+Model+Simulation+Crashes">Climate Model Simulation Crashes</a> </td></tr></table></td>\n\t\t\t<!-- <td> Given Latin hypercube samples of 18 climate model input parameter values, predict climate model simulation crashes and determine the parameter value combinations that cause the failures. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 540 </td>\n\t\t\t<td> 18 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/MicroMass"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/MicroMass">MicroMass</a> </td></tr></table></td>\n\t\t\t<!-- <td> A dataset to explore machine learning approaches for the identification of microorganisms from mass-spectrometry data. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 931 </td>\n\t\t\t<td> 1300 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/QSAR+biodegradation"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/QSAR+biodegradation">QSAR biodegradation</a> </td></tr></table></td>\n\t\t\t<!-- <td> Data set containing values for 41 attributes (molecular descriptors) used to classify 1055 chemicals into 2 classes (ready and not ready biodegradable). </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 1055 </td>\n\t\t\t<td> 41 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/BLOGGER"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/BLOGGER">BLOGGER</a> </td></tr></table></td>\n\t\t\t<!-- <td> In this paper, we look for to recognize the causes of users tend\r\nto cyber space in Kohkiloye and Boyer Ahmad Province in\r\nIran </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 100 </td>\n\t\t\t<td> 6 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Daily+and+Sports+Activities"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Daily+and+Sports+Activities">Daily and Sports Activities</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset comprises motion sensor data of 19 daily and sports activities each performed by 8 subjects in their own style for 5 minutes. Five Xsens MTx units are used on the torso, arms, and legs.\r\n </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 9120 </td>\n\t\t\t<td> 5625 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/User+Knowledge+Modeling"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/User+Knowledge+Modeling">User Knowledge Modeling</a> </td></tr></table></td>\n\t\t\t<!-- <td> It is the real dataset about the students\' knowledge status about the subject of Electrical DC Machines. The dataset had been obtained from Ph.D. Thesis. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td><p

class="normal">Integer </td>\n\t\t\t<td> 403 </td>\n\t\t\t<td> 5 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection">Reuters RCV1 RCV2 Multilingual, Multiview Text Categorization Test collection</a> </td></tr></table></td>\n\t\t\t<!-- <td> This test collection contains feature characteristics of documents originally written in five different languages and their translations, over a common set of 6 categories.   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 111740 </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/NYSK"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/NYSK">NYSK</a> </td></tr></table></td>\n\t\t\t<!-- <td> NYSK (New York v. Strauss-Kahn) is a collection of English news articles about the case relating to allegations of sexual assault against the former IMF director Dominique Strauss-Kahn (May 2011). </td> -->\n\t\t\t<td> Multivariate, Sequential, Text </td>\n\t\t\t<td> Clustering </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 10421 </td>\n\t\t\t<td> 7 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Social </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Turkiye+Student+Evaluation"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Turkiye+Student+Evaluation">Turkiye Student Evaluation</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data set contains a total 5820 evaluation scores provided by students from Gazi University in Ankara (Turkey). There is a total of 28 course specific questions and additional 5 attributes. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 5820 </td>\n\t\t\t<td> 33 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machines%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machines%29">ser Knowledge Modeling Data (Students\' Knowledge Levels on DC Electrical Machines)</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset is about the users\' learning activities and knowledge levels on subjects of DC Electrical Machines. The dataset had been obtained from online web-courses and reported in my Ph.D. Thesis. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 403 </td>\n\t\t\t<td> 5 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/EEG+Eye+State"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/EEG+Eye+State">EEG Eye State</a> </td></tr></table></td>\n\t\t\t<!-- <td> The data set consists of 14 EEG values and a value indicating the eye state. </td> -->\n\t\t\t<td> Multivariate, Sequential, Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 14980 </td>\n\t\t\t<td> 15 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure">Physicochemical Properties of Protein Tertiary Structure</a> </td></tr></table></td>\n\t\t\t<!-- <td> This is a data set of Physicochemical Properties of Protein Tertiary Structure. The data set is taken from CASP 5-9. There are 45730 decoys and size varying from 0 to 21 armstrong. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 45730 </td>\n\t\t\t<td> 9 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/seismic-bumps"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/seismic-bumps">seismic-bumps</a> </td></tr></table></td>\n\t\t\t<!-- <td> The data describe the problem of high energy (higher than 10^4 J) seismic bumps forecasting in a coal \r\nmine. Data come from two of longwalls located in a Polish coal mine. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td>

2584 </td>\n\t\t\t<td> 19 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/banknote+authentication"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/banknote+authentication">banknote authentication</a> </td></tr></table></td>\n\t\t\t<!-- <td> Data were extracted from images that were taken for the evaluation of an authentication procedure for banknotes. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 1372 </td>\n\t\t\t<td> 5 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Problem%3A+Pat"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Problem%3A+Pat">USPTO Algorithm Challenge, run by NASA-Harvard Tournament Lab and TopCoder Problem: Pat</a> </td></tr></table></td>\n\t\t\t<!-- <td> Data used for USPTO Algorithm Competition. Contains drawing pages from US patents with manually labeled figure and part labels. </td> -->\n\t\t\t<td> Domain-Theory </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 306 </td>\n\t\t\t<td> 5 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/YouTube+Multiview+Video+Games+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/YouTube+Multiview+Video+Games+Dataset">YouTube Multiview Video Games Dataset</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset contains about 120k instances, each described by 13 feature types, with class information, specially useful for exploring multiview topics (cotraining, ensembles, clustering,..). </td> -->\n\t\t\t<td> Multivariate, Text </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 120000 </td>\n\t\t\t<td> 1000000 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations">Gas Sensor Array Drift Dataset at Different Concentrations</a> </td></tr></table></td>\n\t\t\t<!-- <td> This archive contains 13910 measurements from 16 chemical sensors exposed to 6 different gases at various concentration levels. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification, Regression, Clustering, Causa </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 13910 </td>\n\t\t\t<td> 129 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors">Activities of Daily Living (ADLs) Recognition Using Binary Sensors</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset comprises information regarding the ADLs performed by two users on a daily basis in their \r\nown homes.   </td> -->\n\t\t\t<td> Multivariate, Sequential, Time-Series </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2747 </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/SkillCraft1+Master+Table+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/SkillCraft1+Master+Table+Dataset">SkillCraft1 Master Table Dataset</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data was used in Thompson et al. (2013). A list of possible game actions is discussed in Thompson, Blair, Chen, & Henrey (2013). </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 3395 </td>\n\t\t\t<td> 20 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Game </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units">Weight Lifting Exercises monitored with Inertial Measurement Units</a> </td></tr></table></td>\n\t\t\t<!-- <td> Six young health subjects were asked to perform 5 variations of the biceps curl weight lifting exercise. One of the variations is the one predicted by the health professional. </td> -->\n\t\t\t<td> Multivariate </td>\n\

t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 39242 </td>\n\t\t\t<td> 152 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/SML2010"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/SML2010">SML2010</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset is collected from a monitor system mounted in a domotic house. It corresponds to approximately 40 days of monitoring data. </td> -->\n\t\t\t<td> Multivariate, Sequential, Time-Series, Text </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 4137 </td>\n\t\t\t<td> 24 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Bike+Sharing+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Bike+Sharing+Dataset">Bike Sharing Dataset</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset contains the hourly and daily count of rental bikes between years 2011 and 2012 in Capital bikeshare system with the corresponding weather and seasonal information. </td> -->\n\t\t\t<td> Univariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 17389 </td>\n\t\t\t<td> 16 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Social </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Predict+keywords+activities+in+a+online+social+media"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Predict+keywords+activities+in+a+online+social+media">Predict keywords activities in a online social media</a> </td></tr></table></td>\n\t\t\t<!-- <td> The data from Twitter was collected during 360 consecutive days. It was done by querying 1497 English keywords sampled from Wikipedia. This dataset is proposed in a Learning to rank setting. </td> -->\n\t\t\t<td> Multivariate, Sequential, Time-Series </td>\n\t\t\t<td>   </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 51 </td>\n\t\t\t<td> 35 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Thoracic+Surgery+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Thoracic+Surgery+Data">Thoracic Surgery Data</a> </td></tr></table></td>\n\t\t\t<!-- <td> The data is dedicated to classification problem related to the post-operative life expectancy in the lung cancer patients: class 1 - death within one year after surgery, class 2 - survival. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 470 </td>\n\t\t\t<td> 17 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/EMG+dataset+in+Lower+Limb"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/EMG+dataset+in+Lower+Limb">EMG dataset in Lower Limb</a> </td></tr></table></td>\n\t\t\t<!-- <td> 3 different exercises: sitting, standing and walking in the muscles: biceps femoris, vastus medialis, rectus femoris and semitendinosus addition to goniometry in the exercises. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td>   </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 132 </td>\n\t\t\t<td> 5 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/SUSY"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/SUSY">SUSY</a> </td></tr></table></td>\n\t\t\t<!-- <td> This is a classification problem to distinguish between a signal process which produces supersymmetric particles and a background process which does not. </td> -->\n\t\t\t<td>   </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 5000000 </td>\n\t\t\t<td> 18 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/HIGGS"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/HIGGS">HIGGS</a> </td></tr></table></td>\n\t\t\t<!-- <td> This is a classification problem to distinguish between a signal process which produces Higgs bosons and a background process which does not.   </td> -->\n\t\t\t<td>   </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 11000000 </td>\n\t\t\t<td> 28 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Qualitative_Bankruptcy"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Qualitative_Ba

nkruptcy">Qualitative_Bankruptcy</a> </td></tr></table></td>\n\t\t\t<!-- <td> Predict the Bankruptcy from Qualitative parameters from experts. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 250 </td>\n\t\t\t<td> 7 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/LSVT+Voice+Rehabilitation"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/LSVT+Voice+Rehabilitation">LSVT Voice Rehabilitation</a> </td></tr></table></td>\n\t\t\t<!-- <td> 126 samples from 14 participants, 309 features. Aim: assess whether voice rehabilitation treatment lead to phonations considered \'acceptable\' or \'unacceptable\' (binary class classification problem). </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 126 </td>\n\t\t\t<td> 309 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer">Dataset for ADL Recognition with Wrist-worn Accelerometer</a> </td></tr></table></td>\n\t\t\t<!-- <td> Recordings of 16 volunteers performing 14 Activities of Daily Living (ADL) while carrying a single wrist-worn tri-axial accelerometer. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 3 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Wilt"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Wilt">Wilt</a> </td></tr></table></td>\n\t\t\t<!-- <td> High-resolution Remote Sensing data set (Quickbird). Small number of training samples of diseased trees, large number for other land cover. Testing data set from stratified random sample of image. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 4889 </td>\n\t\t\t<td> 6 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/User+Identification+From+Walking+Activity"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/User+Identification+From+Walking+Activity">User Identification From Walking Activity</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset collects data from an Android smartphone positioned in the chest pocket from 22 participants walking in the wild over a predefined path. \r\n </td> -->\n\t\t\t<td> Univariate, Sequential, Time-Series </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer">Activity Recognition from Single Chest-Mounted Accelerometer</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset collects data from a wearable accelerometer mounted on the chest. The dataset is intended for Activity Recognition research purposes. </td> -->\n\t\t\t<td> Univariate, Sequential, Time-Series </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Leaf"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Leaf">Leaf</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset consists in a collection of shape and texture features extracted from digital images of leaf specimens originating from a total of 40 different plant species. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 340 </td>\n\t\t\t<td> 16 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Dresses_Attribute_Sales"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Dresses_Attribute_Sales">Dresses_Attribute_Sales</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset contain Attributes of dresses and their recommendations according to their sales.Sales are monitor on the basis of alternate days.   </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td>

  </td>\n\t\t\t<td> 501 </td>\n\t\t\t<td> 13 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Tamilnadu+Electricity+Board+Hourly+Readings"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Tamilnadu+Electricity+Board+Hourly+Readings">Tamilnadu Electricity Board Hourly Readings</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data can be effectively produced the result to fewer parameter of the Load profile can be reduced in the Database   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Regression, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 45781 </td>\n\t\t\t<td> 5 </td>\n\t\t\t<td> 2013 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Airfoil+Self-Noise"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Airfoil+Self-Noise">Airfoil Self-Noise</a> </td></tr></table></td>\n\t\t\t<!-- <td> NASA data set, obtained from a series of aerodynamic and acoustic tests of two and three-dimensional airfoil blade sections conducted in an anechoic wind tunnel. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 1503 </td>\n\t\t\t<td> 6 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Wholesale+customers"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Wholesale+customers">Wholesale customers</a> </td></tr></table></td>\n\t\t\t<!-- <td> The data set refers to clients of a wholesale distributor. It includes the annual spending in monetary units (m.u.) on diverse product categories </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 440 </td>\n\t\t\t<td> 8 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Business </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis">Twitter Data set for Arabic Sentiment Analysis</a> </td></tr></table></td>\n\t\t\t<!-- <td> This problem of Sentiment Analysis (SA) has been studied well on the English language but not Arabic one. Two main approaches have been devised: corpus-based and lexicon-based.   </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2000 </td>\n\t\t\t<td> 2 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Social </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Combined+Cycle+Power+Plant"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Combined+Cycle+Power+Plant">Combined Cycle Power Plant</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset contains 9568 data points collected from a Combined Cycle Power Plant over 6 years (2006-2011), when the plant was set to work with full load.   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 9568 </td>\n\t\t\t<td> 4 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Urban+Land+Cover"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Urban+Land+Cover">Urban Land Cover</a> </td></tr></table></td>\n\t\t\t<!-- <td> Classification of urban land cover using high resolution aerial imagery. Intended to assist sustainable urban planning efforts. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 168 </td>\n\t\t\t<td> 148 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Diabetes+130-US+hospitals+for+years+1999-2008"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Diabetes+130-US+hospitals+for+years+1999-2008">Diabetes 130-US hospitals for years 1999-2008</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data has been prepared to analyze factors related to readmission as well as other \r\n\r\noutcomes pertaining to patients with diabetes. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 100000 </td>\n\t\t\t<td> 55 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Bach+Choral+Harmony"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Bach+Choral+Harmony">Bach Choral Harmony</a> </td></tr></table></td>\n\t\t\t<!-- <td> The data set consists of 5665 chorales from the composer Johann Sebastian Bach (1685-1750). Each chorale is represented by 60 events, and each event is represented by 4 features (integers).   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 5665 </td>\n\t\t\t<td> 4 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Online+Retail"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Online+Retail">Online Retail</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data set contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Clustering </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 541909 </td>\n\t\t\t<td> 8 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Business </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Dota2+Games+Results"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Dota2+Games+Results">Dota2 Games Results</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data set contains results of matches played in a period of 7 days. Matches played in unbalanced game mode were removed. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 102944 </td>\n\t\t\t<td> 116 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Game </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Gas+Sensor+Array+Drift+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Gas+Sensor+Array+Drift+Dataset">Gas Sensor Array Drift Dataset</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset contains 13910 measurements from 16 chemical sensors utilized in simulations for drift compensation in a discrimination task of 6 gases at various levels of concentrations. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 13910 </td>\n\t\t\t<td> 128 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations">Gas Sensor Array Drift Dataset at Different Concentrations</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset contains 13910 measurements from 16 chemical sensors utilized in simulations for drift compensation in a discrimination task of 6 gases at various levels of concentrations. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 13910 </td>\n\t\t\t<td> 128 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations">Gas Sensor Array Drift Dataset at Different Concentrations

mony">Bach Choral Harmony</a> </td></tr></table></td>\n\t\t\t<!-- <td> The data se t is composed of 60 chorales (5665 events) by J.S. Bach (1675-1750).\r\nEach event of each chorale is labelled using 1 among 101 chord labels and described\r\nthrough 14 features. </td> -->\n\t\t\t<td> Sequential </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td><p cla ss="normal"> </td>\n\t\t\t<td> 5665 </td>\n\t\t\t<td> 1 7 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Other&n bsp; </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/StoneFlakes"><img src="assets/ML images/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Stone Flakes">StoneFlakes</a> </td></tr></table></td>\n\t\t\t<!-- <td> Stone flakes are wa ste products of the stone tool production in\r\nthe prehistoric era. The variables are means of geometric and\r\n stylistic features of the flakes contained in different inventories. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Clustering, Causal-Discovery </ p></td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 79 </td>\n\t\t\ t<td> 8 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td><p c lass="normal">Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="d atasets/Tennis+Major+Tournament+Match+Statistics"><img src="assets/MLimages/SmallLargedefault.jpg" bord er=1 /></a> </td><td> <a href="datasets/Tennis+Major+Tournament+Match+Statis tics">Tennis Major Tournament Match Statistics</a> </td></tr></table></td>\n\t\t\t<!-- <td> This is a collection of 8 files containing the match statistics for both women and men at the four major ten nis tournaments of the year 2013. Each file has 42 columns and a minimum of 76 rows. </td> -->\n\t\ t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Regression, Clustering </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 127 </td>\n\t\t\t<td> 42 </td>\n\t\t\t<td> 2014&nb sp; </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td ><a href="datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings"><img src="asse ts/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/ Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings">Parkinson Speech Dataset with M ultiple Types of Sound Recordings</a> </td></tr></table></td>\n\t\t\t<!-- <td> The tra ining data belongs to 20 Parkinson\'s Disease (PD) patients and 20 healthy subjects. From all subjects, multiple types of sound recordings (26) are taken. </td> -->\n\t\t\t<td> Multivariate </ p></td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Int eger, Real </td>\n\t\t\t<td> 1040 </td>\n\t\t\t<td> 26& nbsp; </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Life < /p></td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Gesture+Phase+Seg mentation"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Gesture+Phase+Segmentation">Gesture Phase Segmentation</a> </td></t r></table></td>\n\t\t\t<!-- <td> The dataset is composed by features extracted from 7 videos with people gesticulating, aiming at studying Gesture Phase Segmentation. It contains 50 attributes divided into two files for each video. </td> -->\n\t\t\t<td> Multivariate, Sequential, Time-Series& nbsp; </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 9900 </td>\n\t\t\t<td> 50&n bsp; </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Perfume+Data"><img src="assets/MLim ages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Perfume +Data">Perfume Data</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data consists o f odors of 20 different perfumes. Data was obtained by using a handheld odor meter (OMX-GR sensor) per sec ond for 28 seconds period. </td> -->\n\t\t\t<td> Univariate, Domain-Theory < /p></td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Int eger </td>\n\t\t\t<td> 560 </td>\n\t\t\t<td> 2 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/BlogFeedback"><img s rc="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="d atasets/BlogFeedback">BlogFeedback</a> </td></tr></table></td>\n\t\t\t<!-- <td> Ins tances in this dataset contain features extracted from blog posts. The task associated with the data is to predict how many comments the post will receive. </td> -->\n\t\t\t<td> Multivariate < /p></td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Integer, Real&n bsp; </td>\n\t\t\t<td> 60021 </td>\n\t\t\t<td> 281 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Social </td > -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/REALDISP+Activity+Recognition+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href=" datasets/REALDISP+Activity+Recognition+Dataset">REALDISP Activity Recognition Dataset</a> </td><
/tr></table></td>\n\t\t\t<!-- <td> The REALDISP dataset is devised to evaluate techniques do

</tr></table></td>\n\t\t\t<!-- <td> The READISP dataset is devised to evaluate techniques de aling with the effects of sensor displacement in wearable activity recognition as well as to benchmark general a ctivity recognition algorithms   </td> -->\n\t\t\t<td> Multivariate, Time-Series < /p></td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 1419 </td>\n\t\t\t<td> 120 </td>\n\t\t \t<td> 2014 </td>\n\t\t\t<!-- <td> Computer </td> -->\ n\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Newspaper+and+magazine+imag es+segmentation+dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td > <a href="datasets/Newspaper+and+magazine+images+segmentation+dataset">Newsp aper and magazine images segmentation dataset</a> </td></tr></table></td>\n\t\t\t<!-- <td> Dataset is well suited for segmentation tasks. It contains 101 scanned pages from different newspapers and magazines in Russian with ground truth pixel-based masks. </td> -->\n\t\t\t<td>   </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> &nb sp; </td>\n\t\t\t<td> 101 </td>\n\t\t\t<td>   </td>\n\t \t\t<td> 2014 </td>\n\t\t\t<!-- <td> Computer </td> --> \n\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/AAAI+2014+Accepted+Papers"><img src="assets/M Limages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/AAAI +2014+Accepted+Papers">AAAI 2014 Accepted Papers</a> </td></tr></table></td>\n\t\t\t<!-- <td><p cl ass="normal">This data set compromises the metadata for the 2014 AAAI conference\'s accepted papers, inclu ding paper titles, authors, abstracts, and keywords of varying granularity. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Clustering </td>\n\t\t\t<td><p cla ss="normal"> </td>\n\t\t\t<td> 399 </td>\n\t\t\t<td> 6& nbsp; </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Computer& nbsp; </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Gas+sensor+a rray+under+flow+modulation"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td> <td> <a href="datasets/Gas+sensor+array+under+flow+modulation">Gas sensor array u nder flow modulation</a> </td></tr></table></td>\n\t\t\t<!-- <td> The data set contain s 58 time series acquired from 16 chemical sensors under gas flow modulation conditions. The sensors were e xposed to different gaseous binary mixtures of acetone and ethanol. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification, Regression < /p></td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 58 </td>\n\t\t \t<td> 120432 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- < td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/ Gas+sensor+array+exposed+to+turbulent+gas+mixtures"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Gas+sensor+array+exposed+to+turb ulent+gas+mixtures">Gas sensor array exposed to turbulent gas mixtures</a> </td></tr></table></td>\ n\t\t\t<!-- <td> A chemical detection platform composed of 8 chemoresistive gas sensors was exposed to turbulent gas mixtures generated naturally in a wind tunnel. The acquired time series of the sensors are provided. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t< td> Classification, Regression </td>\n\t\t\t<td> Real < /td>\n\t\t\t<td> 180 </td>\n\t\t\t<td> 150000 </td>\n\t\t \t<td> 2014 </td>\n\t\t\t<!-- <td> Computer </td> -->\ n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/UJIIndoorLoc"><img src="assets/ MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/UJ IIndoorLoc">UJIIndoorLoc</a> </td></tr></table></td>\n\t\t\t<!-- <td> The UJIIndoorL oc is a Multi-Building Multi-Floor indoor localization database to test Indoor Positioning System that rely on WLA N/WiFi fingerprint. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td><p cla ss="normal">Classification, Regression </td>\n\t\t\t<td> Integer, Real </ td>\n\t\t\t<td> 21048 </td>\n\t\t\t<td> 529 </td>\n\t\t\t< <td> 2014 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\ t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Sentence+Classification"><img src="assets/MLimages/ SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Sentence+Cla ssification">Sentence Classification</a> </td></tr></table></td>\n\t\t\t<!-- <td> Contai ns sentences from the abstract and introduction of 30 articles annotated with a modified Argumentative Zones annotation scheme. These articles come from biology, machine learning and psychology. </td> -->\n\ t\t\t<td> Text </td>\n\t\t\t<td> Classification </td>\n\t\t \t<td> Integer </td>\n\t\t\t<td>   </td>\n\t\t\t<td><p cla ss="normal"> </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Dow+Jones+Index"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Dow+Jones+Index">Dow Jones Index</a> </td></tr></table></td>\n\t\t \t<!-- <td> This dataset contains weekly data for the Dow Jones Industrial Index. It has been

used in computational investing research. </td> -->\n\t\t\t<td> Time-Series </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 750 </td>\n\t\t\t<td> 16 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Business </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/sEMG+for+Basic+Hand+movements"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/sEMG+for+Basic+Hand+movements">sEMG for Basic Hand movements</a> </td></tr></table></td>\n\t\t\t<!-- <td> The “sEMG for Basic Hand movements” includes 2 databases of surface electromyographic signals of 6 hand movements using Delsys\' EMG System. Healthy subjects conducted six daily life grasps. </td> -->\n\t\t\t<td> Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 3000 </td>\n\t\t\t<td> 2500 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/AAAI+2013+Accepted+Papers"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/AAAI+2013+Accepted+Papers">AAAI 2013 Accepted Papers</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data set compromises the metadata for the 2013 AAAI conference\'s accepted papers (main track only), including paper titles, abstracts, and keywords of varying granularity. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Clustering </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 150 </td>\n\t\t\t<td> 5 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Geographical+Original+of+Music"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Geographical+Original+of+Music">Geographical Original of Music</a> </td></tr></table></td>\n\t\t\t<!-- <td> Instances in this dataset contain audio features extracted from 1059 wave files. The task associated with the data is to predict the geographical origin of music.\r\n </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 1059 </td>\n\t\t\t<td> 68 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants">Condition Based Maintenance of Naval Propulsion Plants</a> </td></tr></table></td>\n\t\t\t<!-- <td> Data have been generated from a sophisticated simulator of a Gas Turbines (GT), mounted on a Frigate characterized by a COmbined Diesel eLectric And Gas (CODLAG) propulsion plant type. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 11934 </td>\n\t\t\t<td> 16 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Grammatical+Facial+Expressions"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Grammatical+Facial+Expressions">Grammatical Facial Expressions</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset supports the development of models that make possible to interpret Grammatical Facial Expressions from Brazilian Sign Language (Libras). </td> -->\n\t\t\t<td> Multivariate, Sequential </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 27965 </td>\n\t\t\t<td> 100 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/NoisyOffice"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/NoisyOffice">NoisyOffice</a> </td></tr></table></td>\n\t\t\t<!-- <td> Corpus intended to do cleaning (or binarization) and enhancement of noisy grayscale printed text images using supervised learning methods. Noisy images and their corresponding ground truth provided. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 216 </td>\n\t\t\t<td> 216 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/MHEALTH+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/MHEALTH+Dataset">MHEALTH Dataset</a> </td></tr></table></td>\n\t\t\t<!-- <td> The MHEALTH (Mobile Health) dataset is devised to benchmark techniques dealing with human behavior analysis based on multimodal body sensing. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td><p class="

="normal">120 </td>\n\t\t\t<td> 23 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Student+Performance"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Student+Performance">Student Performance</a> </td></tr></table></td>\n\t\t\t<!-- <td> Predict student performance in secondary education (high school).   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 649 </td>\n\t\t\t<td> 33 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Social </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/ElectricityLoadDiagrams20112014"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/ElectricityLoadDiagrams20112014">ElectricityLoadDiagrams20112014</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data set contains electricity consumption of 370 points/clients.\r\n </td> -->\n\t\t\t<td> Time-Series </td>\n\t\t\t<td> Regression, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 370 </td>\n\t\t\t<td> 140256 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Gas+sensor+array+under+dynamic+gas+mixtures"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Gas+sensor+array+under+dynamic+gas+mixtures">Gas sensor array under dynamic gas mixtures</a> </td></tr></table></td>\n\t\t\t<!-- <td> The data set contains the recordings of 16 chemical sensors exposed to two dynamic gas mixtures at varying concentrations. For each mixture, signals were acquired continuously during 12 hours. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 4178504 </td>\n\t\t\t<td> 19 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/microblogPCU"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/microblogPCU">microblogPCU</a> </td></tr></table></td>\n\t\t\t<!-- <td> MicroblogPCU data is crawled from sina weibo microblog[http://weibo.com/]. This data can be used to study machine learning methods as well as do some social network research.   </td> -->\n\t\t\t<td> Multivariate, Univariate, Sequential, Text </td>\n\t\t\t<td> Classification, Causal-Discovery </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 221579 </td>\n\t\t\t<td> 20 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Firm-Teacher_Clave-Direction_Classification"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Firm-Teacher_Clave-Direction_Classification">Firm-Teacher_Clave-Direction_Classification</a> </td></tr></table></td>\n\t\t\t<!-- <td> The data are binary attack-point vectors and their clave-direction class(es) according to the partido-alto-based paradigm. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 10800 </td>\n\t\t\t<td> 20 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Dataset+for+Sensorless+Drive+Diagnosis"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Dataset+for+Sensorless+Drive+Diagnosis">Dataset for Sensorless Drive Diagnosis</a> </td></tr></table></td>\n\t\t\t<!-- <td> Features are extracted from motor current. The motor has intact and defective components. This results in 11 different classes with different conditions.   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 58509 </td>\n\t\t\t<td> 49 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/TV+News+Channel+Commercial+Detection+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/TV+News+Channel+Commercial+Detection+Dataset">TV News Channel Commercial Detection Dataset</a> </td></tr></table></td>\n\t\t\t<!-- <td> TV Commercials data set consists of standard audio-visual features of video shots extracted from 150 hours of TV news broadcast of 3 Indian and 2 international news channels (30 Hours each).   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 129685 </td>\n\t\t\t<td> 12 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Phishing+Websites"><img src="assets/MLimages/SmallLargedefau

lt.jpg" border=1 /></a> </td><td> <a href="datasets/Phishing+Websites">Phishing Websites</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset collected mainly from: PhishTank archive, MillerSmiles archive, Google’s searching operators. </td> -->\n\t\t\t<td>   </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 2456 </td>\n\t\t\t<td> 30 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Computer Security </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Greenhouse+Gas+Observing+Network"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Greenhouse+Gas+Observing+Network">Greenhouse Gas Observing Network</a> </td></tr></table></td>\n\t\t\t<!-- <td> Design an observing network to monitor emissions of a greenhouse gas (GHG) in California given time series of synthetic observations and tracers from weather model simulations.\r\n </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 2921 </td>\n\t\t\t<td> 5232 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Diabetic+Retinopathy+Debrecen+Data+Set"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Diabetic+Retinopathy+Debrecen+Data+Set">Diabetic Retinopathy Debrecen Data Set</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset contains features extracted from the Messidor image set to predict whether an image contains signs of diabetic retinopathy or not.   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 1151 </td>\n\t\t\t<td> 20 </td>\n\t\t\t<td> 2014 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/HIV-1+protease+cleavage"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/HIV-1+protease+cleavage">HIV-1 protease cleavage</a> </td></tr></table></td>\n\t\t\t<!-- <td> The data contains lists of octamers (8 amino acids) and a flag (-1 or 1) depending on whether HIV-1 protease will cleave in the central position (between amino acids 4 and 5). </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td> 6590 </td>\n\t\t\t<td> 1 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Sentiment+Labelled+Sentences"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Sentiment+Labelled+Sentences">Sentiment Labelled Sentences</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset contains sentences labelled with positive or negative sentiment. </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 3000 </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Online+News+Popularity"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Online+News+Popularity">Online News Popularity</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset summarizes a heterogeneous set of features about articles published by Mashable in a period of two years. The goal is to predict the number of shares in social networks (popularity). </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 39797 </td>\n\t\t\t<td> 61 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Business </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Forest+type+mapping"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Forest+type+mapping">Forest type mapping</a> </td></tr></table></td>\n\t\t\t<!-- <td> Multi-temporal remote sensing data of a forested area in Japan. The goal is to map different forest types using spectral data. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 326 </td>\n\t\t\t<td> 27 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/wiki4HE"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/wiki4HE">wiki4HE</a> </td></tr></table></td>\n\t\t\t<!-- <td> Survey of faculty members from two Spanish universities on teaching uses of Wikipedia </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression, Clustering, Causal-Discovery </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 913 </td>\n\t\t\t<td> 53 </td>\n\t\t\t<td>

2015 </td>\n\t\t\t<!-- <td> Social </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset">Online Video Characteristics and Transcoding Time Dataset</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset contains a million randomly sampled video instances listing 10 fundamental video characteristics along with the YouTube video ID.   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 168286 </td>\n\t\t\t<td> 11 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Chronic_Kidney_Disease"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Chronic_Kidney_Disease">Chronic_Kidney_Disease</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset can be used to predict the chronic kidney disease and it can be collected from the hospital nearly 2 months of period. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 400 </td>\n\t\t\t<td> 25 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014">Machine Learning based ZZ Alpha Ltd. Stock Recommendations 2012-2014</a> </td></tr></table></td>\n\t\t\t<!-- <td> The data here are the ZZAlpha® machine learning recommendations made for various US traded stock portfolios the morning of each day during the 3 year period Jan 1, 2012 - Dec 31, 2014.   </td> -->\n\t\t\t<td> Sequential, Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 314080 </td>\n\t\t\t<td> 0 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Business </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Folio"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Folio">Folio</a> </td></tr></table></td>\n\t\t\t<!-- <td> 20 photos of leaves for each of 32 different species. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 637 </td>\n\t\t\t<td> 20 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015">Taxi Service Trajectory - Prediction Challenge, ECML PKDD 2015</a> </td></tr></table></td>\n\t\t\t<!-- <td> An accurate dataset describing trajectories performed by all the 442 taxis running in the city of Porto, in Portugal.\r\n </td> -->\n\t\t\t<td> Multivariate, Sequential, Time-Series, Domain-Theory </td>\n\t\t\t<td> Clustering, Causal-Discovery </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 1710671 </td>\n\t\t\t<td> 9 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Cuff-Less+Blood+Pressure+Estimation"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Cuff-Less+Blood+Pressure+Estimation">Cuff-Less Blood Pressure Estimation</a> </td></tr></table></td>\n\t\t\t<!-- <td> This Data set provides preprocessed and cleaned vital signals which can be used in designing algorithms for cuff-less estimation of the blood pressure. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 12000 </td>\n\t\t\t<td> 3 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions">Smartphone-Based Recognition of Human Activities and Postural Transitions</a> </td></tr></table></td>\n\t\t\t<!-- <td> Activity recognition data set built from the recordings of 30 subjects performing basic activities and postural transitions while carrying a waist-mounted smartphone with embedded inertial sensors.\r\n </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 10929 </td>\n\t\t\t<td> 561 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td>

Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Mice+Protein+Expression"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Mice+Protein+Expression">Mice Protein Expression</a> </td></tr></table></td>\n\t\t\t<!-- <td> Expression levels of 77 proteins measured in the cerebral cortex of 8 classes of control and Down syndrome mice exposed to context fear conditioning, a task used to assess associative learning. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 1080 </td>\n\t\t\t<td> 82 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/UJIIndoorLoc-Mag"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/UJIIndoorLoc-Mag">UJIIndoorLoc-Mag</a> </td></tr></table></td>\n\t\t\t<!-- <td> The UJIIndoorLoc-Mag is an indoor localization database to test Indoor Positioning System that rely on Earth\'s magnetic field variations. </td> -->\n\t\t\t<td> Multivariate, Sequential, Time-Series </td>\n\t\t\t<td> Classification, Regression, Clustering </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 40000 </td>\n\t\t\t<td> 13 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Heterogeneity+Activity+Recognition"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Heterogeneity+Activity+Recognition">Heterogeneity Activity Recognition</a> </td></tr></table></td>\n\t\t\t<!-- <td> The Heterogeneity Human Activity Recognition (HHAR) dataset from Smartphones and Smartwatches is a dataset devised to benchmark human activity recognition algorithms (classification, automatic data segmentation, sensor fusion, feature extraction, etc.) in real-world contexts; specifically, the dataset is gathered with a variety of different device models and use-scenarios, in order to reflect sensing heterogeneities to be expected in real deployments. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 43930257 </td>\n\t\t\t<td> 16 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set">Educational Process Mining (EPM): A Learning Analytics Data Set</a> </td></tr></table></td>\n\t\t\t<!-- <td> Educational Process Mining data set is built from the recordings of 115 subjects\' activities through a logging application while learning with an educational simulator. </td> -->\n\t\t\t<td> Multivariate, Sequential, Time-Series </td>\n\t\t\t<td> Classification, Regression, Clustering </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 230318 </td>\n\t\t\t<td> 13 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/HEPMASS"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/HEPMASS">HEPMASS</a> </td></tr></table></td>\n\t\t\t<!-- <td> The search for exotic particles requires sorting through a large number of collisions to find the events of interest. This data set challenges one to detect a new particle of unknown mass. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 10500000 </td>\n\t\t\t<td> 28 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Indoor+User+Movement+Prediction+from+RSS+data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Indoor+User+Movement+Prediction+from+RSS+data">Indoor User Movement Prediction from RSS data</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset contains temporal data from a Wireless Sensor Network deployed in real-world office environments. The task is intended as real-life benchmark in the area of Ambient Assisted Living. </td> -->\n\t\t\t<td> Multivariate, Sequential, Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 13197 </td>\n\t\t\t<td> 4 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Open+University+Learning+Analytics+dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Open+University+Learning+Analytics+dataset">Open University Learning Analytics dataset</a> </td></tr></table></td>\n\t\t\t<!-- <td> Open University Learning Analytics Dataset contains data about courses, students and their interactions with Virtual Learning Environment for seven selected courses and more than 30000 students. </td> -->\n\t\t\t<td> Multivariate, Sequential, Time-Se

ries </td>\n\t\t\t<td> Classification, Regression, Clustering </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/default+of+credit+card+clients"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/default+of+credit+card+clients">default of credit card clients</a> </td></tr></table></td>\n\t\t\t<!-- <td> This research aimed at the case of customers' default payments in Taiwan and compares the predictive accuracy of probability of default among six data mining methods. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 30000 </td>\n\t\t\t<td> 24 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Business </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Mesothelioma%E2%80%99s+disease+data+set+"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Mesothelioma%E2%80%99s+disease+data+set+">Mesothelioma's disease data set </a> </td></tr></table></td>\n\t\t\t<!-- <td> Mesothelioma's disease data set were prepared at Dicle University Faculty of Medicine in Turkey.\r\nThree hundred and twenty-four Mesothelioma patient data. In the dataset, all samples have 34 features. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 324 </td>\n\t\t\t<td> 34 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Online+Retail"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Online+Retail">Online Retail</a> </td></tr></table></td>\n\t\t\t<!-- <td> This is a transnational data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail. </td> -->\n\t\t\t<td> Multivariate, Sequential, Time-Series </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 541909 </td>\n\t\t\t<td> 8 </td>\n\t\t\t<td> 2015 </td>\n\t\t\t<!-- <td> Business </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/SIFT10M"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/SIFT10M">SIFT10M</a> </td></tr></table></td>\n\t\t\t<!-- <td> In SIFT10M, each data point is a SIFT feature which is extracted from Caltech-256 by the open source VLFeat library. The corresponding patches of the SIFT features are provided. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Causal-Discovery </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 11164866 </td>\n\t\t\t<td> 128 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/GPS+Trajectories"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/GPS+Trajectories">GPS Trajectories</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset has been feed by Android app called Go!Track. It is available at Goolge Play Store(https://play.google.com/store/apps/details?id=com.go.router).   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 163 </td>\n\t\t\t<td> 15 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Detect+Malacious+Executable%28AntiVirus%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Detect+Malacious+Executable%28AntiVirus%29">Detect Malacious Executable(AntiVirus)</a> </td></tr></table></td>\n\t\t\t<!-- <td> I extract features from malacious and non-malacious and create and training dataset to teach svm classifier.Dataset made of unknown executable to detect if it is virus or normal safe executable. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 373 </td>\n\t\t\t<td> 513 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Occupancy+Detection+"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Occupancy+Detection+">Occupancy Detection </a> </td></tr></table></td>\n\t\t\t<!-- <td> Experimental data used for binary classification (room occupancy) from Temperature,Humidity,Light and CO2. Ground-truth occupancy was obtained from time stamped pictures that were taken every minute. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 20560 </td>\n\t\t\t<td> 7 </td>\n\t\t\t<td> 2016

></td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Parkinson%E2%80%99s+Disease"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Parkinson%E2%80%99s+Disease">Improved Spiral Test Using Digitized Graphics Tablet for Monitoring Parkinson’s Disease</a> </td></tr></table></td>\n\t\t\t<!-- <td> Handwriting database consists of 25 PWP(People with Parkinson) and 15 healthy individuals.Three types of recordings (Static Spiral Test, Dynamic Spiral Test and Stability Test) are taken. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Regression, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 40 </td>\n\t\t\t<td> 7 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/News+Aggregator"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/News+Aggregator">News Aggregator</a> </td></tr></table></td>\n\t\t\t<!-- <td> References to news pages collected from an web aggregator in the period from 10-March-2014 to 10-August-2014. The resources are grouped into clusters that represent pages discussing the same story.   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 422937 </td>\n\t\t\t<td> 5 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Air+Quality"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Air+Quality">Air Quality</a> </td></tr></table></td>\n\t\t\t<!-- <td> Contains the responses of a gas multisensor device deployed on the field in an Italian city. Hourly responses averages are recorded along with gas concentrations references from a certified analyzer.   </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 9358 </td>\n\t\t\t<td> 15 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Twin+gas+sensor+arrays"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Twin+gas+sensor+arrays">Twin gas sensor arrays</a> </td></tr></table></td>\n\t\t\t<!-- <td> 5 replicates of an 8-MOX gas sensor array were exposed to different gas conditions (4 volatiles at 10 concentration levels each). </td> -->\n\t\t\t<td> Multivariate, Time-Series, Domain-Theory </td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 640 </td>\n\t\t\t<td> 480000 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Gas+sensors+for+home+activity+monitoring"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Gas+sensors+for+home+activity+monitoring">Gas sensors for home activity monitoring</a> </td></tr></table></td>\n\t\t\t<!-- <td> 100 recordings of a sensor array under different conditions in a home setting: background, wine and banana presentations. The array includes 8 MOX gas sensors, and humidity and temperature sensors.\r\n </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 919438 </td>\n\t\t\t<td> 11 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Facebook+Comment+Volume+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Facebook+Comment+Volume+Dataset">Facebook Comment Volume Dataset</a> </td></tr></table></td>\n\t\t\t<!-- <td> Instances in this dataset contain features extracted from facebook posts. The task associated with the data is to predict how many comments the post will receive. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 40949 </td>\n\t\t\t<td> 54 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Living+%28AAL%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Living+%28AAL%29">Smartphone Dataset for Human Activity Recognition (HAR) in Ambient Assisted Living (AAL)</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data is an addition to an existing dataset on UCI. We collected more data to improve the accuracy of our human activity recognition algorithms applied in the domain of Ambient Assisted Living.   </td> -->\n\t\t\t<td> Time-Series </td>\n\t\t\t<td> Cla

ssification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 5744 </td>\n\t\t\t<td> 561 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Polish+companies+bankruptcy+data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Polish+companies+bankruptcy+data">Polish companies bankruptcy data</a> </td></tr></table></td>\n\t\t\t<!-- <td> The data set is about bankruptcy prediction of Polish companies.The bankrupt companies were analyzed in the period 2000-2012, while the still operating companies were evaluated from 2007 to 2013. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 10503 </td>\n\t\t\t<td> 64 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Business </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29">Activity Recognition system based on Multisensor data fusion (AReM)</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset contains temporal data from a Wireless Sensor Network worn by an actor performing the activities: bending, cycling, lying down, sitting, standing, walking. </td> -->\n\t\t\t<td> Multivariate, Sequential, Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 42240 </td>\n\t\t\t<td> 6 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Dota2+Games+Results"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Dota2+Games+Results">Dota2 Games Results</a> </td></tr></table></td>\n\t\t\t<!-- <td> Dota 2 is a popular computer game with two teams of 5 players. At the start of the game each player chooses a unique hero with different strengths and weaknesses. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 102944 </td>\n\t\t\t<td> 116 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Game </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Facebook+metrics"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Facebook+metrics">Facebook metrics</a> </td></tr></table></td>\n\t\t\t<!-- <td> Facebook performance metrics of a renowned cosmetic\'s brand Facebook page. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 500 </td>\n\t\t\t<td> 19 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Business </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/UbiqLog+%28smartphone+lifelogging%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/UbiqLog+%28smartphone+lifelogging%29">UbiqLog (smartphone lifelogging)</a> </td></tr></table></td>\n\t\t\t<!-- <td> UbiqLog is the smartphone lifelogging tool that runs on the smartphone of 35 users for about 2 months. \r\n </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Causal-Discovery </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 9782222 </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/NIPS+Conference+Papers+1987-2015"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/NIPS+Conference+Papers+1987-2015">NIPS Conference Papers 1987-2015</a> </td></tr></table></td>\n\t\t\t<!-- <td> This data set contains the distribution of words in the full text of the NIPS conference papers published from 1987 to 2015. </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td> Clustering </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 11463 </td>\n\t\t\t<td> 5812 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/HTRU2"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/HTRU2">HTRU2</a> </td></tr></table></td>\n\t\t\t<!-- <td> Pulsar candidates collected during the HTRU survey. Pulsars are a type of star, of considerable scientific interest. Candidates must be classified in to pulsar and non-pulsar classes to aid discovery. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 17898 </td>\n\t\t\t<td> 9 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Drug+consumption+%28quantified%29"><img src="assets/

EEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Drug+consumption+%28quantified%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Drug+consumption+%28quantified%29">Drug consumption (quantified)</a> </td></tr></table></td>\n\t\t\t<!-- <td> Classify type of drug consumer by personality data </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 1885 </td>\n\t\t\t<td> 32 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Social </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Appliances+energy+prediction"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Appliances+energy+prediction">Appliances energy prediction</a> </td></tr></table></td>\n\t\t\t<!-- <td> Experimental data used to create regression models of appliances energy use in a low energy building. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 19735 </td>\n\t\t\t<td> 29 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Miskolc+IIS+Hybrid+IPS"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Miskolc+IIS+Hybrid+IPS">Miskolc IIS Hybrid IPS</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset was created for the comparison and evaluation of hybrid indoor positioning methods. The dataset presented contains data from W-LAN and Bluetooth interfaces, and Magnetometer.   </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td> Classification, Clustering, Causal-Discovery </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 1540 </td>\n\t\t\t<td> 67 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/KDC-4007+dataset+Collection"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/KDC-4007+dataset+Collection">KDC-4007 dataset Collection</a> </td></tr></table></td>\n\t\t\t<!-- <td> KDC-4007 dataset Collection is the Kurdish Documents Classification text used in categories regarding Kurdish Sorani news and articles. </td> -->\n\t\t\t<td> Multivariate, Text </td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 4007 </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Geo-Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Geo-Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone">Geo-Magnetic field and WLAN dataset for indoor localisation from wristband and smartphone</a> </td></tr></table></td>\n\t\t\t<!-- <td> A multisource and multivariate dataset for indoor localisation methods based on WLAN and Geo-Magnetic field fingerprinting </td> -->\n\t\t\t<td> Multivariate, Sequential, Time-Series </td>\n\t\t\t<td> Classification, Regression, Clustering </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 153540 </td>\n\t\t\t<td> 25 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/DrivFace"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/DrivFace">DrivFace</a> </td></tr></table></td>\n\t\t\t<!-- <td> The DrivFace contains images sequences of subjects while driving in real scenarios. It is composed of 606 samples of 640×480, acquired over different days from 4 drivers with several facial features. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Regression, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 606 </td>\n\t\t\t<td> 6400 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Website+Phishing"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Website+Phishing">Website Phishing</a> </td></tr></table></td>\n\t\t\t<!-- <td> \r\n\r\n </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 1353 </td>\n\t\t\t<td> 10 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/YouTube+Spam+Collection"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/YouTube+Spam+Collection">YouTube Spam Collection</a> </td></tr></table></td>\n\t\t\t<!-- <td> It is a public set of comments collected for spam research. It has five datasets composed by 1,956 real messages extracted from five videos that were among the 10 most viewed on the collecti

on period. </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 1956 </td>\n\t\t\t<td> 5 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Beijing+PM2.5+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Beijing+PM2.5+Data">Beijing PM2.5 Data</a> </td></tr></table></td>\n\t\t\t<!-- <td> This hourly data set contains the PM2.5 data of US Embassy in Beijing. Meanwhile, meteorological data from Beijing Capital International Airport are also included.   </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 43824 </td>\n\t\t\t<td> 13 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Cargo+2000+Freight+Tracking+and+Tracing"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Cargo+2000+Freight+Tracking+and+Tracing">Cargo 2000 Freight Tracking and Tracing</a> </td></tr></table></td>\n\t\t\t<!-- <td> Sanitized and anonymized Cargo 2000 (C2K) airfreight tracking and tracing events, covering five months of business execution (3,942 process instances, 7,932 transport legs, 56,082 activities).   </td> -->\n\t\t\t<td> Multivariate, Sequential </td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 3942 </td>\n\t\t\t<td> 98 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Business </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Cervical+cancer+%28Risk+Factors%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Cervical+cancer+%28Risk+Factors%29">Cervical cancer (Risk Factors)</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset focuses on the prediction of indicators/diagnosis of cervical cancer. The features cover demographic information, habits, and historic medical records. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 858 </td>\n\t\t\t<td> 36 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Quality+Assessment+of+Digital+Colposcopies"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Quality+Assessment+of+Digital+Colposcopies">Quality Assessment of Digital Colposcopies</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset explores the subjective quality assessment of digital colposcopies. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 287 </td>\n\t\t\t<td> 69 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/KASANDR"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/KASANDR">KASANDR</a> </td></tr></table></td>\n\t\t\t<!-- <td> KASANDR is a novel, publicly available collection for recommendation systems that records the behavior of customers of the European leader in e-Commerce advertising, Kelkoo.   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Causal-Discovery </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 17764280 </td>\n\t\t\t<td> 2158859 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/FMA%3A+A+Dataset+For+Music+Analysis"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/FMA%3A+A+Dataset+For+Music+Analysis">FMA: A Dataset For Music Analysis</a> </td></tr></table></td>\n\t\t\t<!-- <td> FMA features 106,574 tracks and includes song title, album, artist, genres; play counts, favorites, comments; description, biography, tags; together with audio (343 days, 917 GiB) and features. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 106574 </td>\n\t\t\t<td> 518 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Air+quality"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Air+quality">Air quality</a> </td></tr></table></td>\n\t\t\t<!-- <td> Contains the responses of a gas multisensor device deployed on the field in an Italian city.   </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 9358 </td>\n\t\t\t<td> 15 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><

tr><td><a href="datasets/Epileptic+Seizure+Recognition"><img src="assets/MLimages/SmallLargedefault.jpg" b
order=1 /></a> </td><td> <a href="datasets/Epileptic+Seizure+Recognition">Epilep
tic Seizure Recognition</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset is a p
re-processed and re-structured/reshaped version of a very commonly used dataset featuring epileptic seizure d
etection.   </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td><
p class="normal">Classification, Clustering </td>\n\t\t\t<td> Integer, Real </p
></td>\n\t\t\t<td> 11500 </td>\n\t\t\t<td> 179 </td>\n\t
\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t
</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Devanagari+Handwritten+Character+D
ataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"
><a href="datasets/Devanagari+Handwritten+Character+Dataset">Devanagari Handwritten Character Data
set</a> </td></tr></table></td>\n\t\t\t<!-- <td> This is an image database of Handwrit
ten Devanagari characters. There are 46 classes of characters with 2000 examples each. The dataset is split in
to training set(85%) and testing set(15%).   </td> -->\n\t\t\t<td>   </td>\n\t
\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\
t\t\t<td> 92000 </td>\n\t\t\t<td>   </td>\n\t\t\t<td><p cl
ass="normal">2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><
tr>\n\t\t\t<td><table><tr><td><a href="datasets/Stock+portfolio+performance"><img src="assets/MLimages/Sma
llLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Stock+portfolio+pe
rformance">Stock portfolio performance</a> </td></tr></table></td>\n\t\t\t<!-- <td> T
he data set of performances of weighted scoring stock portfolios are obtained with mixture design from the US
stock market historical database. </td> -->\n\t\t\t<td> Multivariate </td>\
n\t\t\t<td> Regression </td>\n\t\t\t<td> Real </td>\n\t\t
\t<td> 315 </td>\n\t\t\t<td> 12 </td>\n\t\t\t<td><p clas
s="normal">2016 </td>\n\t\t\t<!-- <td> Business </td> -->\n\t\t\t</tr><tr b
gcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/MoCap+Hand+Postures"><img src="assets/MLi
mages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/MoCa
p+Hand+Postures">MoCap Hand Postures</a> </td></tr></table></td>\n\t\t\t<!-- <td><p class="normal"
>5 types of hand postures from 12 users were recorded using unlabeled markers attached to fingers of a glove
in a motion capture environment. Due to resolution and occlusion, missing values are common. </td>
-->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Cluster
ing </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 780
95 </td>\n\t\t\t<td> 38 </td>\n\t\t\t<td> 2016 </
p></td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td>
<a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="
datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech">Early biomark
ers of Parkinson&#65533;s disease based on natural connected speech</a> </td></tr></table></td>\n\t\t\t<!--
<td> Predict a pattern of neurodegeneration in the dataset of speech features obtained from
patients with early untreated Parkinson’s disease and patients at high risk developing Parkinson’s disease.&nbs
p; </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classifica
tion, Regression </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td><p class=
"normal">130 </td>\n\t\t\t<td> 65 </td>\n\t\t\t<td> 20
17 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\
n\t\t\t<td><table><tr><td><a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Educ
ation+Setting"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="n
ormal"><a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting"
>Data for Software Engineering Teamwork Assessment in Education Setting</a> </td></tr></table></td
>\n\t\t\t<!-- <td> Data include over 100 Team Activity Measures and outcomes (ML classes)
obtained from activities of 74 student teams during the creation of final class project in SW Eng. classes at SFS
U, Fulda, FAU </td> -->\n\t\t\t<td> Sequential, Time-Series </td>\n\t\t\t<
td> Classification </td>\n\t\t\t<td> Integer, Real </td>\
n\t\t\t<td> 74 </td>\n\t\t\t<td> 102 </td>\n\t\t\t<td><p
class="normal">2017 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr>
<tr>\n\t\t\t<td><table><tr><td><a href="datasets/PM2.5+Data+of+Five+Chinese+Cities"><img src="assets/MLim
ages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/PM2.5+
Data+of+Five+Chinese+Cities">PM2.5 Data of Five Chinese Cities</a> </td></tr></table></td>\n\t\t\t<!--
<td> This hourly data set contains the PM2.5 data in Beijing, Shanghai, Guangzhou, Chengd
u and Shenyang. Meanwhile, meteorological data for each city are also included. </td> -->\n\t\t\t<td>
Multivariate, Time-Series </td>\n\t\t\t<td> Regression </
p></td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 52854 </
p></td>\n\t\t\t<td> 86 </td>\n\t\t\t<td> 2017 </td>\n\t\t

\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet">Parkinson Disease Spiral Drawings Using Digitized Graphics Tablet</a> </td></tr></table></td>\n\t\t\t<!-- <td> Handwriting database consists of 62 PWP(People with Parkinson) and 15 healthy individuals. Three types of recordings (Static Spiral Test, Dynamic Spiral Test and Stability Test) are taken. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Regression, Clustering </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 77 </td>\n\t\t\t<td> 7 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Sales_Transactions_Dataset_Weekly"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly</a> </td></tr></table></td>\n\t\t\t<!-- <td> Contains weekly purchased quantities of 800 over products over 52 weeks. Normalised values are provided too. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Clustering </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 811 </td>\n\t\t\t<td> 53 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td>   </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Las+Vegas+Strip"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Las+Vegas+Strip">Las Vegas Strip</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset includes quantitative and categorical features from online reviews from 21 hotels located in Las Vegas Strip, extracted from TripAdvisor (http://www.tripadvisor.com). </td> -->\n\t\t\t<td>   </td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 504 </td>\n\t\t\t<td> 20 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Business </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Eco-hotel"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Eco-hotel">Eco-hotel</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset includes Online Textual Reviews from both online (e.g., TripAdvisor) and offline (e.g., Guests\' book) sources from the Areias do Seixo Eco-Resort. </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 401 </td>\n\t\t\t<td> 1 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Business </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/MEU-Mobile+KSD"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/MEU-Mobile+KSD">MEU-Mobile KSD</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset contains keystroke dynamics data collected on a touch mobile device (Nexus 7). The dataset contains 2856 records, 51 records per subject for 56 subjects.   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 2856 </td>\n\t\t\t<td> 71 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Crowdsourced+Mapping"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Crowdsourced+Mapping">Crowdsourced Mapping</a> </td></tr></table></td>\n\t\t\t<!-- <td> Crowdsourced data from OpenStreetMap is used to automate the classification of satellite images into different land cover classes (impervious, farm, forest, grass, orchard, water).   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 10546 </td>\n\t\t\t<td> 29 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/gene+expression+cancer+RNA-Seq"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/gene+expression+cancer+RNA-Seq">gene expression cancer RNA-Seq</a> </td></tr></table></td>\n\t\t\t<!-- <td> This collection of data is part of the RNA-Seq (HiSeq) PANCAN data set, it is a random extraction of gene expressions of patients having different types of tumor: BRCA, KIRC, COAD, LUAD and PRAD. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 801 </td>\n\t\t\t<td> 20531 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Hybrid+Indoor+Positioning+D

ataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer">Hybrid Indoor Positioning Dataset from WiFi RSSI, Bluetooth and magnetometer</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset was created for the comparison and evaluation of hybrid indoor positioning methods. The dataset presented contains data from W-LAN and Bluetooth interfaces, and Magnetometer.   </td> -->\n\t\t\t<td> Multivariate, Sequential, Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 1540 </td>\n\t\t\t<td> 65 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/chestnut+%E2%80%93+LARVIC"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/chestnut+%E2%80%93+LARVIC">chestnut – LARVIC</a> </td></tr></table></td>\n\t\t\t<!-- <td> The research project presents this database, shows the images of chestnuts that will be processed to determine the presence or absence of defects </td> -->\n\t\t\t<td>   </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 1451 </td>\n\t\t\t<td> 3 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%28OBS%29+Network"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%28OBS%29+Network">Burst Header Packet (BHP) flooding attack on Optical Burst Switching (OBS) Network</a> </td></tr></table></td>\n\t\t\t<!-- <td> One of the primary challenges in identifying the risks of the Burst Header Packet (BHP) flood attacks in Optical Burst Switching networks (OBS) is the scarcity of reliable historical data.   </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 1075 </td>\n\t\t\t<td> 22 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Motion+Capture+Hand+Postures"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Motion+Capture+Hand+Postures">Motion Capture Hand Postures</a> </td></tr></table></td>\n\t\t\t<!-- <td> 5 types of hand postures from 12 users were recorded using unlabeled markers on fingers of a glove in a motion capture environment. Due to resolution and occlusion, missing values are common. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 78095 </td>\n\t\t\t<td> 38 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Anuran+Calls+%28MFCCs%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Anuran+Calls+%28MFCCs%29">Anuran Calls (MFCCs)</a> </td></tr></table></td>\n\t\t\t<!-- <td> Acoustic features extracted from syllables of anuran (frogs) calls, including the family, the genus, and the species labels (multilabel).   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 7195 </td>\n\t\t\t<td> 22 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization">TTC-3600: Benchmark dataset for Turkish text categorization</a> </td></tr></table></td>\n\t\t\t<!-- <td> The TTC-3600 data set is a collection of Turkish news and articles including categorized 3,600 documents from 6 well-known portals in Turkey. It has 4 different forms in ARFF Weka format. </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 3600 </td>\n\t\t\t<td> 4814 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy">Gastrointestinal Lesions in Regular Colonoscopy</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset contains features extracted from colonoscopy videos used to detect gastrointestinal lesions. It contains 76 lesions: 15 serrated adenomas, 21 hyperplastic lesions and 40 adenoma.   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 76 </td>\n\t\t\t<td> 698 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="da

tasets/Daily+Demand+Forecasting+Orders"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Daily+Demand+Forecasting+Orders">Daily Demand Forecasting Orders</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset was collected during 60 days, this is a real database of a brazilian logistics company. </td> -->\n\t\t\t<td> Time-Series </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 60 </td>\n\t\t\t<td> 13 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Business </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Paper+Reviews"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Paper+Reviews">Paper Reviews</a> </td></tr></table></td>\n\t\t\t<!-- <td> This sentiment analysis data set contains scientific paper reviews from an international conference on computing and informatics. The task is to predict the orientation or the evaluation of a review. </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 405 </td>\n\t\t\t<td> 10 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/extention+of+Z-Alizadeh+sani+dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/extention+of+Z-Alizadeh+sani+dataset">extention of Z-Alizadeh sani dataset</a> </td></tr></table></td>\n\t\t\t<!-- <td> It was collected for CAD diagnosis. </td> -->\n\t\t\t<td>   </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 303 </td>\n\t\t\t<td> 59 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Z-Alizadeh+Sani"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Z-Alizadeh+Sani">Z-Alizadeh Sani</a> </td></tr></table></td>\n\t\t\t<!-- <td> It was collected for CAD diagnosis. </td> -->\n\t\t\t<td>   </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 303 </td>\n\t\t\t<td> 56 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Dynamic+Features+of+VirusShare+Executables"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Dynamic+Features+of+VirusShare+Executables">Dynamic Features of VirusShare Executables</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset contains the dynamic features of 107,888 executables, collected by VirusShare from Nov/2010 to Jul/2014. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 107888 </td>\n\t\t\t<td> 482 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/IDA2016Challenge"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/IDA2016Challenge">IDA2016Challenge</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset consists of data collected from heavy Scania trucks in everyday usage.   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 76000 </td>\n\t\t\t<td> 171 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/DSRC+Vehicle+Communications"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/DSRC+Vehicle+Communications">DSRC Vehicle Communications</a> </td></tr></table></td>\n\t\t\t<!-- <td> This set Provides data regarding wireless communications between vehicles and road side units. two separate data sets are provided (normal scenario) and in the presence of attacker (jammer). </td> -->\n\t\t\t<td> Sequential, Text </td>\n\t\t\t<td> Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 10000 </td>\n\t\t\t<td> 5 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Mturk+User-Perceived+Clusters+over+Images"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Mturk+User-Perceived+Clusters+over+Images">Mturk User-Perceived Clusters over Images</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset was collected by Shan-Hung Wu and DataLab members at NTHU, Taiwan. There\'re 325 user-perceived clusters from 100 users and their corresponding descriptions. </td> -->\n\t\t\t<td> Multivariate, Text </td>\n\t\t\t<td> Clustering </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 180 </td>\n\t\t\t<td> 500 </td>\n\t\t\t<td><p

class="normal">2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Character+Font+Images"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Character+Font+Images">Character Font Images</a> </td></tr></table></td>\n\t\t\t<!-- <td> Character images from scanned and computer generated fonts. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 745000 </td>\n\t\t\t<td> 411 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Instance+Labels"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Instance+Labels">DeliciousMIL: A Data Set for Multi-Label Multi-Instance Learning with Instance Labels</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset includes 1) 12234 documents (8251 training, 3983 test) extracted from DeliciousT140 dataset, 2) class labels for all documents, 3) labels for a subset of sentences of the test documents. </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 12234 </td>\n\t\t\t<td> 8519 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++">Autistic Spectrum Disorder Screening Data for Children </a> </td></tr></table></td>\n\t\t\t<!-- <td> Children screening data for autism suitable for classification and predictive tasks   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 292 </td>\n\t\t\t<td> 21 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++">Autistic Spectrum Disorder Screening Data for Adolescent </a> </td></tr></table></td>\n\t\t\t<!-- <td> Autistic Spectrum Disorder Screening Data for Adolescent. This dataset is related to classification and predictive tasks. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 104 </td>\n\t\t\t<td> 21 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/APS+Failure+at+Scania+Trucks"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/APS+Failure+at+Scania+Trucks">APS Failure at Scania Trucks</a> </td></tr></table></td>\n\t\t\t<!-- <td> The datasets\' positive class consists of component failures for a specific component of the APS system. The negative class consists of trucks with failures for components not related to the APS. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 60000 </td>\n\t\t\t<td> 171 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Wireless+Indoor+Localization"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Wireless+Indoor+Localization">Wireless Indoor Localization</a> </td></tr></table></td>\n\t\t\t<!-- <td> Collected in indoor space by observing signal strengths of seven WiFi signals visible on a smartphone. The decision variable is one of the four rooms.   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 2000 </td>\n\t\t\t<td> 7 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/HCC+Survival"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/HCC+Survival">HCC Survival</a> </td></tr></table></td>\n\t\t\t<!-- <td> Hepatocellular Carcinoma dataset (HCC dataset) was collected at a University Hospital in Portugal. It contains real clinical data of 165 patients diagnosed with HCC. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 165 </td>\n\t\t\t<td> 49 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015"><img src="assets/MLimages/

SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015">CSM (Conventional and Social Media Movies) Dataset 2014 and 2015</a> </td></tr></table></td>\n\t\t\t<!-- <td> 12 features categorized as conventional and social media features. Both conventional features, collected from movies databases on Web as well as social media features(YouTube,Twitter). </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 217 </td>\n\t\t\t<td> 12 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29">University of Tehran Question Dataset 2016 (UTQD.2016)</a> </td></tr></table></td>\n\t\t\t<!-- <td> Persian questions gathered from a jeopardy game broadcasted on Iranian national television.   </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 1175 </td>\n\t\t\t<td> 3 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Autism+Screening+Adult"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Autism+Screening+Adult">Autism Screening Adult</a> </td></tr></table></td>\n\t\t\t<!-- <td> Autistic Spectrum Disorder Screening Data for Adult. This dataset is related to classification and predictive tasks. </td> -->\n\t\t\t<td>   </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 704 </td>\n\t\t\t<td> 21 </td>\n\t\t\t<td> 2017 </td>\n\t\t\t<!-- <td> Social </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor">Activity recognition with healthy older people using a batteryless wearable sensor</a> </td></tr></table></td>\n\t\t\t<!-- <td> Sequential motion data from 14 healthy older people aged 66 to 86 years old using a batteryless, wearable sensor on top of their clothing for the recognition of activities in clinical environments. </td> -->\n\t\t\t<td> Sequential </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 75128 </td>\n\t\t\t<td> 9 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Immunotherapy+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Immunotherapy+Dataset">Immunotherapy Dataset</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset contains information about wart treatment results of 90 patients using immunotherapy. </td> -->\n\t\t\t<td> Univariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 90 </td>\n\t\t\t<td> 8 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Cryotherapy+Dataset+"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Cryotherapy+Dataset+">Cryotherapy Dataset </a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset contains information about wart treatment results of 90 patients using cryotherapy. </td> -->\n\t\t\t<td> Univariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 90 </td>\n\t\t\t<td> 7 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes">OCT data & Color Fundus Images of Left & Right Eyes</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset contains OCT data (in mat format) and color fundus data (in jpg format) of left & right eyes of 50 healthy persons. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 50 </td>\n\t\t\t<td> 2 </td>\n\t\t\t<td> 2016 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Discrete+Tone+Image+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Discrete+Tone+Image+Dataset">Discrete Tone Image Dataset</a> </td></tr></table></td>\n\t\t\t<!-- <td> Discrete Tone Images (DTI) are available which needs to b

d></tr></table></td>\n\t\t\t<!-- <td> Discrete Tone Images(DTI)are available which needs to b e analyzed in detail. Here, we created this dataset for those who do research in DTI. \r\n </td> -->\n\t \t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </t d>\n\t\t\t<td>   </td>\n\t\t\t<td> 71 </td>\n\t\t\t<td><p c lass="normal">11 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/News+Popularity+in+Multiple+Social+Media+Platforms"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/News+Popularity+in+Multiple+Social+Media+Platforms" >News Popularity in Multiple Social Media Platforms</a> </td></tr></table></td>\n\t\t\t<!-- <td> Large data set of news items and their respective social feedback on multiple platforms: Facebook, G oogle+ and LinkedIn. </td> -->\n\t\t\t<td> Multivariate, Time-Series, Text </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Integer, Real&nbs p; </td>\n\t\t\t<td> 93239 </td>\n\t\t\t<td> 11 </td >\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </td > -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Ultrasonic+flowmeter+diagn ostics"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Ultrasonic+flowmeter+diagnostics">Ultrasonic flowmeter diagnostics</a> </td></ tr></table></td>\n\t\t\t<!-- <td> Fault diagnosis of four liquid ultrasonic flowmeters </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification&n bsp; </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 540 </t d>\n\t\t\t<td> 173 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!- - <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datase ts/ICMLA+2014+Accepted+Papers+Data+Set"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </a> </td><td> <a href="datasets/ICMLA+2014+Accepted+Papers+Data+Set">ICM LA 2014 Accepted Papers Data Set</a> </td></tr></table></td>\n\t\t\t<!-- <td> This d ata set compromises the metadata for the 2014 ICMLA conference\'s accepted papers, including ID, paper title s, author\'s keywords, abstracts and sessions in which they were exposed. </td> -->\n\t\t\t<td><p clas s="normal">Multivariate </td>\n\t\t\t<td> Classification, Clustering </td>\ n\t\t\t<td>   </td>\n\t\t\t<td> 105 </td>\n\t\t\t<td><p cla ss="normal">5 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/B LE+RSSI+Dataset+for+Indoor+localization+and+Navigation"><img src="assets/MLimages/SmallLargedefault.jp g" border=1 /></a> </td><td> <a href="datasets/BLE+RSSI+Dataset+for+Indoor+lo calization+and+Navigation">BLE RSSI Dataset for Indoor localization and Navigation</a> </td></tr></ta ble></td>\n\t\t\t<!-- <td> This dataset contains RSSI readings gathered from an array of Bluet ooth Low Energy (BLE) iBeacons in a real-world and operational indoor environment for localization and naviga tion purposes. </td> -->\n\t\t\t<td> Multivariate, Sequential, Time-Series </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Integ er </td>\n\t\t\t<td> 6611 </td>\n\t\t\t<td> 15 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Container+Crane+Controller+Data+Set"><i mg src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a hr ef="datasets/Container+Crane+Controller+Data+Set">Container Crane Controller Data Set</a> </td></t r></table></td>\n\t\t\t<!-- <td> A container crane has the function of transporting containers fr om one point to another point. </td> -->\n\t\t\t<td> Univariate, Domain-Theory&nb sp; </td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 15 </td>\n\t\t\t<td> 3 </ p></td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </ p></td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Residential+Building+ Data+Set"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Residential+Building+Data+Set">Residential Building Data Set</a> </td></tr> </table></td>\n\t\t\t<!-- <td> Data set includes construction cost, sale prices, project variables , and economic variables corresponding to real estate single-family residential apartments in Tehran, Iran. &nb sp; </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regress ion </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 372 < /p></td>\n\t\t\t<td> 105 </td>\n\t\t\t<td> 2018 </td>\n\ t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="d atasets/Health+News+in+Twitter"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </ td><td> <a href="datasets/Health+News+in+Twitter">Health News in Twitter</a> </td></tr></table></td>\n\t\t\t<!-- <td> The data was collected in 2015 using Twitter API. This dataset contains health news from more than 15 major health news agencies such as BBC, CNN, and NYT. &n bsp; </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td> Clustering&nb

sp; </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 58000 </td>\n\t\t\t<td> 25000 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/chipseq"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/chipseq">chipseq</a> </td></tr></table></td>\n\t\t\t<!-- <td> ChIP-seq experiments characterize protein modifications or binding at\r\nspecific genomic locations in specific samples. The machine learning\r\nproblem in these data is structured binary classification. </td> -->\n\t\t\t<td> Sequential </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 4960 </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/SGEMM+GPU+kernel+performance"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/SGEMM+GPU+kernel+performance">SGEMM GPU kernel performance</a> </td></tr></table></td>\n\t\t\t<!-- <td> Running times for multiplying two 2048 x 2048 matrices using a GPU OpenCL SGEMM kernel with varying parameters (using the library \'CLTune\'). </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 241600 </td>\n\t\t\t<td> 18 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Repeat+Consumption+Matrices"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Repeat+Consumption+Matrices">Repeat Consumption Matrices</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset contains 7 datasets of User - Item matrices, where each entry represents how many times a user consumed an item. Item is used as an umbrella term for various categories. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 130000 </td>\n\t\t\t<td> 21000 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/detection_of_IoT_botnet_attacks_N_BaIoT"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/detection_of_IoT_botnet_attacks_N_BaIoT">detection_of_IoT_botnet_attacks_N_BaIoT</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset addresses the lack of public botnet datasets, especially for the IoT. It suggests *real* traffic data, gathered from 9 commercial IoT devices authentically infected by Mirai and BASHLITE. </td> -->\n\t\t\t<td> Multivariate, Sequential </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 7062606 </td>\n\t\t\t<td> 115 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Absenteeism+at+work"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Absenteeism+at+work">Absenteeism at work</a> </td></tr></table></td>\n\t\t\t<!-- <td> The database was created with records of absenteeism at work from July 2007 to July 2010 at a courier company in Brazil. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 740 </td>\n\t\t\t<td> 21 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Business </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/SCADI"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/SCADI">SCADI</a> </td></tr></table></td>\n\t\t\t<!-- <td> First self-care activities dataset based on ICF-CY. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 70 </td>\n\t\t\t<td> 206 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Condition+monitoring+of+hydraulic+systems"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Condition+monitoring+of+hydraulic+systems">Condition monitoring of hydraulic systems</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset addresses the condition assessment of a hydraulic test rig based on multi sensor data. Four fault types are superimposed with several severity grades impeding selective quantification. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 2205 </td>\n\t\t\t<td> 43680 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Carbon+Nanotubes"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td>

<td> <a href="datasets/Carbon+Nanotubes">Carbon Nanotubes</a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset contains 10721 initial and calculated atomic coordinates of carbon nanotubes. </td> -->\n\t\t\t<td> Univariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 10721 </td>\n\t\t\t<td> 8 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Optical+Interconnection+Network+"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Optical+Interconnection+Network+">Optical Interconnection Network </a> </td></tr></table></td>\n\t\t\t<!-- <td> This dataset contains 640 performance measurements from a simulation of 2-Dimensional Multiprocessor Optical Interconnection Network.   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 640 </td>\n\t\t\t<td> 10 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Sports+articles+for+objectivity+analysis"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Sports+articles+for+objectivity+analysis">Sports articles for objectivity analysis</a> </td></tr></table></td>\n\t\t\t<!-- <td> 1000 sports articles were labeled using Amazon Mechanical Turk as objective or subjective. The raw texts, extracted features, and the URLs from which the articles were retrieved are provided. </td> -->\n\t\t\t<td> Multivariate, Text </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 1000 </td>\n\t\t\t<td> 59 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Social </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Breast+Cancer+Coimbra"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Breast+Cancer+Coimbra">Breast Cancer Coimbra</a> </td></tr></table></td>\n\t\t\t<!-- <td> Clinical features were observed or measured for 64 patients with breast cancer and 52 healthy controls.   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 116 </td>\n\t\t\t<td> 10 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data">GNFUV Unmanned Surface Vehicles Sensor Data</a> </td></tr></table></td>\n\t\t\t<!-- <td> The data-set contains four (4) sets of mobile sensor readings data (humidity, temperature) corresponding to a swarm of four (4) Unmanned Surface Vehicles (USVs) in a test-bed in Athens (Greece).   </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 1672 </td>\n\t\t\t<td> 5 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Dishonest+Internet+users+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Dishonest+Internet+users+Dataset">Dishonest Internet users Dataset</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset was used to test an architecture based on a trust model capable to cope with the evaluation of the trustworthiness of users interacting in pervasive environments. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 322 </td>\n\t\t\t<td> 5 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Victorian+Era+Authorship+Attribution"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Victorian+Era+Authorship+Attribution">Victorian Era Authorship Attribution</a> </td></tr></table></td>\n\t\t\t<!-- <td> To create the largest authorship attribution dataset, we extracted works of 50 well-known authors. To have a non-exhaustive learning, in training there are 45 authors whereas, in the testing, it\'s 50 </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 93600 </td>\n\t\t\t<td> 1000 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set">Simulated Falls and Daily Living Activities Data Set</a> </td></tr></table></td>\n\t\t\t<!-- <td> 20 falls and 16 daily living activities were performed by 17 volunteers with 5 repetitions while wearing

6 sensors (3.060 instances) that attached to their head, chest, waist, wrist, thigh and ankle. </td> -->\n\t\t\t<td> Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 3060 </td>\n\t\t\t<td> 138 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing">Multimodal Damage Identification for Humanitarian Computing</a> </td></tr></table></td>\n\t\t\t<!-- <td> 5879 captioned images (image and text) from social media related to damage during natural disasters/wars, and belong to 6 classes: Fires, Floods, Natural landscape, Infrastructural, Human, Non-damage. </td> -->\n\t\t\t<td> Multivariate, Text </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 5879 </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Social </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals">EEG Steady-State Visual Evoked Potential Signals</a> </td></tr></table></td>\n\t\t\t<!-- <td> This database consists on 30 subjects performing Brain Computer Interface for Steady State Visual Evoked Potentials (BCI-SSVEP).   </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 9200 </td>\n\t\t\t<td> 16 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Roman+Urdu+Data+Set"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Roman+Urdu+Data+Set">Roman Urdu Data Set</a> </td></tr></table></td>\n\t\t\t<!-- <td> Roman Urdu (the scripting style for Urdu language) is one of the limited resource languages.A data corpus comprising of more than 20000 records was collected. </td> -->\n\t\t\t<td> Text </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 20000 </td>\n\t\t\t<td> 2 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Avila"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Avila">Avila</a> </td></tr></table></td>\n\t\t\t<!-- <td> The Avila data set has been extracted from 800 images of the \'Avila Bible\', an XII century giant Latin copy of the Bible. The prediction task consists in associating each pattern to a copyist. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 20867 </td>\n\t\t\t<td> 10 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/PANDOR"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/PANDOR">PANDOR</a> </td></tr></table></td>\n\t\t\t<!-- <td> PANDOR is a novel and publicly available dataset for online recommendation provided by Purch (http://www.purch.com/).   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Recommendation </td>\n\t\t\t<td> Categorical </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Drug+Review+Dataset+%28Druglib.com%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Drug+Review+Dataset+%28Druglib.com%29">Drug Review Dataset (Druglib.com)</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset provides patient reviews on specific drugs along with related conditions. Reviews and ratings are grouped into reports on the three aspects benefits, side effects and overall comment. </td> -->\n\t\t\t<td> Multivariate, Text </td>\n\t\t\t<td> Classification, Regression, Clustering </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 4143 </td>\n\t\t\t<td> 8 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td>   </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Drug+Review+Dataset+%28Drugs.com%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Drug+Review+Dataset+%28Drugs.com%29">Drug Review Dataset (Drugs.com)</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset provides patient reviews on specific drugs along with related conditions and a 10 star patient rating reflecting overall patient satisfaction. </td> -->\n\t\t\t<td> Multivariate, Text </td>\n\t\t\t<td> Classification, Regression, Clustering </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td><p cla

ss="normal">215063 </td>\n\t\t\t<td> 6 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Physical+Unclonable+Functions"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Physical+Unclonable+Functions">Physical Unclonable Functions</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset is generated from Physical Unclonable Functions (PUFs) simulation, specifically XOR Arbiter PUFs. PUFs are used for authentication purposes. For more info, refer to our paper below. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 6000000 </td>\n\t\t\t<td> 129 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Superconductivty+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Superconductivty+Data">Superconductivty Data</a> </td></tr></table></td>\n\t\t\t<!-- <td> Two file s contain data on 21263 superconductors and their relevant features. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 21263 </td>\n\t\t\t<td> 81 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29">WESAD (Wearable Stress and Affect Detection)</a> </td></tr></table></td>\n\t\t\t<!-- <td> WESAD (Wearable Stress and Affect Detection) contains data of 15 subjects during a stress-affect lab study, while wearing physiological and motion sensors. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 63000000 </td>\n\t\t\t<td> 12 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2">GNFUV Unmanned Surface Vehicles Sensor Data Set 2</a> </td></tr></table></td>\n\t\t\t<!-- <td> The data-set contains eight (2x4) data-sets of mobile sensor readings data (humidity, temperature) corresponding to a swarm of four Unmanned Surface Vehicles (USVs) in a test-bed, Athens, Greece. </td> -->\n\t\t\t<td> Multivariate, Sequential, Time-Series </td>\n\t\t\t<td> Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 10190 </td>\n\t\t\t<td> 6 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Student+Academics+Performance"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Student+Academics+Performance">Student Academics Performance</a> </td></tr></table></td>\n\t\t\t<!-- <td> The dataset tried to find the end semester percentage prediction based on different social, economic and academic attributes.   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 300 </td>\n\t\t\t<td> 22 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Online+Shoppers+Purchasing+Intention+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Online+Shoppers+Purchasing+Intention+Dataset">Online Shoppers Purchasing Intention Dataset</a> </td></tr></table></td>\n\t\t\t<!-- <td> Of the 12,330 sessions in the dataset,\r\n84.5% (10,422) were negative class samples that did not\r\nend with shopping, and the rest (1908) were positive class\r\nsamples ending with shopping. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 12330 </td>\n\t\t\t<td> 18 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Business </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/PMU-UD"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/PMU-UD">PMU-UD</a> </td></tr></table></td>\n\t\t\t<!-- <td> The handwritten dataset was collected from 170 participants with a total of 5,180 numeral patterns. The dataset is named Prince Mohammad Bin Fahd University - Urdu/Arabic Database (PMU-UD).   </td> -->\n\t\t\t<td> Univariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 5180 </td>\n\t\t\t<td> 9 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr>

<tr>\n\t\t\t<td><table><tr><td><a href="datasets/Parkinson%27s+Disease+Classification"><img src="assets/MLi
mages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Parkin
son%27s+Disease+Classification">Parkinson\'s Disease Classification</a> </td></tr></table></td>\n\t\t\t
<!-- <td> The data used in this study were gathered from 188 patients with PD (107 men and
81 women) with ages ranging from 33 to 87 (65.1±10.9). </td> -->\n\t\t\t<td> Multi
variate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td><p class="normal"
>Integer, Real </td>\n\t\t\t<td> 756 </td>\n\t\t\t<td> 7
54 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Comput
er </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Electrical+
Grid+Stability+Simulated+Data+"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </t
d><td> <a href="datasets/Electrical+Grid+Stability+Simulated+Data+">Electrical Grid Sta
bility Simulated Data </a> </td></tr></table></td>\n\t\t\t<!-- <td> The local stability an
alysis of the 4-node star system (electricity producer is in the center) implementing Decentral Smart Grid Contr
ol concept.   </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td><p class="n
ormal">Classification, Regression </td>\n\t\t\t<td> Real </td>\n\t\t\t<td>
10000 </td>\n\t\t\t<td> 14 </td>\n\t\t\t<td><p class="
normal">2018 </td>\n\t\t\t<!-- <td> Physical </td> -->\n\t\t\t</tr><tr>\n\t\t\
t<td><table><tr><td><a href="datasets/Caesarian+Section+Classification+Dataset"><img src="assets/MLimage
s/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Caesarian+
Section+Classification+Dataset">Caesarian Section Classification Dataset</a> </td></tr></table></td>\n
\t\t\t<!-- <td> This dataset contains information about caesarian section results of 80 pregna
nt women with the most important characteristics of delivery problems in the medical field. </td> -->\n
\t\t\t<td> Univariate </td>\n\t\t\t<td> Classification </t
d>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 80 </td>\n\t\t\t<t
d> 5 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td><p cla
ss="normal">Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datas
ets/BAUM-1"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="no
rmal"><a href="datasets/BAUM-1">BAUM-1</a> </td></tr></table></td>\n\t\t\t<!-- <td><p class="nor
mal">BAUM-1 dataset contains 1184 multimodal facial video clips collected from 31 subjects. The 1184 video cl
ips contain spontaneous facial expressions and speech of 13 emotional and mental states. </td> -->\
n\t\t\t<td> Time-Series </td>\n\t\t\t<td> Classification
</td>\n\t\t\t<td>   </td>\n\t\t\t<td> 1184 </td>\n\t\t\t<td
>   </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td><p class
="normal">Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/BAUM-2"><
img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a h
ref="datasets/BAUM-2">BAUM-2</a> </td></tr></table></td>\n\t\t\t<!-- <td> A multilin
gual audio-visual affective face database consisting of 1047 video clips of 286 subjects.   </td> -->\n\t\
t\t<td> Time-Series </td>\n\t\t\t<td> Classification </t
d>\n\t\t\t<td>   </td>\n\t\t\t<td> 1047 </td>\n\t\t\t<td><
p class="normal"> </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td><p class="n
ormal">Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="data
sets/Audit+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class
="normal"><a href="datasets/Audit+Data">Audit Data</a> </td></tr></table></td>\n\t\t\t<!-- <td><p c
lass="normal">Exhaustive one year non-confidential data in the year 2015 to 2016 of firms is collected from the
Auditor Office of India to build a predictor for classifying suspicious firms. </td> -->\n\t\t\t<td><p class
="normal">Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td><p
class="normal">Real </td>\n\t\t\t<td> 777 </td>\n\t\t\t<td><p class="nor
mal">18 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Ot
her </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/BuddyMove+Data+Set"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="
datasets/BuddyMove+Data+Set">BuddyMove Data Set</a> </td></tr></table></td>\n\t\t\t<!-- <td><p cla
ss="normal">User interest information extracted from user reviews published in holidayiq.com about various typ
es of point of interests in South India </td> -->\n\t\t\t<td> Multivariate, Text <
/p></td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Re
al </td>\n\t\t\t<td> 249 </td>\n\t\t\t<td> 7 <
/td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Other </td>
-->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Real+estate+valuation+data+s
et"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <b
><a href="datasets/Real+estate+valuation+data+set">Real estate valuation data set</a> </td></tr></ta
ble></td>\n\t\t\t<!-- <td> The “real estate valuation” is a regression problem. The market hist
orical data set of real estate valuation are collected from Sindian Dist., New Taipei City, Taiwan.   </td
> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Regression <

/p></td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 414 </td>\n\t\t\t<td> 7 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Business </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connected+speech+Data+Set+"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connected+speech+Data+Set+">Early biomarkers of Parkinson's disease based on natural connected speech Data Set </a> </td></tr></table></td>\n\t\t\t<!-- <td> . </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td>   </td>\n\t\t\t<td>   </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Somerville+Happiness+Survey"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Somerville+Happiness+Survey">Somerville Happiness Survey</a> </td></tr></table></td>\n\t\t\t<!-- <td> A data extract of a non-federal dataset posted here https://catalog.data.gov/dataset/somerville-happiness-survey-responses-2011-2013-2015 </td> -->\n\t\t\t<td>   </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Integer </td>\n\t\t\t<td> 143 </td>\n\t\t\t<td> 7 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/2.4+GHZ+Indoor+Channel+Measurements"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/2.4+GHZ+Indoor+Channel+Measurements">2.4 GHZ Indoor Channel Measurements</a> </td></tr></table></td>\n\t\t\t<!-- <td> Measurement of the S21,consists of 10 sweeps, each sweep contains 601 frequency points with spacing of 0.167MHz to cover a 100MHz band centered at 2.4GHz. </td> -->\n\t\t\t<td> Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 7840 </td>\n\t\t\t<td> 5 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/EMG+data+for+gestures"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/EMG+data+for+gestures">EMG data for gestures</a> </td></tr></table></td>\n\t\t\t<!-- <td> These are files of raw EMG data recorded by MYO Thalmic bracelet </td> -->\n\t\t\t<td> Time-Series </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 30000 </td>\n\t\t\t<td> 6 </td>\n\t\t\t<td> 2019 </td>\n\t\t\t<!-- <td> Life </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Parking+Birmingham"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Parking+Birmingham">Parking Birmingham</a> </td></tr></table></td>\n\t\t\t<!-- <td> Data collected from car parks in Birmingham that are operated by NCP from \r\nBirmingham City Council. UK Open Government Licence (OGL).\r\nhttps://data.birmingham.gov.uk/dataset/birmingham-parking </td> -->\n\t\t\t<td> Multivariate, Univariate, Sequential, Time-Series </td>\n\t\t\t<td> Classification, Regression, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 35717 </td>\n\t\t\t<td> 4 </td>\n\t\t\t<td> 2019 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Behavior+of+the+urban+traffic+of+the+city+of+Sao+Paulo+in+Brazil"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Behavior+of+the+urban+traffic+of+the+city+of+Sao+Paulo+in+Brazil">Behavior of the urban traffic of the city of Sao Paulo in Brazil</a> </td></tr></table></td>\n\t\t\t<!-- <td> The database was created with records of behavior of the urban traffic of the city of Sao Paulo in Brazil. </td> -->\n\t\t\t<td> Multivariate, Time-Series </td>\n\t\t\t<td> Classification, Regression </td>\n\t\t\t<td> Integer, Real </td>\n\t\t\t<td> 135 </td>\n\t\t\t<td> 18 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Computer </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="datasets/Travel+Reviews"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td> <a href="datasets/Travel+Reviews">Travel Reviews</a> </td></tr></table></td>\n\t\t\t<!-- <td> Reviews on destinations in 10 categories mentioned across East Asia. Each traveler rating is mapped as Excellent(4), Very Good(3), Average(2), Poor(1), and Terrible(0) and average rating is used. </td> -->\n\t\t\t<td> Multivariate, Text </td>\n\t\t\t<td> Classification, Clustering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 980 </td>\n\t\t\t<td> 11 </td>\n\t\t\t<td> 2018 </td>\n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr bgcolor="DDEEFF">\n\t\t\t<td><table><tr><td><a href="datasets/Tarvel+Review+

Ratings"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal
"><a href="datasets/Tarvel+Review+Ratings">Tarvel Review Ratings</a> </td></tr></table></td>\n\
t\t\t<!-- <td> Google reviews on attractions from 24 categories across Europe are considered
. Google user rating ranges from 1 to 5 and average user rating per category is calculated. </td> -->\
n\t\t\t<td> Multivariate, Text </td>\n\t\t\t<td> Classification, Clust
ering </td>\n\t\t\t<td> Real </td>\n\t\t\t<td> 5456&nbs
p; </td>\n\t\t\t<td> 25 </td>\n\t\t\t<td> 2018 </td>\
n\t\t\t<!-- <td> Other </td> -->\n\t\t\t</tr><tr>\n\t\t\t<td><table><tr><td><a href="dat
asets/Rice+Leaf+Diseases"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><t
d> <a href="datasets/Rice+Leaf+Diseases">Rice Leaf Diseases</a> </td></tr></t
able></td>\n\t\t\t<!-- <td> There are three classes/diseases: Bacterial leaf blight, Brown spot,
and Leaf smut, each having 40 images. The format of all images is jpg.   </td> -->\n\t\t\t<td><p class
="normal">Multivariate </td>\n\t\t\t<td> Classification </td>\n\t\t\t<td><p
class="normal">Integer </td>\n\t\t\t<td> 120 </td>\n\t\t\t<td><p class="n
ormal"> </td>\n\t\t\t<td> 2019 </td>\n\t\t\t<!-- <td> Co
mputer </td> -->\n\t\t\t</tr></table>\n\t</td>\n</tr>\n</table>\n\n\t</td>\n</tr>\n</table>\n\n\n\n<table
cellpadding=5 align=center><tr valign=center>\n\t\t<td> Supported By: </td>\n <td><i
mg src="assets/nsfe.gif" height=60 /> </td>\n <td> In Collaboration With: </td>
\n <td><img src="assets/rexaSmall.jpg" /></td>\n</tr></table>\n\n \n\n<a hre
f="about.html">About</a> || \n<a href="citation_policy.html">Citation Policy</a> &nbs
p;|| \n<a href="donation_policy.html">Donation Policy</a> || \n<a href="contact.html">
Contact</a> || \n<a href="http://cml.ics.uci.edu">CML</a>\n\n \n\n</body>\
n</html>\n'

In [22]:

```
soup = BeautifulSoup(html)
```

In [23]:

```
soup
```

Out[23]:

```
<!DOCTYPE HTML>
<html><body><p>"-//W3C//DTD HTML 4.01 Transitional//EN\"&gt;

</p>
<title>UCI Machine Learning Repository: Data Sets</title>
<!-- Stylesheet link -->
<link href="assets/ml.css" rel="stylesheet" type="text/css"/>
<script language="JavaScript" type="text/javascript">
<!--
function checkform ( form )
{
  // see http://www.thesitewizard.com/archive/validation.shtml
  // for an explanation of this script and how to use it on your
  // own website

  // ** START **
  if (form.q.value == "")
  {
    alert( "Please enter search terms." );
    form.q.focus();
    return false ;
  }

  if (getCheckedValue(form.sitesearch) == "ics.uci.edu" && form.q.value.indexOf("site:archive.ics.uci.edu/ml") ==
-1)
  {
    form.q.value = form.q.value + " site:archive.ics.uci.edu/ml";
  }
```

```
  // ** END **
  return true ;
}

// return the value of the radio button that is checked
// return an empty string if none are checked, or
// there are no radio buttons
function getCheckedValue(radioObj) {
 if(!radioObj)
  return "";
 var radioLength = radioObj.length;
 if(radioLength == undefined)
  if(radioObj.checked)
   return radioObj.value;
  else
   return "";
 for(var i = 0; i < radioLength; i++) {
  if(radioObj[i].checked) {
   return radioObj[i].value;
  }
 }
 return "";
}




//-->
</script>
<!-- SITE HEADER (INCLUDES LOGO AND SEARCH BOX) -->
<!-- SITE HEADER (INCLUDES LOGO AND SEARCH BOX) -->
<table bgcolor="#003366" width="100%">
<tr>
<td>
<span class="normal"><a alt="Home" href="index.html"><img border="0" src="assets/logo.gif"/></a><br/>
<a href="http://cml.ics.uci.edu"><font color="FFDD33">Center for Machine Learning and Intelligent Systems</fo
nt></a></span>
</td>
<td align="right" valign="top" width="100%">
<span class="whitetext">
<a href="about.html">About</a>
  <a href="citation_policy.html">Citation Policy</a>
  <a href="donation_policy.html">Donate a Data Set</a>
  <a href="contact.html">Contact</a>
</span>
<br/>
<br/>
<!-- Search Google -->
<form action="http://www.google.com/custom" method="GET" onsubmit="return checkform(this);">
<input maxlength="255" name="q" size="30" type="text" value=""/>
<input name="sa" type="submit" value="Search"/>
<input name="cof" type="hidden" value="AH:center;LH:130;L:http://archive.ics.uci.edu/assets/logo.gif;LW:384;A
WFID:869c0b2eaa8d518e;"/>
<input name="domains" type="hidden" value="ics.uci.edu"/>
<br/>
<input checked="" name="sitesearch" type="radio" value="ics.uci.edu"/> <span class="whitetext"><font size="1"
>Repository</font></span>
<input name="sitesearch" type="radio" value=""/> <span class="whitetext"><font size="1">Web</font></span>

  <a href="http://www.google.com/search"><img align="middle" alt="Google" border="0" height="27" src="http://
```

www.google.com/logos/Logo_25blk.gif"/></a>

</form>
<!-- Search Google -->
<a href="datasets.php"><font color="#FFDD33" size="3">View ALL Data Sets<
/font></a>

</td>
</tr>
</table>

<table cellpadding="3">
<tr>
<td valign="top">
<table><tr><td> Browse Through: </td></tr></table>
<table border="2" cellpadding="2" width="150">
<tr> <td bgcolor="#003366"> Default Task
</td>
</tr>
<tr>
<td valign="top"> <a href="datasets.php?format=&task=cla&att=&area=&
numAtt=&numIns=&type=&sort=nameUp&view=table">Classification</a> <font color="red">
(350)</font>
<a href="datasets.php?format=&task=reg&att=&area=&numAtt=&nu
mIns=&type=&sort=nameUp&view=table">Regression</a> <font color="red">(96)</font>
<a
href="datasets.php?format=&task=clu&att=&area=&numAtt=&numIns=&type=&a
mp;sort=nameUp&view=table">Clustering</a> <font color="red">(84)</font>
<a href="datasets.php?fo
rmat=&task=other&att=&area=&numAtt=&numIns=&type=&sort=nameUp&a
mp;view=table">Other</a> <font color="red">(55)</font>
</td>
</tr>
<tr> <td bgcolor="#003366"> Attribute Type
</td>
</tr>
<tr>
<td valign="top"> <a href="datasets.php?format=&task=&att=cat&area=&
numAtt=&numIns=&type=&sort=nameUp&view=table">Categorical</a> <font color="red">(3
8)</font>
<a href="datasets.php?format=&task=&att=num&area=&numAtt=&numI
ns=&type=&sort=nameUp&view=table">Numerical</a> <font color="red">(307)</font>
<a hr
ef="datasets.php?format=&task=&att=mix&area=&numAtt=&numIns=&type=&am
p;sort=nameUp&view=table">Mixed</a> <font color="red">(55)</font>
</td>
</tr>
<tr> <td bgcolor="#003366"> Data Type
</td>
</tr>
<tr>
<td valign="top"> <a href="datasets.php?format=&task=&att=&area=&nu
mAtt=&numIns=&type=mvar&sort=nameUp&view=table">Multivariate</a> <font color="red"
>(357)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numI
ns=&type=uvar&sort=nameUp&view=table">Univariate</a> <font color="red">(23)</font>
<a
href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=seq&a
mp;sort=nameUp&view=table">Sequential</a> <font color="red">(47)</font>
<a href="datasets.php?f
ormat=&task=&att=&area=&numAtt=&numIns=&type=ts&sort=nameUp&amp
;view=table">Time-Series</a> <font color="red">(91)</font>
<a href="datasets.php?format=&task=&a
mp;att=&area=&numAtt=&numIns=&type=text&sort=nameUp&view=table">Text</
a> <font color="red">(53)</font>
<a href="datasets.php?format=&task=&att=&area=&n
umAtt=&numIns=&type=dt&sort=nameUp&view=table">Domain-Theory</a> <font color="re
d">(23)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=&num
Ins=&type=other&sort=nameUp&view=table">Other</a> <font color="red">(21)</font>
</td>
</tr>
<tr><td bgcolor="#003366"> Area

</td>
</tr>
<tr>
<td valign="top"> <a href="datasets.php?format=&task=&att=&area=life&numAtt=&numIns=&type=&sort=nameUp&view=table">Life Sciences</a> <font color="red">(107)</font>
<a href="datasets.php?format=&task=&att=&area=phys&numAtt=&numIns=&type=&sort=nameUp&view=table">Physical Sciences</a> <font color="red">(49)</font>
<a href="datasets.php?format=&task=&att=&area=comp&numAtt=&numIns=&type=&sort=nameUp&view=table">CS / Engineering</a> <font color="red">(170)</font>
<a href="datasets.php?format=&task=&att=&area=soc&numAtt=&numIns=&type=&sort=nameUp&view=table">Social Sciences</a> <font color="red">(26)</font>
<a href="datasets.php?format=&task=&att=&area=bus&numAtt=&numIns=&type=&sort=nameUp&view=table">Business</a> <font color="red">(29)</font>
<a href="datasets.php?format=&task=&att=&area=game&numAtt=&numIns=&type=&sort=nameUp&view=table">Game</a> <font color="red">(10)</font>
<a href="datasets.php?format=&task=&att=&area=other&numAtt=&numIns=&type=&sort=nameUp&view=table">Other</a> <font color="red">(73)</font>
</td>
</tr>
<tr><td bgcolor="#003366"> # Attributes
</td>
</tr>
<tr>
<td valign="top"> <a href="datasets.php?format=&task=&att=&area=&numAtt=less10&numIns=&type=&sort=nameUp&view=table">Less than 10</a> <font color="red">(113)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=10to100&numIns=&type=&sort=nameUp&view=table">10 to 100</a> <font color="red">(210)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=greater100&numIns=&type=&sort=nameUp&view=table">Greater than 100</a> <font color="red">(84)</font>
</td>
</tr>
<tr><td bgcolor="#003366"> # Instances </td>
</tr>
<tr>
<td valign="top">
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=less100&type=&sort=nameUp&view=table">Less than 100</a> <font color="red">(27)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=100to1000&type=&sort=nameUp&view=table">100 to 1000</a> <font color="red">(162)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=greater1000&type=&sort=nameUp&view=table">Greater than 1000</a> <font color="red">(246)</font>
</td>
</tr>
<tr><td bgcolor="#003366"> Format Type
</td>
</tr>
<tr>
<td valign="top"> <a href="datasets.php?format=mat&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table">Matrix</a> <font color="red">(324)</font>
<a href="datasets.php?format=nonmat&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table">Non-Matrix</a> <font color="red">(145)</font>
</td>
</tr></table>
</td>
<td valign="top">
<table width="100%">
<tr>
<td> 469 Data Sets </td>
<td align="right"> <font color="gray">Table View</font> <a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list">List View</a> </td>
</tr>

</table><table border="1" cellpadding="5"><tr bgcolor="#003366">
<td class="normal, whitetext"> <a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameDown&view=table">Name</a> </td>
<!-- <td> Abstract </td> -->
<td> <a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=typeUp&view=table">Data Types</a> </td>
<td> <a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=taskUp&view=table">Default Task</a> </td>
<td> <a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=attTypeUp&view=table">Attribute Types</a> </td>
<td> <a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=instUp&view=table"># Instances</a> </td>
<td> <a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=attUp&view=table"># Attributes</a> </td>
<td> <a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=dateUp&view=table">Year</a> </td>
<!-- <td> Area </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Abalone"><img border="1" src="assets/MLimages/SmallLarge1.jpg"/></a> </td><td> <a href="datasets/Abalone">Abalone</a> </td></tr></table></td>
<!-- <td> Predict the age of abalone from physical measurements </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer, Real </td>
<td> 4177 </td>
<td> 8 </td>
<td> 1995 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Adult"><img border="1" src="assets/MLimages/SmallLarge2.jpg"/></a> </td><td> <a href="datasets/Adult">Adult</a> </td></tr></table></td>
<!-- <td> Predict whether income exceeds $50K/yr based on census data. Also known as "Census Income" dataset. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 48842 </td>
<td> 14 </td>
<td> 1996 </td>
<!-- <td> Social </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Annealing"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Annealing">Annealing</a> </td></tr></table></td>
<!-- <td> Steel annealing data </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer, Real </td>
<td> 798 </td>
<td> 38 </td>
<td> </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Anonymous+Microsoft+Web+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Anonymous+Microsoft+Web+Data">Anonymous Microsoft Web Data</a> </td></tr></table></td>
<!-- <td> Log of anonymous users of www.microsoft.com; predict areas of the web site a user visited based on data on other areas the user visited. </td> -->
<td> </td>
<td> Recommender-Systems </td>
<td> Categorical </td>

<td> 37711 </td>
<td> 294 </td>
<td> 1998 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Arrhythmia"><img border="1" src="assets/MLimages/SmallLarge5.jpg"/>
</a> </td><td> <a href="datasets/Arrhythmia">Arrhythmia</a> </td></tr></table>
</td>
<!-- <td> Distinguish between the presence and absence of cardiac arrhythmia and classify it in one of the 16 groups. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer, Real </td>
<td> 452 </td>
<td> 279 </td>
<td> 1998 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Artificial+Characters"><img border="1" src="assets/MLimages/SmallLarge6.jpg"/></a> </td><td> <a href="datasets/Artificial+Characters">Artificial Characters</a>
</td></tr></table></td>
<!-- <td> Dataset artificially generated by using first order theory which describes structure of ten capital letters of English alphabet </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer, Real </td>
<td> 6000 </td>
<td> 7 </td>
<td> 1992 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Audiology+%28Original%29"><img border="1" src="assets/MLimages/SmallLarge7.jpg"/></a> </td><td> <a href="datasets/Audiology+%28Original%29">Audiology (Original)</a> </td></tr></table></td>
<!-- <td> Nominal audiology dataset from Baylor </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 226 </td>
<td> </td>
<td> 1987 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Audiology+%28Standardized%29"><img border="1" src="assets/MLimages/SmallLarge7.jpg"/></a> </td><td> <a href="datasets/Audiology+%28Standardized%29">Audiology (Standardized)</a> </td></tr></table></td>
<!-- <td> Standardized version of the original audiology database </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 226 </td>
<td> 69 </td>
<td> 1992 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Auto+MPG"><img border="1" src="assets/MLimages/SmallLarge9.jpg"/>
</a> </td><td> <a href="datasets/Auto+MPG">Auto MPG</a> </td></tr></table>
</td>
<!-- <td> Revised from CMU StatLib library, data concerns city-cycle fuel consumption
</td> -->
<td> Multivariate </td>
<td> Regression </td>

<td> Regression </td>
<td> Categorical, Real </td>
<td> 398 </td>
<td> 8 </td>
<td> 1993 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Automobile"><img border="1" src="assets/MLimages/SmallLarge10.jpg"/
></a> </td><td> <a href="datasets/Automobile">Automobile</a> </td></tr></table
></td>
<!-- <td> From 1985 Ward's Automotive Yearbook </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Categorical, Integer, Real </td>
<td> 205 </td>
<td> 26 </td>
<td> 1987 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Badges"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"
/></a> </td><td> <a href="datasets/Badges">Badges</a> </td></tr></table></td>
<!-- <td> Badges labeled with a "+" or "-" as a function of a person's name </td> --
>
<td> Univariate, Text </td>
<td> Classification </td>
<td> </td>
<td> 294 </td>
<td> 1 </td>
<td> 1994 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Balance+Scale"><img border="1" src="assets/MLimages/SmallLarge12.j
pg"/></a> </td><td> <a href="datasets/Balance+Scale">Balance Scale</a> </td>
</tr></table></td>
<!-- <td> Balance scale weight & distance database </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 625 </td>
<td> 4 </td>
<td> 1994 </td>
<!-- <td> Social </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Balloons"><img border="1" src="assets/MLimages/SmallLarge13.jpg"/></
a> </td><td> <a href="datasets/Balloons">Balloons</a> </td></tr></table></td>
<!-- <td> Data previously used in cognitive psychology experiment; 4 data sets represent diff
erent conditions of an experiment </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 16 </td>
<td> 4 </td>
<td> </td>
<!-- <td> Social </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Breast+Cancer"><img border="1" src="assets/MLimages/SmallLarge14.j
pg"/></a> </td><td> <a href="datasets/Breast+Cancer">Breast Cancer</a> </td>
</tr></table></td>
<!-- <td> Breast Cancer Data (Restricted Access) </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 286 </td>

<td> 286 </td>
<td> 9 </td>
<td> 1988 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Breast+Cancer+Wisconsin+%28Original%29"><img border="1" src="ass
ets/MLimages/SmallLarge14.jpg"/></a> </td><td> <a href="datasets/Breast+Cancer+Wis
consin+%28Original%29">Breast Cancer Wisconsin (Original)</a> </td></tr></table></td>
<!-- <td> Original Wisconsin Breast Cancer Database </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 699 </td>
<td> 10 </td>
<td> 1992 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Breast+Cancer+Wisconsin+%28Prognostic%29"><img border="1" src="a
ssets/MLimages/SmallLarge14.jpg"/></a> </td><td> <a href="datasets/Breast+Cancer+W
isconsin+%28Prognostic%29">Breast Cancer Wisconsin (Prognostic)</a> </td></tr></table></td>
<!-- <td> Prognostic Wisconsin Breast Cancer Database </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Real </td>
<td> 198 </td>
<td> 34 </td>
<td> 1995 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29"><img border="1" src="a
ssets/MLimages/SmallLarge14.jpg"/></a> </td><td> <a href="datasets/Breast+Cancer+W
isconsin+%28Diagnostic%29">Breast Cancer Wisconsin (Diagnostic)</a> </td></tr></table></td>
<!-- <td> Diagnostic Wisconsin Breast Cancer Database </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 569 </td>
<td> 32 </td>
<td> 1995 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Pittsburgh+Bridges"><img border="1" src="assets/MLimages/SmallLarge
18.jpg"/></a> </td><td> <a href="datasets/Pittsburgh+Bridges">Pittsburgh Bridges</a></
b> </td></tr></table></td>
<!-- <td> Bridges database that has original and numeric-discretized datasets </td
> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 108 </td>
<td> 13 </td>
<td> 1990 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Car+Evaluation"><img border="1" src="assets/MLimages/SmallLarge19.j
pg"/></a> </td><td> <a href="datasets/Car+Evaluation">Car Evaluation</a> </td
></tr></table></td>
<!-- <td> Derived from simple hierarchical decision model, this database may be useful for te
sting constructive induction and structure discovery methods. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
</tr><tr bgcolor="DDEEFF">
</tr><tr>
</tr><tr bgcolor="DDEEFF">
</tr><tr>
</tr><tr bgcolor="DDEEFF">
</tr>

<td> 1728 </td>
<td> 6 </td>
<td> 1997 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Census+Income"><img border="1" src="assets/MLimages/SmallLarge2.jpg"/></a> </td><td> <a href="datasets/Census+Income">Census Income</a> </td></tr></table></td>
<!-- <td> Predict whether income exceeds $50K/yr based on census data. Also known as "Adult" dataset. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 48842 </td>
<td> 14 </td>
<td> 1996 </td>
<!-- <td> Social </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Chess+%28King-Rook+vs.+King-Knight%29"><img border="1" src="assets/MLimages/SmallLarge24.jpg"/></a> </td><td> <a href="datasets/Chess+%28King-Rook+vs.+King-Knight%29">Chess (King-Rook vs. King-Knight)</a> </td></tr></table></td>
<!-- <td> Knight Pin Chess End-Game Database Creator </td> -->
<td> Multivariate, Data-Generator </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> </td>
<td> 22 </td>
<td> 1988 </td>
<!-- <td> Game </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Chess+%28King-Rook+vs.+King-Pawn%29"><img border="1" src="assets/MLimages/SmallLarge24.jpg"/></a> </td><td> <a href="datasets/Chess+%28King-Rook+vs.+King-Pawn%29">Chess (King-Rook vs. King-Pawn)</a> </td></tr></table></td>
<!-- <td> King+Rook versus King+Pawn on a7 (usually abbreviated KRKPA7). </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 3196 </td>
<td> 36 </td>
<td> 1989 </td>
<!-- <td> Game </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Chess+%28King-Rook+vs.+King%29"><img border="1" src="assets/MLimages/SmallLarge24.jpg"/></a> </td><td> <a href="datasets/Chess+%28King-Rook+vs.+King%29">Chess (King-Rook vs. King)</a> </td></tr></table></td>
<!-- <td> Chess Endgame Database for White King and Rook against Black King (KRK). </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 28056 </td>
<td> 6 </td>
<td> 1994 </td>
<!-- <td> Game </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Chess+%28Domain+Theories%29"><img border="1" src="assets/MLimages/SmallLarge24.jpg"/></a> </td><td> <a href="datasets/Chess+%28Domain+Theories%29">Chess (Domain Theories)</a> </td></tr></table></td>
<!-- <td> 6 different domain theories for generating legal moves of chess </td> -->
<td> Domain-Theory </td>
<td> </td>

<td> </td>
<td> </td>
<td> </td>
<td> </td>
<!-- <td> Game </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Bach+Chorales"><img border="1" src="assets/MLimages/SmallLarge25.jpg"/></a> </td><td> <a href="datasets/Bach+Chorales">Bach Chorales</a> </td></tr></table></td>
<!-- <td> Time-series data based on chorales; challenge is to learn generative grammar; data in Lisp </td> -->
<td> Univariate, Time-Series </td>
<td> </td>
<td> Categorical, Integer </td>
<td> 100 </td>
<td> 6 </td>
<td> </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Connect-4"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Connect-4">Connect-4</a> </td></tr></table></td>
<!-- <td> Contains connect-4 positions </td> -->
<td> Multivariate, Spatial </td>
<td> Classification </td>
<td> Categorical </td>
<td> 67557 </td>
<td> 42 </td>
<td> 1995 </td>
<!-- <td> Game </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Credit+Approval"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Credit+Approval">Credit Approval</a> </td></tr></table></td>
<!-- <td> This data concerns credit card applications; good mix of attributes </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer, Real </td>
<td> 690 </td>
<td> 15 </td>
<td> </td>
<!-- <td> Financial </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Japanese+Credit+Screening"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Japanese+Credit+Screening">Japanese Credit Screening</a> </td></tr></table></td>
<!-- <td> Includes domain theory (generated by talking to Japanese domain experts); data in Lisp </td> -->
<td> Multivariate, Domain-Theory </td>
<td> Classification </td>
<td> Categorical, Real, Integer </td>
<td> 125 </td>
<td> </td>
<td> 1992 </td>
<!-- <td> Financial </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Computer+Hardware"><img border="1" src="assets/MLimages/SmallLarge29.jpg"/></a> </td><td> <a href="datasets/Computer+Hardware">Computer Hardware</a> </td></tr></table></td>
<!-- <td> Relative CPU Performance Data, described in terms of its cycle time, memory size, etc. </td> -->

<td> Multivariate </td>
<td> Regression </td>
<td> Integer </td>
<td> 209 </td>
<td> 9 </td>
<td> 1987 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Contraceptive+Method+Choice"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Contraceptive+Method+Choice">Contraceptive Method Choice</a> </td></tr></table></td>
<!-- <td> Dataset is a subset of the 1987 National Indonesia Contraceptive Prevalence Survey. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 1473 </td>
<td> 9 </td>
<td> 1997 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Covertype"><img border="1" src="assets/MLimages/SmallLarge31.jpg"/></a> </td><td> <a href="datasets/Covertype">Covertype</a> </td></tr></table></td>
<!-- <td> Forest CoverType dataset </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 581012 </td>
<td> 54 </td>
<td> 1998 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Cylinder+Bands"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Cylinder+Bands">Cylinder Bands</a> </td></tr></table></td>
<!-- <td> Used in decision tree induction for mitigating process delays known as "cylinder bands" in rotogravure printing </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer, Real </td>
<td> 512 </td>
<td> 39 </td>
<td> 1995 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Dermatology"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Dermatology">Dermatology</a> </td></tr></table></td>
<!-- <td> Aim for this dataset is to determine the type of Eryhemato-Squamous Disease. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 366 </td>
<td> 33 </td>
<td> 1998 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Diabetes"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Diabetes">Diabetes</a> </td></tr></table></td>

<!-- <td> This diabetes dataset is from AIM '94 </td> -->
<td> Multivariate, Time-Series </td>
<td> </td>
<td> Categorical, Integer </td>
<td> </td>
<td> 20 </td>
<td> </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/DGP2+-+The+Second+Data+Generation+Program"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/DGP2+-+The+Second+Data+Generation+Program">DGP2 - The Second Data Generation Program</a> </td></tr></table></td>
<!-- <td> Generates application domains based on specific parameters, number of features, and proportion of positive to negative examples </td> -->
<td> Data-Generator </td>
<td> </td>
<td> Real </td>
<td> </td>
<td> </td>
<td> </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Document+Understanding"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Document+Understanding">Document Understanding</a> </td></tr></table></td>
<!-- <td> Five concepts, expressed as predicates, to be learned </td> -->
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> 1994 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/EBL+Domain+Theories"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/EBL+Domain+Theories">EBL Domain Theories</a> </td></tr></table></td>
<!-- <td> Assorted small-scale domain theories </td> -->
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Echocardiogram"><img border="1" src="assets/MLimages/SmallLarge38.jpg"/></a> </td><td> <a href="datasets/Echocardiogram">Echocardiogram</a> </td></tr></table></td>
<!-- <td> Data for classifying if patients will survive for at least one year after a heart attack </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer, Real </td>
<td> 132 </td>
<td> 12 </td>
<td> 1989 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Ecoli"><img border="1" src="assets/MLimages/SmallLarge120.jpg"/></a> </td><td> <a href="datasets/Ecoli">Ecoli</a> </td></tr></table></td>

<!-- <td> This data contains protein localization sites </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 336 </td>
<td> 8 </td>
<td> 1996 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Flags"><img border="1" src="assets/MLimages/SmallLarge40.jpg"/></a>
</td><td> <a href="datasets/Flags">Flags</a> </td></tr></table></td>
<!-- <td> From Collins Gem Guide to Flags, 1986 </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 194 </td>
<td> 30 </td>
<td> 1990 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Function+Finding"><img border="1" src="assets/MLimages/SmallLarge4
1.jpg"/></a> </td><td> <a href="datasets/Function+Finding">Function Finding</a></
p></td></tr></table></td>
<!-- <td> Cases collected mostly from investigations in physical science; intention is to evalua
te function-finding algorithms </td> -->
<td> </td>
<td> Function-Learning </td>
<td> Real </td>
<td> 352 </td>
<td> </td>
<td> 1990 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Glass+Identification"><img border="1" src="assets/MLimages/SmallLarg
e42.jpg"/></a> </td><td> <a href="datasets/Glass+Identification">Glass Identification</a>
</td></tr></table></td>
<!-- <td> From USA Forensic Science Service; 6 types of glass; defined in terms of their oxid
e content (i.e. Na, Fe, K, etc) </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 214 </td>
<td> 10 </td>
<td> 1987 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Haberman%27s+Survival"><img border="1" src="assets/MLimages/Smal
lLargedefault.jpg"/></a> </td><td> <a href="datasets/Haberman%27s+Survival">Haberm
an's Survival</a> </td></tr></table></td>
<!-- <td> Dataset contains cases from study conducted on the survival of patients who had u
ndergone surgery for breast cancer </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 306 </td>
<td> 3 </td>
<td> 1999 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Hayes-Roth"><img border="1" src="assets/MLimages/SmallLargedefault.
jpg"/></a> </td><td> <a href="datasets/Hayes-Roth">Hayes-Roth</a> </td></tr>
</table></td>

<!-- <td> Topic: human subjects study </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 160 </td>
<td> 5 </td>
<td> 1989 </td>
<!-- <td> Social </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Heart+Disease">
</a> </td><td> <a href="datasets/Heart+Disease">Heart Disease</a> </td>
</tr></table></td>
<!-- <td> 4 databases: Cleveland, Hungary, Switzerland, and the VA Long Beach
</td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer, Real </td>
<td> 303 </td>
<td> 75 </td>
<td> 1988 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Hepatitis">
</a> </td><td> <a href="datasets/Hepatitis">Hepatitis</a> </td></tr></table></td>
<!-- <td> From G.Gong: CMU; Mostly Boolean or numeric-valued attribute types; Includes co
st data (donated by Peter Turney) </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer, Real </td>
<td> 155 </td>
<td> 19 </td>
<td> 1988 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Horse+Colic">
</a> </td><td> <a href="datasets/Horse+Colic">Horse Colic</a> </td></tr></table></td>
<!-- <td> Well documented attributes; 368 instances with 28 attributes (continuous, discrete,
and nominal); 30% missing values </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer, Real </td>
<td> 368 </td>
<td> 27 </td>
<td> 1989 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/ICU">
</a> </td><td> <a href="datasets/ICU">ICU</a> </td></tr></table></td>
<!-- <td> Data set prepared for the use of participants for the 1994 AAAI Spring Symposium
on Artificial Intelligence in Medicine. </td> -->
<td> Multivariate, Time-Series </td>
<td> </td>
<td> Real </td>
<td> </td>
<td> </td>
<td> </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Image+Segmentation">
</a> </td><td> <a href="datasets/Image+Segmentation">Image Segment
ation</a> </td></tr></table></td>

<!-- <td> Image data described by high-level numeric-valued attributes, 7 classes
</td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 2310 </td>
<td> 19 </td>
<td> 1990 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Internet+Advertisements"><img border="1" src="assets/MLimages/Small
Large51.jpg"/></a> </td><td> <a href="datasets/Internet+Advertisements">Internet Adver
tisements</a> </td></tr></table></td>
<!-- <td> This dataset represents a set of possible advertisements on Internet pages.
</td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer, Real </td>
<td> 3279 </td>
<td> 1558 </td>
<td> 1998 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Ionosphere"><img border="1" src="assets/MLimages/SmallLarge52.jpg"/
></a> </td><td> <a href="datasets/Ionosphere">Ionosphere</a> </td></tr></tabl
e></td>
<!-- <td> Classification of radar returns from the ionosphere </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 351 </td>
<td> 34 </td>
<td> 1989 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Iris"><img border="1" src="assets/MLimages/SmallLarge53.jpg"/></a> </
td><td> <a href="datasets/Iris">Iris</a> </td></tr></table></td>
<!-- <td> Famous database; from Fisher, 1936 </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 150 </td>
<td> 4 </td>
<td> 1988 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/ISOLET"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"
/></a> </td><td> <a href="datasets/ISOLET">ISOLET</a> </td></tr></table></td
>
<!-- <td> Goal: Predict which letter-name was spoken--a simple classification task. </p
></td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 7797 </td>
<td> 617 </td>
<td> 1994 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Kinship"><img border="1" src="assets/MLimages/SmallLarge55.jpg"/></a
> </td><td> <a href="datasets/Kinship">Kinship</a> </td></tr></table></td>
<!-- <td> Relational dataset </td> -->
</tr></table>
</body>
</html>

<!-- <td> Relational dataset </td> -->
<td> Relational </td>
<td> Relational-Learning </td>
<td> Categorical </td>
<td> 104 </td>
<td> 12 </td>
<td> 1990 </td>
<!-- <td> Social </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Labor+Relations"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Labor+Relations">Labor Relations</a> </td></tr></table></td>
<!-- <td> From Collective Bargaining Review </td> -->
<td> Multivariate </td>
<td> </td>
<td> Categorical, Integer, Real </td>
<td> 57 </td>
<td> 16 </td>
<td> 1988 </td>
<!-- <td> Social </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/LED+Display+Domain"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/LED+Display+Domain">LED Display Domain</a> </td></tr></table></td>
<!-- <td> From Classification and Regression Trees book; We provide here 2 C programs for generating sample databases </td> -->
<td> Multivariate, Data-Generator </td>
<td> Classification </td>
<td> Categorical </td>
<td> </td>
<td> 7 </td>
<td> 1988 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Lenses"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Lenses">Lenses</a> </td></tr></table></td>
<!-- <td> Database for fitting contact lenses </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 24 </td>
<td> 4 </td>
<td> 1990 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Letter+Recognition"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Letter+Recognition">Letter Recognition</a> </td></tr></table></td>
<!-- <td> Database of character image features; try to identify the letter </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 20000 </td>
<td> 16 </td>
<td> 1991 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Liver+Disorders"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Liver+Disorders">Liver Disorders</a> </td></tr></table></td>
<!-- <td> BUPA Medical Research Ltd. database donated by Richard S. Forsyth </td> -->
<td> Multivariate </td>

<td> Multivariate </td>
<td> </td>
<td> Categorical, Integer, Real </td>
<td> 345 </td>
<td> 7 </td>
<td> 1990 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Logic+Theorist"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Logic+Theorist">Logic Theorist</a> </td></tr></table></td>
<!-- <td> All code for Logic Theorist </td> -->
<td> Domain-Theory </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Lung+Cancer"><img border="1" src="assets/MLimages/SmallLarge62.jpg"/></a> </td><td> <a href="datasets/Lung+Cancer">Lung Cancer</a> </td></tr></table></td>
<!-- <td> Lung cancer data; no attribute definitions </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 32 </td>
<td> 56 </td>
<td> 1992 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Lymphography"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Lymphography">Lymphography</a> </td></tr></table></td>
<!-- <td> This lymphography domain was obtained from the University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia. (Restricted access) </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 148 </td>
<td> 18 </td>
<td> 1988 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Mechanical+Analysis"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Mechanical+Analysis">Mechanical Analysis</a> </td></tr></table></td>
<!-- <td> Fault diagnosis problem of electromechanical devices; also PUMPS DATA SET is newer version with domain theory and results </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer, Real </td>
<td> 209 </td>
<td> 8 </td>
<td> 1990 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Meta-data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Meta-data">Meta-data</a> </td></tr></table></td>
<!-- <td> Meta-Data was used in order to give advice about which classification method is ap

propriate for a particular dataset (taken from results of Statlog project). </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer, Real </td>
<td> 528 </td>
<td> 22 </td>
<td> 1996 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Mobile+Robots"><img border="1" src="assets/MLimages/SmallLarge66.jpg"/></a> </td><td> <a href="datasets/Mobile+Robots">Mobile Robots</a> </td></tr></table></td>
<!-- <td> Learning concepts from sensor data of a mobile robot; set of data sets </td> -->
<td> Domain-Theory </td>
<td> </td>
<td> Categorical, Integer, Real </td>
<td> </td>
<td> </td>
<td> 1995 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29"><img border="1" src="assets/MLimages/SmallLarge67.jpg"/></a> </td><td> <a href="datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29">Molecular Biology (Promoter Gene Sequences)</a> </td></tr></table></td>
<!-- <td> E. Coli promoter gene sequences (DNA) with partial domain theory </td> -->
<td> Sequential, Domain-Theory </td>
<td> Classification </td>
<td> Categorical </td>
<td> 106 </td>
<td> 58 </td>
<td> 1990 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%29"><img border="1" src="assets/MLimages/SmallLarge67.jpg"/></a> </td><td> <a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular Biology (Protein Secondary Structure)</a> </td></tr></table></td>
<!-- <td> From CMU connectionist bench repository; Classifies secondary structure of certain globular proteins </td> -->
<td> Sequential </td>
<td> Classification </td>
<td> Categorical </td>
<td> 128 </td>
<td> </td>
<td> </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29"><img border="1" src="assets/MLimages/SmallLarge67.jpg"/></a> </td><td> <a href="datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29">Molecular Biology (Splice-junction Gene Sequences)</a> </td></tr></table></td>
<!-- <td> Primate splice-junction gene sequences (DNA) with associated imperfect domain theory </td> -->
<td> Sequential, Domain-Theory </td>
<td> Classification </td>
<td> Categorical </td>
<td> 3190 </td>
<td> 61 </td>
<td> 1992 </td>

<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/MONK%27s+Problems"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/MONK%27s+Problems">MONK's Problems</a> </td></tr></table></td>
<!-- <td> A set of three artificial domains over the same attribute space; Used to test a wide range of induction algorithms </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 432 </td>
<td> 7 </td>
<td> 1992 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Moral+Reasoner"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Moral+Reasoner">Moral Reasoner</a> </td></tr></table></td>
<!-- <td> Horn-clause model that qualitatively simulates moral reasoning; Theory includes negated literals </td> -->
<td> Domain-Theory </td>
<td> </td>
<td> </td>
<td> 202 </td>
<td> </td>
<td> 1994 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Multiple+Features"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Multiple+Features">Multiple Features</a> </td></tr></table></td>
<!-- <td> This dataset consists of features of handwritten numerals (`0'--`9') extracted from a collection of Dutch utility maps </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 2000 </td>
<td> 649 </td>
<td> </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Mushroom"><img border="1" src="assets/MLimages/SmallLarge73.jpg"/></a> </td><td> <a href="datasets/Mushroom">Mushroom</a> </td></tr></table></td>
<!-- <td> From Audobon Society Field Guide; mushrooms described in terms of physical characteristics; classification: poisonous or edible </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 8124 </td>
<td> 22 </td>
<td> 1987 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Musk+%28Version+1%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Musk+%28Version+1%29">Musk (Version 1)</a> </td></tr></table></td>
<!-- <td> The goal is to learn to predict whether new molecules will be musks or non-musks </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>

<td> 476 </td>
<td> 168 </td>
<td> 1994 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Musk+%28Version+2%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Musk+%28Version+2%29">Musk (Version 2)</a> </td></tr></table></td>
<!-- <td> The goal is to learn to predict whether new molecules will be musks or non-musks </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 6598 </td>
<td> 168 </td>
<td> 1994 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Nursery"><img border="1" src="assets/MLimages/SmallLargedefault.jpg" /></a> </td><td> <a href="datasets/Nursery">Nursery</a> </td></tr></table></td>
<!-- <td> Nursery Database was derived from a hierarchical decision model originally developed to rank applications for nursery schools. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 12960 </td>
<td> 8 </td>
<td> 1997 </td>
<!-- <td> Social </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Othello+Domain+Theory"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Othello+Domain+Theory">Othello Domain Theory</a> </td></tr></table></td>
<!-- <td> Used in research to generate features for an inductive learning system </td> -->
<td> Domain-Theory </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> 1991 </td>
<!-- <td> Game </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Page+Blocks+Classification"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Page+Blocks+Classification">Page Blocks Classification</a> </td></tr></table></td>
<!-- <td> The problem consists of classifying all the blocks of the page layout of a document that has been detected by a segmentation process. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 5473 </td>
<td> 10 </td>
<td> 1995 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Optical+Recognition+of+Handwritten+Digits"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Optical+Recognition+of+Handwritten+Digits">Optical Recognition of Handwritten Digits</a> </td></tr></table></td>
<!-- <td> Two versions of this database available; see folder </td> -->
<td> Multivariate </td>

<td> Classification </td>
<td> Integer </td>
<td> 5620 </td>
<td> 64 </td>
<td> 1998 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits">Pen-Based Recognition of Handwritten Digits</a> </td></tr></table></td>
<!-- <td> Digit database of 250 samples from 44 writers </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 10992 </td>
<td> 16 </td>
<td> 1998 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Post-Operative+Patient"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Post-Operative+Patient">Post-Operative Patient</a> </td></tr></table></td>
<!-- <td> Dataset of patient features </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 90 </td>
<td> 8 </td>
<td> 1993 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Primary+Tumor"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Primary+Tumor">Primary Tumor</a> </td></tr></table></td>
<!-- <td> From Ljubljana Oncology Institute </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 339 </td>
<td> 17 </td>
<td> 1988 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Prodigy"><img border="1" src="assets/MLimages/SmallLargedefault.jpg" /></a> </td><td> <a href="datasets/Prodigy">Prodigy</a> </td></tr></table></td>
<!-- <td> Assorted domains like blocksworld, eightpuzzle, and schedworld. </td> -->
<td> Domain-Theory </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Qualitative+Structure+Activity+Relationships"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Qualitative+Structure+Activity+Relationships">Qualitative Structure Activity Relationships</a> </td></tr></table></td>
<!-- <td> Two sets of datasets are given: pyrimidines and triazines </td> -->
<td> Domain-Theory </td>
<td> </td>

<td> </td>
<td> </td>
<td> </td>
<td> </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Quadruped+Mammals"><img border="1" src="assets/MLimages/SmallLarge86.jpg"/></a> </td><td> <a href="datasets/Quadruped+Mammals">Quadruped Mammals</a> </td></tr></table></td>
<!-- <td> The file animals.c is a data generator of structured instances representing quadruped animals </td> -->
<td> Multivariate, Data-Generator </td>
<td> Classification </td>
<td> Real </td>
<td> </td>
<td> 72 </td>
<td> 1992 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Servo"><img border="1" src="assets/MLimages/SmallLarge87.jpg"/></a> </td><td> <a href="datasets/Servo">Servo</a> </td></tr></table></td>
<!-- <td> Data was from a simulation of a servo system </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Categorical, Integer </td>
<td> 167 </td>
<td> 4 </td>
<td> 1993 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Shuttle+Landing+Control"><img border="1" src="assets/MLimages/SmallLarge92.jpg"/></a> </td><td> <a href="datasets/Shuttle+Landing+Control">Shuttle Landing Control</a> </td></tr></table></td>
<!-- <td> Tiny database; all nominal values </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 15 </td>
<td> 6 </td>
<td> 1988 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Solar+Flare"><img border="1" src="assets/MLimages/SmallLarge89.jpg"/></a> </td><td> <a href="datasets/Solar+Flare">Solar Flare</a> </td></tr></table></td>
<!-- <td> Each class attribute counts the number of solar flares of a certain class that occur in a 24 hour period </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Categorical </td>
<td> 1389 </td>
<td> 10 </td>
<td> 1989 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Soybean+%28Large%29"><img border="1" src="assets/MLimages/SmallLarge90.jpg"/></a> </td><td> <a href="datasets/Soybean+%28Large%29">Soybean (Large)</a> </td></tr></table></td>
<!-- <td> Michalski's famous soybean disease database </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>

<td> 307 </td>
<td> 35 </td>
<td> 1988 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Soybean+%28Small%29"><img border="1" src="assets/MLimages/SmallLarge90.jpg"/></a> </td><td> <a href="datasets/Soybean+%28Small%29">Soybean (Small)</a> </td></tr></table></td>
<!-- <td> Michalski's famous soybean disease database </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 47 </td>
<td> 35 </td>
<td> 1987 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Challenger+USA+Space+Shuttle+O-Ring"><img border="1" src="assets/MLimages/SmallLarge92.jpg"/></a> </td><td> <a href="datasets/Challenger+USA+Space+Shuttle+O-Ring">Challenger USA Space Shuttle O-Ring</a> </td></tr></table></td>
<!-- <td> Task: predict the number of O-rings that experience thermal distress on a flight at 31 degrees F given data on the previous 23 shuttle flights </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Integer </td>
<td> 23 </td>
<td> 4 </td>
<td> 1993 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Low+Resolution+Spectrometer"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Low+Resolution+Spectrometer">Low Resolution Spectrometer</a> </td></tr></table></td>
<!-- <td> From IRAS data -- NASA Ames Research Center </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 531 </td>
<td> 102 </td>
<td> 1988 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Spambase"><img border="1" src="assets/MLimages/SmallLarge94.jpg"/></a> </td><td> <a href="datasets/Spambase">Spambase</a> </td></tr></table></td>
<!-- <td> Classifying Email as Spam or Non-Spam </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 4601 </td>
<td> 57 </td>
<td> 1999 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/SPECT+Heart"><img border="1" src="assets/MLimages/SmallLarge45.jpg"/></a> </td><td> <a href="datasets/SPECT+Heart">SPECT Heart</a> </td></tr></table></td>
<!-- <td> Data on cardiac Single Proton Emission Computed Tomography (SPECT) images. Each patient classified into two categories: normal and abnormal. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>

<td> Categorical </td>
<td> 267 </td>
<td> 22 </td>
<td> 2001 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/SPECTF+Heart"><img border="1" src="assets/MLimages/SmallLarge45.jpg"/></a> </td><td> <a href="datasets/SPECTF+Heart">SPECTF Heart</a> </td></tr></table></td>
<!-- <td> Data on cardiac Single Proton Emission Computed Tomography (SPECT) images. Each patient classified into two categories: normal and abnormal. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 267 </td>
<td> 44 </td>
<td> 2001 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Sponge"><img border="1" src="assets/MLimages/SmallLarge97.jpg"/></a> </td><td> <a href="datasets/Sponge">Sponge</a> </td></tr></table></td>
<!-- <td> Data on sponges; Attributes in spanish </td> -->
<td> Multivariate </td>
<td> Clustering </td>
<td> Categorical, Integer </td>
<td> 76 </td>
<td> 45 </td>
<td> </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Statlog+Project"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Statlog+Project">Statlog Project</a> </td></tr></table></td>
<!-- <td> Various Databases: Vehicle silhouttes, Landsat Sattelite, Shuttle, Australian Credit Approval, Heart Disease, Image Segmentation, German Credit </td> -->
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> 1992 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Student+Loan+Relational"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Student+Loan+Relational">Student Loan Relational</a> </td></tr></table></td>
<!-- <td> Student Loan Relational Domain </td> -->
<td> Domain-Theory </td>
<td> </td>
<td> </td>
<td> 1000 </td>
<td> </td>
<td> 1993 </td>
<!-- <td> Social </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Teaching+Assistant+Evaluation"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Teaching+Assistant+Evaluation">Teaching Assistant Evaluation</a> </td></tr></table></td>
<!-- <td> The data consist of evaluations of teaching performance; scores are "low", "medium", or "high" </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>

<td> Categorical, Integer </td>
<td> 151 </td>
<td> 5 </td>
<td> 1997 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Tic-Tac-Toe+Endgame"><img border="1" src="assets/MLimages/SmallLarge101.jpg"/></a> </td><td> <a href="datasets/Tic-Tac-Toe+Endgame">Tic-Tac-Toe Endgame</a> </td></tr></table></td>
<!-- <td> Binary classification task on possible configurations of tic-tac-toe game </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 958 </td>
<td> 9 </td>
<td> 1991 </td>
<!-- <td> Game </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Thyroid+Disease"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Thyroid+Disease">Thyroid Disease</a> </td></tr></table></td>
<!-- <td> 10 separate databases from Garavan Institute </td> -->
<td> Multivariate, Domain-Theory </td>
<td> Classification </td>
<td> Categorical, Real </td>
<td> 7200 </td>
<td> 21 </td>
<td> 1987 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Trains"><img border="1" src="assets/MLimages/SmallLarge103.jpg"/></a> </td><td> <a href="datasets/Trains">Trains</a> </td></tr></table></td>
<!-- <td> 2 data formats (structured, one-instance-per-line) </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 10 </td>
<td> 32 </td>
<td> 1994 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/University"><img border="1" src="assets/MLimages/SmallLarge104.jpg"/></a> </td><td> <a href="datasets/University">University</a> </td></tr></table></td>
<!-- <td> Data in original (LISP-readable) form </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 285 </td>
<td> 17 </td>
<td> 1988 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Congressional+Voting+Records"><img border="1" src="assets/MLimages/SmallLarge105.jpg"/></a> </td><td> <a href="datasets/Congressional+Voting+Records">Congressional Voting Records</a> </td></tr></table></td>
<!-- <td> 1984 United Stated Congressional Voting Records; Classify as Republican or Democrat </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
</tr><tr bgcolor="DDEEFF">
</tr><tr>
</tr><tr bgcolor="DDEEFF">
</tr><tr>
</tr><tr bgcolor="DDEEFF">
</tr>

<td> 435 </td>
<td> 16 </td>
<td> 1987 </td>
<!-- <td> Social </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Water+Treatment+Plant"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Water+Treatment+Plant">Water Treatment Plant</a> </td></tr></table></td>
<!-- <td> Multiple classes predict plant state </td> -->
<td> Multivariate </td>
<td> Clustering </td>
<td> Integer, Real </td>
<td> 527 </td>
<td> 38 </td>
<td> 1993 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Waveform+Database+Generator+%28Version+1%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Waveform+Database+Generator+%28Version+1%29">Waveform Database Generator (Version 1)</a> </td></tr></table></td>
<!-- <td> CART book's waveform domains </td> -->
<td> Multivariate, Data-Generator </td>
<td> Classification </td>
<td> Real </td>
<td> 5000 </td>
<td> 21 </td>
<td> 1988 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Waveform+Database+Generator+%28Version+2%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Waveform+Database+Generator+%28Version+2%29">Waveform Database Generator (Version 2)</a> </td></tr></table></td>
<!-- <td> CART book's waveform domains </td> -->
<td> Multivariate, Data-Generator </td>
<td> Classification </td>
<td> Real </td>
<td> 5000 </td>
<td> 40 </td>
<td> 1988 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Wine"><img border="1" src="assets/MLimages/SmallLarge109.jpg"/></a> </td><td> <a href="datasets/Wine">Wine</a> </td></tr></table></td>
<!-- <td> Using chemical analysis determine the origin of wines </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 178 </td>
<td> 13 </td>
<td> 1991 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Yeast"><img border="1" src="assets/MLimages/SmallLarge110.jpg"/></a> </td><td> <a href="datasets/Yeast">Yeast</a> </td></tr></table></td>
<!-- <td> Predicting the Cellular Localization Sites of Proteins </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 1484 </td>
<td> 8 </td>

<td> 1996 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Zoo"><img border="1" src="assets/MLimages/SmallLarge111.jpg"/></a>
</td><td> <a href="datasets/Zoo">Zoo</a> </td></tr></table></td>
<!-- <td> Artificial, 7 classes of animals </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 101 </td>
<td> 17 </td>
<td> 1990 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Undocumented"><img border="1" src="assets/MLimages/SmallLargedef
ault.jpg"/></a> </td><td> <a href="datasets/Undocumented">Undocumented</a></p
></td></tr></table></td>
<!-- <td> Various datasets without documentation (feel free to explore!) </td> -->
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Twenty+Newsgroups"><img border="1" src="assets/MLimages/SmallLar
gedefault.jpg"/></a> </td><td> <a href="datasets/Twenty+Newsgroups">Twenty Newsgro
ups</a> </td></tr></table></td>
<!-- <td> This data set consists of 20000 messages taken from 20 newsgroups. </
td> -->
<td> Text </td>
<td> </td>
<td> </td>
<td> 20000 </td>
<td> </td>
<td> 1999 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Australian+Sign+Language+signs"><img border="1" src="assets/MLimag
es/SmallLarge114.jpg"/></a> </td><td> <a href="datasets/Australian+Sign+Language+si
gns">Australian Sign Language signs</a> </td></tr></table></td>
<!-- <td> This data consists of sample of Auslan (Australian Sign Language) signs. Examples
of 95 signs were collected from five signers with a total of 6650 sign samples. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Categorical, Real </td>
<td> 6650 </td>
<td> 15 </td>
<td> 1999 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Australian+Sign+Language+signs+%28High+Quality%29"><img border="
1" src="assets/MLimages/SmallLarge114.jpg"/></a> </td><td> <a href="datasets/Australi
an+Sign+Language+signs+%28High+Quality%29">Australian Sign Language signs (High Quality)</a>
</td></tr></table></td>
<!-- <td> This data consists of sample of Auslan (Australian Sign Language) signs. 27 examp
les of each of 95 Auslan signs were captured from a native signer using high-quality position trackers </p
></td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Real </td>

<td> 2565 </td>
<td> 22 </td>
<td> 2002 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/US+Census+Data+%281990%29"><img border="1" src="assets/MLimages/SmallLarge2.jpg"/></a> </td><td> <a href="datasets/US+Census+Data+%281990%29">US Census Data (1990)</a> </td></tr></table></td>
<!-- <td> The USCensus1990raw data set contains a one percent sample of the Public Use Microdata Samples (PUMS) person records drawn from the full 1990 census sample. </td> -->
<td> Multivariate </td>
<td> Clustering </td>
<td> Categorical </td>
<td> 2458285 </td>
<td> 68 </td>
<td> </td>
<!-- <td> Social </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Census-Income+%28KDD%29"><img border="1" src="assets/MLimages/SmallLarge2.jpg"/></a> </td><td> <a href="datasets/Census-Income+%28KDD%29">Census-Income (KDD)</a> </td></tr></table></td>
<!-- <td> This data set contains weighted census data extracted from the 1994 and 1995 current population surveys conducted by the U.S. Census Bureau. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 299285 </td>
<td> 40 </td>
<td> 2000 </td>
<!-- <td> Social </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Coil+1999+Competition+Data"><img border="1" src="assets/MLimages/SmallLarge118.jpg"/></a> </td><td> <a href="datasets/Coil+1999+Competition+Data">Coil 1999 Competition Data</a> </td></tr></table></td>
<!-- <td> This data set is from the 1999 Computational Intelligence and Learning (COIL) competition. The data contains measurements of river chemical concentrations and algae densities. </td> -->
<td> Multivariate </td>
<td> </td>
<td> Categorical, Real </td>
<td> 340 </td>
<td> 17 </td>
<td> 1999 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Corel+Image+Features"><img border="1" src="assets/MLimages/SmallLarge119.jpg"/></a> </td><td> <a href="datasets/Corel+Image+Features">Corel Image Features</a> </td></tr></table></td>
<!-- <td> This dataset contains image features extracted from a Corel image collection. Four sets of features are available based on the color histogram, color histogram layout, color moments, and co-occurence </td> -->
<td> Multivariate </td>
<td> </td>
<td> Real </td>
<td> 68040 </td>
<td> 89 </td>
<td> 1999 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/E.+Coli+Genes"><img border="1" src="assets/MLimages/SmallLarge120.jpg"/></a> </td><td> <a href="datasets/E.+Coli+Genes">E. Coli Genes</a> </td>
</tr></table></td>

<!-- <td> Data giving characteristics of each ORF (potential gene) in the E. coli genome. Seq uence, homology (similarity to other genes) and structural information, and function (if known) are provided.&nb sp; </td> -->
<td> Relational </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> 2001 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/EEG+Database"><img border="1" src="assets/MLimages/SmallLarge121 .jpg"/></a> </td><td> <a href="datasets/EEG+Database">EEG Database</a> </t d></tr></table></td>
<!-- <td> This data arises from a large study to examine EEG correlates of genetic predisposi tion to alcoholism. It contains measurements from 64 electrodes placed on the scalp sampled at 256 Hz </td> -->
<td> Multivariate, Time-Series </td>
<td> </td>
<td> Categorical, Integer, Real </td>
<td> 122 </td>
<td> 4 </td>
<td> 1999 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/El+Nino"><img border="1" src="assets/MLimages/SmallLargedefault.jpg" /></a> </td><td> <a href="datasets/El+Nino">El Nino</a> </td></tr></table></td>
<!-- <td> The data set contains oceanographic and surface meteorological readings taken fr om a series of buoys positioned throughout the equatorial Pacific. </td> -->
<td> Spatio-temporal </td>
<td> </td>
<td> Integer, Real </td>
<td> 178080 </td>
<td> 12 </td>
<td> 1999 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Entree+Chicago+Recommendation+Data"><img border="1" src="assets/ MLimages/SmallLarge123.jpg"/></a> </td><td> <a href="datasets/Entree+Chicago+Reco mmendation+Data">Entree Chicago Recommendation Data</a> </td></tr></table></td>
<!-- <td> This data contains a record of user interactions with the Entree Chicago restaurant recommendation system. </td> -->
<td> Transactional, Sequential </td>
<td> Recommender-Systems </td>
<td> Categorical </td>
<td> 50672 </td>
<td> </td>
<td> 2000 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/CMU+Face+Images"><img border="1" src="assets/MLimages/SmallLarg e124.jpg"/></a> </td><td> <a href="datasets/CMU+Face+Images">CMU Face Images</a > </td></tr></table></td>
<!-- <td> This data consists of 640 black and white face images of people taken with varying pose (straight, left, right, up), expression (neutral, happy, sad, angry), eyes (wearing sunglasses or not), and si ze </td> -->
<td> Image </td>
<td> Classification </td>
<td> Integer </td>
<td> 640 </td>
<td> </td>
<td> 1999 </td>

<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Insurance+Company+Benchmark+%28COIL+2000%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Insurance+Company+Benchmark+%28COIL+2000%29">Insurance Company Benchmark (COIL 2000)</a> </td></tr></table></td>
<!-- <td> This data set used in the CoIL 2000 Challenge contains information on customers of an insurance company. The data consists of 86 variables and includes product usage data and socio-demographic data </td> -->
<td> Multivariate </td>
<td> Regression, Description </td>
<td> Categorical, Integer </td>
<td> 9000 </td>
<td> 86 </td>
<td> 2000 </td>
<!-- <td> Social </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Internet+Usage+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Internet+Usage+Data">Internet Usage Data</a> </td></tr></table></td>
<!-- <td> This data contains general demographic information on internet users in 1997. </td> -->
<td> Multivariate </td>
<td> </td>
<td> Categorical, Integer </td>
<td> 10104 </td>
<td> 72 </td>
<td> 1999 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/IPUMS+Census+Database"><img border="1" src="assets/MLimages/SmallLarge2.jpg"/></a> </td><td> <a href="datasets/IPUMS+Census+Database">IPUMS Census Database</a> </td></tr></table></td>
<!-- <td> This data set contains unweighted PUMS census data from the Los Angeles and Long Beach areas for the years 1970, 1980, and 1990. </td> -->
<td> Multivariate </td>
<td> </td>
<td> Categorical, Integer </td>
<td> 256932 </td>
<td> 61 </td>
<td> 1999 </td>
<!-- <td> Social </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Japanese+Vowels"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Japanese+Vowels">Japanese Vowels</a> </td></tr></table></td>
<!-- <td> This dataset records 640 time series of 12 LPC cepstrum coefficients taken from nine male speakers. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 640 </td>
<td> 12 </td>
<td> </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/KDD+Cup+1998+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/KDD+Cup+1998+Data">KDD Cup 1998 Data</a> </td></tr></table></td>
<!-- <td> This is the data set used for The Second International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-98 </td> -->
<td> Multivariate </td>

<td> Regression </td>
<td> Categorical, Integer </td>
<td> 191779 </td>
<td> 481 </td>
<td> 1998 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/KDD+Cup+1999+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/KDD+Cup+1999+Data">KDD Cup 1999 Data</a> </td></tr></table></td>
<!-- <td> This is the data set used for The Third International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-99 </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 4000000 </td>
<td> 42 </td>
<td> 1999 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/M.+Tuberculosis+Genes"><img border="1" src="assets/MLimages/SmallLarge131.jpg"/></a> </td><td> <a href="datasets/M.+Tuberculosis+Genes">M. Tuberculosis Genes</a> </td></tr></table></td>
<!-- <td> Data giving characteristics of each ORF (potential gene) in the M. tuberculosis bacterium. Sequence, homology (similarity to other genes) and structural information, and function (if known) are provided </td> -->
<td> Relational </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> 2001 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Movie"><img border="1" src="assets/MLimages/SmallLarge132.jpg"/></a> </td><td> <a href="datasets/Movie">Movie</a> </td></tr></table></td>
<!-- <td> This data set contains a list of over 10000 films including many older, odd, and cult films. There is information on actors, casts, directors, producers, studios, etc. </td> -->
<td> Multivariate, Relational </td>
<td> </td>
<td> </td>
<td> 10000 </td>
<td> </td>
<td> 1999 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/MSNBC.com+Anonymous+Web+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/MSNBC.com+Anonymous+Web+Data">MSNBC.com Anonymous Web Data</a> </td></tr></table></td>
<!-- <td> This data describes the page visits of users who visited msnbc.com on September 28, 1999. Visits are recorded at the level of URL category (see description) and are recorded in time order. </td> -->
<td> Sequential </td>
<td> </td>
<td> Categorical </td>
<td> 989818 </td>
<td> </td>
<td> </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/NSF+Research+Award+Abstracts+1990-2003"><img border="1" src="assets/MLimages/SmallLarge134.jpg"/></a> </td><td> <a href="datasets/NSF+Research+A

ward+Abstracts+1990-2003">NSF Research Award Abstracts 1990-2003</a> </td></tr></table></td>
<!-- <td> This data set consists of (a) 129,000 abstracts describing NSF awards for basic res
earch, (b) bag-of-word data files extracted from the abstracts, (c) a list of words used for indexing the bag-of-w
ord </td> -->
<td> Text </td>
<td> </td>
<td> </td>
<td> 129000 </td>
<td> </td>
<td> 2003 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Pioneer-1+Mobile+Robot+Data"><img border="1" src="assets/MLimages
/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Pioneer-1+Mobile+Robot+Dat
a">Pioneer-1 Mobile Robot Data</a> </td></tr></table></td>
<!-- <td> This dataset contains time series sensor readings of the Pioneer-1 mobile robot. Th
e data is broken into "experiences" in which the robot takes action for some period of time and experiences a c
ontrol </td> -->
<td> Multivariate, Time-Series </td>
<td> </td>
<td> Categorical, Real </td>
<td> </td>
<td> </td>
<td> 1999 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Pseudo+Periodic+Synthetic+Time+Series"><img border="1" src="assets/
MLimages/SmallLarge136.jpg"/></a> </td><td> <a href="datasets/Pseudo+Periodic+Synt
hetic+Time+Series">Pseudo Periodic Synthetic Time Series</a> </td></tr></table></td>
<!-- <td> This data set is designed for testing indexing schemes in time series databases. Th
e data appears highly periodic, but never exactly repeats itself. </td> -->
<td> Univariate, Time-Series </td>
<td> </td>
<td> </td>
<td> 100000 </td>
<td> </td>
<td> 1999 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Reuters-21578+Text+Categorization+Collection"><img border="1" src="a
ssets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Reuters-2157
8+Text+Categorization+Collection">Reuters-21578 Text Categorization Collection</a> </td></tr></table
></td>
<!-- <td> This is a collection of documents that appeared on Reuters newswire in 1987. The
documents were assembled and indexed with categories. </td> -->
<td> Text </td>
<td> Classification </td>
<td> Categorical </td>
<td> 21578 </td>
<td> 5 </td>
<td> 1997 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Robot+Execution+Failures"><img border="1" src="assets/MLimages/Sm
allLarge138.jpg"/></a> </td><td> <a href="datasets/Robot+Execution+Failures">Robot E
xecution Failures</a> </td></tr></table></td>
<!-- <td> This dataset contains force and torque measurements on a robot after failure detec
tion. Each failure is characterized by 15 force/torque samples collected at regular time intervals </td>
-->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Integer </td>

<td> Integer </td>
<td> 463 </td>
<td> 90 </td>
<td> 1999 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Synthetic+Control+Chart+Time+Series"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Synthetic+Control+Chart+Time+Series">Synthetic Control Chart Time Series</a> </td></tr></table></td>
<!-- <td> This data consists of synthetically generated control charts. </td> -->
<td> Time-Series </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 600 </td>
<td> </td>
<td> 1999 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Syskill+and+Webert+Web+Page+Ratings"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Syskill+and+Webert+Web+Page+Ratings">Syskill and Webert Web Page Ratings</a> </td></tr></table></td>
<!-- <td> This database contains HTML source of web pages plus the ratings of a single user on these web pages. Web pages are on four seperate subjects (Bands- recording artists; Goats; Sheep; and BioMedical) </td> -->
<td> Multivariate, Text </td>
<td> Classification </td>
<td> Categorical </td>
<td> 332 </td>
<td> 5 </td>
<td> 1998 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/UNIX+User+Data"><img border="1" src="assets/MLimages/SmallLarge141.jpg"/></a> </td><td> <a href="datasets/UNIX+User+Data">UNIX User Data</a> </td></tr></table></td>
<!-- <td> This file contains 9 sets of sanitized user data drawn from the command histories of 8 UNIX computer users at Purdue over the course of up to 2 years. </td> -->
<td> Text, Sequential </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Volcanoes+on+Venus+-+JARtool+experiment"><img border="1" src="assets/MLimages/SmallLarge142.jpg"/></a> </td><td> <a href="datasets/Volcanoes+on+Venus+-+JARtool+experiment">Volcanoes on Venus - JARtool experiment</a> </td></tr></table></td>
<!-- <td> The JARtool project was a pioneering effort to develop an automatic system for cataloging small volcanoes in the large set of Venus images returned by the Magellan spacecraft. </td> -->
<td> Image </td>
<td> Classification </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Statlog+%28Australian+Credit+Approval%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Statlog+%28Australian+Credit+Approval%29">Statlog (Australian Credit Approval)</a> </td></tr></table></td>
<!-- <td> This file concerns credit card applications. This database exists elsewhere in the re-

<!-- <td> This file concerns credit card applications. This database exists elsewhere in the re
pository (Credit Screening Database) in a slightly different form </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer, Real </td>
<td> 690 </td>
<td> 14 </td>
<td> </td>
<!-- <td> Financial </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Statlog+%28German+Credit+Data%29"><img border="1" src="assets/ML
images/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Statlog+%28German+
Credit+Data%29">Statlog (German Credit Data)</a> </td></tr></table></td>
<!-- <td> This dataset classifies people described by a set of attributes as good or bad credit
risks. Comes in two formats (one all numeric). Also comes with a cost matrix </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 1000 </td>
<td> 20 </td>
<td> 1994 </td>
<!-- <td> Financial </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Statlog+%28Heart%29"><img border="1" src="assets/MLimages/SmallLa
rge45.jpg"/></a> </td><td> <a href="datasets/Statlog+%28Heart%29">Statlog (Heart)</a
> </td></tr></table></td>
<!-- <td> This dataset is a heart disease database similar to a database already present in th
e repository (Heart Disease databases) but in a slightly different form </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Real </td>
<td> 270 </td>
<td> 13 </td>
<td> </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Statlog+%28Landsat+Satellite%29"><img border="1" src="assets/MLima
ges/SmallLarge146.jpg"/></a> </td><td> <a href="datasets/Statlog+%28Landsat+Satellit
e%29">Statlog (Landsat Satellite)</a> </td></tr></table></td>
<!-- <td> Multi-spectral values of pixels in 3x3 neighbourhoods in a satellite image, and the cl
assification associated with the central pixel in each neighbourhood </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 6435 </td>
<td> 36 </td>
<td> 1993 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Statlog+%28Image+Segmentation%29"><img border="1" src="assets/ML
images/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Statlog+%28Image+Se
gmentation%29">Statlog (Image Segmentation)</a> </td></tr></table></td>
<!-- <td> This dataset is an image segmentation database similar to a database already pres
ent in the repository (Image segmentation database) but in a slightly different form. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 2310 </td>
<td> 19 </td>
<td> 1990 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">

<td><table><tr><td><a href="datasets/Statlog+%28Shuttle%29"><img border="1" src="assets/MLimages/SmallLarge92.jpg"/></a> </td><td> <a href="datasets/Statlog+%28Shuttle%29">Statlog (Shuttle)</a> </td></tr></table></td>
<!-- <td> The shuttle dataset contains 9 attributes all of which are numerical. Approximately 80% of the data belongs to class 1 </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 58000 </td>
<td> 9 </td>
<td> </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Statlog+%28Vehicle+Silhouettes%29"><img border="1" src="assets/MLimages/SmallLarge149.jpg"/></a> </td><td> <a href="datasets/Statlog+%28Vehicle+Silhouettes%29">Statlog (Vehicle Silhouettes)</a> </td></tr></table></td>
<!-- <td> 3D objects within a 2D image by application of an ensemble of shape feature extractors to the 2D silhouettes of the objects. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 946 </td>
<td> 18 </td>
<td> </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Connectionist+Bench+%28Nettalk+Corpus%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Connectionist+Bench+%28Nettalk+Corpus%29">Connectionist Bench (Nettalk Corpus)</a> </td></tr></table></td>
<!-- <td> The file "nettalk.data" contains a list of 20,008 English words, along with a phonetic transcription for each word. The task is to train a network to produce the proper phonemes </td> -->
<td> Multivariate </td>
<td> </td>
<td> Categorical </td>
<td> 20008 </td>
<td> 4 </td>
<td> </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29">Connectionist Bench (Sonar, Mines vs. Rocks)</a> </td></tr></table></td>
<!-- <td> The task is to train a network to discriminate between sonar signals bounced off a metal cylinder and those bounced off a roughly cylindrical rock. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 208 </td>
<td> 60 </td>
<td> </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29">Connectionist Bench (Vowel Recognition - Deterding Data)</a> </td></tr></table></td>
<!-- <td> Speaker independent recognition of the eleven steady state vowels of British English using a specified training set of lpc derived log area ratios. </td> -->
<td> </td>
<td> Classification </td>
<td> Real </td>

<td> 528 </td>
<td> 10 </td>
<td> </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Economic+Sanctions">
</a> </td><td> <a href="datasets/Economic+Sanctions">Economic Sanctions</a> </td></tr></table></td>
<!-- <td> Domain Theory on Economic Sanctions; Undocumented </td> -->
<td> Domain-Theory </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<!-- <td> Financial </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Protein+Data">
</a> </td><td> <a href="datasets/Protein+Data">Protein Data</a> </td></tr></table></td>
<!-- <td> Undocumented </td> -->
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Cloud">
</a> </td><td> <a href="datasets/Cloud">Cloud</a> </td></tr></table></td>
<!-- <td> Little Documentation </td> -->
<td> Multivariate </td>
<td> </td>
<td> Real </td>
<td> 1024 </td>
<td> 10 </td>
<td> 1989 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/CalIt2+Building+People+Counts">
</a> </td><td> <a href="datasets/CalIt2+Building+People+Counts">CalIt2 Building People Counts</a> </td></tr></table></td>
<!-- <td> This data comes from the main door of the CalIt2 building at UCI. </td> -->
<td> Multivariate, Time-Series </td>
<td> </td>
<td> Categorical, Integer </td>
<td> 10080 </td>
<td> 4 </td>
<td> 2006 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Dodgers+Loop+Sensor">
</a> </td><td> <a href="datasets/Dodgers+Loop+Sensor">Dodgers Loop Sensor</a> </td></tr></table></td>
<!-- <td> Loop sensor data was collected for the Glendale on ramp for the 101 North freeway in Los Angeles </td> -->
<td> Multivariate, Time-Series </td>
<td> </td>
<td> Categorical, Integer </td>
<td> 50400 </td>

<td> 3 </td>
<td> 2006 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Poker+Hand"><img border="1" src="assets/MLimages/SmallLarge158.jpg"/></a> </td><td> <a href="datasets/Poker+Hand">Poker Hand</a> </td></tr></table></td>
<!-- <td> Purpose is to predict poker hands </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 1025010 </td>
<td> 11 </td>
<td> 2007 </td>
<!-- <td> Game </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/MAGIC+Gamma+Telescope"><img border="1" src="assets/MLimages/SmallLarge159.jpg"/></a> </td><td> <a href="datasets/MAGIC+Gamma+Telescope">MAGIC Gamma Telescope</a> </td></tr></table></td>
<!-- <td> Data are MC generated to simulate registration of high energy gamma particles in an atmospheric Cherenkov telescope </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 19020 </td>
<td> 11 </td>
<td> 2007 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/UJI+Pen+Characters"><img border="1" src="assets/MLimages/SmallLarge160.jpg"/></a> </td><td> <a href="datasets/UJI+Pen+Characters">UJI Pen Characters</a> </td></tr></table></td>
<!-- <td> Data consists of written characters in a UNIPEN-like format </td> -->
<td> Multivariate, Sequential </td>
<td> Classification </td>
<td> Integer </td>
<td> 1364 </td>
<td> </td>
<td> 2007 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Mammographic+Mass"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Mammographic+Mass">Mammographic Mass</a> </td></tr></table></td>
<!-- <td> Discrimination of benign and malignant mammographic masses based on BI-RADS attributes and the patient's age. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 961 </td>
<td> 6 </td>
<td> 2007 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Forest+Fires"><img border="1" src="assets/MLimages/SmallLarge162.jpg"/></a> </td><td> <a href="datasets/Forest+Fires">Forest Fires</a> </td></tr></table></td>
<!-- <td> This is a difficult regression task, where the aim is to predict the burned area of forest fires, in the northeast region of Portugal, by using meteorological and other data (see details at: http://www.dsi.uminho.pt/~pcortez/forestfires). </td> -->
<td> Multivariate </td>
<td> Regression </td>

<td> Real </td>
<td> 517 </td>
<td> 13 </td>
<td> 2008 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Reuters+Transcribed+Subset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Reuters+Transcribed+Subset">Reuters Transcribed Subset</a> </td></tr></table></td>
<!-- <td> This dataset is created by reading out 200 files from the 10 largest Reuters classes and using an Automatic Speech Recognition system to create corresponding transcriptions. </td> -->
<td> Text </td>
<td> Classification </td>
<td> </td>
<td> 200 </td>
<td> </td>
<td> 2008 </td>
<!-- <td> Business </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Bag+of+Words"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Bag+of+Words">Bag of Words</a> </td></tr></table></td>
<!-- <td> This data set contains five text collections in the form of bags-of-words. </td> -->
<td> Text </td>
<td> Clustering </td>
<td> Integer </td>
<td> 8000000 </td>
<td> 100000 </td>
<td> 2008 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Concrete+Compressive+Strength"><img border="1" src="assets/MLimages/SmallLarge165.jpg"/></a> </td><td> <a href="datasets/Concrete+Compressive+Strength">Concrete Compressive Strength</a> </td></tr></table></td>
<!-- <td> Concrete is the most important material in civil engineering. The concrete compressive strength is a highly nonlinear function of age and ingredients.   </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 1030 </td>
<td> 9 </td>
<td> 2007 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Hill-Valley"><img border="1" src="assets/MLimages/SmallLarge166.jpg"/></a> </td><td> <a href="datasets/Hill-Valley">Hill-Valley</a> </td></tr></table></td>
<!-- <td> Each record represents 100 points on a two-dimensional graph. When plotted in order (from 1 through 100) as the Y co-ordinate, the points will create either a Hill (a &#65533;bump&#65533; in the terrain) or a Valley (a &#65533;dip&#65533; in the terrain). </td> -->
<td> Sequential </td>
<td> Classification </td>
<td> Real </td>
<td> 606 </td>
<td> 101 </td>
<td> 2008 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Arcene"><img border="1" src="assets/MLimages/SmallLarge167.jpg"/></a> </td><td> <a href="datasets/Arcene">Arcene</a> </td></tr></table></td>

<!-- <td> ARCENE's task is to distinguish cancer versus normal patterns from mass-spectro metric data. This is a two-class classification problem with continuous input variables. This dataset is one of 5 d atasets of the NIPS 2003 feature selection challenge. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 900 </td>
<td> 10000 </td>
<td> 2008 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Dexter"><img border="1" src="assets/MLimages/SmallLarge168.jpg"/></a> </td><td> <a href="datasets/Dexter">Dexter</a> </td></tr></table></td>
<!-- <td> DEXTER is a text classification problem in a bag-of-word representation. This is a t wo-class classification problem with sparse continuous input variables. This dataset is one of five datasets of th e NIPS 2003 feature selection challenge.
  </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 2600 </td>
<td> 20000 </td>
<td> 2008 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Dorothea"><img border="1" src="assets/MLimages/SmallLarge169.jpg"/></a> </td><td> <a href="datasets/Dorothea">Dorothea</a> </td></tr></table></td>
<!-- <td> DOROTHEA is a drug discovery dataset. Chemical compounds represented by stru ctural molecular features must be classified as active (binding to thrombin) or inactive. This is one of 5 datasets of the NIPS 2003 feature selection challenge. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 1950 </td>
<td> 100000 </td>
<td> 2008 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Gisette"><img border="1" src="assets/MLimages/SmallLarge170.jpg"/></a> </td><td> <a href="datasets/Gisette">Gisette</a> </td></tr></table></td>
<!-- <td> GISETTE is a handwritten digit recognition problem. The problem is to separate the highly confusible digits '4' and '9'. This dataset is one of five datasets of the NIPS 2003 feature selection challen ge.
  </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 13500 </td>
<td> 5000 </td>
<td> 2008 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Madelon"><img border="1" src="assets/MLimages/SmallLarge171.jpg"/></a> </td><td> <a href="datasets/Madelon">Madelon</a> </td></tr></table></td>
<!-- <td> MADELON is an artificial dataset, which was part of the NIPS 2003 feature selection challenge. This is a two-class classification problem with continuous input variables. The difficulty is that the pro blem is multivariate and highly non-linear.   </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 4400 </td>

<td> 500 </td>
<td> 2008 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Ozone+Level+Detection"><img border="1" src="assets/MLimages/SmallLarge172.jpg"/></a> </td><td> <a href="datasets/Ozone+Level+Detection">Ozone Level Detection</a> </td></tr></table></td>
<!-- <td> Two ground ozone level data sets are included in this collection. One is the eight hour peak set (eighthr.data), the other is the one hour peak set (onehr.data). Those data were collected from 1998 to 2004 at the Houston, Galveston and Brazoria area. </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 2536 </td>
<td> 73 </td>
<td> 2008 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Abscisic+Acid+Signaling+Network"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Abscisic+Acid+Signaling+Network">Abscisic Acid Signaling Network</a> </td></tr></table></td>
<!-- <td> The objective is to determine the set of boolean rules that describe the interactions of the nodes within this plant signaling network. The dataset includes 300 separate boolean pseudodynamic simulations using an asynchronous update scheme.   </td> -->
<td> Multivariate </td>
<td> Causal-Discovery </td>
<td> Integer </td>
<td> 300 </td>
<td> 43 </td>
<td> 2008 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Parkinsons"><img border="1" src="assets/MLimages/SmallLarge174.jpg" /></a> </td><td> <a href="datasets/Parkinsons">Parkinsons</a> </td></tr></table></td>
<!-- <td> Oxford Parkinson's Disease Detection Dataset </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 197 </td>
<td> 23 </td>
<td> 2008 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Character+Trajectories"><img border="1" src="assets/MLimages/SmallLarge175.jpg"/></a> </td><td> <a href="datasets/Character+Trajectories">Character Trajectories</a> </td></tr></table></td>
<!-- <td> Multiple, labelled samples of pen tip trajectories recorded whilst writing individual characters. All samples are from the same writer, for the purposes of primitive extraction. Only characters with a single pen-down segment were considered. </td> -->
<td> Time-Series </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 2858 </td>
<td> 3 </td>
<td> 2008 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Blood+Transfusion+Service+Center"><img border="1" src="assets/MLimages/SmallLarge176.jpg"/></a> </td><td> <a href="datasets/Blood+Transfusion+Service+Center">Blood Transfusion Service Center</a> </td></tr></table></td>
<!-- <td> Data taken from the Blood Transfusion Service Center in Hsin-Chu City in Taiwan --

<!-- <td> Data taken from the Blood Transfusion Service Center in Hsin-Chu City in Taiwan this is a classification problem.   </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 748 </td>
<td> 5 </td>
<td> 2008 </td>
<!-- <td> Business </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/UJI+Pen+Characters+%28Version+2%29"><img border="1" src="assets/MLimages/SmallLarge160.jpg"/></a> </td><td> <a href="datasets/UJI+Pen+Characters+%28Version+2%29">UJI Pen Characters (Version 2)</a> </td></tr></table></td>
<!-- <td> A pen-based database with more than 11k isolated handwritten characters </td> -->
<td> Multivariate, Sequential </td>
<td> Classification </td>
<td> Integer </td>
<td> 11640 </td>
<td> </td>
<td> 2009 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Semeion+Handwritten+Digit"><img border="1" src="assets/MLimages/SmallLarge178.jpg"/></a> </td><td> <a href="datasets/Semeion+Handwritten+Digit">Semeion Handwritten Digit</a> </td></tr></table></td>
<!-- <td> 1593 handwritten digits from around 80 persons were scanned, stretched in a rectangular box 16x16 in a gray scale of 256 values. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 1593 </td>
<td> 256 </td>
<td> 2008 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/SECOM"><img border="1" src="assets/MLimages/SmallLarge179.jpg"/></a> </td><td> <a href="datasets/SECOM">SECOM</a> </td></tr></table></td>
<!-- <td> Data from a semi-conductor manufacturing process </td> -->
<td> Multivariate </td>
<td> Classification, Causal-Discovery </td>
<td> Real </td>
<td> 1567 </td>
<td> 591 </td>
<td> 2008 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Plants"><img border="1" src="assets/MLimages/SmallLarge180.jpg"/></a> </td><td> <a href="datasets/Plants">Plants</a> </td></tr></table></td>
<!-- <td> Data has been extracted from the USDA plants database. It contains all plants (species and genera) in the database and the states of USA and Canada where they occur. </td> -->
<td> Multivariate </td>
<td> Clustering </td>
<td> Categorical </td>
<td> 22632 </td>
<td> 70 </td>
<td> 2008 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Libras+Movement"><img border="1" src="assets/MLimages/SmallLarge181.jpg"/></a> </td><td> <a href="datasets/Libras+Movement">Libras Movement</a> </td></tr></table></td>
<!-- <td> The data set contains 15 classes of 24 instances each. Each class references to a

<!-- <td> The data set contains 15 classes of 24 instances each. Each class references to a hand movement type in LIBRAS (Portuguese
name 'L&#65533;ngua BRAsileira de Sinais', oficial brazilian signal language). </td> -->
<td> Multivariate, Sequential </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 360 </td>
<td> 91 </td>
<td> 2009 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Concrete+Slump+Test"><img border="1" src="assets/MLimages/SmallLarge165.jpg"/></a> </td><td> <a href="datasets/Concrete+Slump+Test">Concrete Slump Test</a> </td></tr></table></td>
<!-- <td> Concrete is a highly complex material. The slump flow of concrete is not only determined by the water content, but that is also influenced by other concrete ingredients. </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 103 </td>
<td> 10 </td>
<td> 2009 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Communities+and+Crime"><img border="1" src="assets/MLimages/SmallLarge183.jpg"/></a> </td><td> <a href="datasets/Communities+and+Crime">Communities and Crime</a> </td></tr></table></td>
<!-- <td> Communities within the United States. The data combines socio-economic data from the 1990 US Census, law enforcement data from the 1990 US LEMAS survey, and crime data from the 1995 FBI UCR. </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 1994 </td>
<td> 128 </td>
<td> 2009 </td>
<!-- <td> Social </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Acute+Inflammations"><img border="1" src="assets/MLimages/SmallLarge184.jpg"/></a> </td><td> <a href="datasets/Acute+Inflammations">Acute Inflammations</a> </td></tr></table></td>
<!-- <td> The data was created by a medical expert as a data set to test the expert system, which will perform the presumptive diagnosis of two diseases of the urinary system.
  </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 120 </td>
<td> 6 </td>
<td> 2009 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Wine+Quality"><img border="1" src="assets/MLimages/SmallLarge186.jpg"/></a> </td><td> <a href="datasets/Wine+Quality">Wine Quality</a> </td></tr></table></td>
<!-- <td> Two datasets are included, related to red and white vinho verde wine samples, from the north of Portugal. The goal is to model wine quality based on physicochemical tests (see [Cortez et al., 2009], http://www3.dsi.uminho.pt/pcortez/wine/). </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Real </td>
<td> 4898 </td>
</tr></table></td>
</tr></table></td>
</tr>

<td> 12 </td>
<td> 2009 </td>
<!-- <td> Business </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/URL+Reputation"><img border="1" src="assets/MLimages/SmallLarge187.jpg"/></a> </td><td> <a href="datasets/URL+Reputation">URL Reputation</a> </td></tr></table></td>
<!-- <td> Anonymized 120-day subset of the ICML-09 URL data containing 2.4 million examples and 3.2 million features. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 2396130 </td>
<td> 3231961 </td>
<td> 2009 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/p53+Mutants"><img border="1" src="assets/MLimages/SmallLarge188.jpg"/></a> </td><td> <a href="datasets/p53+Mutants">p53 Mutants</a> </td></tr></table></td>
<!-- <td> The goal is to model mutant p53 transcriptional activity (active vs inactive) based on data extracted from biophysical simulations.
  </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 16772 </td>
<td> 5409 </td>
<td> 2010 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Parkinsons+Telemonitoring"><img border="1" src="assets/MLimages/SmallLarge174.jpg"/></a> </td><td> <a href="datasets/Parkinsons+Telemonitoring">Parkinsons Telemonitoring</a> </td></tr></table></td>
<!-- <td> Oxford Parkinson's Disease Telemonitoring Dataset </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Integer, Real </td>
<td> 5875 </td>
<td> 26 </td>
<td> 2009 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Demospongiae"><img border="1" src="assets/MLimages/SmallLarge190.jpg"/></a> </td><td> <a href="datasets/Demospongiae">Demospongiae</a> </td></tr></table></td>
<!-- <td> Marine sponges of the Demospongiae class classification domain. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 503 </td>
<td> </td>
<td> 2010 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Opinosis+Opinion+%26frasl%3B+Review"><img border="1" src="assets/MLimages/SmallLarge191.jpg"/></a> </td><td> <a href="datasets/Opinosis+Opinion+%26frasl%3B+Review">Opinosis Opinion / Review</a> </td></tr></table></td>
<!-- <td> This dataset contains sentences extracted from user reviews on a given topic. Example topics are “performance of Toyota Camry” and “sound quality of ipod nano”.   </td> -->
<td> Text </td>

<td> </td>
<td> </td>
<td> 51 </td>
<td> </td>
<td> 2010 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Breast+Tissue"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Breast+Tissue">Breast Tissue</a> </td></tr></table></td>
<!-- <td> Dataset with electrical impedance measurements of freshly excised tissue samples from the breast. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 106 </td>
<td> 10 </td>
<td> 2010 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Cardiotocography"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Cardiotocography">Cardiotocography</a> </td></tr></table></td>
<!-- <td> The dataset consists of measurements of fetal heart rate (FHR) and uterine contraction (UC) features on cardiotocograms classified by expert obstetricians. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 2126 </td>
<td> 23 </td>
<td> 2010 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Wall-Following+Robot+Navigation+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Wall-Following+Robot+Navigation+Data">Wall-Following Robot Navigation Data</a> </td></tr></table></td>
<!-- <td> The data were collected as the SCITOS G5 robot navigates through the room following the wall in a clockwise direction, for 4 rounds, using 24 ultrasound sensors arranged circularly around its 'waist'. </td> -->
<td> Multivariate, Sequential </td>
<td> Classification </td>
<td> Real </td>
<td> 5456 </td>
<td> 24 </td>
<td> 2010 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Spoken+Arabic+Digit"><img border="1" src="assets/MLimages/SmallLarge195.jpg"/></a> </td><td> <a href="datasets/Spoken+Arabic+Digit">Spoken Arabic Digit</a> </td></tr></table></td>
<!-- <td> This dataset contains timeseries of mel-frequency cepstrum coefficients (MFCCs) corresponding to spoken Arabic digits. Includes data from 44 male and 44 female native Arabic speakers. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 8800 </td>
<td> 13 </td>
<td> 2010 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Localization+Data+for+Person+Activity"><img border="1" src="assets/ML

images/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Localization+Data+for+Person+Activity">Localization Data for Person Activity</a> </td></tr></table></td>
<!-- <td> Data contains recordings of five people performing different activities. Each person wore four sensors (tags) while performing the same scenario five times.   </td> -->
<td> Univariate, Sequential, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 164860 </td>
<td> 8 </td>
<td> 2010 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/AutoUniv"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/AutoUniv">AutoUniv</a> </td></tr></table></td>
<!-- <td> AutoUniv is an advanced data generator for classifications tasks. The aim is to reflect the nuances and heterogeneity of real data. Data can be generated in .csv, ARFF or C4.5 formats. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer, Real </td>
<td> </td>
<td> </td>
<td> 2010 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Steel+Plates+Faults"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Steel+Plates+Faults">Steel Plates Faults</a> </td></tr></table></td>
<!-- <td> A dataset of steel plates' faults, classified into 7 different types.
The goal was to train machine learning for automatic pattern recognition.
  </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 1941 </td>
<td> 27 </td>
<td> 2010 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/MiniBooNE+particle+identification"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/MiniBooNE+particle+identification">MiniBooNE particle identification</a> </td></tr></table></td>
<!-- <td> This dataset is taken from the MiniBooNE experiment and is used to distinguish electron neutrinos (signal) from muon neutrinos (background). </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 130065 </td>
<td> 50 </td>
<td> 2010 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/YearPredictionMSD"><img border="1" src="assets/MLimages/SmallLarge203.jpg"/></a> </td><td> <a href="datasets/YearPredictionMSD">YearPredictionMSD</a> </td></tr></table></td>
<!-- <td> Prediction of the release year of a song from audio features. Songs are mostly western, commercial tracks ranging from 1922 to 2011, with a peak in the year 2000s. </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 515345 </td>

<td> 90 </td>
<td> 2011 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/PEMS-SF"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/PEMS-SF">PEMS-SF</a> </td></tr></table></td>
<!-- <td> 15 months worth of daily data (440 daily records) that describes the occupancy rate, between 0 and 1, of different car lanes of the San Francisco bay area freeways across time. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 440 </td>
<td> 138672 </td>
<td> 2011 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/OpinRank+Review+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/OpinRank+Review+Dataset">OpinRank Review Dataset</a> </td></tr></table></td>
<!-- <td> This data set contains user reviews of cars and and hotels collected from Tripadvisor (~259,000
reviews) and Edmunds (~42,230 reviews).   </td> -->
<td> Text </td>
<td> </td>
<td> </td>
<td> </td>
<td> </td>
<td> 2011 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Relative+location+of+CT+slices+on+axial+axis"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Relative+location+of+CT+slices+on+axial+axis">Relative location of CT slices on axial axis</a> </td></tr></table></td>
<!-- <td> The dataset consists of 384 features extracted from CT images. The class variable is numeric and denotes the relative location of the CT slice on the axial axis of the human body. </td> -->
<td> Domain-Theory </td>
<td> Regression </td>
<td> Real </td>
<td> 53500 </td>
<td> 386 </td>
<td> 2011 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Online+Handwritten+Assamese+Characters+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Online+Handwritten+Assamese+Characters+Dataset">Online Handwritten Assamese Characters Dataset</a> </td></tr></table></td>
<!-- <td> This is a dataset of 8235 online handwritten assamese characters. The “online” process involves capturing of data as text is written on a digitizing tablet with an electronic pen. </td> -->
<td> Multivariate, Sequential </td>
<td> Classification </td>
<td> Integer </td>
<td> 8235 </td>
<td> </td>
<td> 2011 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/PubChem+Bioassay+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/PubChem+Bioassay+Data">PubC

hem Bioassay Data</a> </td></tr></table></td>
<!-- <td> These highly imbalanced bioassay datasets are from the differing types of screenin g that can be performed using HTS technology. 21 datasets were created from 12 bioassays. </td> -
->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> </td>
<td> </td>
<td> 2011 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Record+Linkage+Comparison+Patterns"><img border="1" src="assets/M Limages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Record+Linkage+Co mparison+Patterns">Record Linkage Comparison Patterns</a> </td></tr></table></td>
<!-- <td> Element-wise comparison of records with personal data from a record linkage settin g. The task is to decide from a comparison pattern whether the underlying records belong to one person. ; </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 5749132 </td>
<td> 12 </td>
<td> 2011 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Communities+and+Crime+Unnormalized"><img border="1" src="assets/ MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Communities+and+ Crime+Unnormalized">Communities and Crime Unnormalized</a> </td></tr></table></td>
<!-- <td> Communities in the US. Data combines socio-economic data from the '90 Census, l aw enforcement data from the 1990 Law Enforcement Management and Admin Stats survey, and crime data fr om the 1995 FBI UCR </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 2215 </td>
<td> 147 </td>
<td> 2011 </td>
<!-- <td> Social </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Vertebral+Column"><img border="1" src="assets/MLimages/SmallLarge default.jpg"/></a> </td><td> <a href="datasets/Vertebral+Column">Vertebral Column</a> </td></tr></table></td>
<!-- <td> Data set containing values for six biomechanical features used to classify orthopae dic patients into 3 classes (normal, disk hernia or spondilolysthesis) or 2 classes (normal or abnormal). </ p></td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 310 </td>
<td> 6 </td>
<td> 2011 </td>
<!-- <td>   </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/EMG+Physical+Action+Data+Set"><img border="1" src="assets/MLimag es/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/EMG+Physical+Action+Data +Set">EMG Physical Action Data Set</a> </td></tr></table></td>
<!-- <td> The Physical Action Data Set includes 10 normal and 10 aggressive physical action s that measure the human activity. The data have been collected by 4 subjects using the Delsys EMG wireless apparatus. </td> -->
<td> Time-Series </td>
<td> Classification </td>

<td> Real </td>
<td> 10000 </td>
<td> 8 </td>
<td> 2011 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Vicon+Physical+Action+Data+Set"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Vicon+Physical+Action+Data+Set">Vicon Physical Action Data Set</a> </td></tr></table></td>
<!-- <td> The Physical Action Data Set includes 10 normal and 10 aggressive physical actions that measure the human activity. The data have been collected by 10 subjects using the Vicon 3D tracker. </td> -->
<td> Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 3000 </td>
<td> 27 </td>
<td> 2011 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Amazon+Commerce+reviews+set"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Amazon+Commerce+reviews+set">Amazon Commerce reviews set</a> </td></tr></table></td>
<!-- <td> The dataset is used for authorship identification in online Writeprint which is a new research field of pattern recognition.   </td> -->
<td> Multivariate, Text, Domain-Theory </td>
<td> Classification </td>
<td> Real </td>
<td> 1500 </td>
<td> 10000 </td>
<td> 2011 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Amazon+Access+Samples"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Amazon+Access+Samples">Amazon Access Samples</a> </td></tr></table></td>
<!-- <td> Amazon's InfoSec is getting smarter about the way Access data is leveraged. This is an anonymized sample of access provisioned within the company. </td> -->
<td> Time-Series, Domain-Theory </td>
<td> Regression, Clustering, Causal-Discovery </td>
<td> </td>
<td> 30000 </td>
<td> 20000 </td>
<td> 2011 </td>
<!-- <td> Business </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Reuter_50_50"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Reuter_50_50">Reuter_50_50</a> </td></tr></table></td>
<!-- <td> The dataset is used for authorship identification in online Writeprint which is a new research field of pattern recognition.   </td> -->
<td> Multivariate, Text, Domain-Theory </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 2500 </td>
<td> 10000 </td>
<td> 2011 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Farm+Ads"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Farm+Ads">Farm Ads</a> </td></tr></table></td>

<!-- <td> This data was collected from text ads found on twelve websites that deal with various farm animal related topics. The binary labels are based on whether or not the content owner approves of the ad. </td> -->
<td> Text </td>
<td> Classification </td>
<td> </td>
<td> 4143 </td>
<td> 54877 </td>
<td> 2011 </td>
<!-- <td> Business </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/DBWorld+e-mails"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/DBWorld+e-mails">DBWorld e-mails</a> </td></tr></table></td>
<!-- <td> It contains 64 e-mails which I have manually collected from DBWorld mailing list. They are classified in: 'announces of conferences' and 'everything else'. </td> -->
<td> Text </td>
<td> Classification </td>
<td> </td>
<td> 64 </td>
<td> 4702 </td>
<td> 2011 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29">KEGG Metabolic Relation Network (Directed)</a> </td></tr></table></td>
<!-- <td> KEGG Metabolic pathways modeled as directed relation network. Variety of graphical features presented. </td> -->
<td> Multivariate, Univariate, Text </td>
<td> Classification, Regression, Clustering </td>
<td> Integer, Real </td>
<td> 53414 </td>
<td> 24 </td>
<td> 2011 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29">KEGG Metabolic Reaction Network (Undirected)</a> </td></tr></table></td>
<!-- <td> KEGG Metabolic pathways modeled as un-directed reaction network. Variety of graphical features presented. </td> -->
<td> Multivariate, Univariate, Text </td>
<td> Classification, Regression, Clustering </td>
<td> Integer, Real </td>
<td> 65554 </td>
<td> 29 </td>
<td> 2011 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Bank+Marketing"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Bank+Marketing">Bank Marketing</a> </td></tr></table></td>
<!-- <td> The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe a term deposit (variable y). </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 45211 </td>

<td> 45211 </td>
<td> 17 </td>
<td> 2012 </td>
<!-- <td> Business </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/YouTube+Comedy+Slam+Preference+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/YouTube+Comedy+Slam+Preference+Data">YouTube Comedy Slam Preference Data</a> </td></tr></table></td>
<!-- <td> This dataset provides user vote data on which video from a pair of videos is funnier collected on YouTube Comedy Slam. The task is to automatically predict this preference based on video metadata. </td> -->
<td> Text </td>
<td> Classification </td>
<td> </td>
<td> 1138562 </td>
<td> 3 </td>
<td> 2012 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Gas+Sensor+Array+Drift+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Gas+Sensor+Array+Drift+Dataset">Gas Sensor Array Drift Dataset</a> </td></tr></table></td>
<!-- <td> This archive contains 13910 measurements from 16 chemical sensors utilized in simulations for drift compensation in a discrimination task of 6 gases at various levels of concentrations. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 13910 </td>
<td> 128 </td>
<td> 2012 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/ILPD+%28Indian+Liver+Patient+Dataset%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/ILPD+%28Indian+Liver+Patient+Dataset%29">ILPD (Indian Liver Patient Dataset)</a> </td></tr></table></td>
<!-- <td> This data set contains 10 variables that are age, gender, total Bilirubin, direct Bilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT and Alkphos. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 583 </td>
<td> 10 </td>
<td> 2012 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/OPPORTUNITY+Activity+Recognition"><img border="1" src="assets/MLimages/SmallLarge226.jpg"/></a> </td><td> <a href="datasets/OPPORTUNITY+Activity+Recognition">OPPORTUNITY Activity Recognition</a> </td></tr></table></td>
<!-- <td> The OPPORTUNITY Dataset for Human Activity Recognition from Wearable, Object, and Ambient Sensors is a dataset devised to benchmark human activity recognition algorithms (classification, automatic data segmentation, sensor fusion, feature extraction, etc). </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 2551 </td>
<td> 242 </td>
<td> 2012 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Nomao"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Nomao">Nomao</a> </td></tr></table></td>
<!-- <td> Nomao collects data about places (name, phone, localization ...) from many sources

<!-- <td> Nomao collects data about places (name, phone, localization...) from many sources
.
Deduplication consists in detecting what data refer to the same place.
Instances in the dataset compare 2 spots. </td> -->
<td> Univariate </td>
<td> Classification </td>
<td> Real </td>
<td> 34465 </td>
<td> 120 </td>
<td> 2012 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/SMS+Spam+Collection"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/SMS+Spam+Collection">SMS Spam Collection</a> </td></tr></table></td>
<!-- <td> The SMS Spam Collection is a public set of SMS labeled messages that have been collected for mobile phone spam research. </td> -->
<td> Multivariate, Text, Domain-Theory </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 5574 </td>
<td> </td>
<td> 2012 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Skin+Segmentation"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Skin+Segmentation">Skin Segmentation</a> </td></tr></table></td>
<!-- <td> The Skin Segmentation dataset is constructed over B, G, R color space. Skin and Nonskin dataset is generated using skin textures from face images of diversity of age, gender, and race people. </td> -->
<td> Univariate </td>
<td> Classification </td>
<td> Real </td>
<td> 245057 </td>
<td> 4 </td>
<td> 2012 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Planning+Relax"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Planning+Relax">Planning Relax</a> </td></tr></table></td>
<!-- <td> The dataset concerns with the classification of two mental stages from recorded EEG signals: Planning (during imagination of motor act) and Relax state.   </td> -->
<td> Univariate </td>
<td> Classification </td>
<td> Real </td>
<td> 182 </td>
<td> 13 </td>
<td> 2012 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/PAMAP2+Physical+Activity+Monitoring"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/PAMAP2+Physical+Activity+Monitoring">PAMAP2 Physical Activity Monitoring</a> </td></tr></table></td>
<!-- <td> The PAMAP2 Physical Activity Monitoring dataset contains data of 18 different physical activities, performed by 9 subjects wearing 3 inertial measurement units and a heart rate monitor. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 3850505 </td>
</tr><tr>

<td> 52 </td>
<td> 2012 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Restaurant+%26+consumer+data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Restaurant+%26+consumer+data">Restaurant & consumer data</a> </td></tr></table></td>
<!-- <td> The dataset was obtained from a recommender system prototype. The task was to generate a top-n list of restaurants according to the consumer preferences.   </td> -->
<td> Multivariate </td>
<td> </td>
<td> </td>
<td> 138 </td>
<td> 47 </td>
<td> 2012 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/CNAE-9"><img border="1" src="assets/MLimages/SmallLargedefault.jpg" /></a> </td><td> <a href="datasets/CNAE-9">CNAE-9</a> </td></tr></table></td>
<!-- <td> This is a data set containing 1080 documents of free text business descriptions of Brazilian companies categorized into a
subset of 9 categories </td> -->
<td> Multivariate, Text </td>
<td> Classification </td>
<td> Integer </td>
<td> 1080 </td>
<td> 857 </td>
<td> 2012 </td>
<!-- <td> Business </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Individual+household+electric+power+consumption"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Individual+household+electric+power+consumption">Individual household electric power consumption</a> </td></tr></table></td>
<!-- <td> Measurements of electric power consumption in one household with a one-minute sampling rate over a period of almost 4 years. Different electrical quantities and some sub-metering values are available. </td> -->
<td> Multivariate, Time-Series </td>
<td> Regression, Clustering </td>
<td> Real </td>
<td> 2075259 </td>
<td> 9 </td>
<td> 2012 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/seeds"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/seeds">seeds</a> </td></tr></table></td>
<!-- <td> Measurements of geometrical properties of kernels belonging to three different varieties of wheat. A soft X-ray technique and GRAINS package were used to construct all seven, real-valued attributes. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 210 </td>
<td> 7 </td>
<td> 2012 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Northix"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Northix">Northix</a> </td></tr></table></td>
<!-- <td> Northix is designed to be a schema matching benchmark problem for data integrati

on of two entity relationship databases.   </td> -->
<td> Multivariate, Univariate, Text </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 115 </td>
<td> 200 </td>
<td> 2012 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/QtyT40I10D100K"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/QtyT40I10D100K">QtyT40I10D100K</a> </td></tr></table></td>
<!-- <td> Since there is no numerical sequential data stream available in standard data sets, this data set is generated from the original T40I10D100K data set </td> -->
<td> Sequential </td>
<td> </td>
<td> Integer </td>
<td> 3960456 </td>
<td> 4 </td>
<td> 2012 </td>
<!-- <td>   </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Legal+Case+Reports"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Legal+Case+Reports">Legal Case Reports</a> </td></tr></table></td>
<!-- <td> A textual corpus of 4000 legal cases for automatic summarization and citation analysis. For each document we collect catchphrases, citations sentences, citation catchphrases and citation classes. </td> -->
<td> Text </td>
<td> Classification </td>
<td> </td>
<td> </td>
<td> </td>
<td> 2012 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Human+Activity+Recognition+Using+Smartphones"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Human+Activity+Recognition+Using+Smartphones">Human Activity Recognition Using Smartphones</a> </td></tr></table></td>
<!-- <td> Human Activity Recognition database built from the recordings of 30 subjects performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Clustering </td>
<td> </td>
<td> 10299 </td>
<td> 561 </td>
<td> 2012 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/One-hundred+plant+species+leaves+data+set"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/One-hundred+plant+species+leaves+data+set">One-hundred plant species leaves data set</a> </td></tr></table></td>
<!-- <td> Sixteen samples of leaf each of one-hundred plant species. For each sample, a shape descriptor, fine scale margin and texture histogram are given. </td> -->
<td> </td>
<td> Classification </td>
<td> Real </td>
<td> 1600 </td>
<td> 64 </td>

<td> 2012 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Energy+efficiency"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Energy+efficiency">Energy efficiency</a> </td></tr></table></td>
<!-- <td> This study looked into assessing the heating load and cooling load requirements of buildings (that is, energy efficiency) as a function of building parameters. </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Integer, Real </td>
<td> 768 </td>
<td> 8 </td>
<td> 2012 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Yacht+Hydrodynamics"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Yacht+Hydrodynamics">Yacht Hydrodynamics</a> </td></tr></table></td>
<!-- <td> Delft data set, used to predict the hydodynamic performance of sailing yachts from dimensions and velocity. </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 308 </td>
<td> 7 </td>
<td> 2013 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Fertility"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Fertility">Fertility</a> </td></tr></table></td>
<!-- <td> 100 volunteers provide a semen sample analyzed according to the WHO 2010 criteria. Sperm concentration are related to socio-demographic data, environmental factors, health status, and life habits </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Real </td>
<td> 100 </td>
<td> 10 </td>
<td> 2013 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Daphnet+Freezing+of+Gait"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Daphnet+Freezing+of+Gait">Daphnet Freezing of Gait</a> </td></tr></table></td>
<!-- <td> This dataset contains the annotated readings of 3 acceleration sensors at the hip and leg of Parkinson's disease patients that experience freezing of gait (FoG) during walking tasks.
  </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 237 </td>
<td> 9 </td>
<td> 2013 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29">3D Road Network (North Jutland, Denmark)</a> </td></tr></table></td>
<!-- <td> 3D road network with highly accurate elevation information (+-20cm) from Denmark used in eco-routing and fuel/Co2-estimation routing algorithms. </td> -->

<td> Sequential, Text </td>
<td> Regression, Clustering </td>
<td> Real </td>
<td> 434874 </td>
<td> 4 </td>
<td> 2013 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/ISTANBUL+STOCK+EXCHANGE"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/ISTANBUL+STOCK+EXCHANGE">ISTANBUL STOCK EXCHANGE</a> </td></tr></table></td>
<!-- <td> Data sets includes returns of Istanbul Stock Exchange with seven other international index; SP, DAX, FTSE, NIKKEI, BOVESPA, MSCE_EU, MSCI_EM from Jun 5, 2009 to Feb 22, 2011. </td> -->
<td> Multivariate, Univariate, Time-Series </td>
<td> Classification, Regression </td>
<td> Real </td>
<td> 536 </td>
<td> 8 </td>
<td> 2013 </td>
<!-- <td> Business </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Buzz+in+social+media+"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Buzz+in+social+media+">Buzz in social media </a> </td></tr></table></td>
<!-- <td> This data-set contains examples of buzz events from two different social networks: Twitter, and Tom's Hardware, a forum network focusing on new technology with more conservative dynamics. </td> -->
<td> Time-Series, Multivariate </td>
<td> Regression, Classification </td>
<td> Integer, Real </td>
<td> 140000 </td>
<td> 77 </td>
<td> 2013 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/First-order+theorem+proving"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/First-order+theorem+proving">First-order theorem proving</a> </td></tr></table></td>
<!-- <td> Given a theorem, predict which of five heuristics will give the fastest proof when used by a first-order prover. A sixth prediction declines to attempt a proof, should the theorem be too difficult. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 6118 </td>
<td> 51 </td>
<td> 2013 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29">Wearable Computing: Classification of Body Postures and Movements (PUC-Rio)</a> </td></tr></table></td>
<!-- <td> A dataset with 5 classes (sitting-down, standing-up, standing, walking, and sitting) collected on 8 hours of activities of 4 healthy subjects. We also established a baseline performance index. </td> -->
<td> Sequential </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 165632 </td>

<td> 18 </td>
<td> 2013 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Gas+sensor+arrays+in+open+sampling+settings"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Gas+sensor+arrays+in+open+sampling+settings">Gas sensor arrays in open sampling settings</a> </td></tr></table></td>
<!-- <td> The dataset contains 18000 time-series recordings from a chemical detection platform at six different locations in a wind tunnel facility in response to ten high-priority chemical gaseous substances </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 18000 </td>
<td> 1950000 </td>
<td> 2013 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Climate+Model+Simulation+Crashes"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Climate+Model+Simulation+Crashes">Climate Model Simulation Crashes</a> </td></tr></table></td>
<!-- <td> Given Latin hypercube samples of 18 climate model input parameter values, predict climate model simulation crashes and determine the parameter value combinations that cause the failures. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 540 </td>
<td> 18 </td>
<td> 2013 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/MicroMass"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/MicroMass">MicroMass</a> </td></tr></table></td>
<!-- <td> A dataset to explore machine learning approaches for the identification of microorganisms from mass-spectrometry data. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 931 </td>
<td> 1300 </td>
<td> 2013 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/QSAR+biodegradation"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/QSAR+biodegradation">QSAR biodegradation</a> </td></tr></table></td>
<!-- <td> Data set containing values for 41 attributes (molecular descriptors) used to classify 1055 chemicals into 2 classes (ready and not ready biodegradable). </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 1055 </td>
<td> 41 </td>
<td> 2013 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/BLOGGER"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/BLOGGER">BLOGGER</a> </td></tr></table></td>

<!-- <td> In this paper, we look for to recognize the causes of users tend
to cyber space in Kohkiloye and Boyer Ahmad Province in
Iran </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> </td>
<td> 100 </td>
<td> 6 </td>
<td> 2013 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Daily+and+Sports+Activities"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Daily+and+Sports+Activities">Daily and Sports Activities</a> </td></tr></table></td>
<!-- <td> The dataset comprises motion sensor data of 19 daily and sports activities each performed by 8 subjects in their own style for 5 minutes. Five Xsens MTx units are used on the torso, arms, and legs.
  </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 9120 </td>
<td> 5625 </td>
<td> 2013 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/User+Knowledge+Modeling"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/User+Knowledge+Modeling">User Knowledge Modeling</a> </td></tr></table></td>
<!-- <td> It is the real dataset about the students' knowledge status about the subject of Electrical DC Machines. The dataset had been obtained from Ph.D. Thesis. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> Integer </td>
<td> 403 </td>
<td> 5 </td>
<td> 2013 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection">Reuters RCV1 RCV2 Multilingual, Multiview Text Categorization Test collection</a> </td></tr></table></td>
<!-- <td> This test collection contains feature characteristics of documents originally written in five different languages and their translations, over a common set of 6 categories.   </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 111740 </td>
<td> </td>
<td> 2013 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/NYSK"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/NYSK">NYSK</a> </td></tr></table></td>
<!-- <td> NYSK (New York v. Strauss-Kahn) is a collection of English news articles about the case relating to allegations of sexual assault against the former IMF director Dominique Strauss-Kahn (May 2011). </td> -->
<td> Multivariate, Sequential, Text </td>
<td> Clustering </td>
<td> </td>

<td> 10421 </td>
<td> 7 </td>
<td> 2013 </td>
<!-- <td> Social </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Turkiye+Student+Evaluation"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Turkiye+Student+Evaluation">Turkiye Student Evaluation</a> </td></tr></table></td>
<!-- <td> This data set contains a total 5820 evaluation scores provided by students from Gazi University in Ankara (Turkey). There is a total of 28 course specific questions and additional 5 attributes. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> </td>
<td> 5820 </td>
<td> 33 </td>
<td> 2013 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machines%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machines%29">ser Knowledge Modeling Data (Students' Knowledge Levels on DC Electrical Machines)</a> </td></tr></table></td>
<!-- <td> The dataset is about the users' learning activities and knowledge levels on subjects of DC Electrical Machines. The dataset had been obtained from online web-courses and reported in my Ph.D. Thesis. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 403 </td>
<td> 5 </td>
<td> 2013 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/EEG+Eye+State"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/EEG+Eye+State">EEG Eye State</a> </td></tr></table></td>
<!-- <td> The data set consists of 14 EEG values and a value indicating the eye state. </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 14980 </td>
<td> 15 </td>
<td> 2013 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure">Physicochemical Properties of Protein Tertiary Structure</a> </td></tr></table></td>
<!-- <td> This is a data set of Physicochemical Properties of Protein Tertiary Structure. The data set is taken from CASP 5-9. There are 45730 decoys and size varying from 0 to 21 armstrong. </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 45730 </td>
<td> 9 </td>
<td> 2013 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Online+Retail"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Online+Retail">Online Retail</a> </td></tr></table></td>
<!-- <td> This is a transnational data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail. </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Clustering </td>
<td> Integer, Real </td>
<td> 541909 </td>
<td> 8 </td>
<td> 2015 </td>
<!-- <td> Business </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Wine+Quality"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Wine+Quality">Wine Quality</a> </td></tr></table></td>
<!-- <td> Two datasets are included, related to red and white vinho verde wine samples, from the north of Portugal. The goal is to model wine quality based on physicochemical tests (see [Cortez et al., 2009], http://www3.dsi.uminho.pt/pcortez/wine5.pdf). </td> -->
<td> Multivariate </td>
<td> Classification, Regression, Clustering </td>
<td> Real </td>
<td> 4898 </td>
<td> 11 </td>
<td> 2009 </td>
<!-- <td> Business </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Yeast"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Yeast">Yeast</a> </td></tr></table></td>
<!-- <td> Predict the localization site of proteins </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 1484 </td>
<td> 8 </td>
<td> 1996 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Yacht+Hydrodynamics"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Yacht+Hydrodynamics">Yacht Hydrodynamics</a> </td></tr></table></td>
<!-- <td> Delft data set used for predicting the hydodynamic performance of sailing yachts from dimensions and boat velocity. </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 308 </td>
<td> 6 </td>
<td> 2013 </td>
<!-- <td> Physics </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Year+Prediction+MSD"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Year+Prediction+MSD">YearPredictionMSD</a> </td></tr></table></td>
<!-- <td> Prediction of the release year of a song from audio features. Songs are mostly western, commercial tracks ranging from 1922 to 2011, with a peak in the 2000s. </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 515345 </td>
<td> 90 </td>
<td> 2011 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Zoo"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Zoo">Zoo</a> </td></tr></table></td>
<!-- <td> A simple database containing 17 Boolean-valued attributes describing animals. The "type" attribute appears to be the class attribute. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> Integer </td>
<td> 101 </td>
<td> 17 </td>
<td> 1990 </td>
<!-- <td> Other </td> -->
</tr></table>
</td></tr></table>
</td></tr></table>
</td>

<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/seismic-bumps"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/seismic-bumps">seismic-bumps</a> </td></tr></table></td>
<!-- <td> The data describe the problem of high energy (higher than 10^4 J) seismic bumps forecasting in a coal
mine. Data come from two of longwalls located in a Polish coal mine. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 2584 </td>
<td> 19 </td>
<td> 2013 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/banknote+authentication"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/banknote+authentication">banknote authentication</a> </td></tr></table></td>
<!-- <td> Data were extracted from images that were taken for the evaluation of an authentication procedure for banknotes. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 1372 </td>
<td> 5 </td>
<td> 2013 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Problem%3A+Pat"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Problem%3A+Pat">USPTO Algorithm Challenge, run by NASA-Harvard Tournament Lab and TopCoder Problem: Pat</a> </td></tr></table></td>
<!-- <td> Data used for USPTO Algorithm Competition. Contains drawing pages from US patents with manually labeled figure and part labels. </td> -->
<td> Domain-Theory </td>
<td> Classification </td>
<td> Integer </td>
<td> 306 </td>
<td> 5 </td>
<td> 2013 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/YouTube+Multiview+Video+Games+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/YouTube+Multiview+Video+Games+Dataset">YouTube Multiview Video Games Dataset</a> </td></tr></table></td>
<!-- <td> This dataset contains about 120k instances, each described by 13 feature types, with class information, specially useful for exploring multiview topics (cotraining, ensembles, clustering,..). </td> -->
<td> Multivariate, Text </td>
<td> Classification, Clustering </td>
<td> Integer, Real </td>
<td> 120000 </td>
<td> 1000000 </td>
<td> 2013 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations">Gas Sensor Array Drift Dataset at Different Concentrations</a> </td></tr></table></td>
</tr><tr>
</tr><tr bgcolor="DDEEFF">
</tr><tr>
</tr><tr bgcolor="DDEEFF">
</tr><tr>
</tr><tr bgcolor="DDEEFF">

<!-- <td> This archive contains 13910 measurements from 16 chemical sensors exposed to 6 different gases at various concentration levels. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Regression, Clustering, Causa </td>
<td> Real </td>
<td> 13910 </td>
<td> 129 </td>
<td> 2013 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors">Activities of Daily Living (ADLs) Recognition Using Binary Sensors</a> </td></tr></table></td>
<!-- <td> This dataset comprises information regarding the ADLs performed by two users on a daily basis in their
own homes.   </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Classification, Clustering </td>
<td> </td>
<td> 2747 </td>
<td> </td>
<td> 2013 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/SkillCraft1+Master+Table+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/SkillCraft1+Master+Table+Dataset">SkillCraft1 Master Table Dataset</a> </td></tr></table></td>
<!-- <td> This data was used in Thompson et al. (2013). A list of possible game actions is discussed in Thompson, Blair, Chen, & Henrey (2013). </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Integer, Real </td>
<td> 3395 </td>
<td> 20 </td>
<td> 2013 </td>
<!-- <td> Game </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units">Weight Lifting Exercises monitored with Inertial Measurement Units</a> </td></tr></table></td>
<!-- <td> Six young health subjects were asked to perform 5 variations of the biceps curl weight lifting exercise. One of the variations is the one predicted by the health professional. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 39242 </td>
<td> 152 </td>
<td> 2013 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/SML2010"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/SML2010">SML2010</a> </td></tr></table></td>
<!-- <td> This dataset is collected from a monitor system mounted in a domotic house. It corresponds to approximately 40 days of monitoring data. </td> -->
<td> Multivariate, Sequential, Time-Series, Text </td>
<td> Regression </td>
<td> Real </td>
<td> 4137 </td>
<td> 24 </td>

<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Bike+Sharing+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Bike+Sharing+Dataset">Bike Sharing Dataset</a> </td></tr></table></td>
<!-- <td> This dataset contains the hourly and daily count of rental bikes between years 2011 and 2012 in Capital bikeshare system with the corresponding weather and seasonal information. </td> -->
<td> Univariate </td>
<td> Regression </td>
<td> Integer, Real </td>
<td> 17389 </td>
<td> 16 </td>
<td> 2013 </td>
<!-- <td> Social </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Predict+keywords+activities+in+a+online+social+media"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Predict+keywords+activities+in+a+online+social+media">Predict keywords activities in a online social media</a> </td></tr></table></td>
<!-- <td> The data from Twitter was collected during 360 consecutive days. It was done by querying 1497 English keywords sampled from Wikipedia. This dataset is proposed in a Learning to rank setting. </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> </td>
<td> Integer, Real </td>
<td> 51 </td>
<td> 35 </td>
<td> 2013 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Thoracic+Surgery+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Thoracic+Surgery+Data">Thoracic Surgery Data</a> </td></tr></table></td>
<!-- <td> The data is dedicated to classification problem related to the post-operative life expectancy in the lung cancer patients: class 1 - death within one year after surgery, class 2 - survival. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 470 </td>
<td> 17 </td>
<td> 2013 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/EMG+dataset+in+Lower+Limb"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/EMG+dataset+in+Lower+Limb">EMG dataset in Lower Limb</a> </td></tr></table></td>
<!-- <td> 3 different exercises: sitting, standing and walking in the muscles: biceps femoris, vastus medialis, rectus femoris and semitendinosus addition to goniometry in the exercises. </td> -->
<td> Multivariate, Time-Series </td>
<td> </td>
<td> Real </td>
<td> 132 </td>
<td> 5 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/SUSY"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/SUSY">SUSY</a> </td></tr></table></td>
<!-- <td> This is a classification problem to distinguish between a signal process which produ

ces supersymmetric particles and a background process which does not. </td> -->
<td> </td>
<td> Classification </td>
<td> Real </td>
<td> 5000000 </td>
<td> 18 </td>
<td> 2014 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/HIGGS"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/
></a> </td><td> <a href="datasets/HIGGS">HIGGS</a> </td></tr></table></td>
<!-- <td> This is a classification problem to distinguish between a signal process which produ
ces Higgs bosons and a background process which does not.   </td> -->
<td> </td>
<td> Classification </td>
<td> Real </td>
<td> 11000000 </td>
<td> 28 </td>
<td> 2014 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Qualitative_Bankruptcy"><img border="1" src="assets/MLimages/SmallL
argedefault.jpg"/></a> </td><td> <a href="datasets/Qualitative_Bankruptcy">Qualitative_
Bankruptcy</a> </td></tr></table></td>
<!-- <td> Predict the Bankruptcy from Qualitative parameters from experts. </td> -
->
<td> Multivariate </td>
<td> Classification </td>
<td> </td>
<td> 250 </td>
<td> 7 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/LSVT+Voice+Rehabilitation"><img border="1" src="assets/MLimages/Sm
allLargedefault.jpg"/></a> </td><td> <a href="datasets/LSVT+Voice+Rehabilitation">LSV
T Voice Rehabilitation</a> </td></tr></table></td>
<!-- <td> 126 samples from 14 participants, 309 features. Aim: assess whether voice rehabilit
ation treatment lead to phonations considered 'acceptable' or 'unacceptable' (binary class classification proble
m). </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 126 </td>
<td> 309 </td>
<td> 2014 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer"><img b
order="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datase
ts/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer">Dataset for ADL Recognition with Wrist-wor
n Accelerometer</a> </td></tr></table></td>
<!-- <td> Recordings of 16 volunteers performing 14 Activities of Daily Living (ADL) while carr
ying a single wrist-worn tri-axial accelerometer. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Clustering </td>
<td> </td>
<td> </td>
<td> 3 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">

<td><table><tr><td><a href="datasets/Wilt"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Wilt">Wilt</a> </td></tr></table></td>
<!-- <td> High-resolution Remote Sensing data set (Quickbird). Small number of training samples of diseased trees, large number for other land cover. Testing data set from stratified random sample of image. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> </td>
<td> 4889 </td>
<td> 6 </td>
<td> 2014 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/User+Identification+From+Walking+Activity"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/User+Identification+From+Walking+Activity">User Identification From Walking Activity</a> </td></tr></table></td>
<!-- <td> The dataset collects data from an Android smartphone positioned in the chest pocket from 22 participants walking in the wild over a predefined path.
  </td> -->
<td> Univariate, Sequential, Time-Series </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> </td>
<td> </td>
<td> 2014 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer">Activity Recognition from Single Chest-Mounted Accelerometer</a> </td></tr></table></td>
<!-- <td> The dataset collects data from a wearable accelerometer mounted on the chest. The dataset is intended for Activity Recognition research purposes. </td> -->
<td> Univariate, Sequential, Time-Series </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> </td>
<td> </td>
<td> 2014 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Leaf"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Leaf">Leaf</a> </td></tr></table></td>
<!-- <td> This dataset consists in a collection of shape and texture features extracted from digital images of leaf specimens originating from a total of 40 different plant species. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 340 </td>
<td> 16 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Dresses_Attribute_Sales"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Dresses_Attribute_Sales">Dresses_Attribute_Sales</a> </td></tr></table></td>
<!-- <td> This dataset contain Attributes of dresses and their recommendations according to their sales.Sales are monitor on the basis of alternate days.   </td> -->
<td> Text </td>
<td> Classification, Clustering </td>
<td> </td>
<td> 501 </td>

<td> 13 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Tamilnadu+Electricity+Board+Hourly+Readings"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Tamilnadu+Electricity+Board+Hourly+Readings">Tamilnadu Electricity Board Hourly Readings</a> </td></tr></table></td>
<!-- <td> This data can be effectively produced the result to fewer parameter of the Load profile can be reduced in the Database   </td> -->
<td> Multivariate </td>
<td> Classification, Regression, Clustering </td>
<td> Real </td>
<td> 45781 </td>
<td> 5 </td>
<td> 2013 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Airfoil+Self-Noise"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Airfoil+Self-Noise">Airfoil Self-Noise</a> </td></tr></table></td>
<!-- <td> NASA data set, obtained from a series of aerodynamic and acoustic tests of two and three-dimensional airfoil blade sections conducted in an anechoic wind tunnel. </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 1503 </td>
<td> 6 </td>
<td> 2014 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Wholesale+customers"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Wholesale+customers">Wholesale customers</a> </td></tr></table></td>
<!-- <td> The data set refers to clients of a wholesale distributor. It includes the annual spending in monetary units (m.u.) on diverse product categories </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> Integer </td>
<td> 440 </td>
<td> 8 </td>
<td> 2014 </td>
<!-- <td> Business </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis">Twitter Data set for Arabic Sentiment Analysis</a> </td></tr></table></td>
<!-- <td> This problem of Sentiment Analysis (SA) has been studied well on the English language but not Arabic one. Two main approaches have been devised: corpus-based and lexicon-based.   </td> -->
<td> Text </td>
<td> Classification </td>
<td> </td>
<td> 2000 </td>
<td> 2 </td>
<td> 2014 </td>
<!-- <td> Social </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Combined+Cycle+Power+Plant"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Combined+Cycle+Power+Plant">Combined Cycle Power Plant</a> </td></tr></table></td>

<!-- <td> The dataset contains 9568 data points collected from a Combined Cycle Power Pla
nt over 6 years (2006-2011), when the plant was set to work with full load.   </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 9568 </td>
<td> 4 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Urban+Land+Cover"><img border="1" src="assets/MLimages/SmallLarg
edefault.jpg"/></a> </td><td> <a href="datasets/Urban+Land+Cover">Urban Land Cover
</a> </td></tr></table></td>
<!-- <td> Classification of urban land cover using high resolution aerial imagery. Intended to
assist sustainable urban planning efforts. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> </td>
<td> 168 </td>
<td> 148 </td>
<td> 2014 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Diabetes+130-US+hospitals+for+years+1999-2008"><img border="1" src
="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Diabetes+
130-US+hospitals+for+years+1999-2008">Diabetes 130-US hospitals for years 1999-2008</a> </td></t
r></table></td>
<!-- <td> This data has been prepared to analyze factors related to readmission as well as ot
her

outcomes pertaining to patients with diabetes. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> Integer </td>
<td> 100000 </td>
<td> 55 </td>
<td> 2014 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Bach+Choral+Harmony"><img border="1" src="assets/MLimages/SmallL
argedefault.jpg"/></a> </td><td> <a href="datasets/Bach+Choral+Harmony">Bach Choral
Harmony</a> </td></tr></table></td>
<!-- <td> The data set is composed of 60 chorales (5665 events) by J.S. Bach (1675-1750).
Each event of each chorale is labelled using 1 among 101 chord labels and described
through 14 features. </td> -->
<td> Sequential </td>
<td> Classification </td>
<td> </td>
<td> 5665 </td>
<td> 17 </td>
<td> 2014 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/StoneFlakes"><img border="1" src="assets/MLimages/SmallLargedefault
.jpg"/></a> </td><td> <a href="datasets/StoneFlakes">StoneFlakes</a> </td></tr
></table></td>
<!-- <td> Stone flakes are waste products of the stone tool production in
the prehistoric era. The variables are means of geometric and
stylistic features of the flakes contained in different inventories. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering, Causal-Discovery </td>
<td> Real </td>

<td> 79 </td>
<td> 8 </td>
<td> 2014 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Tennis+Major+Tournament+Match+Statistics"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Tennis+Major+Tournament+Match+Statistics">Tennis Major Tournament Match Statistics</a> </td></tr></table></td>
<!-- <td> This is a collection of 8 files containing the match statistics for both women and men at the four major tennis tournaments of the year 2013. Each file has 42 columns and a minimum of 76 rows. </td> -->
<td> Multivariate </td>
<td> Classification, Regression, Clustering </td>
<td> Integer, Real </td>
<td> 127 </td>
<td> 42 </td>
<td> 2014 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings">Parkinson Speech Dataset with Multiple Types of Sound Recordings</a> </td></tr></table></td>
<!-- <td> The training data belongs to 20 Parkinson's Disease (PD) patients and 20 healthy subjects. From all subjects, multiple types of sound recordings (26) are taken. </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Integer, Real </td>
<td> 1040 </td>
<td> 26 </td>
<td> 2014 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Gesture+Phase+Segmentation"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Gesture+Phase+Segmentation">Gesture Phase Segmentation</a> </td></tr></table></td>
<!-- <td> The dataset is composed by features extracted from 7 videos with people gesticulating, aiming at studying Gesture Phase Segmentation. It contains 50 attributes divided into two files for each video. </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 9900 </td>
<td> 50 </td>
<td> 2014 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Perfume+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Perfume+Data">Perfume Data</a> </td></tr></table></td>
<!-- <td> This data consists of odors of 20 different perfumes. Data was obtained by using a handheld odor meter (OMX-GR sensor) per second for 28 seconds period. </td> -->
<td> Univariate, Domain-Theory </td>
<td> Classification, Clustering </td>
<td> Integer </td>
<td> 560 </td>
<td> 2 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/BlogFeedback"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/BlogFeedback">BlogFeedback</a> </td></t

ult.jpg "/></a> </td><td> <a href="datasets/BlogFeedback">BlogFeedback</a> </td></tr></table></td>
<!-- <td> Instances in this dataset contain features extracted from blog posts. The task associated with the data is to predict how many comments the post will receive. </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Integer, Real </td>
<td> 60021 </td>
<td> 281 </td>
<td> 2014 </td>
<!-- <td> Social </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/REALDISP+Activity+Recognition+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/REALDISP+Activity+Recognition+Dataset">REALDISP Activity Recognition Dataset</a> </td></tr></table></td>
<!-- <td> The REALDISP dataset is devised to evaluate techniques dealing with the effects of sensor displacement in wearable activity recognition as well as to benchmark general activity recognition algorithms   </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 1419 </td>
<td> 120 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Newspaper+and+magazine+images+segmentation+dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Newspaper+and+magazine+images+segmentation+dataset">Newspaper and magazine images segmentation dataset</a> </td></tr></table></td>
<!-- <td> Dataset is well suited for segmentation tasks. It contains 101 scanned pages from different newspapers and magazines in Russian with ground truth pixel-based masks. </td> -->
<td> </td>
<td> Classification </td>
<td> </td>
<td> 101 </td>
<td> </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/AAAI+2014+Accepted+Papers"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/AAAI+2014+Accepted+Papers">AAAI 2014 Accepted Papers</a> </td></tr></table></td>
<!-- <td> This data set compromises the metadata for the 2014 AAAI conference's accepted papers, including paper titles, authors, abstracts, and keywords of varying granularity. </td> -->
<td> Multivariate </td>
<td> Clustering </td>
<td> </td>
<td> 399 </td>
<td> 6 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Gas+sensor+array+under+flow+modulation"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Gas+sensor+array+under+flow+modulation">Gas sensor array under flow modulation</a> </td></tr></table></td>
<!-- <td> The data set contains 58 time series acquired from 16 chemical sensors under gas flow modulation conditions. The sensors were exposed to different gaseous binary mixtures of acetone and ethanol. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Regression </td>
<td> Real </td>
<td> 58 </td>
</tr></table>
</body>
</html>

<td> 58 </td>
<td> 120432 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures">Gas sensor array exposed to turbulent gas mixtures</a> </td></tr></table></td>
<!-- <td> A chemical detection platform composed of 8 chemoresistive gas sensors was exposed to turbulent gas mixtures generated naturally in a wind tunnel. The acquired time series of the sensors are provided. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Regression </td>
<td> Real </td>
<td> 180 </td>
<td> 150000 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/UJIIndoorLoc"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/UJIIndoorLoc">UJIIndoorLoc</a> </td></tr></table></td>
<!-- <td> The UJIIndoorLoc is a Multi-Building Multi-Floor indoor localization database to test Indoor Positioning System that rely on WLAN/WiFi fingerprint. </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Integer, Real </td>
<td> 21048 </td>
<td> 529 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Sentence+Classification"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Sentence+Classification">Sentence Classification</a> </td></tr></table></td>
<!-- <td> Contains sentences from the abstract and introduction of 30 articles annotated with a modified Argumentative Zones annotation scheme. These articles come from biology, machine learning and psychology. </td> -->
<td> Text </td>
<td> Classification </td>
<td> Integer </td>
<td> </td>
<td> </td>
<td> 2014 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Dow+Jones+Index"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Dow+Jones+Index">Dow Jones Index</a> </td></tr></table></td>
<!-- <td> This dataset contains weekly data for the Dow Jones Industrial Index. It has been used in computational investing research. </td> -->
<td> Time-Series </td>
<td> Classification, Clustering </td>
<td> Integer, Real </td>
<td> 750 </td>
<td> 16 </td>
<td> 2014 </td>
<!-- <td> Business </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/sEMG+for+Basic+Hand+movements"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/sEMG+for+Basic+Hand

+movements">sEMG for Basic Hand movements</a> </td></tr></table></td>
<!-- <td> The “sEMG for Basic Hand movements” includes 2 databases of surface electromyographic signals of 6 hand movements using Delsys' EMG System. Healthy subjects conducted six daily life grasps. </td> -->
<td> Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 3000 </td>
<td> 2500 </td>
<td> 2014 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/AAAI+2013+Accepted+Papers"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/AAAI+2013+Accepted+Papers">AAAI 2013 Accepted Papers</a> </td></tr></table></td>
<!-- <td> This data set compromises the metadata for the 2013 AAAI conference's accepted papers (main track only), including paper titles, abstracts, and keywords of varying granularity. </td> -->
<td> Multivariate </td>
<td> Clustering </td>
<td> </td>
<td> 150 </td>
<td> 5 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Geographical+Original+of+Music"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Geographical+Original+of+Music">Geographical Original of Music</a> </td></tr></table></td>
<!-- <td> Instances in this dataset contain audio features extracted from 1059 wave files. The task associated with the data is to predict the geographical origin of music.
  </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Real </td>
<td> 1059 </td>
<td> 68 </td>
<td> 2014 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants">Condition Based Maintenance of Naval Propulsion Plants</a> </td></tr></table></td>
<!-- <td> Data have been generated from a sophisticated simulator of a Gas Turbines (GT), mounted on a Frigate characterized by a COmbined Diesel eLectric And Gas (CODLAG) propulsion plant type. </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 11934 </td>
<td> 16 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Grammatical+Facial+Expressions"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Grammatical+Facial+Expressions">Grammatical Facial Expressions</a> </td></tr></table></td>
<!-- <td> This dataset supports the development of models that make possible to interpret Grammatical Facial Expressions from Brazilian Sign Language (Libras). </td> -->
<td> Multivariate, Sequential </td>
<td> Classification, Clustering </td>

<td> Real </td>
<td> 27965 </td>
<td> 100 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/NoisyOffice"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/NoisyOffice">NoisyOffice</a> </td></tr></table></td>
<!-- <td> Corpus intended to do cleaning (or binarization) and enhancement of noisy grayscale printed text images using supervised learning methods. Noisy images and their corresponding ground truth provided. </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Real </td>
<td> 216 </td>
<td> 216 </td>
<td> 2015 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/MHEALTH+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/MHEALTH+Dataset">MHEALTH Dataset</a> </td></tr></table></td>
<!-- <td> The MHEALTH (Mobile Health) dataset is devised to benchmark techniques dealing with human behavior analysis based on multimodal body sensing. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 120 </td>
<td> 23 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Student+Performance"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Student+Performance">Student Performance</a> </td></tr></table></td>
<!-- <td> Predict student performance in secondary education (high school).   </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Integer </td>
<td> 649 </td>
<td> 33 </td>
<td> 2014 </td>
<!-- <td> Social </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/ElectricityLoadDiagrams20112014"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/ElectricityLoadDiagrams20112014">ElectricityLoadDiagrams20112014</a> </td></tr></table></td>
<!-- <td> This data set contains electricity consumption of 370 points/clients.
  </td> -->
<td> Time-Series </td>
<td> Regression, Clustering </td>
<td> Real </td>
<td> 370 </td>
<td> 140256 </td>
<td> 2015 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Gas+sensor+array+under+dynamic+gas+mixtures"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Gas+sensor+array+under+dynamic+gas+mixtures">Gas sensor array under dynamic gas mixtures</a> </td></tr>

</table></td>
<!-- <td> The data set contains the recordings of 16 chemical sensors exposed to two dynamic gas mixtures at varying concentrations. For each mixture, signals were acquired continuously during 12 hours. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Regression </td>
<td> Real </td>
<td> 4178504 </td>
<td> 19 </td>
<td> 2015 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/microblogPCU"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/microblogPCU">microblogPCU</a> </td></tr></table></td>
<!-- <td> MicroblogPCU data is crawled from sina weibo microblog[http://weibo.com/]. This data can be used to study machine learning methods as well as do some social network research.   </td> -->
<td> Multivariate, Univariate, Sequential, Text </td>
<td> Classification, Causal-Discovery </td>
<td> Integer, Real </td>
<td> 221579 </td>
<td> 20 </td>
<td> 2015 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Firm-Teacher_Clave-Direction_Classification"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Firm-Teacher_Clave-Direction_Classification">Firm-Teacher_Clave-Direction_Classification</a> </td></tr></table></td>
<!-- <td> The data are binary attack-point vectors and their clave-direction class(es) according to the partido-alto-based paradigm. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> </td>
<td> 10800 </td>
<td> 20 </td>
<td> 2015 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Dataset+for+Sensorless+Drive+Diagnosis"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Dataset+for+Sensorless+Drive+Diagnosis">Dataset for Sensorless Drive Diagnosis</a> </td></tr></table></td>
<!-- <td> Features are extracted from motor current. The motor has intact and defective components. This results in 11 different classes with different conditions.   </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 58509 </td>
<td> 49 </td>
<td> 2015 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/TV+News+Channel+Commercial+Detection+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/TV+News+Channel+Commercial+Detection+Dataset">TV News Channel Commercial Detection Dataset</a> </td></tr></table></td>
<!-- <td> TV Commercials data set consists of standard audio-visual features of video shots extracted from 150 hours of TV news broadcast of 3 Indian and 2 international news channels (30 Hours each) .   </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>

<td> Real </td>
<td> 129685 </td>
<td> 12 </td>
<td> 2015 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Phishing+Websites"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Phishing+Websites">Phishing Websites</a> </td></tr></table></td>
<!-- <td> This dataset collected mainly from: PhishTank archive, MillerSmiles archive, Google ’s searching operators. </td> -->
<td> </td>
<td> Classification </td>
<td> Integer </td>
<td> 2456 </td>
<td> 30 </td>
<td> 2015 </td>
<!-- <td> Computer Security </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Greenhouse+Gas+Observing+Network"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Greenhouse+Gas+Observing+Network">Greenhouse Gas Observing Network</a> </td></tr></table></td>
<!-- <td> Design an observing network to monitor emissions of a greenhouse gas (GHG) in California given time series of synthetic observations and tracers from weather model simulations.   </td> -->
<td> Multivariate, Time-Series </td>
<td> Regression </td>
<td> Real </td>
<td> 2921 </td>
<td> 5232 </td>
<td> 2015 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Diabetic+Retinopathy+Debrecen+Data+Set"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Diabetic+Retinopathy+Debrecen+Data+Set">Diabetic Retinopathy Debrecen Data Set</a> </td></tr></table></td>
<!-- <td> This dataset contains features extracted from the Messidor image set to predict whether an image contains signs of diabetic retinopathy or not.   </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 1151 </td>
<td> 20 </td>
<td> 2014 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/HIV-1+protease+cleavage"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/HIV-1+protease+cleavage">HIV-1 protease cleavage</a> </td></tr></table></td>
<!-- <td> The data contains lists of octamers (8 amino acids) and a flag (-1 or 1) depending on whether HIV-1 protease will cleave in the central position (between amino acids 4 and 5). </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 6590 </td>
<td> 1 </td>
<td> 2015 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Sentiment+Labelled+Sentences"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Sentiment+Labelled+Sentences">Sentiment Labelled Sentences</a> </td></tr></table></td>

<!-- <td> The dataset contains sentences labelled with positive or negative sentiment.
</td> -->
<td> Text </td>
<td> Classification </td>
<td> </td>
<td> 3000 </td>
<td> </td>
<td> 2015 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Online+News+Popularity"><img border="1" src="assets/MLimages/Small
Largedefault.jpg"/></a> </td><td> <a href="datasets/Online+News+Popularity">Online Ne
ws Popularity</a> </td></tr></table></td>
<!-- <td> This dataset summarizes a heterogeneous set of features about articles published
by Mashable in a period of two years. The goal is to predict the number of shares in social networks (popularity
). </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Integer, Real </td>
<td> 39797 </td>
<td> 61 </td>
<td> 2015 </td>
<!-- <td> Business </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Forest+type+mapping"><img border="1" src="assets/MLimages/SmallLa
rgedefault.jpg"/></a> </td><td> <a href="datasets/Forest+type+mapping">Forest type m
apping</a> </td></tr></table></td>
<!-- <td> Multi-temporal remote sensing data of a forested area in Japan. The goal is to map
different forest types using spectral data. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> </td>
<td> 326 </td>
<td> 27 </td>
<td> 2015 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/wiki4HE"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"
/></a> </td><td> <a href="datasets/wiki4HE">wiki4HE</a> </td></tr></table></td
>
<!-- <td> Survey of faculty members from two Spanish universities on teaching uses of Wikip
edia </td> -->
<td> Multivariate </td>
<td> Regression, Clustering, Causal-Discovery </td>
<td> </td>
<td> 913 </td>
<td> 53 </td>
<td> 2015 </td>
<!-- <td> Social </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset"><img b
order="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datase
ts/Online+Video+Characteristics+and+Transcoding+Time+Dataset">Online Video Characteristics and Transcodi
ng Time Dataset</a> </td></tr></table></td>
<!-- <td> The dataset contains a million randomly sampled video instances listing 10 fundam
ental video characteristics along with the YouTube video ID.   </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Integer, Real </td>
<td> 168286 </td>
<td> 11 </td>
<td> 2015 </td>

<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Chronic_Kidney_Disease"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Chronic_Kidney_Disease">Chronic_Kidney_Disease</a> </td></tr></table></td>
<!-- <td> This dataset can be used to predict the chronic kidney disease and it can be collected from the hospital nearly 2 months of period. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 400 </td>
<td> 25 </td>
<td> 2015 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014">Machine Learning based ZZAlpha Ltd. Stock Recommendations 2012-2014</a> </td></tr></table></td>
<!-- <td> The data here are the ZZAlpha® machine learning recommendations made for various US traded stock portfolios the morning of each day during the 3 year period Jan 1, 2012 - Dec 31, 2014.   </td> -->
<td> Sequential, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 314080 </td>
<td> 0 </td>
<td> 2015 </td>
<!-- <td> Business </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Folio"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Folio">Folio</a> </td></tr></table></td>
<!-- <td> 20 photos of leaves for each of 32 different species. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> </td>
<td> 637 </td>
<td> 20 </td>
<td> 2015 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015">Taxi Service Trajectory - Prediction Challenge, ECML PKDD 2015</a> </td></tr></table></td>
<!-- <td> An accurate dataset describing trajectories performed by all the 442 taxis running in the city of Porto, in Portugal.
  </td> -->
<td> Multivariate, Sequential, Time-Series, Domain-Theory </td>
<td> Clustering, Causal-Discovery </td>
<td> Real </td>
<td> 1710671 </td>
<td> 9 </td>
<td> 2015 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Cuff-Less+Blood+Pressure+Estimation"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Cuff-Less+Blood+Pressure+Estimation">Cuff-Less Blood Pressure Estimation</a> </td></tr></table></td>
<!-- <td> This Data set provides preprocessed and cleaned vital signals which can be used in designing algorithms for cuff-less estimation of the blood pressure. </td> -->
<td> Multivariate </td>

<td> Classification, Regression </td>
<td> Real </td>
<td> 12000 </td>
<td> 3 </td>
<td> 2015 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions">Smartphone-Based Recognition of Human Activities and Postural Transitions</a> </td></tr></table></td>
<!-- <td> Activity recognition data set built from the recordings of 30 subjects performing basic activities and postural transitions while carrying a waist-mounted smartphone with embedded inertial sensors. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 10929 </td>
<td> 561 </td>
<td> 2015 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Mice+Protein+Expression"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Mice+Protein+Expression">Mice Protein Expression</a> </td></tr></table></td>
<!-- <td> Expression levels of 77 proteins measured in the cerebral cortex of 8 classes of control and Down syndrome mice exposed to context fear conditioning, a task used to assess associative learning. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 1080 </td>
<td> 82 </td>
<td> 2015 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/UJIIndoorLoc-Mag"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/UJIIndoorLoc-Mag">UJIIndoorLoc-Mag</a> </td></tr></table></td>
<!-- <td> The UJIIndoorLoc-Mag is an indoor localization database to test Indoor Positioning System that rely on Earth's magnetic field variations. </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Classification, Regression, Clustering </td>
<td> Integer, Real </td>
<td> 40000 </td>
<td> 13 </td>
<td> 2015 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Heterogeneity+Activity+Recognition"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Heterogeneity+Activity+Recognition">Heterogeneity Activity Recognition</a> </td></tr></table></td>
<!-- <td> The Heterogeneity Human Activity Recognition (HHAR) dataset from Smartphones and Smartwatches is a dataset devised to benchmark human activity recognition algorithms (classification, automatic data segmentation, sensor fusion, feature extraction, etc.) in real-world contexts; specifically, the dataset is gathered with a variety of different device models and use-scenarios, in order to reflect sensing heterogeneities to be expected in real deployments. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 43930257 </td>
<td> 16 </td>

<td> 16 </td>
<td> 2015 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set">Educational Process Mining (EPM): A Learning Analytics Data Set</a> </td></tr></table></td>
<!-- <td> Educational Process Mining data set is built from the recordings of 115 subjects' activities through a logging application while learning with an educational simulator. </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Classification, Regression, Clustering </td>
<td> Integer </td>
<td> 230318 </td>
<td> 13 </td>
<td> 2015 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/HEPMASS"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/HEPMASS">HEPMASS</a> </td></tr></table></td>
<!-- <td> The search for exotic particles requires sorting through a large number of collisions to find the events of interest. This data set challenges one to detect a new particle of unknown mass. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 10500000 </td>
<td> 28 </td>
<td> 2016 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Indoor+User+Movement+Prediction+from+RSS+data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Indoor+User+Movement+Prediction+from+RSS+data">Indoor User Movement Prediction from RSS data</a> </td></tr></table></td>
<!-- <td> This dataset contains temporal data from a Wireless Sensor Network deployed in real-world office environments. The task is intended as real-life benchmark in the area of Ambient Assisted Living. </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 13197 </td>
<td> 4 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Open+University+Learning+Analytics+dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Open+University+Learning+Analytics+dataset">Open University Learning Analytics dataset</a> </td></tr></table></td>
<!-- <td> Open University Learning Analytics Dataset contains data about courses, students and their interactions with Virtual Learning Environment for seven selected courses and more than 30000 students. </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Classification, Regression, Clustering </td>
<td> Integer </td>
<td> </td>
<td> </td>
<td> 2015 </td>
<!-- <td> Computer </td> -->
</tr><tr>

<td><table><tr><td><a href="datasets/default+of+credit+card+clients"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/default+of+credit+card+clients">default of credit card clients</a> </td></tr></table></td>
<!-- <td> This research aimed at the case of customers' default payments in Taiwan and compares the predictive accuracy of probability of default among six data mining methods. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 30000 </td>
<td> 24 </td>
<td> 2016 </td>
<!-- <td> Business </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Mesothelioma%E2%80%99s+disease+data+set+"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Mesothelioma%E2%80%99s+disease+data+set+">Mesothelioma's disease data set </a> </td></tr></table></td>
<!-- <td> Mesothelioma's disease data set were prepared at Dicle University Faculty of Medicine in Turkey.
Three hundred and twenty-four Mesothelioma patient data. In the dataset, all samples have 34 features. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 324 </td>
<td> 34 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Online+Retail"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Online+Retail">Online Retail</a> </td></tr></table></td>
<!-- <td> This is a transnational data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail. </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Classification, Clustering </td>
<td> Integer, Real </td>
<td> 541909 </td>
<td> 8 </td>
<td> 2015 </td>
<!-- <td> Business </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/SIFT10M"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/SIFT10M">SIFT10M</a> </td></tr></table></td>
<!-- <td> In SIFT10M, each data point is a SIFT feature which is extracted from Caltech-256 by the open source VLFeat library. The corresponding patches of the SIFT features are provided. </td> -->
<td> Multivariate </td>
<td> Causal-Discovery </td>
<td> Integer </td>
<td> 11164866 </td>
<td> 128 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/GPS+Trajectories"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/GPS+Trajectories">GPS Trajectories</a> </td></tr></table></td>
<!-- <td> The dataset has been feed by Android app called Go!Track. It is available at Google Play Store(https://play.google.com/store/apps/details?id=com.go.router).   </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>

<td> Real </td>
<td> 163 </td>
<td> 15 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Detect+Malacious+Executable%28AntiVirus%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Detect+Malacious+Executable%28AntiVirus%29">Detect Malacious Executable(AntiVirus)</a> </td></tr></table></td>
<!-- <td> I extract features from malacious and non-malacious and create and training dataset to teach svm classifier.Dataset made of unknown executable to detect if it is virus or normal safe executable. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 373 </td>
<td> 513 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Occupancy+Detection+"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Occupancy+Detection+">Occupancy Detection </a> </td></tr></table></td>
<!-- <td> Experimental data used for binary classification (room occupancy) from Temperature,Humidity,Light and CO2. Ground-truth occupancy was obtained from time stamped pictures that were taken every minute. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 20560 </td>
<td> 7 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Parkinson%E2%80%99s+Disease"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Parkinson%E2%80%99s+Disease">Improved Spiral Test Using Digitized Graphics Tablet for Monitoring Parkinson’s Disease</a> </td></tr></table></td>
<!-- <td> Handwriting database consists of 25 PWP(People with Parkinson) and 15 healthy individuals.Three types of recordings (Static Spiral Test, Dynamic Spiral Test and Stability Test) are taken. </td> -->
<td> Multivariate </td>
<td> Classification, Regression, Clustering </td>
<td> Real </td>
<td> 40 </td>
<td> 7 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/News+Aggregator"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/News+Aggregator">News Aggregator</a> </td></tr></table></td>
<!-- <td> References to news pages collected from an web aggregator in the period from 10-March-2014 to 10-August-2014. The resources are grouped into clusters that represent pages discussing the same story. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> </td>
<td> 422937 </td>
<td> 5 </td>

<td> 2016 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Air+Quality"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Air+Quality">Air Quality</a> </td></tr></table></td>
<!-- <td> Contains the responses of a gas multisensor device deployed on the field in an Italian city. Hourly responses averages are recorded along with gas concentrations references from a certified analyzer.   </td> -->
<td> Multivariate, Time-Series </td>
<td> Regression </td>
<td> Real </td>
<td> 9358 </td>
<td> 15 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Twin+gas+sensor+arrays"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Twin+gas+sensor+arrays">Twin gas sensor arrays</a> </td></tr></table></td>
<!-- <td> 5 replicates of an 8-MOX gas sensor array were exposed to different gas conditions (4 volatiles at 10 concentration levels each). </td> -->
<td> Multivariate, Time-Series, Domain-Theory </td>
<td> Classification, Regression </td>
<td> Real </td>
<td> 640 </td>
<td> 480000 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Gas+sensors+for+home+activity+monitoring"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Gas+sensors+for+home+activity+monitoring">Gas sensors for home activity monitoring</a> </td></tr></table></td>
<!-- <td> 100 recordings of a sensor array under different conditions in a home setting: background, wine and banana presentations. The array includes 8 MOX gas sensors, and humidity and temperature sensors.
  </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 919438 </td>
<td> 11 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Facebook+Comment+Volume+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Facebook+Comment+Volume+Dataset">Facebook Comment Volume Dataset</a> </td></tr></table></td>
<!-- <td> Instances in this dataset contain features extracted from facebook posts. The task associated with the data is to predict how many comments the post will receive. </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Integer, Real </td>
<td> 40949 </td>
<td> 54 </td>
<td> 2016 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Living+%28AAL%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Living+%28AAL%29">Smartphone Dataset for Human Activity Recognition (

HAR) in Ambient Assisted Living (AAL)</a> </td></tr></table></td>
<!-- <td> This data is an addition to an existing dataset on UCI. We collected more data to im prove the accuracy of our human activity recognition algorithms applied in the domain of Ambient Assisted Livin g.   </td> -->
<td> Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 5744 </td>
<td> 561 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Polish+companies+bankruptcy+data"><img border="1" src="assets/MLim ages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Polish+companies+bankr uptcy+data">Polish companies bankruptcy data</a> </td></tr></table></td>
<!-- <td> The dataset is about bankruptcy prediction of Polish companies.The bankrupt comp anies were analyzed in the period 2000-2012, while the still operating companies were evaluated from 2007 to 2013. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 10503 </td>
<td> 64 </td>
<td> 2016 </td>
<!-- <td> Business </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28A ReM%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29">Activit y Recognition system based on Multisensor data fusion (AReM)</a> </td></tr></table></td>
<!-- <td> This dataset contains temporal data from a Wireless Sensor Network worn by an ac tor performing the activities: bending, cycling, lying down, sitting, standing, walking. </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 42240 </td>
<td> 6 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Dota2+Games+Results"><img border="1" src="assets/MLimages/SmallL argedefault.jpg"/></a> </td><td> <a href="datasets/Dota2+Games+Results">Dota2 Gam es Results</a> </td></tr></table></td>
<!-- <td> Dota 2 is a popular computer game with two teams of 5 players. At the start of the g ame each player chooses a unique hero with different strengths and weaknesses. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> </td>
<td> 102944 </td>
<td> 116 </td>
<td> 2016 </td>
<!-- <td> Game </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Facebook+metrics"><img border="1" src="assets/MLimages/SmallLarge default.jpg"/></a> </td><td> <a href="datasets/Facebook+metrics">Facebook metrics</a > </td></tr></table></td>
<!-- <td> Facebook performance metrics of a renowned cosmetic's brand Facebook page.&n bsp; </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Integer </td>
<td> 500 </td>

<td> 19 </td>
<td> 2016 </td>
<!-- <td> Business </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/UbiqLog+%28smartphone+lifelogging%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/UbiqLog+%28smartphone+lifelogging%29">UbiqLog (smartphone lifelogging)</a> </td></tr></table></td>
<!-- <td> UbiqLog is the smartphone lifelogging tool that runs on the smartphone of 35 users for about 2 months.
  </td> -->
<td> Multivariate </td>
<td> Causal-Discovery </td>
<td> </td>
<td> 9782222 </td>
<td> </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/NIPS+Conference+Papers+1987-2015"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/NIPS+Conference+Papers+1987-2015">NIPS Conference Papers 1987-2015</a> </td></tr></table></td>
<!-- <td> This data set contains the distribution of words in the full text of the NIPS conference papers published from 1987 to 2015. </td> -->
<td> Text </td>
<td> Clustering </td>
<td> Integer </td>
<td> 11463 </td>
<td> 5812 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/HTRU2"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/HTRU2">HTRU2</a> </td></tr></table></td>
<!-- <td> Pulsar candidates collected during the HTRU survey. Pulsars are a type of star, of considerable scientific interest. Candidates must be classified in to pulsar and non-pulsar classes to aid discovery. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 17898 </td>
<td> 9 </td>
<td> 2017 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Drug+consumption+%28quantified%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Drug+consumption+%28quantified%29">Drug consumption (quantified)</a> </td></tr></table></td>
<!-- <td> Classify type of drug consumer by personality data </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 1885 </td>
<td> 32 </td>
<td> 2016 </td>
<!-- <td> Social </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Appliances+energy+prediction"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Appliances+energy+prediction">Appliances energy prediction</a> </td></tr></table></td>
<!-- <td> Experimental data used to create regression models of appliances energy use in a low energy building. </td> -->
<td> Multivariate, Time-Series </td>

<td> Regression </td>
<td> Real </td>
<td> 19735 </td>
<td> 29 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Miskolc+IIS+Hybrid+IPS"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Miskolc+IIS+Hybrid+IPS">Miskolc IIS Hybrid IPS</a> </td></tr></table></td>
<!-- <td> The dataset was created for the comparison and evaluation of hybrid indoor positioning methods. The dataset presented contains data from W-LAN and Bluetooth interfaces, and Magnetometer. </td> -->
<td> Text </td>
<td> Classification, Clustering, Causal-Discovery </td>
<td> Integer </td>
<td> 1540 </td>
<td> 67 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/KDC-4007+dataset+Collection"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/KDC-4007+dataset+Collection">KDC-4007 dataset Collection</a> </td></tr></table></td>
<!-- <td> KDC-4007 dataset Collection is the Kurdish Documents Classification text used in categories regarding Kurdish Sorani news and articles. </td> -->
<td> Multivariate, Text </td>
<td> Classification, Regression </td>
<td> Integer </td>
<td> 4007 </td>
<td> </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Geo-Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Geo-Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone">Geo-Magnetic field and WLAN dataset for indoor localisation from wristband and smartphone</a> </td></tr></table></td>
<!-- <td> A multisource and multivariate dataset for indoor localisation methods based on WLAN and Geo-Magnetic field fingerprinting </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Classification, Regression, Clustering </td>
<td> Integer, Real </td>
<td> 153540 </td>
<td> 25 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/DrivFace"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/DrivFace">DrivFace</a> </td></tr></table></td>
<!-- <td> The DrivFace contains images sequences of subjects while driving in real scenarios. It is composed of 606 samples of 640×480, acquired over different days from 4 drivers with several facial features. </td> -->
<td> Multivariate </td>
<td> Classification, Regression, Clustering </td>
<td> Real </td>
<td> 606 </td>
<td> 6400 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
</tr></table>
</body>
</html>

</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Website+Phishing"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Website+Phishing">Website Phishing</a> </td></tr></table></td>
<!-- <td>

  </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 1353 </td>
<td> 10 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/YouTube+Spam+Collection"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/YouTube+Spam+Collection">YouTube Spam Collection</a> </td></tr></table></td>
<!-- <td> It is a public set of comments collected for spam research. It has five datasets composed by 1,956 real messages extracted from five videos that were among the 10 most viewed on the collection period. </td> -->
<td> Text </td>
<td> Classification </td>
<td> </td>
<td> 1956 </td>
<td> 5 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Beijing+PM2.5+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Beijing+PM2.5+Data">Beijing PM2.5 Data</a> </td></tr></table></td>
<!-- <td> This hourly data set contains the PM2.5 data of US Embassy in Beijing. Meanwhile, meteorological data from Beijing Capital International Airport are also included.   </td> -->
<td> Multivariate, Time-Series </td>
<td> Regression </td>
<td> Integer, Real </td>
<td> 43824 </td>
<td> 13 </td>
<td> 2017 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Cargo+2000+Freight+Tracking+and+Tracing"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Cargo+2000+Freight+Tracking+and+Tracing">Cargo 2000 Freight Tracking and Tracing</a> </td></tr></table></td>
<!-- <td> Sanitized and anonymized Cargo 2000 (C2K) airfreight tracking and tracing events, covering five months of business execution (3,942 process instances, 7,932 transport legs, 56,082 activities).   </td> -->
<td> Multivariate, Sequential </td>
<td> Classification, Regression </td>
<td> Integer </td>
<td> 3942 </td>
<td> 98 </td>
<td> 2016 </td>
<!-- <td> Business </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Cervical+cancer+%28Risk+Factors%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Cervical+cancer+%28Risk+Factors%29">Cervical cancer (Risk Factors)</a> </td></tr></table></td>
<!-- <td> This dataset focuses on the prediction of indicators/diagnosis of cervical cancer. The features cover demographic information, habits, and historic medical records. </td> -->
<td> Multivariate </td>

<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 858 </td>
<td> 36 </td>
<td> 2017 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Quality+Assessment+of+Digital+Colposcopies"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Quality+Assessment+of+Digital+Colposcopies">Quality Assessment of Digital Colposcopies</a> </td></tr></table></td>
<!-- <td> This dataset explores the subjective quality assessment of digital colposcopies. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 287 </td>
<td> 69 </td>
<td> 2017 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/KASANDR"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/KASANDR">KASANDR</a> </td></tr></table></td>
<!-- <td> KASANDR is a novel, publicly available collection for recommendation systems that records the behavior of customers of the European leader in e-Commerce advertising, Kelkoo.   </td> -->
<td> Multivariate </td>
<td> Causal-Discovery </td>
<td> Integer </td>
<td> 17764280 </td>
<td> 2158859 </td>
<td> 2017 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/FMA%3A+A+Dataset+For+Music+Analysis"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/FMA%3A+A+Dataset+For+Music+Analysis">FMA: A Dataset For Music Analysis</a> </td></tr></table></td>
<!-- <td> FMA features 106,574 tracks and includes song title, album, artist, genres; play counts, favorites, comments; description, biography, tags; together with audio (343 days, 917 GiB) and features. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 106574 </td>
<td> 518 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Air+quality"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Air+quality">Air quality</a> </td></tr></table></td>
<!-- <td> Contains the responses of a gas multisensor device deployed on the field in an Italian city.   </td> -->
<td> Multivariate, Time-Series </td>
<td> Regression </td>
<td> Real </td>
<td> 9358 </td>
<td> 15 </td>
<td> 2016 </td>
<!-- <td> Other </td> -->
</tr><tr>

</tr><tr>
<td><table><tr><td><a href="datasets/Epileptic+Seizure+Recognition"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Epileptic+Seizure+Recognition">Epileptic Seizure Recognition</a> </td></tr></table></td>
<!-- <td> This dataset is a pre-processed and re-structured/reshaped version of a very commonly used dataset featuring epileptic seizure detection.   </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Clustering </td>
<td> Integer, Real </td>
<td> 11500 </td>
<td> 179 </td>
<td> 2017 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Devanagari+Handwritten+Character+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Devanagari+Handwritten+Character+Dataset">Devanagari Handwritten Character Dataset</a> </td></tr></table></td>
<!-- <td> This is an image database of Handwritten Devanagari characters. There are 46 classes of characters with 2000 examples each. The dataset is split into training set(85%) and testing set(15%).   </td> -->
<td> </td>
<td> Classification </td>
<td> Integer </td>
<td> 92000 </td>
<td> </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Stock+portfolio+performance"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Stock+portfolio+performance">Stock portfolio performance</a> </td></tr></table></td>
<!-- <td> The data set of performances of weighted scoring stock portfolios are obtained with mixture design from the US stock market historical database. </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 315 </td>
<td> 12 </td>
<td> 2016 </td>
<!-- <td> Business </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/MoCap+Hand+Postures"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/MoCap+Hand+Postures">MoCap Hand Postures</a> </td></tr></table></td>
<!-- <td> 5 types of hand postures from 12 users were recorded using unlabeled markers attached to fingers of a glove in a motion capture environment. Due to resolution and occlusion, missing values are common. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> Integer, Real </td>
<td> 78095 </td>
<td> 38 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech">Early biomarkers of Parkinson&#65533;s disease based on natural connected speech</a> </td></tr></table></td>
<!-- <td> Predict a pattern of neurodegeneration in the dataset of speech features obtained from patients with early untreated Parkinson’s disease and patients at high risk developing Parkinson’s disease.

  </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Integer, Real </td>
<td> 130 </td>
<td> 65 </td>
<td> 2017 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting">Data for Software Engineering Teamwork Assessment in Education Setting</a> </td></tr></table></td>
<!-- <td> Data include over 100 Team Activity Measures and outcomes (ML classes) obtained from activities of 74 student teams during the creation of final class project in SW Eng. classes at SFSU, Fulda, FAU </td> -->
<td> Sequential, Time-Series </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 74 </td>
<td> 102 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/PM2.5+Data+of+Five+Chinese+Cities"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/PM2.5+Data+of+Five+Chinese+Cities">PM2.5 Data of Five Chinese Cities</a> </td></tr></table></td>
<!-- <td> This hourly data set contains the PM2.5 data in Beijing, Shanghai, Guangzhou, Chengdu and Shenyang. Meanwhile, meteorological data for each city are also included. </td> -->
<td> Multivariate, Time-Series </td>
<td> Regression </td>
<td> Integer, Real </td>
<td> 52854 </td>
<td> 86 </td>
<td> 2017 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet">Parkinson Disease Spiral Drawings Using Digitized Graphics Tablet</a> </td></tr></table></td>
<!-- <td> Handwriting database consists of 62 PWP(People with Parkinson) and 15 healthy individuals. Three types of recordings (Static Spiral Test, Dynamic Spiral Test and Stability Test) are taken. </td> -->
<td> Multivariate </td>
<td> Classification, Regression, Clustering </td>
<td> Integer </td>
<td> 77 </td>
<td> 7 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Sales_Transactions_Dataset_Weekly"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly</a> </td></tr></table></td>
<!-- <td> Contains weekly purchased quantities of 800 over products over 52 weeks. Normalised values are provided too. </td> -->
<td> Multivariate, Time-Series </td>
<td> Clustering </td>
<td> Integer, Real </td>
<td> 811 </td>
<td> 53 </td>

<td> 2017 </td>
<!-- <td>   </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Las+Vegas+Strip"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Las+Vegas+Strip">Las Vegas Strip</a> </td></tr></table></td>
<!-- <td> This dataset includes quantitative and categorical features from online reviews from 21 hotels located in Las Vegas Strip, extracted from TripAdvisor (http://www.tripadvisor.com). </td> -->
<td> </td>
<td> Classification, Regression </td>
<td> Integer </td>
<td> 504 </td>
<td> 20 </td>
<td> 2017 </td>
<!-- <td> Business </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Eco-hotel"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Eco-hotel">Eco-hotel</a> </td></tr></table></td>
<!-- <td> This dataset includes Online Textual Reviews from both online (e.g., TripAdvisor) and offline (e.g., Guests' book) sources from the Areias do Seixo Eco-Resort. </td> -->
<td> Text </td>
<td> </td>
<td> </td>
<td> 401 </td>
<td> 1 </td>
<td> 2017 </td>
<!-- <td> Business </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/MEU-Mobile+KSD"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/MEU-Mobile+KSD">MEU-Mobile KSD</a> </td></tr></table></td>
<!-- <td> This dataset contains keystroke dynamics data collected on a touch mobile device (Nexus 7). The dataset contains 2856 records, 51 records per subject for 56 subjects.   </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 2856 </td>
<td> 71 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Crowdsourced+Mapping"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Crowdsourced+Mapping">Crowdsourced Mapping</a> </td></tr></table></td>
<!-- <td> Crowdsourced data from OpenStreetMap is used to automate the classification of satellite images into different land cover classes (impervious, farm, forest, grass, orchard, water).   </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> </td>
<td> 10546 </td>
<td> 29 </td>
<td> 2016 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/gene+expression+cancer+RNA-Seq"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/gene+expression+cancer+RNA-Seq">gene expression cancer RNA-Seq</a> </td></tr></table></td>
<!-- <td> This collection of data is part of the RNA-Seq (HiSeq) PANCAN data set, it is a random extraction of gene expressions of patients having different types of tumor: BRCA, KIRC, COAD, LUAD and

PRAD. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 801 </td>
<td> 20531 </td>
<td> 2016 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer">Hybrid Indoor Positioning Dataset from WiFi RSSI, Bluetooth and magnetometer</a> </td></tr></table></td>
<!-- <td> The dataset was created for the comparison and evaluation of hybrid indoor positioning methods. The dataset presented contains data from W-LAN and Bluetooth interfaces, and Magnetometer. </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 1540 </td>
<td> 65 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/chestnut+%E2%80%93+LARVIC"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/chestnut+%E2%80%93+LARVIC">chestnut – LARVIC</a> </td></tr></table></td>
<!-- <td> The research project presents this database, shows the images of chestnuts that will be processed to determine the presence or absence of defects </td> -->
<td> </td>
<td> Classification, Clustering </td>
<td> </td>
<td> 1451 </td>
<td> 3 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%28OBS%29+Network"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td>
<td> <a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%28OBS%29+Network">Burst Header Packet (BHP) flooding attack on Optical Burst Switching (OBS) Network</a> </td></tr></table></td>
<!-- <td> One of the primary challenges in identifying the risks of the Burst Header Packet (BHP) flood attacks in Optical Burst Switching networks (OBS) is the scarcity of reliable historical data.   </td> -->
<td> Text </td>
<td> Classification </td>
<td> Integer </td>
<td> 1075 </td>
<td> 22 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Motion+Capture+Hand+Postures"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Motion+Capture+Hand+Postures">Motion Capture Hand Postures</a> </td></tr></table></td>
<!-- <td> 5 types of hand postures from 12 users were recorded using unlabeled markers on fingers of a glove in a motion capture environment. Due to resolution and occlusion, missing values are common. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>

<td> Real </td>
<td> 78095 </td>
<td> 38 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Anuran+Calls+%28MFCCs%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Anuran+Calls+%28MFCCs%29">Anuran Calls (MFCCs)</a> </td></tr></table></td>
<!-- <td> Acoustic features extracted from syllables of anuran (frogs) calls, including the family, the genus, and the species labels (multilabel).   </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 7195 </td>
<td> 22 </td>
<td> 2017 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization">TTC-3600: Benchmark dataset for Turkish text categorization</a> </td></tr></table></td>
<!-- <td> The TTC-3600 data set is a collection of Turkish news and articles including categorized 3,600 documents from 6 well-known portals in Turkey. It has 4 different forms in ARFF Weka format. </td> -->
<td> Text </td>
<td> Classification, Clustering </td>
<td> Integer </td>
<td> 3600 </td>
<td> 4814 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy">Gastrointestinal Lesions in Regular Colonoscopy</a> </td></tr></table></td>
<!-- <td> This dataset contains features extracted from colonoscopy videos used to detect gastrointestinal lesions. It contains 76 lesions: 15 serrated adenomas, 21 hyperplastic lesions and 40 adenoma.   </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 76 </td>
<td> 698 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Daily+Demand+Forecasting+Orders"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Daily+Demand+Forecasting+Orders">Daily Demand Forecasting Orders</a> </td></tr></table></td>
<!-- <td> The dataset was collected during 60 days, this is a real database of a brazilian logistics company. </td> -->
<td> Time-Series </td>
<td> Regression </td>
<td> Integer </td>
<td> 60 </td>
<td> 13 </td>
<td> 2017 </td>
<!-- <td> Business </td> -->
</tr><tr>

<td><table><tr><td><a href="datasets/Paper+Reviews"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Paper+Reviews">Paper Reviews</a> </td></tr></table></td>
<!-- <td> This sentiment analysis data set contains scientific paper reviews from an international conference on computing and informatics. The task is to predict the orientation or the evaluation of a review. </td> -->
<td> Text </td>
<td> Classification, Regression </td>
<td> Integer </td>
<td> 405 </td>
<td> 10 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/extention+of+Z-Alizadeh+sani+dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/extention+of+Z-Alizadeh+sani+dataset">extention of Z-Alizadeh sani dataset</a> </td></tr></table></td>
<!-- <td> It was collected for CAD diagnosis. </td> -->
<td> </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 303 </td>
<td> 59 </td>
<td> 2017 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Z-Alizadeh+Sani"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Z-Alizadeh+Sani">Z-Alizadeh Sani</a> </td></tr></table></td>
<!-- <td> It was collected for CAD diagnosis. </td> -->
<td> </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 303 </td>
<td> 56 </td>
<td> 2017 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Dynamic+Features+of+VirusShare+Executables"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Dynamic+Features+of+VirusShare+Executables">Dynamic Features of VirusShare Executables</a> </td></tr></table></td>
<!-- <td> This dataset contains the dynamic features of 107,888 executables, collected by VirusShare from Nov/2010 to Jul/2014. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Regression </td>
<td> Integer </td>
<td> 107888 </td>
<td> 482 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/IDA2016Challenge"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/IDA2016Challenge">IDA2016Challenge</a> </td></tr></table></td>
<!-- <td> The dataset consists of data collected from heavy Scania trucks in everyday usage. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 76000 </td>
<td> 171 </td>

<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/DSRC+Vehicle+Communications"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/DSRC+Vehicle+Communications">DSRC Vehicle Communications</a> </td></tr></table></td>
<!-- <td> This set Provides data regarding wireless communications between vehicles and road side units. two separate data sets are provided (normal scenario) and in the presence of attacker (jammer). </td> -->
<td> Sequential, Text </td>
<td> Clustering </td>
<td> Real </td>
<td> 10000 </td>
<td> 5 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Mturk+User-Perceived+Clusters+over+Images"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Mturk+User-Perceived+Clusters+over+Images">Mturk User-Perceived Clusters over Images</a> </td></tr></table></td>
<!-- <td> This dataset was collected by Shan-Hung Wu and DataLab members at NTHU, Taiwan. There're 325 user-perceived clusters from 100 users and their corresponding descriptions. </td> -->
<td> Multivariate, Text </td>
<td> Clustering </td>
<td> Integer </td>
<td> 180 </td>
<td> 500 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Character+Font+Images"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Character+Font+Images">Character Font Images</a> </td></tr></table></td>
<!-- <td> Character images from scanned and computer generated fonts. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 745000 </td>
<td> 411 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Instance+Labels"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Instance+Labels">DeliciousMIL: A Data Set for Multi-Label Multi-Instance Learning with Instance Labels</a> </td></tr></table></td>
<!-- <td> This dataset includes 1) 12234 documents (8251 training, 3983 test) extracted from DeliciousT140 dataset, 2) class labels for all documents, 3) labels for a subset of sentences of the test documents. </td> -->
<td> Text </td>
<td> Classification </td>
<td> Integer </td>
<td> 12234 </td>
<td> 8519 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children+"><img bor

<td><table><tr><td><a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++">Autistic Spectrum Disorder Screening Data for Children </a> </td></tr></table></td>
<!-- <td> Children screening data for autism suitable for classification and predictive tasks   </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 292 </td>
<td> 21 </td>
<td> 2017 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++">Autistic Spectrum Disorder Screening Data for Adolescent </a> </td></tr></table></td>
<!-- <td> Autistic Spectrum Disorder Screening Data for Adolescent. This dataset is related to classification and predictive tasks. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 104 </td>
<td> 21 </td>
<td> 2017 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/APS+Failure+at+Scania+Trucks"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/APS+Failure+at+Scania+Trucks">APS Failure at Scania Trucks</a> </td></tr></table></td>
<!-- <td> The datasets' positive class consists of component failures for a specific component of the APS system. The negative class consists of trucks with failures for components not related to the APS. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 60000 </td>
<td> 171 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Wireless+Indoor+Localization"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Wireless+Indoor+Localization">Wireless Indoor Localization</a> </td></tr></table></td>
<!-- <td> Collected in indoor space by observing signal strengths of seven WiFi signals visible on a smartphone. The decision variable is one of the four rooms.   </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 2000 </td>
<td> 7 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/HCC+Survival"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/HCC+Survival">HCC Survival</a> </td></tr></table></td>
<!-- <td> Hepatocellular Carcinoma dataset (HCC dataset) was collected at a University Hospital in Portugal. It contains real clinical data of 165 patients diagnosed with HCC. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>

<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">165 </p></td>
<td><p class="normal">49 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Life </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td><p class="normal"><b><a href="datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015">CSM (Conventional and Social Media Movies) Dataset 2014 and 2015</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">12 features categorized as conventional and social media features. Both conventional features, collected from movies databases on Web as well as social media features(YouTube,Twitter). </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">217 </p></td>
<td><p class="normal">12 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td><p class="normal"><b><a href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29">University of Tehran Question Dataset 2016 (UTQD.2016)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Persian questions gathered from a jeopardy game broadcasted on Iranian national television. </p></td> -->
<td><p class="normal">Text </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">1175 </p></td>
<td><p class="normal">3 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Other </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Autism+Screening+Adult"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td><p class="normal"><b><a href="datasets/Autism+Screening+Adult">Autism Screening Adult</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Autistic Spectrum Disorder Screening Data for Adult. This dataset is related to classification and predictive tasks. </p></td> -->
<td><p class="normal"> </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">704 </p></td>
<td><p class="normal">21 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Social </p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td><p class="normal"><b><a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor">Activity recognition with healthy older people using a batteryless wearable sensor</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Sequential motion data from 14 healthy older people aged 66 to 86 years old using a batteryless, wearable sensor on top of their clothing for the recognition of activities in clinical environments. </p></td> -->
<td><p class="normal">Sequential </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">75128 </p></td>
<td><p class="normal">9 </p></td>
<td><p class="normal">2016 </p></td>

<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Immunotherapy+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Immunotherapy+Dataset">Immunotherapy Dataset</a> </td></tr></table></td>
<!-- <td> This dataset contains information about wart treatment results of 90 patients using immunotherapy. </td> -->
<td> Univariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 90 </td>
<td> 8 </td>
<td> 2018 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Cryotherapy+Dataset+"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Cryotherapy+Dataset+">Cryotherapy Dataset </a> </td></tr></table></td>
<!-- <td> This dataset contains information about wart treatment results of 90 patients using cryotherapy. </td> -->
<td> Univariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 90 </td>
<td> 7 </td>
<td> 2018 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes">OCT data & Color Fundus Images of Left & Right Eyes</a> </td></tr></table></td>
<!-- <td> This dataset contains OCT data (in mat format) and color fundus data (in jpg format) of left & right eyes of 50 healthy persons. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 50 </td>
<td> 2 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Discrete+Tone+Image+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Discrete+Tone+Image+Dataset">Discrete Tone Image Dataset</a> </td></tr></table></td>
<!-- <td> Discrete Tone Images(DTI)are available which needs to be analyzed in detail. Here, we created this dataset for those who do research in DTI.
  </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> </td>
<td> 71 </td>
<td> 11 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/News+Popularity+in+Multiple+Social+Media+Platforms"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/News+Popularity+in+Multiple+Social+Media+Platforms">News Popularity in Multiple Social Media Platforms</a> </td></tr></table></td>
<!-- <td> Large data set of news items and their respective social feedback on multiple platforms: Facebook, Google+ and LinkedIn. </td> -->

<td> Multivariate, Time-Series, Text </td>
<td> Regression </td>
<td> Integer, Real </td>
<td> 93239 </td>
<td> 11 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Ultrasonic+flowmeter+diagnostics"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Ultrasonic+flowmeter+diagnostics">Ultrasonic flowmeter diagnostics</a> </td></tr></table></td>
<!-- <td> Fault diagnosis of four liquid ultrasonic flowmeters </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 540 </td>
<td> 173 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/ICMLA+2014+Accepted+Papers+Data+Set"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/ICMLA+2014+Accepted+Papers+Data+Set">ICMLA 2014 Accepted Papers Data Set</a> </td></tr></table></td>
<!-- <td> This data set compromises the metadata for the 2014 ICMLA conference's accepted papers, including ID, paper titles, author's keywords, abstracts and sessions in which they were exposed. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> </td>
<td> 105 </td>
<td> 5 </td>
<td> 2018 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation">BLE RSSI Dataset for Indoor localization and Navigation</a> </td></tr></table></td>
<!-- <td> This dataset contains RSSI readings gathered from an array of Bluetooth Low Energy (BLE) iBeacons in a real-world and operational indoor environment for localization and navigation purposes. </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Classification, Clustering </td>
<td> Integer </td>
<td> 6611 </td>
<td> 15 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Container+Crane+Controller+Data+Set"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Container+Crane+Controller+Data+Set">Container Crane Controller Data Set</a> </td></tr></table></td>
<!-- <td> A container crane has the function of transporting containers from one point to another point. </td> -->
<td> Univariate, Domain-Theory </td>
<td> Classification, Regression </td>
<td> Real </td>
<td> 15 </td>
<td> 3 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">

<td><table><tr><td><a href="datasets/Residential+Building+Data+Set"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Residential+Building+Data+Set">Residential Building Data Set</a> </td></tr></table></td>
<!-- <td> Data set includes construction cost, sale prices, project variables, and economic variables corresponding to real estate single-family residential apartments in Tehran, Iran.   </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 372 </td>
<td> 105 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Health+News+in+Twitter"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Health+News+in+Twitter">Health News in Twitter</a> </td></tr></table></td>
<!-- <td> The data was collected in 2015 using Twitter API. This dataset contains health news from more than 15 major health news agencies such as BBC, CNN, and NYT.   </td> -->
<td> Text </td>
<td> Clustering </td>
<td> Real </td>
<td> 58000 </td>
<td> 25000 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/chipseq"><img border="1" src="assets/MLimages/SmallLargedefault.jpg" /></a> </td><td> <a href="datasets/chipseq">chipseq</a> </td></tr></table></td>
<!-- <td> ChIP-seq experiments characterize protein modifications or binding at
specific genomic locations in specific samples. The machine learning
problem in these data is structured binary classification. </td> -->
<td> Sequential </td>
<td> Classification </td>
<td> Integer </td>
<td> 4960 </td>
<td> </td>
<td> 2018 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/SGEMM+GPU+kernel+performance"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/SGEMM+GPU+kernel+performance">SGEMM GPU kernel performance</a> </td></tr></table></td>
<!-- <td> Running times for multiplying two 2048 x 2048 matrices using a GPU OpenCL SGEMM kernel with varying parameters (using the library 'CLTune'). </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Integer </td>
<td> 241600 </td>
<td> 18 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Repeat+Consumption+Matrices"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Repeat+Consumption+Matrices">Repeat Consumption Matrices</a> </td></tr></table></td>
<!-- <td> The dataset contains 7 datasets of User - Item matrices, where each entry represents how many times a user consumed an item. Item is used as an umbrella term for various categories. </td> -->
<td> Multivariate </td>
<td> Clustering </td>
<td> Real </td>
<td> 130000 </td>

<td> 21000 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/detection_of_IoT_botnet_attacks_N_BaIoT"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/detection_of_IoT_botnet_attacks_N_BaIoT">detection_of_IoT_botnet_attacks_N_BaIoT</a> </td></tr></table></td>
<!-- <td> This dataset addresses the lack of public botnet datasets, especially for the IoT. It suggests *real* traffic data, gathered from 9 commercial IoT devices authentically infected by Mirai and BASHLITE. </td> -->
<td> Multivariate, Sequential </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 7062606 </td>
<td> 115 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Absenteeism+at+work"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Absenteeism+at+work">Absenteeism at work</a> </td></tr></table></td>
<!-- <td> The database was created with records of absenteeism at work from July 2007 to July 2010 at a courier company in Brazil. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Clustering </td>
<td> Integer, Real </td>
<td> 740 </td>
<td> 21 </td>
<td> 2018 </td>
<!-- <td> Business </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/SCADI"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/SCADI">SCADI</a> </td></tr></table></td>
<!-- <td> First self-care activities dataset based on ICF-CY. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> </td>
<td> 70 </td>
<td> 206 </td>
<td> 2018 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Condition+monitoring+of+hydraulic+systems"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Condition+monitoring+of+hydraulic+systems">Condition monitoring of hydraulic systems</a> </td></tr></table></td>
<!-- <td> The data set addresses the condition assessment of a hydraulic test rig based on multi sensor data. Four fault types are superimposed with several severity grades impeding selective quantification. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Regression </td>
<td> Real </td>
<td> 2205 </td>
<td> 43680 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Carbon+Nanotubes"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Carbon+Nanotubes">Carbon Nanotubes</a> </td></tr></table></td>
<!-- <td> This dataset contains 10721 initial and calculated atomic coordinates of carbon nanotubes. </td> -->
<td> Univariate </td>

<td> Regression </td>
<td> Real </td>
<td> 10721 </td>
<td> 8 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Optical+Interconnection+Network+"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Optical+Interconnection+Network+">Optical Interconnection Network </a> </td></tr></table></td>
<!-- <td> This dataset contains 640 performance measurements from a simulation of 2-Dimensional Multiprocessor Optical Interconnection Network.   </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Integer, Real </td>
<td> 640 </td>
<td> 10 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Sports+articles+for+objectivity+analysis"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Sports+articles+for+objectivity+analysis">Sports articles for objectivity analysis</a> </td></tr></table></td>
<!-- <td> 1000 sports articles were labeled using Amazon Mechanical Turk as objective or subjective. The raw texts, extracted features, and the URLs from which the articles were retrieved are provided. </td> -->
<td> Multivariate, Text </td>
<td> Classification </td>
<td> Integer </td>
<td> 1000 </td>
<td> 59 </td>
<td> 2018 </td>
<!-- <td> Social </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Breast+Cancer+Coimbra"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Breast+Cancer+Coimbra">Breast Cancer Coimbra</a> </td></tr></table></td>
<!-- <td> Clinical features were observed or measured for 64 patients with breast cancer and 52 healthy controls.   </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 116 </td>
<td> 10 </td>
<td> 2018 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data">GNFUV Unmanned Surface Vehicles Sensor Data</a> </td></tr></table></td>
<!-- <td> The data-set contains four (4) sets of mobile sensor readings data (humidity, temperature) corresponding to a swarm of four (4) Unmanned Surface Vehicles (USVs) in a test-bed in Athens (Greece).   </td> -->
<td> Multivariate, Time-Series </td>
<td> Regression </td>
<td> Real </td>
<td> 1672 </td>
<td> 5 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">

<td><table><tr><td><a href="datasets/Dishonest+Internet+users+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Dishonest+Internet+users+Dataset">Dishonest Internet users Dataset</a> </td></tr></table></td>
<!-- <td> The dataset was used to test an architecture based on a trust model capable to cope with the evaluation of the trustworthiness of users interacting in pervasive environments. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> </td>
<td> 322 </td>
<td> 5 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Victorian+Era+Authorship+Attribution"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Victorian+Era+Authorship+Attribution">Victorian Era Authorship Attribution</a> </td></tr></table></td>
<!-- <td> To create the largest authorship attribution dataset, we extracted works of 50 well-known authors. To have a non-exhaustive learning, in training there are 45 authors whereas, in the testing, it's 50 </td> -->
<td> Text </td>
<td> Classification </td>
<td> </td>
<td> 93600 </td>
<td> 1000 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set">Simulated Falls and Daily Living Activities Data Set</a> </td></tr></table></td>
<!-- <td> 20 falls and 16 daily living activities were performed by 17 volunteers with 5 repetitions while wearing 6 sensors (3.060 instances) that attached to their head, chest, waist, wrist, thigh and ankle. </td> -->
<td> Time-Series </td>
<td> Classification </td>
<td> Integer </td>
<td> 3060 </td>
<td> 138 </td>
<td> 2018 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing">Multimodal Damage Identification for Humanitarian Computing</a> </td></tr></table></td>
<!-- <td> 5879 captioned images (image and text) from social media related to damage during natural disasters/wars, and belong to 6 classes: Fires, Floods, Natural landscape, Infrastructural, Human, Non-damage. </td> -->
<td> Multivariate, Text </td>
<td> Classification </td>
<td> Integer </td>
<td> 5879 </td>
<td> </td>
<td> 2018 </td>
<!-- <td> Social </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals">EEG Steady-State Visual Evoked Potential Signals</a> </td></tr></table></td>
<!-- <td> This database consists on 30 subjects performing Brain Computer Interface for Ste

<!-- <td> This database consists of 30 subjects performing Brain Computer Interface for Steady State Visual Evoked Potentials (BCI-SSVEP).   </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Regression </td>
<td> Integer </td>
<td> 9200 </td>
<td> 16 </td>
<td> 2018 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Roman+Urdu+Data+Set"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Roman+Urdu+Data+Set">Roman Urdu Data Set</a> </td></tr></table></td>
<!-- <td> Roman Urdu (the scripting style for Urdu language) is one of the limited resource languages.A data corpus comprising of more than 20000 records was collected. </td> -->
<td> Text </td>
<td> Classification </td>
<td> </td>
<td> 20000 </td>
<td> 2 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Avila"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Avila">Avila</a> </td></tr></table></td>
<!-- <td> The Avila data set has been extracted from 800 images of the 'Avila Bible', an XII century giant Latin copy of the Bible. The prediction task consists in associating each pattern to a copyist. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 20867 </td>
<td> 10 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/PANDOR"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/PANDOR">PANDOR</a> </td></tr></table></td>
<!-- <td> PANDOR is a novel and publicly available dataset for online recommendation provided by Purch (http://www.purch.com/).   </td> -->
<td> Multivariate </td>
<td> Recommendation </td>
<td> Categorical </td>
<td> </td>
<td> </td>
<td> 2018 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Drug+Review+Dataset+%28Druglib.com%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Drug+Review+Dataset+%28Druglib.com%29">Drug Review Dataset (Druglib.com)</a> </td></tr></table></td>
<!-- <td> The dataset provides patient reviews on specific drugs along with related conditions. Reviews and ratings are grouped into reports on the three aspects benefits, side effects and overall comment.   </td> -->
<td> Multivariate, Text </td>
<td> Classification, Regression, Clustering </td>
<td> Integer </td>
<td> 4143 </td>
<td> 8 </td>
<td> 2018 </td>
<!-- <td>   </td> -->
</tr><tr>

</tr><tr>
<td><table><tr><td><a href="datasets/Drug+Review+Dataset+%28Drugs.com%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Drug+Review+Dataset+%28Drugs.com%29">Drug Review Dataset (Drugs.com)</a> </td></tr></table></td>
<!-- <td> The dataset provides patient reviews on specific drugs along with related conditions and a 10 star patient rating reflecting overall patient satisfaction. </td> -->
<td> Multivariate, Text </td>
<td> Classification, Regression, Clustering </td>
<td> Integer </td>
<td> 215063 </td>
<td> 6 </td>
<td> 2018 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Physical+Unclonable+Functions"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Physical+Unclonable+Functions">Physical Unclonable Functions</a> </td></tr></table></td>
<!-- <td> The dataset is generated from Physical Unclonable Functions (PUFs) simulation, specifically XOR Arbiter PUFs. PUFs are used for authentication purposes. For more info, refer to our paper below. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 6000000 </td>
<td> 129 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Superconductivty+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Superconductivty+Data">Superconductivty Data</a> </td></tr></table></td>
<!-- <td> Two file s contain data on 21263 superconductors and their relevant features. </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 21263 </td>
<td> 81 </td>
<td> 2018 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29">WESAD (Wearable Stress and Affect Detection)</a> </td></tr></table></td>
<!-- <td> WESAD (Wearable Stress and Affect Detection) contains data of 15 subjects during a stress-affect lab study, while wearing physiological and motion sensors. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Regression </td>
<td> Real </td>
<td> 63000000 </td>
<td> 12 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2">GNFUV Unmanned Surface Vehicles Sensor Data Set 2</a> </td></tr></table></td>
<!-- <td> The data-set contains eight (2x4) data-sets of mobile sensor readings data (humidity, temperature) corresponding to a swarm of four Unmanned Surface Vehicles (USVs) in a test-bed, Athens, Greece. </td> -->
<td> Multivariate, Sequential, Time-Series </td>

<td> Multivariate, Sequential, Time-Series </td>
<td> Regression </td>
<td> Real </td>
<td> 10190 </td>
<td> 6 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Student+Academics+Performance"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Student+Academics+Performance">Student Academics Performance</a> </td></tr></table></td>
<!-- <td> The dataset tried to find the end semester percentage prediction based on different social, economic and academic attributes.   </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> </td>
<td> 300 </td>
<td> 22 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Online+Shoppers+Purchasing+Intention+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Online+Shoppers+Purchasing+Intention+Dataset">Online Shoppers Purchasing Intention Dataset</a> </td></tr></table></td>
<!-- <td> Of the 12,330 sessions in the dataset,
84.5% (10,422) were negative class samples that did not
end with shopping, and the rest (1908) were positive class
samples ending with shopping. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> Integer, Real </td>
<td> 12330 </td>
<td> 18 </td>
<td> 2018 </td>
<!-- <td> Business </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/PMU-UD"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/PMU-UD">PMU-UD</a> </td></tr></table></td>
<!-- <td> The handwritten dataset was collected from 170 participants with a total of 5,180 numeral patterns. The dataset is named Prince Mohammad Bin Fahd University - Urdu/Arabic Database (PMU-UD).   </td> -->
<td> Univariate </td>
<td> Classification </td>
<td> </td>
<td> 5180 </td>
<td> 9 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Parkinson%27s+Disease+Classification"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Parkinson%27s+Disease+Classification">Parkinson's Disease Classification</a> </td></tr></table></td>
<!-- <td> The data used in this study were gathered from 188 patients with PD (107 men and 81 women) with ages ranging from 33 to 87 (65.1±10.9). </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 756 </td>
<td> 754 </td>
<td> 2018 </td>

<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Electrical+Grid+Stability+Simulated+Data+"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Electrical+Grid+Stability+Simulated+Data+">Electrical Grid Stability Simulated Data </a> </td></tr></table></td>
<!-- <td> The local stability analysis of the 4-node star system (electricity producer is in the center) implementing Decentral Smart Grid Control concept.   </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Real </td>
<td> 10000 </td>
<td> 14 </td>
<td> 2018 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Caesarian+Section+Classification+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Caesarian+Section+Classification+Dataset">Caesarian Section Classification Dataset</a> </td></tr></table></td>
<!-- <td> This dataset contains information about caesarian section results of 80 pregnant women with the most important characteristics of delivery problems in the medical field. </td> -->
<td> Univariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 80 </td>
<td> 5 </td>
<td> 2018 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/BAUM-1"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/BAUM-1">BAUM-1</a> </td></tr></table></td>
<!-- <td> BAUM-1 dataset contains 1184 multimodal facial video clips collected from 31 subjects. The 1184 video clips contain spontaneous facial expressions and speech of 13 emotional and mental states. </td> -->
<td> Time-Series </td>
<td> Classification </td>
<td> </td>
<td> 1184 </td>
<td> </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/BAUM-2"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/BAUM-2">BAUM-2</a> </td></tr></table></td>
<!-- <td> A multilingual audio-visual affective face database consisting of 1047 video clips of 286 subjects.   </td> -->
<td> Time-Series </td>
<td> Classification </td>
<td> </td>
<td> 1047 </td>
<td> </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Audit+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Audit+Data">Audit Data</a> </td></tr></table></td>
<!-- <td> Exhaustive one year non-confidential data in the year 2015 to 2016 of firms is collected from the Auditor Office of India to build a predictor for classifying suspicious firms. </td> -->
<td> Multivariate </td>
<td> Classification </td>

<td> Real </td>
<td> 777 </td>
<td> 18 </td>
<td> 2018 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/BuddyMove+Data+Set"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/BuddyMove+Data+Set">BuddyMove Data Set</a> </td></tr></table></td>
<!-- <td> User interest information extracted from user reviews published in holidayiq.com about various types of point of interests in South India </td> -->
<td> Multivariate, Text </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 249 </td>
<td> 7 </td>
<td> 2018 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Real+estate+valuation+data+set"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Real+estate+valuation+data+set">Real estate valuation data set</a> </td></tr></table></td>
<!-- <td> The “real estate valuation” is a regression problem. The market historical data set of real estate valuation are collected from Sindian Dist., New Taipei City, Taiwan.   </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Integer, Real </td>
<td> 414 </td>
<td> 7 </td>
<td> 2018 </td>
<!-- <td> Business </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connected+speech+Data+Set+"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td>
<td> <a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connected+speech+Data+Set+">Early biomarkers of Parkinson’s disease based on natural connected speech Data Set </a> </td></tr></table></td>
<!-- <td> . </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> </td>
<td> </td>
<td> 2018 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Somerville+Happiness+Survey"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Somerville+Happiness+Survey">Somerville Happiness Survey</a> </td></tr></table></td>
<!-- <td> A data extract of a non-federal dataset posted here https://catalog.data.gov/dataset/somerville-happiness-survey-responses-2011-2013-2015 </td> -->
<td> </td>
<td> Classification </td>
<td> Integer </td>
<td> 143 </td>
<td> 7 </td>
<td> 2018 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/2.4+GHZ+Indoor+Channel+Measurements"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/2.4+GHZ+Indoor+Channel+Measurements">2.4 GHZ Indoor Channel Measurements</a> </td></tr></table></td>

<!-- <td> Measurement of the S21,consists of 10 sweeps, each sweep contains 601 frequenc y points with spacing of 0.167MHz to cover a 100MHz band centered at 2.4GHz. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 7840 </td>
<td> 5 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/EMG+data+for+gestures"><img border="1" src="assets/MLimages/Small Largedefault.jpg"/></a> </td><td> <a href="datasets/EMG+data+for+gestures">EMG dat a for gestures</a> </td></tr></table></td>
<!-- <td> These are files of raw EMG data recorded by MYO Thalmic bracelet </td > -->
<td> Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 30000 </td>
<td> 6 </td>
<td> 2019 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Parking+Birmingham"><img border="1" src="assets/MLimages/SmallLar gedefault.jpg"/></a> </td><td> <a href="datasets/Parking+Birmingham">Parking Birming ham</a> </td></tr></table></td>
<!-- <td> Data collected from car parks in Birmingham that are operated by NCP from Birmingham City Council. UK Open Government Licence (OGL).
https://data.birmingham.gov.uk/dataset/birmingham-parking </td> -->
<td> Multivariate, Univariate, Sequential, Time-Series </td>
<td> Classification, Regression, Clustering </td>
<td> Real </td>
<td> 35717 </td>
<td> 4 </td>
<td> 2019 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Behavior+of+the+urban+traffic+of+the+city+of+Sao+Paulo+in+Brazil"><i mg border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="d atasets/Behavior+of+the+urban+traffic+of+the+city+of+Sao+Paulo+in+Brazil">Behavior of the urban traffic of th e city of Sao Paulo in Brazil</a> </td></tr></table></td>
<!-- <td> The database was created with records of behavior of the urban traffic of the city of Sao Paulo in Brazil. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Regression </td>
<td> Integer, Real </td>
<td> 135 </td>
<td> 18 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Travel+Reviews"><img border="1" src="assets/MLimages/SmallLargedef ault.jpg"/></a> </td><td> <a href="datasets/Travel+Reviews">Travel Reviews</a> </td></tr></table></td>
<!-- <td> Reviews on destinations in 10 categories mentioned across East Asia. Each travele r rating is mapped as Excellent(4), Very Good(3), Average(2), Poor(1), and Terrible(0) and average rating is us ed. </td> -->
<td> Multivariate, Text </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 980 </td>
<td> 11 </td>

<td> 2018 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Tarvel+Review+Ratings"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Tarvel+Review+Ratings">Tarvel Review Ratings</a> </td></tr></table></td>
<!-- <td> Google reviews on attractions from 24 categories across Europe are considered. Google user rating ranges from 1 to 5 and average user rating per category is calculated. </td> -->
<td> Multivariate, Text </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 5456 </td>
<td> 25 </td>
<td> 2018 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Rice+Leaf+Diseases"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td> <a href="datasets/Rice+Leaf+Diseases">Rice Leaf Diseases</a> </td></tr></table></td>
<!-- <td> There are three classes/diseases: Bacterial leaf blight, Brown spot, and Leaf smut, each having 40 images. The format of all images is jpg.   </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 120 </td>
<td> </td>
<td> 2019 </td>
<!-- <td> Computer </td> -->
</tr></table>
</td>
</tr>
</table>
<table align="center" cellpadding="5"><tr valign="center">
<td> Supported By: </td>
<td><img height="60" src="assets/nsfe.gif"/> </td>
<td> In Collaboration With: </td>
<td><img src="assets/rexaSmall.jpg"/></td>
</tr></table>

<a href="about.html">About</a> ||
<a href="citation_policy.html">Citation Policy</a> ||
<a href="donation_policy.html">Donation Policy</a> ||
<a href="contact.html">Contact</a> ||
<a href="http://cml.ics.uci.edu">CML</a>

</body></html>

In [24]:

```
print(soup.prettify())
```

```
<!DOCTYPE HTML>
<html>
 <body>
  <p>
   "-//W3C//DTD HTML 4.01 Transitional//EN\"&gt;
  </p>
  <title>
   UCI Machine Learning Repository: Data Sets
  </title>
```

```
  <!-- Stylesheet link -->
  <link href="assets/ml.css" rel="stylesheet" type="text/css"/>
  <script language="JavaScript" type="text/javascript">
   <!--
function checkform ( form )
{
  // see http://www.thesitewizard.com/archive/validation.shtml
  // for an explanation of this script and how to use it on your
  // own website

  // ** START **
  if (form.q.value == "")
  {
    alert( "Please enter search terms." );
    form.q.focus();
    return false ;
  }

  if (getCheckedValue(form.sitesearch) == "ics.uci.edu" && form.q.value.indexOf("site:archive.ics.uci.edu/ml") ==
-1)
  {
    form.q.value = form.q.value + " site:archive.ics.uci.edu/ml";
  }

  // ** END **
  return true ;
}

// return the value of the radio button that is checked
// return an empty string if none are checked, or
// there are no radio buttons
function getCheckedValue(radioObj) {
 if(!radioObj)
  return "";
 var radioLength = radioObj.length;
 if(radioLength == undefined)
  if(radioObj.checked)
   return radioObj.value;
  else
   return "";
 for(var i = 0; i < radioLength; i++) {
  if(radioObj[i].checked) {
   return radioObj[i].value;
  }
 }
 return "";
}




//-->
  </script>
  <!-- SITE HEADER (INCLUDES LOGO AND SEARCH BOX) -->
  <!-- SITE HEADER (INCLUDES LOGO AND SEARCH BOX) -->
  <table bgcolor="#003366" width="100%">
   <tr>
    <td>
     <span class="normal">
      <a alt="Home" href="index.html">
       <img border="0" src="assets/logo.gif"/>
```

```
      </a>
      <br/>
      <a href="http://cml.ics.uci.edu">
       <font color="FFDD33">
        Center for Machine Learning and Intelligent Systems
       </font>
      </a>
     </span>
    </td>
    <td align="right" valign="top" width="100%">
     <span class="whitetext">
      <a href="about.html">
       About
      </a>
      <a href="citation_policy.html">
       Citation Policy
      </a>
      <a href="donation_policy.html">
       Donate a Data Set
      </a>
      <a href="contact.html">
       Contact
      </a>
     </span>
     <br/>
     <br/>
     <!-- Search Google -->
     <form action="http://www.google.com/custom" method="GET" onsubmit="return checkform(this);">
      <input maxlength="255" name="q" size="30" type="text" value=""/>
      <input name="sa" type="submit" value="Search"/>
      <input name="cof" type="hidden" value="AH:center;LH:130;L:http://archive.ics.uci.edu/assets/logo.gif;LW:38
4;AWFID:869c0b2eaa8d518e;"/>
      <input name="domains" type="hidden" value="ics.uci.edu"/>
      <br/>
      <input checked="" name="sitesearch" type="radio" value="ics.uci.edu"/>
      <span class="whitetext">
       <font size="1">
        Repository
       </font>
      </span>
      <input name="sitesearch" type="radio" value=""/>
      <span class="whitetext">
       <font size="1">
        Web
       </font>
      </span>
      <a href="http://www.google.com/search">
       <img align="middle" alt="Google" border="0" height="27" src="http://www.google.com/logos/Logo_25blk.gif"
/>
      </a>
      <br/>
     </form>
     <!-- Search Google -->
     <span class="whitetext">
      <a href="datasets.php">
       <font color="#FFDD33" size="3">
        <b>
         View ALL Data Sets
        </b>
       </font>
      </a>
     </span>
```


</td>
</tr>
</table>

<table cellpadding="3">
<tr>
<td valign="top">
<table>
<tr>
<td>

Browse Through:

</td>
</tr>
</table>
<table border="2" cellpadding="2" width="150">
<tr>
<td bgcolor="#003366">

Default Task

</td>
</tr>
<tr>
<td valign="top">

<a href="datasets.php?format=&task=cla&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table">
Classification
</a>
<font color="red">
(350)
</font>

<a href="datasets.php?format=&task=reg&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table">
Regression
</a>
<font color="red">
(96)
</font>

<a href="datasets.php?format=&task=clu&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table">
Clustering
</a>
<font color="red">
(84)
</font>

<a href="datasets.php?format=&task=other&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table">
Other
</a>
<font color="red">
(55)
</font>

</td>
</tr>
<tr>
<td bgcolor="#003366">

Attribute Type

</td>
</tr>
<tr>
<td valign="top">

<a href="datasets.php?format=&task=&att=cat&area=&numAtt=&numIns=&type=&sort=nameUp&view=table">
Categorical
</a>
<font color="red">
(38)
</font>

<a href="datasets.php?format=&task=&att=num&area=&numAtt=&numIns=&type=&sort=nameUp&view=table">
Numerical
</a>
<font color="red">
(307)
</font>

<a href="datasets.php?format=&task=&att=mix&area=&numAtt=&numIns=&type=&sort=nameUp&view=table">
Mixed
</a>
<font color="red">
(55)
</font>

</td>
</tr>
<tr>
<td bgcolor="#003366">

Data Type

</td>
</tr>
<tr>
<td valign="top">

<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=mvar&sort=nameUp&view=table">
Multivariate
</a>
<font color="red">
(357)
</font>

<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=uvar&sort=nameUp&view=table">

e=uval&sort=nameUp&view=table">
Univariate
</a>
<font color="red">
(23)
</font>

<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=seq&sort=nameUp&view=table">
Sequential
</a>
<font color="red">
(47)
</font>

<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=ts&sort=nameUp&view=table">
Time-Series
</a>
<font color="red">
(91)
</font>

<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=text&sort=nameUp&view=table">
Text
</a>
<font color="red">
(53)
</font>

<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=dt&sort=nameUp&view=table">
Domain-Theory
</a>
<font color="red">
(23)
</font>

<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=other&sort=nameUp&view=table">
Other
</a>
<font color="red">
(21)
</font>

</td>
</tr>
<tr>
<td bgcolor="#003366">

Area

</td>
</tr>
<tr>
<td valign="top">

<a href="datasets.php?format=&task=&att=&area=life&numAtt=&numIns=&t

<a href="datasets.php?format=&task=&att=&area=life&numAtt=&numIns=&type=&sort=nameUp&view=table">
Life Sciences
</a>
<font color="red">
(107)
</font>

<a href="datasets.php?format=&task=&att=&area=phys&numAtt=&numIns=&type=&sort=nameUp&view=table">
Physical Sciences
</a>
<font color="red">
(49)
</font>

<a href="datasets.php?format=&task=&att=&area=comp&numAtt=&numIns=&type=&sort=nameUp&view=table">
CS / Engineering
</a>
<font color="red">
(170)
</font>

<a href="datasets.php?format=&task=&att=&area=soc&numAtt=&numIns=&type=&sort=nameUp&view=table">
Social Sciences
</a>
<font color="red">
(26)
</font>

<a href="datasets.php?format=&task=&att=&area=bus&numAtt=&numIns=&type=&sort=nameUp&view=table">
Business
</a>
<font color="red">
(29)
</font>

<a href="datasets.php?format=&task=&att=&area=game&numAtt=&numIns=&type=&sort=nameUp&view=table">
Game
</a>
<font color="red">
(10)
</font>

<a href="datasets.php?format=&task=&att=&area=other&numAtt=&numIns=&type=&sort=nameUp&view=table">
Other
</a>
<font color="red">
(73)
</font>

</td>
</tr>
<tr>
<td bgcolor="#003366">

<b>
Attributes

```
        # Attributes
        </b>
       </p>
      </td>
     </tr>
     <tr>
      <td valign="top">
       <p class="normal">
        <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=less10&amp;numIns=&a
mp;type=&amp;sort=nameUp&amp;view=table">
         Less than 10
        </a>
        <font color="red">
         (113)
        </font>
        <br/>
        <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=10to100&amp;numIns=&
amp;type=&amp;sort=nameUp&amp;view=table">
         10 to 100
        </a>
        <font color="red">
         (210)
        </font>
        <br/>
        <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=greater100&amp;numIns
=&amp;type=&amp;sort=nameUp&amp;view=table">
         Greater than 100
        </a>
        <font color="red">
         (84)
        </font>
       </p>
      </td>
     </tr>
     <tr>
      <td bgcolor="#003366">
       <p class="whitetext">
        <b>
         # Instances
        </b>
       </p>
      </td>
     </tr>
     <tr>
      <td valign="top">
       <p class="normal">
        <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=less100&
amp;type=&amp;sort=nameUp&amp;view=table">
         Less than 100
        </a>
        <font color="red">
         (27)
        </font>
        <br/>
        <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=100to100
0&amp;type=&amp;sort=nameUp&amp;view=table">
         100 to 1000
        </a>
        <font color="red">
         (162)
        </font>
        <br/>
```

<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=greater1000&type=&sort=nameUp&view=table">
Greater than 1000
</a>
<font color="red">
(246)
</font>

</td>
</tr>
<tr>
<td bgcolor="#003366">

Format Type

</td>
</tr>
<tr>
<td valign="top">

<a href="datasets.php?format=mat&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table">
Matrix
</a>
<font color="red">
(324)
</font>

<a href="datasets.php?format=nonmat&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table">
Non-Matrix
</a>
<font color="red">
(145)
</font>

</td>
</tr>
</table>
</td>
<td valign="top">
<table width="100%">
<tr>
<td>

469

Data Sets

</td>
<td align="right">

<font color="gray">
Table View
</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list">
List View
</a>

```
      </p>
     </td>
    </tr>
   </table>
   <table border="1" cellpadding="5">
    <tr bgcolor="#003366">
     <td class="normal, whitetext">
      <p class="normal, whitetext">
       <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=&amp;sort=nameDown&amp;view=table">
        <b>
         Name
        </b>
       </a>
      </p>
     </td>
     <!-- <td><p class="normal, whitetext"><b>Abstract</b></p></td> -->
     <td>
      <p class="normal, whitetext">
       <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=&amp;sort=typeUp&amp;view=table">
        <b>
         Data Types
        </b>
       </a>
      </p>
     </td>
     <td>
      <p class="normal, whitetext">
       <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=&amp;sort=taskUp&amp;view=table">
        <b>
         Default Task
        </b>
       </a>
      </p>
     </td>
     <td>
      <p class="normal, whitetext">
       <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=&amp;sort=attTypeUp&amp;view=table">
        <b>
         Attribute Types
        </b>
       </a>
      </p>
     </td>
     <td>
      <p class="normal, whitetext">
       <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=&amp;sort=instUp&amp;view=table">
        <b>
         # Instances
        </b>
       </a>
      </p>
     </td>
     <td>
      <p class="normal, whitetext">
       <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=&amp;sort=attUp&amp;view=table">
        <b>
```

```
          # Attributes
         </b>
        </a>
       </p>
      </td>
      <td>
       <p class="normal, whitetext">
        <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=&amp;sort=dateUp&amp;view=table">
         <b>
          Year
         </b>
        </a>
       </p>
      </td>
      <!-- <td><p class="normal, whitetext"><b>Area</b></p></td> -->
     </tr>
     <tr>
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/Abalone">
           <img border="1" src="assets/MLimages/SmallLarge1.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Abalone">
             Abalone
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">Predict the age of abalone from physical measurements </p></td> -->
      <td>
       <p class="normal">
        Multivariate
       </p>
      </td>
      <td>
       <p class="normal">
        Classification
       </p>
      </td>
      <td>
       <p class="normal">
        Categorical, Integer, Real
       </p>
      </td>
      <td>
       <p class="normal">
        4177
       </p>
      </td>
      <td>
       <p class="normal">
        8
```

```
 </p>
</td>
<td>
 <p class="normal">
  1995
 </p>
</td>
<!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
 <table>
  <tr>
   <td>
    <a href="datasets/Adult">
     <img border="1" src="assets/MLimages/SmallLarge2.jpg"/>
    </a>
   </td>
   <td>
    <p class="normal">
     <b>
      <a href="datasets/Adult">
       Adult
      </a>
     </b>
    </p>
   </td>
  </tr>
 </table>
</td>
<!-- <td><p class="normal">Predict whether income exceeds $50K/yr based on census data.  Also known a
s "Census Income" dataset.&amp;nbsp;</p></td> -->
<td>
 <p class="normal">
  Multivariate
 </p>
</td>
<td>
 <p class="normal">
  Classification
 </p>
</td>
<td>
 <p class="normal">
  Categorical, Integer
 </p>
</td>
<td>
 <p class="normal">
  48842
 </p>
</td>
<td>
 <p class="normal">
  14
 </p>
</td>
<td>
 <p class="normal">
  1996
 </p>
</td>
```

<!-- <td> Social </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Annealing">

</a>
</td>
<td>

<a href="datasets/Annealing">
Annealing
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Steel annealing data </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer, Real

</td>
<td>

798

</td>
<td>

38

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Anonymous+Microsoft+Web+Data">

</a>
</td>
<td>

<a href="datasets/Anonymous+Microsoft+Web+Data">
Anonymous Microsoft Web Data
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Web data </td> -->
<td>

Multivariate

</td>
<td>

Recommender-Systems

</td>
<td>

Categorical

</td>
<td>

37711

</td>
<td>

294

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Arrhythmia">

</a>
</td>
<td>

<a href="datasets/Arrhythmia">
Arrhythmia
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Medical </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer, Real

</td>
<td>

452

</td>
<td>

279

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Artificial+Characters">

</a>
</td>
<td>

<a href="datasets/Artificial+Characters">
Artificial Characters
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Artificial </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer, Real

</td>
<td>

6000

</td>
<td>

7

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Audiology+(Standardized)">

</a>
</td>
<td>

<a href="datasets/Audiology+(Standardized)">
Audiology (Standardized)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Medical </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical

</td>
<td>

226

</td>
<td>

69

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Auto+MPG">

</a>
</td>
<td>

<a href="datasets/Auto+MPG">
Auto MPG
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Physical </td> -->
<td>

Multivariate

</td>
<td>

Regression

</td>
<td>

Categorical, Real

</td>
<td>

398

</td>
<td>

8

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Automobile">

</a>
</td>
<td>

<a href="datasets/Automobile">
Automobile
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Physical </td> -->
<td>

Multivariate

</td>
<td>

Regression

</td>
<td>

Categorical, Integer, Real

</td>
<td>

205

</td>
<td>

26

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Automobile+Data">

</a>
</td>
<td>

<a href="datasets/Automobile+Data">
Automobile Data
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Physical </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer, Real

</td>
<td>

205

</td>
<td>

26

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Bach+Chorales">

</a>
</td>
<td>

<a href="datasets/Bach+Chorales">
Bach Chorales
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Music </td> -->
<td>

Sequential

</td>
<td>

Classification

</td>
<td>

Categorical

</td>
<td>

5665

</td>
<td>

1

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Balance+Scale">

</a>
</td>
<td>

<a href="datasets/Balance+Scale">
Balance Scale
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Social </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical

</td>
<td>

625

</td>
<td>

4

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Balloons">

</a>
</td>
<td>

<a href="datasets/Balloons">
Balloons
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Artificial </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical

</td>
<td>

16

</td>
<td>

4

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Banana">

</a>
</td>
<td>

<a href="datasets/Banana">
Banana
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Artificial </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

5300

</td>
<td>

2

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Bank+Marketing">

</a>
</td>
<td>

<a href="datasets/Bank+Marketing">
Bank Marketing
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Social </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer, Real

</td>
<td>

45211

</td>
<td>

17

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Banknote+Authentication">

</a>
</td>
<td>

<a href="datasets/Banknote+Authentication">
Banknote Authentication
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Business </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

1372

</td>
<td>

4

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Bar+Crawl:+Detecting+Heavy+Drinking">

</a>
</td>
<td>

<a href="datasets/Bar+Crawl:+Detecting+Heavy+Drinking">
Bar Crawl: Detecting Heavy Drinking
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Social </td> -->
<td>

Time-Series

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

2351

</td>
<td>

23

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Beijing+Multi-Site+Air-Quality+Data">

</a>
</td>
<td>

<a href="datasets/Beijing+Multi-Site+Air-Quality+Data">
Beijing Multi-Site Air-Quality Data
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Physical </td> -->
<td>

Multivariate

</td>
<td>

Time-Series

</td>
<td>

Integer, Real

</td>
<td>

420768

</td>
<td>

18

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Bike+Sharing+Dataset">

</a>
</td>
<td>

<a href="datasets/Bike+Sharing+Dataset">
Bike Sharing Dataset
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Social </td> -->
<td>

Multivariate

</td>
<td>

Regression

</td>
<td>

Integer

</td>
<td>

17389

</td>
<td>

16

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Blood+Transfusion+Service+Center">

</a>
</td>
<td>

<a href="datasets/Blood+Transfusion+Service+Center">
Blood Transfusion Service Center
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Business </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

748

</td>
<td>

4

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Breast+Cancer">

</a>
</td>
<td>

<a href="datasets/Breast+Cancer">
Breast Cancer
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Medical </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical

</td>
<td>

286

</td>
<td>

9

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Breast+Cancer+Coimbra">

</a>
</td>
<td>

<a href="datasets/Breast+Cancer+Coimbra">
Breast Cancer Coimbra
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Medical </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer, Real

</td>
<td>

116

</td>
<td>

9

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Breast+Cancer+Wisconsin+(Diagnostic)">

</a>
</td>
<td>

<a href="datasets/Breast+Cancer+Wisconsin+(Diagnostic)">
Breast Cancer Wisconsin (Diagnostic)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Medical </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

569

</td>
<td>

30

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Breast+Cancer+Wisconsin+(Original)">

</a>
</td>
<td>

<a href="datasets/Breast+Cancer+Wisconsin+(Original)">
Breast Cancer Wisconsin (Original)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Medical </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

699

</td>
<td>

9

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Breast+Cancer+Wisconsin+(Prognostic)">

</a>
</td>
<td>

<a href="datasets/Breast+Cancer+Wisconsin+(Prognostic)">
Breast Cancer Wisconsin (Prognostic)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Medical </td> -->
<td>

Multivariate

</td>
<td>

Regression

</td>
<td>

Real

</td>
<td>

198

</td>
<td>

34

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Breast+Tissue">

</a>
</td>
<td>

<a href="datasets/Breast+Tissue">
Breast Tissue
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Medical </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

106

</td>
<td>

9

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Breast+Tissue">

</a>
</td>
<td>

<a href="datasets/Breast+Tissue">
Breast Tissue
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Medical </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

106

</td>
<td>

9

<img border="1" src="assets/MLimages/SmallLargeDefault.jpg" />
</a>
</td>
<td>

<a href="datasets/Anonymous+Microsoft+Web+Data">
Anonymous Microsoft Web Data
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Log of anonymous users of www.microsoft.com; predict areas of the web site a user visited based on data on other areas the user visited. </td> -->
<td>

</td>
<td>

Recommender-Systems

</td>
<td>

Categorical

</td>
<td>

37711

</td>
<td>

294

</td>
<td>

1998

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Arrhythmia">

</a>
</td>
<td>

<a href="datasets/Arrhythmia">
Arrhythmia
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Cardiac Arrhythmia Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

452

</td>
<td>

279

</td>
<td>

1998

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Audiology+(Standardized)">

</a>
</td>
<td>

<a href="datasets/Audiology+(Standardized)">
Audiology (Standardized)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Standardized version of audiology data set. </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

226

</td>
<td>

69

</td>
<td>

1992

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Balance+Scale">

</a>
</td>
<td>

<a href="datasets/Balance+Scale">
Balance Scale
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Balance Scale Weight & Distance Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

625

</td>
<td>

4

</td>
<td>

1994

</td>
<!-- <td> Social Science </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Balloons">

</a>
</td>
<td>

<a href="datasets/Balloons">
Balloons
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Balloons Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

16

</td>
<td>

4

</td>
<td>

1990

</td>
<!-- <td> Physics </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Bank+Marketing">

</a>
</td>
<td>

<a href="datasets/Bank+Marketing">
Bank Marketing
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe a term deposit (variable y). </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

45211

</td>
<td>

16

</td>
<td>

2012

</td>
<!-- <td> Business </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Banknote+Authentication">

</a>
</td>
<td>

<a href="datasets/Banknote+Authentication">
Banknote Authentication
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Data were extracted from images that were taken from genuine and forged banknote-like specimens. For digitization, an industrial camera usually used for print inspection was used. The final images have 400x 400 pixels. Due to the object lens and distance to the investigated object gray-scale pictures with a resolution of about 660 dpi were gained. Wavelet Transform tool were used to extract features from images. </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

1372

</td>
<td>

4

</td>
<td>

2012

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Blood+Transfusion+Service+Center">

</a>
</td>
<td>

<a href="datasets/Blood+Transfusion+Service+Center">
Blood Transfusion Service Center
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Data taken from the Blood Transfusion Service Center in Hsin-Chu City in Taiwan </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

748

</td>
<td>

4

</td>
<td>

2008

</td>
<!-- <td> Business </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Breast+Cancer+Wisconsin+(Diagnostic)">

</a>
</td>
<td>

<a href="datasets/Breast+Cancer+Wisconsin+(Diagnostic)">
Breast Cancer Wisconsin (Diagnostic)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Diagnostic Wisconsin Breast Cancer Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

569

</td>
<td>

30

</td>
<td>

1995

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Breast+Cancer+Wisconsin+(Original)">

</a>
</td>
<td>

<a href="datasets/Breast+Cancer+Wisconsin+(Original)">
Breast Cancer Wisconsin (Original)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Original Wisconsin Breast Cancer Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

699

</td>
<td>

9

</td>
<td>

1992

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Breast+Cancer+Wisconsin+(Prognostic)">

</a>
</td>
<td>

<a href="datasets/Breast+Cancer+Wisconsin+(Prognostic)">
Breast Cancer Wisconsin (Prognostic)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Prognostic Wisconsin Breast Cancer Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

198

</td>
<td>

33

</td>
<td>

1995

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Breast+Tissue">

</a>
</td>
<td>

<a href="datasets/Breast+Tissue">
Breast Tissue
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Electrical Impedance Measurements of Freshly Excised Tissue Samples from the Breast </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

106

</td>
<td>

9

</td>
<td>

2010

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Car+Evaluation">

</a>
</td>
<td>

<a href="datasets/Car+Evaluation">
Car Evaluation
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Car Evaluation Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

1728

</td>
<td>

6

</td>
<td>

1997

</td>
<!-- <td> Business </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(King-Rook+vs.+King)">

</a>
</td>
<td>

<a href="datasets/Chess+(King-Rook+vs.+King)">
Chess (King-Rook vs. King)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (King-Rook vs. King) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

28056

</td>
<td>

6

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(King-Rook+vs.+Pawn)">

</a>
</td>
<td>

<a href="datasets/Chess+(King-Rook+vs.+Pawn)">
Chess (King-Rook vs. Pawn)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (King-Rook vs. Pawn) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

3196

</td>
<td>

36

</td>
<td>

1989

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(Queen-Rook+vs.+Queen)">

</a>
</td>
<td>

<a href="datasets/Chess+(Queen-Rook+vs.+Queen)">
Chess (Queen-Rook vs. Queen)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (Queen-Rook vs. Queen) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

2000

</td>
<td>

6

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(Queen-Rook+vs.+Rook)">

</a>
</td>
<td>

<a href="datasets/Chess+(Queen-Rook+vs.+Rook)">
Chess (Queen-Rook vs. Rook)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (Queen-Rook vs. Rook) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

2000

</td>
<td>

6

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(Queen-Rook+vs.+Two+Knights)">

</a>
</td>
<td>

<a href="datasets/Chess+(Queen-Rook+vs.+Two+Knights)">
Chess (Queen-Rook vs. Two Knights)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (Queen-Rook vs. Two Knights) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

2000

</td>
<td>

6

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(Queen-Rook+vs.+Two+Pawns)">

</a>
</td>
<td>

<a href="datasets/Chess+(Queen-Rook+vs.+Two+Pawns)">
Chess (Queen-Rook vs. Two Pawns)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (Queen-Rook vs. Two Pawns) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

3196

</td>
<td>

36

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks)">

</a>
</td>
<td>

<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks)">
Chess (Queen-Rook vs. Two Rooks)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (Queen-Rook vs. Two Rooks) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

2000

</td>
<td>

6

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Reduced))">

</a>
</td>
<td>

<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Reduced))">
Chess (Queen-Rook vs. Two Rooks (Reduced))
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (Queen-Rook vs. Two Rooks (Reduced)) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

2000

</td>
<td>

6

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Reduced+Noisy))">

</a>
</td>
<td>

<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Reduced+Noisy))">
Chess (Queen-Rook vs. Two Rooks (Reduced Noisy))
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (Queen-Rook vs. Two Rooks (Reduced Noisy)) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

2000

</td>
<td>

6

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Reduced+Noisy+Imbalanced))">

</a>
</td>
<td>

<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Reduced+Noisy+Imbalanced))">
Chess (Queen-Rook vs. Two Rooks (Reduced Noisy Imbalanced))
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (Queen-Rook vs. Two Rooks (Reduced Noisy Imbalanced)) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

2000

</td>
<td>

6

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Reduced+Noisy+Imbalanced+Scaled))">

</a>
</td>
<td>

<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Reduced+Noisy+Imbalanced+Scaled))">
Chess (Queen-Rook vs. Two Rooks (Reduced Noisy Imbalanced Scaled))
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (Queen-Rook vs. Two Rooks (Reduced Noisy Imbalanced Scaled)) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

2000

</td>
<td>

6

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Reduced+Noisy+Scaled))">

</a>
</td>
<td>

<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Reduced+Noisy+Scaled))">
Chess (Queen-Rook vs. Two Rooks (Reduced Noisy Scaled))
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (Queen-Rook vs. Two Rooks (Reduced Noisy Scaled)) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

2000

</td>
<td>

6

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Reduced+Scaled))">

</a>
</td>
<td>

<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Reduced+Scaled))">
Chess (Queen-Rook vs. Two Rooks (Reduced Scaled))
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (Queen-Rook vs. Two Rooks (Reduced Scaled)) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

2000

</td>
<td>

6

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Scaled))">

</a>
</td>
<td>

<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Scaled))">
Chess (Queen-Rook vs. Two Rooks (Scaled))
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (Queen-Rook vs. Two Rooks (Scaled)) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

2000

</td>
<td>

6

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Scaled+Imbalanced))">

</a>
</td>
<td>

<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Scaled+Imbalanced))">
Chess (Queen-Rook vs. Two Rooks (Scaled Imbalanced))
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (Queen-Rook vs. Two Rooks (Scaled Imbalanced)) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

2000

</td>
<td>

6

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Scaled+Noisy))">

</a>
</td>
<td>

<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Scaled+Noisy))">
Chess (Queen-Rook vs. Two Rooks (Scaled Noisy))
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (Queen-Rook vs. Two Rooks (Scaled Noisy)) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

2000

</td>
<td>

6

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Scaled+Noisy+Imbalanced))">

</a>
</td>
<td>

<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Scaled+Noisy+Imbalanced))">
Chess (Queen-Rook vs. Two Rooks (Scaled Noisy Imbalanced))
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (Queen-Rook vs. Two Rooks (Scaled Noisy Imbalanced)) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

2000

</td>
<td>

6

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Scaled+Noisy+Imbalanced+Scaled))">

</a>
</td>
<td>

<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Scaled+Noisy+Imbalanced+Scaled))">
Chess (Queen-Rook vs. Two Rooks (Scaled Noisy Imbalanced Scaled))
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (Queen-Rook vs. Two Rooks (Scaled Noisy Imbalanced Scaled)) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

2000

</td>
<td>

6

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Scaled+Noisy+Scaled))">

</a>
</td>
<td>

<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Scaled+Noisy+Scaled))">
Chess (Queen-Rook vs. Two Rooks (Scaled Noisy Scaled))
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (Queen-Rook vs. Two Rooks (Scaled Noisy Scaled)) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

2000

</td>
<td>

6

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Scaled+Scaled))">

</a>
</td>
<td>

<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Scaled+Scaled))">
Chess (Queen-Rook vs. Two Rooks (Scaled Scaled))
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (Queen-Rook vs. Two Rooks (Scaled Scaled)) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

2000

</td>
<td>

6

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Scaled+Scaled+Imbalanced))">

</a>
</td>
<td>

<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Scaled+Scaled+Imbalanced))">
Chess (Queen-Rook vs. Two Rooks (Scaled Scaled Imbalanced))
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (Queen-Rook vs. Two Rooks (Scaled Scaled Imbalanced)) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

2000

</td>
<td>

6

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Scaled+Scaled+Noisy))">

</a>
</td>
<td>

<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Scaled+Scaled+Noisy))">
Chess (Queen-Rook vs. Two Rooks (Scaled Scaled Noisy))
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (Queen-Rook vs. Two Rooks (Scaled Scaled Noisy)) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

2000

</td>
<td>

6

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Scaled+Scaled+Noisy+Imbalanced))">

</a>
</td>
<td>

<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Scaled+Scaled+Noisy+Imbalanced))">
Chess (Queen-Rook vs. Two Rooks (Scaled Scaled Noisy Imbalanced))
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (Queen-Rook vs. Two Rooks (Scaled Scaled Noisy Imbalanced)) Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Multivariate

</td>
<td>

2000

</td>
<td>

6

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Scaled+Scaled+Noisy+Imbalanced+Scaled))">

</a>
</td>
<td>

<a href="datasets/Chess+(Queen-Rook+vs.+Two+Rooks+(Scaled+Scaled+Noisy+Imbalanced+Scaled))">
Chess (Queen-Rook vs. Two Rooks (Scaled Scaled Noisy Imbalanced Scaled))
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess (Queen-Rook vs. Two Rooks (Sc

```
        </a>
       </b>
      </p>
     </td>
    </tr>
   </table>
  </td>
  <!-- <td><p class="normal">Distinguish between the presence and absence of cardiac arrhythmia and clas
sify it in one of the 16 groups.&amp;nbsp;</p></td> -->
  <td>
   <p class="normal">
    Multivariate
   </p>
  </td>
  <td>
   <p class="normal">
    Classification
   </p>
  </td>
  <td>
   <p class="normal">
    Categorical, Integer, Real
   </p>
  </td>
  <td>
   <p class="normal">
    452
   </p>
  </td>
  <td>
   <p class="normal">
    279
   </p>
  </td>
  <td>
   <p class="normal">
    1998
   </p>
  </td>
  <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
 </tr>
 <tr bgcolor="DDEEFF">
  <td>
   <table>
    <tr>
     <td>
      <a href="datasets/Artificial+Characters">
       <img border="1" src="assets/MLimages/SmallLarge6.jpg"/>
      </a>
     </td>
     <td>
      <p class="normal">
       <b>
        <a href="datasets/Artificial+Characters">
         Artificial Characters
        </a>
       </b>
      </p>
     </td>
    </tr>
   </table>
  </td>
  <!-- <td><p class="normal">Dataset artificially generated by using first-order theory which describes struct
```

<!-- <td> Dataset artificially generated by using first order theory which describes structu
re of ten capital letters of English alphabet </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer, Real

</td>
<td>

6000

</td>
<td>

7

</td>
<td>

1992

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Audiology+%28Original%29">

</a>
</td>
<td>

<a href="datasets/Audiology+%28Original%29">
Audiology (Original)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Nominal audiology dataset from Baylor </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical

</td>
<td>

226

</td>
<td>

</td>
<td>

1987

</td>
<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Audiology+%28Standardized%29">
<img border="1" src="assets/MLimages/SmallLarge7.jpg"/>
</a>
</td>
<td>

<a href="datasets/Audiology+%28Standardized%29">
Audiology (Standardized)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Standardized version of the original audiology database </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical

</td>

<td>

226

</td>
<td>

69

</td>
<td>

1992

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Auto+MPG">

</a>
</td>
<td>

<a href="datasets/Auto+MPG">
Auto MPG
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Revised from CMU StatLib library, data concerns city-cycle fuel consumption&n
bsp; </td> -->
<td>

Multivariate

</td>
<td>

Regression

</td>
<td>

Categorical, Real

</td>
<td>

398

</td>
<td>

```
  8
 </p>
</td>
<td>
 <p class="normal">
  1993
 </p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
 <table>
  <tr>
   <td>
    <a href="datasets/Automobile">
     <img border="1" src="assets/MLimages/SmallLarge10.jpg"/>
    </a>
   </td>
   <td>
    <p class="normal">
     <b>
      <a href="datasets/Automobile">
       Automobile
      </a>
     </b>
    </p>
   </td>
  </tr>
 </table>
</td>
<!-- <td><p class="normal">From 1985 Ward's Automotive Yearbook </p></td> -->
<td>
 <p class="normal">
  Multivariate
 </p>
</td>
<td>
 <p class="normal">
  Regression
 </p>
</td>
<td>
 <p class="normal">
  Categorical, Integer, Real
 </p>
</td>
<td>
 <p class="normal">
  205
 </p>
</td>
<td>
 <p class="normal">
  26
 </p>
</td>
<td>
 <p class="normal">
  1987
 </p>
</td>
```

```
 <!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Badges">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Badges">
        Badges
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Badges labeled with a "+" or "-" as a function of a person's name </p></t
d> -->
 <td>
  <p class="normal">
   Univariate, Text
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
  </p>
 </td>
 <td>
  <p class="normal">
   294
  </p>
 </td>
 <td>
  <p class="normal">
   1
  </p>
 </td>
 <td>
  <p class="normal">
   1994
  </p>
 </td>
 <!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Balance+Scale">
```

```
      <img border="1" src="assets/MLimages/SmallLarge12.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Balance+Scale">
        Balance Scale
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Balance scale weight & distance database </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Categorical
  </p>
 </td>
 <td>
  <p class="normal">
   625
  </p>
 </td>
 <td>
  <p class="normal">
   4
  </p>
 </td>
 <td>
  <p class="normal">
   1994
  </p>
 </td>
 <!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Balloons">
      <img border="1" src="assets/MLimages/SmallLarge13.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Balloons">
        Balloons
```

Balloons
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Data previously used in cognitive psychology experiment; 4 data sets represent different conditions of an experiment </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical

</td>
<td>

16

</td>
<td>

4

</td>
<td>

</td>
<!-- <td> Social </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Breast+Cancer">

</a>
</td>
<td>

<a href="datasets/Breast+Cancer">
Breast Cancer
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Breast Cancer Data (Restricted Access) </td> -->

<!-- <td> Breast Cancer Data (Restricted Access) </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical

</td>
<td>

286

</td>
<td>

9

</td>
<td>

1988

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Breast+Cancer+Wisconsin+%28Original%29">

</a>
</td>
<td>

<a href="datasets/Breast+Cancer+Wisconsin+%28Original%29">
Breast Cancer Wisconsin (Original)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Original Wisconsin Breast Cancer Database </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>

```
 Classification
 </p>
</td>
<td>
 <p class="normal">
  Integer
 </p>
</td>
<td>
 <p class="normal">
  699
 </p>
</td>
<td>
 <p class="normal">
  10
 </p>
</td>
<td>
 <p class="normal">
  1992
 </p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
 <table>
  <tr>
   <td>
    <a href="datasets/Breast+Cancer+Wisconsin+%28Prognostic%29">
     <img border="1" src="assets/MLimages/SmallLarge14.jpg"/>
    </a>
   </td>
   <td>
    <p class="normal">
     <b>
      <a href="datasets/Breast+Cancer+Wisconsin+%28Prognostic%29">
       Breast Cancer Wisconsin (Prognostic)
      </a>
     </b>
    </p>
   </td>
  </tr>
 </table>
</td>
<!-- <td><p class="normal">Prognostic Wisconsin Breast Cancer Database </p></td> -->
<td>
 <p class="normal">
  Multivariate
 </p>
</td>
<td>
 <p class="normal">
  Classification, Regression
 </p>
</td>
<td>
 <p class="normal">
  Real
 </p>
</td>
```

```
 <td>
  <p class="normal">
   198
  </p>
 </td>
 <td>
  <p class="normal">
   34
  </p>
 </td>
 <td>
  <p class="normal">
   1995
  </p>
 </td>
 <!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29">
      <img border="1" src="assets/MLimages/SmallLarge14.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29">
        Breast Cancer Wisconsin (Diagnostic)
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Diagnostic Wisconsin Breast Cancer Database </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
 </td>
 <td>
  <p class="normal">
   569
  </p>
 </td>
 <td>
  <p class="normal">
   32
```

```
      </p>
     </td>
     <td>
      <p class="normal">
       1995
      </p>
     </td>
     <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
    </tr>
    <tr bgcolor="DDEEFF">
     <td>
      <table>
       <tr>
        <td>
         <a href="datasets/Pittsburgh+Bridges">
          <img border="1" src="assets/MLimages/SmallLarge18.jpg"/>
         </a>
        </td>
        <td>
         <p class="normal">
          <b>
           <a href="datasets/Pittsburgh+Bridges">
            Pittsburgh Bridges
           </a>
          </b>
         </p>
        </td>
       </tr>
      </table>
     </td>
     <!-- <td><p class="normal">Bridges database that has original and numeric-discretized datasets&amp;nbsp;</p
></td> -->
     <td>
      <p class="normal">
       Multivariate
      </p>
     </td>
     <td>
      <p class="normal">
       Classification
      </p>
     </td>
     <td>
      <p class="normal">
       Categorical, Integer
      </p>
     </td>
     <td>
      <p class="normal">
       108
      </p>
     </td>
     <td>
      <p class="normal">
       13
      </p>
     </td>
     <td>
      <p class="normal">
       1990
      </p>
     </td>
```

<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Car+Evaluation">

</a>
</td>
<td>

<a href="datasets/Car+Evaluation">
Car Evaluation
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Derived from simple hierarchical decision model, this database may be useful f
or testing constructive induction and structure discovery methods. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical

</td>
<td>

1728

</td>
<td>

6

</td>
<td>

1997

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

<a href="datasets/Census+Income">
<img border="1" src="assets/MLimages/SmallLarge2.jpg"/>
</a>
</td>
<td>

<a href="datasets/Census+Income">
Census Income
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Predict whether income exceeds $50K/yr based on census data. Also known a
s "Adult" dataset. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer

</td>
<td>

48842

</td>
<td>

14

</td>
<td>

1996

</td>
<!-- <td> Social </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+%28King-Rook+vs.+King-Knight%29">
<img border="1" src="assets/MLimages/SmallLarge24.jpg"/>
</a>
</td>
<td>

<a href="datasets/Chess+%28King-Rook+vs.+King-Knight%29">
Chess (King-Rook vs. King-Knight)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Knight Pin Chess End-Game Database Creator </td> -->
<td>

Multivariate, Data-Generator

</td>
<td>

Classification

</td>
<td>

Categorical, Integer

</td>
<td>

</td>
<td>

22

</td>
<td>

1988

</td>
<!-- <td> Game </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+%28King-Rook+vs.+King-Pawn%29">

</a>
</td>
<td>

<a href="datasets/Chess+%28King-Rook+vs.+King-Pawn%29">
Chess (King-Rook vs. King-Pawn)
</a>

</td>
</tr>
</table>
</td>

<!-- <td> King+Rook versus King+Pawn on a7 (usually abbreviated KRKPA7).
</td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical

</td>
<td>

3196

</td>
<td>

36

</td>
<td>

1989

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+%28King-Rook+vs.+King%29">

</a>
</td>
<td>

<a href="datasets/Chess+%28King-Rook+vs.+King%29">
Chess (King-Rook vs. King)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Chess Endgame Database for White King and Rook against Black King (KRK).
  </td> -->
<td>

Multivariate

</td>

<td>

Classification

</td>
<td>

Categorical, Integer

</td>
<td>

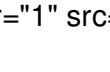
28056

</td>
<td>

6

</td>
<td>

1994

</td>
<!-- <td> Game </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Chess+%28Domain+Theories%29">

</a>
</td>
<td>

<a href="datasets/Chess+%28Domain+Theories%29">
Chess (Domain Theories)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> 6 different domain theories for generating legal moves of chess </td
> -->
<td>

Domain-Theory

</td>
<td>

</td>
<td>

</td>
<td>

</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

```
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <!-- <td><p class="normal">Game </p></td> -->
     </tr>
     <tr>
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/Bach+Chorales">
           <img border="1" src="assets/MLimages/SmallLarge25.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Bach+Chorales">
             Bach Chorales
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">Time-series data based on chorales; challenge is to learn generative grammar;
data in Lisp </p></td> -->
      <td>
       <p class="normal">
        Univariate, Time-Series
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
        Categorical, Integer
       </p>
      </td>
      <td>
       <p class="normal">
        100
       </p>
      </td>
      <td>
       <p class="normal">
        6
       </p>
       </td>
```

```
 </td>
 <td>
  <p class="normal">
  </p>
 </td>
 <!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Connect-4">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Connect-4">
        Connect-4
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Contains connect-4 positions </p></td> -->
 <td>
  <p class="normal">
   Multivariate, Spatial
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Categorical
  </p>
 </td>
 <td>
  <p class="normal">
   67557
  </p>
 </td>
 <td>
  <p class="normal">
   42
  </p>
 </td>
 <td>
  <p class="normal">
   1995
  </p>
 </td>
 <!-- <td><p class="normal">Game </p></td> -->
</tr>
<tr>
```

```
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/Credit+Approval">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Credit+Approval">
             Credit Approval
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">This data concerns credit card applications; good mix of attributes </p></
td> -->
      <td>
       <p class="normal">
        Multivariate
       </p>
      </td>
      <td>
       <p class="normal">
        Classification
       </p>
      </td>
      <td>
       <p class="normal">
        Categorical, Integer, Real
       </p>
      </td>
      <td>
       <p class="normal">
        690
       </p>
      </td>
      <td>
       <p class="normal">
        15
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <!-- <td><p class="normal">Financial </p></td> -->
     </tr>
     <tr bgcolor="DDEEFF">
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/Japanese+Credit+Screening">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
```

```
          <td>
           <p class="normal">
            <b>
             <a href="datasets/Japanese+Credit+Screening">
              Japanese Credit Screening
             </a>
            </b>
           </p>
          </td>
         </tr>
        </table>
       </td>
       <!-- <td><p class="normal">Includes domain theory (generated by talking to Japanese domain experts); da
ta in Lisp&amp;nbsp;</p></td> -->
       <td>
        <p class="normal">
         Multivariate, Domain-Theory
        </p>
       </td>
       <td>
        <p class="normal">
         Classification
        </p>
       </td>
       <td>
        <p class="normal">
         Categorical, Real, Integer
        </p>
       </td>
       <td>
        <p class="normal">
         125
        </p>
       </td>
       <td>
        <p class="normal">
        </p>
       </td>
       <td>
        <p class="normal">
         1992
        </p>
       </td>
       <!-- <td><p class="normal">Financial&amp;nbsp;</p></td> -->
      </tr>
      <tr>
       <td>
        <table>
         <tr>
          <td>
           <a href="datasets/Computer+Hardware">
            <img border="1" src="assets/MLimages/SmallLarge29.jpg"/>
           </a>
          </td>
          <td>
           <p class="normal">
            <b>
             <a href="datasets/Computer+Hardware">
              Computer Hardware
             </a>
            </b>
           </p>
```

```
        </td>
       </tr>
      </table>
     </td>
     <!-- <td><p class="normal">Relative CPU Performance Data, described in terms of its cycle time, memory s
ize, etc.&amp;nbsp;</p></td> -->
     <td>
      <p class="normal">
       Multivariate
      </p>
     </td>
     <td>
      <p class="normal">
       Regression
      </p>
     </td>
     <td>
      <p class="normal">
       Integer
      </p>
     </td>
     <td>
      <p class="normal">
       209
      </p>
     </td>
     <td>
      <p class="normal">
       9
      </p>
     </td>
     <td>
      <p class="normal">
       1987
      </p>
     </td>
     <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
    </tr>
    <tr bgcolor="DDEEFF">
     <td>
      <table>
       <tr>
        <td>
         <a href="datasets/Contraceptive+Method+Choice">
          <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
         </a>
        </td>
        <td>
         <p class="normal">
          <b>
           <a href="datasets/Contraceptive+Method+Choice">
            Contraceptive Method Choice
           </a>
          </b>
         </p>
        </td>
       </tr>
      </table>
     </td>
     <!-- <td><p class="normal">Dataset is a subset of the 1987 National Indonesia Contraceptive Prevalence S
urvey.&amp;nbsp;</p></td> -->
     <td>
```

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer

</td>
<td>

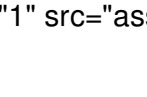
1473

</td>
<td>

9

</td>
<td>

1997

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Covertype">

</a>
</td>
<td>

<a href="datasets/Covertype">
Covertype
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Forest CoverType dataset </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer

</td>
<td>

581012

</td>
<td>

54

</td>
<td>

1998

</td>
<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Cylinder+Bands">

</a>
</td>
<td>

<a href="datasets/Cylinder+Bands">
Cylinder Bands
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Used in decision tree induction for mitigating process delays known as "cylinder bands" in rotogravure printing </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer, Real

</td>
<td>

512

</td>
<td>

39

</td>
<td>

1995

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Dermatology">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Dermatology">
Dermatology
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Aim for this dataset is to determine the type of Eryhemato-Squamous Disease.
  </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer

</td>
<td>

366

</td>
<td>

33

</td>
</tr>
</table>
</body>
</html>

```
  </p>
 </td>
 <td>
  <p class="normal">
   1998
  </p>
 </td>
 <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Diabetes">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Diabetes">
        Diabetes
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">This diabetes dataset is from AIM '94&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate, Time-Series
  </p>
 </td>
 <td>
  <p class="normal">
  </p>
 </td>
 <td>
  <p class="normal">
   Categorical, Integer
  </p>
 </td>
 <td>
  <p class="normal">
  </p>
 </td>
 <td>
  <p class="normal">
   20
  </p>
 </td>
 <td>
  <p class="normal">
  </p>
 </td>
 <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
</tr>
<tr>
 <td>
```

<td>
<table>
<tr>
<td>
<a href="datasets/DGP2+-+The+Second+Data+Generation+Program">

</a>
</td>
<td>

<a href="datasets/DGP2+-+The+Second+Data+Generation+Program">
DGP2 - The Second Data Generation Program
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Generates application domains based on specific parameters, number of featu
res, and proportion of positive to negative examples </td> -->
<td>

Data-Generator

</td>
<td>

</td>
<td>

Real

</td>
<td>

</td>
<td>

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Document+Understanding">

</a>
</td>
<td>

<a href="datasets/Document+Understanding">
Document Understanding
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Five concepts, expressed as predicates, to be learned </td> -->
<td>

</td>
<td>

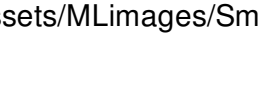
</td>
<td>

</td>
<td>

</td>
<td>

</td>
<td>

1994

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/EBL+Domain+Theories">

</a>
</td>
<td>

<a href="datasets/EBL+Domain+Theories">
EBL Domain Theories
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Assorted small-scale domain theories </td> -->
<td>

```
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <!-- <td><p class="normal">Computer </p></td> -->
     </tr>
     <tr bgcolor="DDEEFF">
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/Echocardiogram">
           <img border="1" src="assets/MLimages/SmallLarge38.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Echocardiogram">
             Echocardiogram
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">Data for classifying if patients will survive for at least one year after a heart atta
ck </p></td> -->
      <td>
       <p class="normal">
        Multivariate
       </p>
      </td>
      <td>
       <p class="normal">
        Classification
       </p>
      </td>
      <td>
       <p class="normal">
        Categorical, Integer, Real
       </p>
      </td>
      <td>
```

132

</td>
<td>

12

</td>
<td>

1989

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Ecoli">

</a>
</td>
<td>

<a href="datasets/Ecoli">
Ecoli
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This data contains protein localization sites </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

336

</td>
<td>

8

```
 </td>
 <td>
  <p class="normal">
   1996
  </p>
 </td>
 <!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Flags">
      <img border="1" src="assets/MLimages/SmallLarge40.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Flags">
        Flags
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">From Collins Gem Guide to Flags, 1986 </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Categorical, Integer
  </p>
 </td>
 <td>
  <p class="normal">
   194
  </p>
 </td>
 <td>
  <p class="normal">
   30
  </p>
 </td>
 <td>
  <p class="normal">
   1990
  </p>
 </td>
 <!-- <td><p class="normal">Other </p></td> -->
</tr>
```

```
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Function+Finding">
      <img border="1" src="assets/MLimages/SmallLarge41.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Function+Finding">
        Function Finding
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Cases collected mostly from investigations in physical science; intention is to ev
aluate function-finding algorithms&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
  </p>
 </td>
 <td>
  <p class="normal">
   Function-Learning
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
 </td>
 <td>
  <p class="normal">
   352
  </p>
 </td>
 <td>
  <p class="normal">
  </p>
 </td>
 <td>
  <p class="normal">
   1990
  </p>
 </td>
 <!-- <td><p class="normal">Physical&amp;nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Glass+Identification">
      <img border="1" src="assets/MLimages/SmallLarge42.jpg"/>
     </a>
    </td>
```

<td>

<a href="datasets/Glass+Identification">
Glass Identification
</a>

</td>
</tr>
</table>
</td>
<!-- <td> From USA Forensic Science Service; 6 types of glass; defined in terms of their oxide content (i.e. Na, Fe, K, etc) </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

214

</td>
<td>

10

</td>
<td>

1987

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Haberman%27s+Survival">

</a>
</td>
<td>

<a href="datasets/Haberman%27s+Survival">
Haberman's Survival
</a>

</td>
</tr>
</table>
</td>
<!-- <td> From University of Chicago's Billings Hospital on the survival of patients who had undergone surgery for breast cancer </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

306

</td>
<td>

3

</td>
<td>

1999

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Heart+Disease">

</a>
</td>
<td>

<a href="datasets/Heart+Disease">
Heart Disease
</a>

</td>
</tr>
</table>
</td>
<!-- <td> 4 databases: Cleveland, Hungarian, Switzerland, and the VA Long Beach </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer, Real

</td>
<td>

303

</td>
<td>

75

</td>
<td>

1988

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Hepatitis">

</a>
</td>
<td>

<a href="datasets/Hepatitis">
Hepatitis
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Hepatitis domain </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer, Real

</td>
<td>

155

</td>
<td>

19

</td>
<td>

1988

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Horse+Colic">

</a>
</td>
<td>

<a href="datasets/Horse+Colic">
Horse Colic
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Surgical lesions of horses </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer, Real

</td>
<td>

368

</td>
<td>

27

</td>
<td>

1989

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Housing">

</a>
</td>
<td>

<a href="datasets/Housing">
Housing
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Housing data for 506 census tracts of Boston from the 1970 census </td> -->
<td>

Multivariate

</td>
<td>

Regression

</td>
<td>

Categorical, Integer, Real

</td>
<td>

506

</td>
<td>

14

</td>
<td>

1993

</td>
<!-- <td> Social </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Image+Segmentation">

</a>
</td>
<td>

<a href="datasets/Image+Segmentation">
Image Segmentation
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Instance Based Image Segmentation </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer, Real

</td>
<td>

2310

</td>
<td>

19

</td>
<td>

1990

</td>
<!-- <td> Computer Science </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Ionosphere">

</a>
</td>
<td>

<a href="datasets/Ionosphere">
Ionosphere
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Classification of radar returns from the ionosphere </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer, Real

</td>
<td>

351

</td>
<td>

34

</td>
<td>

1989

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Iris">

</a>
</td>
<td>

<a href="datasets/Iris">
Iris
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Famous database; from Fisher, 1936 </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

150

</td>
<td>

4

</td>
<td>

1988

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Isolet">

</a>
</td>
<td>

<a href="datasets/Isolet">
ISOLET
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Isolated Letter Speech Recognition </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

7797

</td>
<td>

617

</td>
<td>

1994

</td>
<!-- <td> Computer Science </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Letter+Recognition">

</a>
</td>
<td>

<a href="datasets/Letter+Recognition">
Letter Recognition
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Letter Image Recognition Data </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

20000

</td>
<td>

16

</td>
<td>

1991

</td>
<!-- <td> Computer Science </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Liver+Disorders">

</a>
</td>
<td>

<a href="datasets/Liver+Disorders">
Liver Disorders
</a>

</td>
</tr>
</table>
</td>
<!-- <td> BUPA liver disorders </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer, Real

</td>
<td>

345

</td>
<td>

6

</td>
<td>

1990

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Low+Resolution+Spectrometer">

</a>
</td>
<td>

<a href="datasets/Low+Resolution+Spectrometer">
Low Resolution Spectrometer
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Low Resolution Spectrometer </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer, Real

</td>
<td>

531

</td>
<td>

101

</td>
<td>

1991

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/MAGIC+Gamma+Telescope">

</a>
</td>
<td>

<a href="datasets/MAGIC+Gamma+Telescope">
MAGIC Gamma Telescope
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Simulated registration of high energy gamma particles in a ground-based atmospheric Cherenkov gamma telescope using imaging technique </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

19020

</td>
<td>

11

</td>
<td>

2007

```
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Dataset contains cases from study conducted on the survival of patients who h
ad undergone surgery for breast cancer </p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Integer
</p>
</td>
<td>
<p class="normal">
306
</p>
</td>
<td>
<p class="normal">
3
</p>
</td>
<td>
<p class="normal">
1999
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Hayes-Roth">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Hayes-Roth">
Hayes-Roth
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Topic: human subjects study </p></td> -->
<td>
```

<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical

</td>
<td>

160

</td>
<td>

5

</td>
<td>

1989

</td>
<!-- <td> Social </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Heart+Disease">

</a>
</td>
<td>

<a href="datasets/Heart+Disease">
Heart Disease
</a>

</td>
</tr>
</table>
</td>
<!-- <td> 4 databases: Cleveland, Hungary, Switzerland, and the VA Long Beach
</td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer, Real

</td>
<td>

303

</td>
<td>

75

</td>
<td>

1988

</td>
<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Hepatitis">

</a>
</td>
<td>

<a href="datasets/Hepatitis">
Hepatitis
</a>

</td>
</tr>
</table>
</td>
<!-- <td> From G.Gong: CMU; Mostly Boolean or numeric-valued attribute types; Include
s cost data (donated by Peter Turney) </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer, Real

</td>

<td>
</td>
<td>

155

</td>
<td>

19

</td>
<td>

1988

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Horse+Colic">
<img border="1" src="assets/MLimages/SmallLarge47.jpg"/>
</a>
</td>
<td>

<a href="datasets/Horse+Colic">
Horse Colic
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Well documented attributes; 368 instances with 28 attributes (continuous, discr
ete, and nominal); 30% missing values </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer, Real

</td>
<td>

368

</td>
<td>

27

</td>
<td>

1989

</td>
<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/ICU">

</a>
</td>
<td>

<a href="datasets/ICU">
ICU
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Data set prepared for the use of participants for the 1994 AAAI Spring Symposi
um on Artificial Intelligence in Medicine. </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

</td>
<td>

Real

</td>
<td>

</td>
<td>

</td>
<td>

</td>
<!-- <td> Life </td> -->
</tr>

```
      <tr>
       <td>
        <table>
         <tr>
          <td>
           <a href="datasets/Image+Segmentation">
            <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
           </a>
          </td>
          <td>
           <p class="normal">
            <b>
             <a href="datasets/Image+Segmentation">
              Image Segmentation
             </a>
            </b>
           </p>
          </td>
         </tr>
        </table>
       </td>
       <!-- <td><p class="normal">Image data described by high-level numeric-valued attributes, 7 classes 
</p></td> -->
       <td>
        <p class="normal">
         Multivariate
        </p>
       </td>
       <td>
        <p class="normal">
         Classification
        </p>
       </td>
       <td>
        <p class="normal">
         Real
        </p>
       </td>
       <td>
        <p class="normal">
         2310
        </p>
       </td>
       <td>
        <p class="normal">
         19
        </p>
       </td>
       <td>
        <p class="normal">
         1990
        </p>
       </td>
       <!-- <td><p class="normal">Other </p></td> -->
      </tr>
      <tr bgcolor="DDEEFF">
       <td>
        <table>
         <tr>
          <td>
           <a href="datasets/Internet+Advertisements">
            <img border="1" src="assets/MLimages/SmallLarge51.jpg"/>
```

```
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Internet+Advertisements">
             Internet Advertisements
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">This dataset represents a set of possible advertisements on Internet pages.&n
bsp;</p></td> -->
      <td>
       <p class="normal">
        Multivariate
       </p>
      </td>
      <td>
       <p class="normal">
        Classification
       </p>
      </td>
      <td>
       <p class="normal">
        Categorical, Integer, Real
       </p>
      </td>
      <td>
       <p class="normal">
        3279
       </p>
      </td>
      <td>
       <p class="normal">
        1558
       </p>
      </td>
      <td>
       <p class="normal">
        1998
       </p>
      </td>
      <!-- <td><p class="normal">Computer </p></td> -->
     </tr>
     <tr>
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/Ionosphere">
           <img border="1" src="assets/MLimages/SmallLarge52.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Ionosphere">
             Ionosphere
```

```
      </a>
     </b>
    </p>
   </td>
  </tr>
 </table>
</td>
<!-- <td><p class="normal">Classification of radar returns from the ionosphere </p></td> -->
<td>
 <p class="normal">
  Multivariate
 </p>
</td>
<td>
 <p class="normal">
  Classification
 </p>
</td>
<td>
 <p class="normal">
  Integer, Real
 </p>
</td>
<td>
 <p class="normal">
  351
 </p>
</td>
<td>
 <p class="normal">
  34
 </p>
</td>
<td>
 <p class="normal">
  1989
 </p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Iris">
      <img border="1" src="assets/MLimages/SmallLarge53.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Iris">
        Iris
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Famous database; from Fisher, 1936 </p></td> -->
```

<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

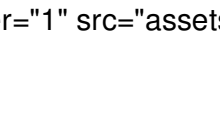
150

</td>
<td>

4

</td>
<td>

1988

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/ISOLET">

</a>
</td>
<td>

<a href="datasets/ISOLET">
ISOLET
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Goal: Predict which letter-name was spoken--a simple classification task.&nbs
p; </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

7797

</td>
<td>

617

</td>
<td>

1994

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Kinship">

</a>
</td>
<td>

<a href="datasets/Kinship">
Kinship
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Relational dataset </td> -->
<td>

Relational

</td>
<td>

Relational-Learning

</td>
<td>

Categorical

</td>

```
 </td>
 <td>
  <p class="normal">
   104
  </p>
 </td>
 <td>
  <p class="normal">
   12
  </p>
 </td>
 <td>
  <p class="normal">
   1990
  </p>
 </td>
 <!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Labor+Relations">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Labor+Relations">
        Labor Relations
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">From Collective Bargaining Review </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
  </p>
 </td>
 <td>
  <p class="normal">
   Categorical, Integer, Real
  </p>
 </td>
 <td>
  <p class="normal">
   57
  </p>
 </td>
 <td>
  <p class="normal">
   16
```

```
 </p>
</td>
<td>
 <p class="normal">
  1988
 </p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
 <table>
  <tr>
   <td>
    <a href="datasets/LED+Display+Domain">
     <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   </td>
   <td>
    <p class="normal">
     <b>
      <a href="datasets/LED+Display+Domain">
       LED Display Domain
      </a>
     </b>
    </p>
   </td>
  </tr>
 </table>
</td>
<!-- <td><p class="normal">From Classification and Regression Trees book; We provide here 2 C program
s for generating sample databases </p></td> -->
<td>
 <p class="normal">
  Multivariate, Data-Generator
 </p>
</td>
<td>
 <p class="normal">
  Classification
 </p>
</td>
<td>
 <p class="normal">
  Categorical
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
  7
 </p>
</td>
<td>
 <p class="normal">
  1988
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
```

```
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Lenses">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Lenses">
        Lenses
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Database for fitting contact lenses </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Categorical
  </p>
 </td>
 <td>
  <p class="normal">
   24
  </p>
 </td>
 <td>
  <p class="normal">
   4
  </p>
 </td>
 <td>
  <p class="normal">
   1990
  </p>
 </td>
 <!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Letter+Recognition">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
```

```
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Letter+Recognition">
             Letter Recognition
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">Database of character image features; try to identify the letter&amp;nbsp;</p></td>
-->
      <td>
       <p class="normal">
        Multivariate
       </p>
      </td>
      <td>
       <p class="normal">
        Classification
       </p>
      </td>
      <td>
       <p class="normal">
        Integer
       </p>
      </td>
      <td>
       <p class="normal">
        20000
       </p>
      </td>
      <td>
       <p class="normal">
        16
       </p>
      </td>
      <td>
       <p class="normal">
        1991
       </p>
      </td>
      <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
     </tr>
     <tr>
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/Liver+Disorders">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Liver+Disorders">
             Liver Disorders
```

```
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">BUPA Medical Research Ltd. database donated by Richard S. Forsyth </
p></td> -->
    <td>
     <p class="normal">
      Multivariate
     </p>
    </td>
    <td>
     <p class="normal">
     </p>
    </td>
    <td>
     <p class="normal">
      Categorical, Integer, Real
     </p>
    </td>
    <td>
     <p class="normal">
      345
     </p>
    </td>
    <td>
     <p class="normal">
      7
     </p>
    </td>
    <td>
     <p class="normal">
      1990
     </p>
    </td>
    <!-- <td><p class="normal">Life </p></td> -->
   </tr>
   <tr bgcolor="DDEEFF">
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/Logic+Theorist">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Logic+Theorist">
           Logic Theorist
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">All code for Logic Theorist </p></td> -->
```

<td>

Domain-Theory

</td>
<td>

</td>
<td>

</td>
<td>

</td>
<td>

</td>
<td>

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Lung+Cancer">

</a>
</td>
<td>

<a href="datasets/Lung+Cancer">
Lung Cancer
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Lung cancer data; no attribute definitions </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>

```
 </p>
</td>
<td>
 <p class="normal">
  32
 </p>
</td>
<td>
 <p class="normal">
  56
 </p>
</td>
<td>
 <p class="normal">
  1992
 </p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Lymphography">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Lymphography">
        Lymphography
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">This lymphography domain was obtained from the University Medical Centre, I
nstitute of Oncology, Ljubljana, Yugoslavia.  (Restricted access) </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Categorical
  </p>
 </td>
 <td>
  <p class="normal">
   148
  </p>
 </td>
```

<td>

18

</td>
<td>

1988

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Mechanical+Analysis">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Mechanical+Analysis">
Mechanical Analysis
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Fault diagnosis problem of electromechanical devices; also PUMPS DATA SET is newer version with domain theory and results </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer, Real

</td>
<td>

209

</td>
<td>

8

</td>
<td>

1990

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Meta-data">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Meta-data">
Meta-data
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Meta-Data was used in order to give advice about which classification method i
s appropriate for a particular dataset (taken from results of Statlog project). </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer, Real

</td>
<td>

528

</td>
<td>

22

</td>
<td>

1996

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>

```
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/Mobile+Robots">
           <img border="1" src="assets/MLimages/SmallLarge66.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Mobile+Robots">
             Mobile Robots
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">Learning concepts from sensor data of a mobile robot; set of data sets </
p></td> -->
      <td>
       <p class="normal">
        Domain-Theory
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
        Categorical, Integer, Real
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
        1995
       </p>
      </td>
      <!-- <td><p class="normal">Computer </p></td> -->
     </tr>
     <tr bgcolor="DDEEFF">
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29">
           <img border="1" src="assets/MLimages/SmallLarge67.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
```

```
          <b>
           <a href="datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29">
            Molecular Biology (Promoter Gene Sequences)
           </a>
          </b>
         </p>
        </td>
       </tr>
      </table>
     </td>
     <!-- <td><p class="normal">E. Coli promoter gene sequences (DNA) with partial domain theory </p>
</td> -->
     <td>
      <p class="normal">
       Sequential, Domain-Theory
      </p>
     </td>
     <td>
      <p class="normal">
       Classification
      </p>
     </td>
     <td>
      <p class="normal">
       Categorical
      </p>
     </td>
     <td>
      <p class="normal">
       106
      </p>
     </td>
     <td>
      <p class="normal">
       58
      </p>
     </td>
     <td>
      <p class="normal">
       1990
      </p>
     </td>
     <!-- <td><p class="normal">Life </p></td> -->
    </tr>
    <tr>
     <td>
      <table>
       <tr>
        <td>
         <a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%29">
          <img border="1" src="assets/MLimages/SmallLarge67.jpg"/>
         </a>
        </td>
        <td>
         <p class="normal">
          <b>
           <a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%29">
            Molecular Biology (Protein Secondary Structure)
           </a>
          </b>
         </p>
        </td>
```

```
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">From CMU connectionist bench repository; Classifies secondary structure of ce
rtain globular proteins </p></td> -->
    <td>
     <p class="normal">
      Sequential
     </p>
    </td>
    <td>
     <p class="normal">
      Classification
     </p>
    </td>
    <td>
     <p class="normal">
      Categorical
     </p>
    </td>
    <td>
     <p class="normal">
      128
     </p>
    </td>
    <td>
     <p class="normal">
     </p>
    </td>
    <td>
     <p class="normal">
     </p>
    </td>
    <td>
     <p class="normal">
     </p>
    </td>
    <!-- <td><p class="normal">Life </p></td> -->
   </tr>
   <tr bgcolor="DDEEFF">
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29">
         <img border="1" src="assets/MLimages/SmallLarge67.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29">
           Molecular Biology (Splice-junction Gene Sequences)
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">Primate splice-junction gene sequences (DNA) with associated imperfect doma
in theory </p></td> -->
    <td>
     <p class="normal">
      Sequential, Domain-Theory
     </p>
```

```
      </td>
      <td>
       <p class="normal">
        Classification
       </p>
      </td>
      <td>
       <p class="normal">
        Categorical
       </p>
      </td>
      <td>
       <p class="normal">
        3190
       </p>
      </td>
      <td>
       <p class="normal">
        61
       </p>
      </td>
      <td>
       <p class="normal">
        1992
       </p>
      </td>
      <!-- <td><p class="normal">Life </p></td> -->
     </tr>
     <tr>
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/MONK%27s+Problems">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/MONK%27s+Problems">
             MONK's Problems
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">A set of three artificial domains over the same attribute space; Used to test a wi
de range of induction algorithms </p></td> -->
      <td>
       <p class="normal">
        Multivariate
       </p>
      </td>
      <td>
       <p class="normal">
        Classification
       </p>
      </td>
      <td>
```

Categorical

</td>
<td>

432

</td>
<td>

7

</td>
<td>

1992

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Moral+Reasoner">

</a>
</td>
<td>

<a href="datasets/Moral+Reasoner">
Moral Reasoner
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Horn-clause model that qualitatively simulates moral reasoning; Theory include
s negated literals </td> -->
<td>

Domain-Theory

</td>
<td>

</td>
<td>

</td>
<td>

202

</td>

```
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
        1994
       </p>
      </td>
      <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
     </tr>
     <tr>
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/Multiple+Features">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Multiple+Features">
             Multiple Features
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">This dataset consists of features of handwritten numerals (`0'--`9') extracted fr
om a collection of Dutch utility maps&amp;nbsp;</p></td> -->
      <td>
       <p class="normal">
        Multivariate
       </p>
      </td>
      <td>
       <p class="normal">
        Classification
       </p>
      </td>
      <td>
       <p class="normal">
        Integer, Real
       </p>
      </td>
      <td>
       <p class="normal">
        2000
       </p>
      </td>
      <td>
       <p class="normal">
        649
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
```

```
    </td>
    <!-- <td><p class="normal">Computer </p></td> -->
   </tr>
   <tr bgcolor="DDEEFF">
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/Mushroom">
         <img border="1" src="assets/MLimages/SmallLarge73.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Mushroom">
           Mushroom
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">From Audobon Society Field Guide; mushrooms described in terms of physical characteristics; classification: poisonous or edible </p></td> -->
    <td>
     <p class="normal">
      Multivariate
     </p>
    </td>
    <td>
     <p class="normal">
      Classification
     </p>
    </td>
    <td>
     <p class="normal">
      Categorical
     </p>
    </td>
    <td>
     <p class="normal">
      8124
     </p>
    </td>
    <td>
     <p class="normal">
      22
     </p>
    </td>
    <td>
     <p class="normal">
      1987
     </p>
    </td>
    <!-- <td><p class="normal">Life </p></td> -->
   </tr>
   <tr>
    <td>
     <table>
```

<tr>
<td>
<a href="datasets/Musk+%28Version+1%29">

</a>
</td>
<td>

<a href="datasets/Musk+%28Version+1%29">
Musk (Version 1)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The goal is to learn to predict whether new molecules will be musks or non-mu
sks </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

476

</td>
<td>

168

</td>
<td>

1994

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Musk+%28Version+2%29">

</a>
</td>
<td>

<a href="datasets/Musk+%28Version+2%29">
Musk (Version 2)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The goal is to learn to predict whether new molecules will be musks or non-mu
sks </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

6598

</td>
<td>

168

</td>
<td>

1994

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Nursery">

</a>
</td>
<td>

<a href="datasets/Nursery">
Nursery
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

12960

</td>
<td>

8

</td>
<td>

1997

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Optical+Recognition+of+Handwritten+Digits">

</a>
</td>
<td>

<a href="datasets/Optical+Recognition+of+Handwritten+Digits">
Optical Recognition of Handwritten Digits
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

5620

</td>
<td>

64

</td>
<td>

1998

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits">

</a>
</td>
<td>

<a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits">
Pen-Based Recognition of Handwritten Digits
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

10992

</td>
<td>

16

</td>
<td>

1998

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Phishing+Websites">

</a>
</td>
<td>

<a href="datasets/Phishing+Websites">
Phishing Websites
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

11055

</td>
<td>

30

</td>
<td>

2012

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Polish+companies+bankruptcy+data">

</a>
</td>
<td>

<a href="datasets/Polish+companies+bankruptcy+data">
Polish companies bankruptcy data
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

43405

</td>
<td>

64

</td>
<td>

2016

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Polish+companies+bankruptcy+data">

</a>
</td>
<td>

<a href="datasets/Polish+companies+bankruptcy+data">
Polish companies bankruptcy data
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

10503

</td>
<td>

64

</td>
<td>

2016

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Polish+companies+bankruptcy+data">

</a>
</td>
<td>

<a href="datasets/Polish+companies+bankruptcy+data">
Polish companies bankruptcy data
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

5910

</td>
<td>

64

</td>
<td>

2016

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Polish+companies+bankruptcy+data">

</a>
</td>
<td>

<a href="datasets/Polish+companies+bankruptcy+data">
Polish companies bankruptcy data
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

2091

</td>
<td>

64

</td>
<td>

2016

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Polish+companies+bankruptcy+data">

</a>
</td>
<td>

<a href="datasets/Polish+companies+bankruptcy+data">
Polish companies bankruptcy data
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

495

</td>
<td>

64

</td>
<td>

2016

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Post-Operative+Patient">

</a>
</td>
<td>

<a href="datasets/Post-Operative+Patient">
Post-Operative Patient
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

90

</td>
<td>

8

</td>
<td>

1998

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Primary+Tumor">

</a>
</td>
<td>

<a href="datasets/Primary+Tumor">
Primary Tumor
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

339

</td>
<td>

17

</td>
<td>

1988

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Qualitative_Bankruptcy">

</a>
</td>
<td>

<a href="datasets/Qualitative_Bankruptcy">
Qualitative_Bankruptcy
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical

</td>
<td>

250

</td>
<td>

6

</td>
<td>

2008

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Quality+Assessment+of+Digital+Colposcopies">

</a>
</td>
<td>

<a href="datasets/Quality+Assessment+of+Digital+Colposcopies">
Quality Assessment of Digital Colposcopies
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

287

</td>
<td>

62

</td>
<td>

2017

</td>
</tr>
</table>
</td>
<!-- <td> Nursery Database was derived from a hierarchical decision model originally de
veloped to rank applications for nursery schools. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical

</td>
<td>

12960

</td>
<td>

8

</td>
<td>

1997

</td>
<!-- <td> Social </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Othello+Domain+Theory">

</a>
</td>
<td>

<a href="datasets/Othello+Domain+Theory">
Othello Domain Theory
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Used in research to generate features for an inductive learning system <
/p></td> -->
<td>

Domain-Theory

</td>
<td>

</td>
<td>

</td>
<td>

</td>
<td>

</td>
<td>

1991

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Page+Blocks+Classification">

</a>
</td>
<td>

<a href="datasets/Page+Blocks+Classification">
Page Blocks Classification
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The problem consists of classifying all the blocks of the page layout of a docum
ent that has been detected by a segmentation process. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

```
  Integer, Real
 </p>
</td>
<td>
 <p class="normal">
  5473
 </p>
</td>
<td>
 <p class="normal">
  10
 </p>
</td>
<td>
 <p class="normal">
  1995
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
 <table>
  <tr>
   <td>
    <a href="datasets/Optical+Recognition+of+Handwritten+Digits">
     <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   </td>
   <td>
    <p class="normal">
     <b>
      <a href="datasets/Optical+Recognition+of+Handwritten+Digits">
       Optical Recognition of Handwritten Digits
      </a>
     </b>
    </p>
   </td>
  </tr>
 </table>
</td>
<!-- <td><p class="normal">Two versions of this database available; see folder </p></td> -->
<td>
 <p class="normal">
  Multivariate
 </p>
</td>
<td>
 <p class="normal">
  Classification
 </p>
</td>
<td>
 <p class="normal">
  Integer
 </p>
</td>
<td>
 <p class="normal">
  5620
 </p>
</td>
```

<td>

64

</td>
<td>

1998

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits">

</a>
</td>
<td>

<a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits">
Pen-Based Recognition of Handwritten Digits
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Digit database of 250 samples from 44 writers </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

10992

</td>
<td>

16

</td>
<td>

1998

</td>
</tr>
</table>
</body>
</html>

```
  </p>
 </td>
 <!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Post-Operative+Patient">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Post-Operative+Patient">
        Post-Operative Patient
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Dataset of patient features </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Categorical, Integer
  </p>
 </td>
 <td>
  <p class="normal">
   90
  </p>
 </td>
 <td>
  <p class="normal">
   8
  </p>
 </td>
 <td>
  <p class="normal">
   1993
  </p>
 </td>
 <!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
```

```
    <td>
     <a href="datasets/Primary+Tumor">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Primary+Tumor">
        Primary Tumor
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">From Ljubljana Oncology Institute </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Categorical
  </p>
 </td>
 <td>
  <p class="normal">
   339
  </p>
 </td>
 <td>
  <p class="normal">
   17
  </p>
 </td>
 <td>
  <p class="normal">
   1988
  </p>
 </td>
 <!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Prodigy">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
```

<a href="datasets/Prodigy">
Prodigy
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Assorted domains like blocksworld, eightpuzzle, and schedworld. </td> -->
<td>

Domain-Theory

</td>
<td>

</td>
<td>

</td>
<td>

</td>
<td>

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Qualitative+Structure+Activity+Relationships">

</a>
</td>
<td>

<a href="datasets/Qualitative+Structure+Activity+Relationships">
Qualitative Structure Activity Relationships
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Two sets of datasets are given: pyrimidines and triazines </td> -->
<td>

Domain-Theory

</td>
<td>

</td>
<td>

</td>
<td>

</td>
<td>

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Quadruped+Mammals">

</a>
</td>
<td>

<a href="datasets/Quadruped+Mammals">
Quadruped Mammals
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The file animals.c is a data generator of structured instances representing qua
druped animals </td> -->
<td>

Multivariate, Data-Generator

</td>
<td>

Classification

</td>
<td>

Real

</td>

```
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
  72
 </p>
</td>
<td>
 <p class="normal">
  1992
 </p>
</td>
<!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
</tr>
<tr>
<td>
 <table>
  <tr>
   <td>
    <a href="datasets/Servo">
     <img border="1" src="assets/MLimages/SmallLarge87.jpg"/>
    </a>
   </td>
   <td>
    <p class="normal">
     <b>
      <a href="datasets/Servo">
       Servo
      </a>
     </b>
    </p>
   </td>
  </tr>
 </table>
</td>
<!-- <td><p class="normal">Data was from a simulation of a servo system&amp;nbsp;</p></td> -->
<td>
 <p class="normal">
  Multivariate
 </p>
</td>
<td>
 <p class="normal">
  Regression
 </p>
</td>
<td>
 <p class="normal">
  Categorical, Integer
 </p>
</td>
<td>
 <p class="normal">
  167
 </p>
</td>
<td>
 <p class="normal">
```

```
  4
 </p>
</td>
<td>
 <p class="normal">
  1993
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
 <table>
  <tr>
   <td>
    <a href="datasets/Shuttle+Landing+Control">
     <img border="1" src="assets/MLimages/SmallLarge92.jpg"/>
    </a>
   </td>
   <td>
    <p class="normal">
     <b>
      <a href="datasets/Shuttle+Landing+Control">
       Shuttle Landing Control
      </a>
     </b>
    </p>
   </td>
  </tr>
 </table>
</td>
<!-- <td><p class="normal">Tiny database; all nominal values </p></td> -->
<td>
 <p class="normal">
  Multivariate
 </p>
</td>
<td>
 <p class="normal">
  Classification
 </p>
</td>
<td>
 <p class="normal">
  Categorical
 </p>
</td>
<td>
 <p class="normal">
  15
 </p>
</td>
<td>
 <p class="normal">
  6
 </p>
</td>
<td>
 <p class="normal">
  1988
 </p>
</td>
```

<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Solar+Flare">

</a>
</td>
<td>

<a href="datasets/Solar+Flare">
Solar Flare
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Each class attribute counts the number of solar flares of a certain class that oc
cur in a 24 hour period </td> -->
<td>

Multivariate

</td>
<td>

Regression

</td>
<td>

Categorical

</td>
<td>

1389

</td>
<td>

10

</td>
<td>

1989

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

<a href="datasets/Soybean+%28Large%29">
<img border="1" src="assets/MLimages/SmallLarge90.jpg"/>
</a>
</td>
<td>

<a href="datasets/Soybean+%28Large%29">
Soybean (Large)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Michalski's famous soybean disease database </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical

</td>
<td>

307

</td>
<td>

35

</td>
<td>

1988

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Soybean+%28Small%29">
<img border="1" src="assets/MLimages/SmallLarge90.jpg"/>
</a>
</td>
<td>

<a href="datasets/Soybean+%28Small%29">

```
        Soybean (Small)
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Michalski's famous soybean disease database&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Categorical
  </p>
 </td>
 <td>
  <p class="normal">
   47
  </p>
 </td>
 <td>
  <p class="normal">
   35
  </p>
 </td>
 <td>
  <p class="normal">
   1987
  </p>
 </td>
 <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Challenger+USA+Space+Shuttle+O-Ring">
      <img border="1" src="assets/MLimages/SmallLarge92.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Challenger+USA+Space+Shuttle+O-Ring">
        Challenger USA Space Shuttle O-Ring
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
```

```
<!-- <td><p class="normal">Task: predict the number of O-rings that experience thermal distress on a flight at 31 degrees F given data on the previous 23 shuttle flights </p></td> -->
<td>
 <p class="normal">
  Multivariate
 </p>
</td>
<td>
 <p class="normal">
  Regression
 </p>
</td>
<td>
 <p class="normal">
  Integer
 </p>
</td>
<td>
 <p class="normal">
  23
 </p>
</td>
<td>
 <p class="normal">
  4
 </p>
</td>
<td>
 <p class="normal">
  1993
 </p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Low+Resolution+Spectrometer">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Low+Resolution+Spectrometer">
        Low Resolution Spectrometer
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">From IRAS data -- NASA Ames Research Center </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
```

Classification

</td>
<td>

Integer, Real

</td>
<td>

531

</td>
<td>

102

</td>
<td>

1988

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Spambase">

</a>
</td>
<td>

<a href="datasets/Spambase">
Spambase
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Classifying Email as Spam or Non-Spam </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer, Real

</td>

```
</td>
<td>
<p class="normal">
4601
</p>
</td>
<td>
<p class="normal">
57
</p>
</td>
<td>
<p class="normal">
1999
</p>
</td>
<!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/SPECT+Heart">
<img border="1" src="assets/MLimages/SmallLarge45.jpg"/>
</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/SPECT+Heart">
SPECT Heart
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Data on cardiac Single Proton Emission Computed Tomography (SPECT) imag
es. Each patient classified into two categories: normal and abnormal.&amp;nbsp;</p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Categorical
</p>
</td>
<td>
<p class="normal">
267
</p>
</td>
<td>
```

<td>

22

</td>
<td>

2001

</td>
<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/SPECTF+Heart">
<img border="1" src="assets/MLimages/SmallLarge45.jpg"/>
</a>
</td>
<td>

<a href="datasets/SPECTF+Heart">
SPECTF Heart
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Data on cardiac Single Proton Emission Computed Tomography (SPECT) imag
es. Each patient classified into two categories: normal and abnormal. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

267

</td>
<td>

44

</td>
<td>

```
   2001
  </p>
 </td>
 <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Sponge">
      <img border="1" src="assets/MLimages/SmallLarge97.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Sponge">
        Sponge
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Data on sponges; Attributes in spanish&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Clustering
  </p>
 </td>
 <td>
  <p class="normal">
   Categorical, Integer
  </p>
 </td>
 <td>
  <p class="normal">
   76
  </p>
 </td>
 <td>
  <p class="normal">
   45
  </p>
 </td>
 <td>
  <p class="normal">
  </p>
 </td>
 <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
```

<td>
<a href="datasets/Statlog+Project">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Statlog+Project">
Statlog Project
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Various Databases: Vehicle silhouttes, Landsat Sattelite, Shuttle, Australian Credit Approval, Heart Disease, Image Segmentation, German Credit </td> -->
<td>

</td>
<td>

</td>
<td>

</td>
<td>

</td>
<td>

</td>
<td>

1992

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Student+Loan+Relational">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Student+Loan+Relational">
Student Loan Relational
</a>

```
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">Student Loan Relational Domain </p></td> -->
      <td>
       <p class="normal">
        Domain-Theory
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
        1000
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
        1993
       </p>
      </td>
      <!-- <td><p class="normal">Social </p></td> -->
     </tr>
     <tr bgcolor="DDEEFF">
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/Teaching+Assistant+Evaluation">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Teaching+Assistant+Evaluation">
             Teaching Assistant Evaluation
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">The data consist of evaluations of teaching performance; scores are "low", "me
dium", or "high" </p></td> -->
      <td>
       <p class="normal">
        Multivariate
       </p>
```

```
      </td>
      <td>
       <p class="normal">
        Classification
       </p>
      </td>
      <td>
       <p class="normal">
        Categorical, Integer
       </p>
      </td>
      <td>
       <p class="normal">
        151
       </p>
      </td>
      <td>
       <p class="normal">
        5
       </p>
      </td>
      <td>
       <p class="normal">
        1997
       </p>
      </td>
      <!-- <td><p class="normal">Other </p></td> -->
     </tr>
     <tr>
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/Tic-Tac-Toe+Endgame">
           <img border="1" src="assets/MLimages/SmallLarge101.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Tic-Tac-Toe+Endgame">
             Tic-Tac-Toe Endgame
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">Binary classification task on possible configurations of tic-tac-toe game <
/p></td> -->
      <td>
       <p class="normal">
        Multivariate
       </p>
      </td>
      <td>
       <p class="normal">
        Classification
       </p>
      </td>
      <td>
```

Categorical

</td>
<td>

958

</td>
<td>

9

</td>
<td>

1991

</td>
<!-- <td> Game </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Thyroid+Disease">

</a>
</td>
<td>

<a href="datasets/Thyroid+Disease">
Thyroid Disease
</a>

</td>
</tr>
</table>
</td>
<!-- <td> 10 separate databases from Garavan Institute </td> -->
<td>

Multivariate, Domain-Theory

</td>
<td>

Classification

</td>
<td>

Categorical, Real

</td>
<td>

7200

</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

```
 </td>
 <td>
  <p class="normal">
   21
  </p>
 </td>
 <td>
  <p class="normal">
   1987
  </p>
 </td>
 <!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Trains">
      <img border="1" src="assets/MLimages/SmallLarge103.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Trains">
        Trains
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">2 data formats (structured, one-instance-per-line) </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Categorical
  </p>
 </td>
 <td>
  <p class="normal">
   10
  </p>
 </td>
 <td>
  <p class="normal">
   32
  </p>
 </td>
 <td>
  <p class="normal">
```

1994

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/University">

</a>
</td>
<td>

<a href="datasets/University">
University
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Data in original (LISP-readable) form </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer

</td>
<td>

285

</td>
<td>

17

</td>
<td>

1988

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
</table>
</td>
</tr>
</table>
</td>
</tr>
</table>
</td>

<table>
<tr>
<td>
<a href="datasets/Congressional+Voting+Records">

</a>
</td>
<td>

<a href="datasets/Congressional+Voting+Records">
Congressional Voting Records
</a>

</td>
</tr>
</table>
</td>
<!-- <td> 1984 United Stated Congressional Voting Records; Classify as Republican or D
emocrat </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical

</td>
<td>

435

</td>
<td>

16

</td>
<td>

1987

</td>
<!-- <td> Social </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Water+Treatment+Plant">

</a>
</td>
<td>

<td>

<a href="datasets/Water+Treatment+Plant">
Water Treatment Plant
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Multiple classes predict plant state </td> -->
<td>

Multivariate

</td>
<td>

Clustering

</td>
<td>

Integer, Real

</td>
<td>

527

</td>
<td>

38

</td>
<td>

1993

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Waveform+Database+Generator+%28Version+1%29">

</a>
</td>
<td>

<a href="datasets/Waveform+Database+Generator+%28Version+1%29">
Waveform Database Generator (Version 1)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Multiple classes predict plant state </td> -->
<td>

Multivariate

</td>
<td>

Clustering

</td>
<td>

Integer, Real

</td>
<td>

527

</td>
<td>

38

</td>
<td>

1993

</td>
<!-- <td> Physical </td> -->
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>
</table>
</td>

```
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">CART book's waveform domains </p></td> -->
 <td>
  <p class="normal">
   Multivariate, Data-Generator
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
 </td>
 <td>
  <p class="normal">
   5000
  </p>
 </td>
 <td>
  <p class="normal">
   21
  </p>
 </td>
 <td>
  <p class="normal">
   1988
  </p>
 </td>
 <!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Waveform+Database+Generator+%28Version+2%29">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Waveform+Database+Generator+%28Version+2%29">
        Waveform Database Generator (Version 2)
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">CART book's waveform domains </p></td> -->
 <td>
  <p class="normal">
   Multivariate, Data-Generator
```

```
 </p>
</td>
<td>
 <p class="normal">
  Classification
 </p>
</td>
<td>
 <p class="normal">
  Real
 </p>
</td>
<td>
 <p class="normal">
  5000
 </p>
</td>
<td>
 <p class="normal">
  40
 </p>
</td>
<td>
 <p class="normal">
  1988
 </p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Wine">
      <img border="1" src="assets/MLimages/SmallLarge109.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Wine">
        Wine
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Using chemical analysis determine the origin of wines </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
```

Integer, Real

</td>
<td>

178

</td>
<td>

13

</td>
<td>

1991

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Yeast">

</a>
</td>
<td>

<a href="datasets/Yeast">
Yeast
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Predicting the Cellular Localization Sites of Proteins </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

1484

```
 </td>
 <td>
  <p class="normal">
   8
  </p>
 </td>
 <td>
  <p class="normal">
   1996
  </p>
 </td>
 <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Zoo">
      <img border="1" src="assets/MLimages/SmallLarge111.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Zoo">
        Zoo
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Artificial, 7 classes of animals&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Categorical, Integer
  </p>
 </td>
 <td>
  <p class="normal">
   101
  </p>
 </td>
 <td>
  <p class="normal">
   17
  </p>
 </td>
 <td>
  <p class="normal">
```

```
      1990
     </p>
    </td>
    <!-- <td><p class="normal">Life </p></td> -->
   </tr>
   <tr bgcolor="DDEEFF">
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/Undocumented">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Undocumented">
           Undocumented
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">Various datasets without documentation (feel free to explore!) </p></td>
-->
    <td>
     <p class="normal">
     </p>
    </td>
    <td>
     <p class="normal">
     </p>
    </td>
    <td>
     <p class="normal">
     </p>
    </td>
    <td>
     <p class="normal">
     </p>
    </td>
    <td>
     <p class="normal">
     </p>
    </td>
    <td>
     <p class="normal">
     </p>
    </td>
    <td>
     <p class="normal">
     </p>
    </td>
    <!-- <td><p class="normal">Other </p></td> -->
   </tr>
   <tr>
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/Twenty+Newsgroups">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
```

```
        </td>
        <td>
         <p class="normal">
          <b>
           <a href="datasets/Twenty+Newsgroups">
            Twenty Newsgroups
           </a>
          </b>
         </p>
        </td>
       </tr>
      </table>
     </td>
     <!-- <td><p class="normal">This data set consists of 20000 messages taken from 20 newsgroups. </
p></td> -->
     <td>
      <p class="normal">
       Text
      </p>
     </td>
     <td>
      <p class="normal">
      </p>
     </td>
     <td>
      <p class="normal">
      </p>
     </td>
     <td>
      <p class="normal">
       20000
      </p>
     </td>
     <td>
      <p class="normal">
      </p>
     </td>
     <td>
      <p class="normal">
       1999
      </p>
     </td>
     <!-- <td><p class="normal">Other </p></td> -->
    </tr>
    <tr bgcolor="DDEEFF">
     <td>
      <table>
       <tr>
        <td>
         <a href="datasets/Australian+Sign+Language+signs">
          <img border="1" src="assets/MLimages/SmallLarge114.jpg"/>
         </a>
        </td>
        <td>
         <p class="normal">
          <b>
           <a href="datasets/Australian+Sign+Language+signs">
            Australian Sign Language signs
           </a>
          </b>
         </p>
        </td>
```

```
    </tr>
   </table>
  </td>
  <!-- <td><p class="normal">This data consists of sample of Auslan (Australian Sign Language) signs. Exam
ples of 95 signs were collected from five signers with a total of 6650 sign samples. </p></td> -->
  <td>
   <p class="normal">
    Multivariate, Time-Series
   </p>
  </td>
  <td>
   <p class="normal">
    Classification
   </p>
  </td>
  <td>
   <p class="normal">
    Categorical, Real
   </p>
  </td>
  <td>
   <p class="normal">
    6650
   </p>
  </td>
  <td>
   <p class="normal">
    15
   </p>
  </td>
  <td>
   <p class="normal">
    1999
   </p>
  </td>
  <!-- <td><p class="normal">Other </p></td> -->
 </tr>
 <tr>
  <td>
   <table>
    <tr>
     <td>
      <a href="datasets/Australian+Sign+Language+signs+%28High+Quality%29">
       <img border="1" src="assets/MLimages/SmallLarge114.jpg"/>
      </a>
     </td>
     <td>
      <p class="normal">
       <b>
        <a href="datasets/Australian+Sign+Language+signs+%28High+Quality%29">
         Australian Sign Language signs (High Quality)
        </a>
       </b>
      </p>
     </td>
    </tr>
   </table>
  </td>
  <!-- <td><p class="normal">This data consists of sample of Auslan (Australian Sign Language) signs. 27 ex
amples of each of 95 Auslan signs were captured from a native signer using high-quality position trackers 
;</p></td> -->
  <td>
```

<td>

Multivariate, Time-Series

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

2565

</td>
<td>

22

</td>
<td>

2002

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/US+Census+Data+%281990%29">

</a>
</td>
<td>

<a href="datasets/US+Census+Data+%281990%29">
US Census Data (1990)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The USCensus1990raw data set contains a one percent sample of the Public U
se Microdata Samples (PUMS) person records drawn from the full 1990 census sample. </td> -->
<td>

Multivariate

</td>
<td>

```
Clustering
</p>
</td>
<td>
<p class="normal">
Categorical
</p>
</td>
<td>
<p class="normal">
2458285
</p>
</td>
<td>
<p class="normal">
68
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Census-Income+%28KDD%29">
<img border="1" src="assets/MLimages/SmallLarge2.jpg"/>
</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Census-Income+%28KDD%29">
Census-Income (KDD)
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This data set contains weighted census data extracted from the 1994 and 1995 current population surveys conducted by the U.S. Census Bureau. </p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Categorical, Integer
</p>
</td>
```

<td>

299285

</td>
<td>

40

</td>
<td>

2000

</td>
<!-- <td> Social </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Coil+1999+Competition+Data">

</a>
</td>
<td>

<a href="datasets/Coil+1999+Competition+Data">
Coil 1999 Competition Data
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This data set is from the 1999 Computational Intelligence and Learning (COIL) competition. The data contains measurements of river chemical concentrations and algae densities. </td> -->
<td>

Multivariate

</td>
<td>

</td>
<td>

Categorical, Real

</td>
<td>

340

</td>
<td>

```
 17
 </p>
</td>
<td>
 <p class="normal">
  1999
 </p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Corel+Image+Features">
      <img border="1" src="assets/MLimages/SmallLarge119.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Corel+Image+Features">
        Corel Image Features
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">This dataset contains image features extracted from a Corel image collection.
Four sets of features are available based on the color histogram, color histogram layout, color moments, and c
o-occurence </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
 </td>
 <td>
  <p class="normal">
   68040
  </p>
 </td>
 <td>
  <p class="normal">
   89
  </p>
 </td>
 <td>
  <p class="normal">
   1999
  </p>
```

```
      </td>
      <!-- <td><p class="normal">Other </p></td> -->
     </tr>
     <tr bgcolor="DDEEFF">
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/E.+Coli+Genes">
           <img border="1" src="assets/MLimages/SmallLarge120.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/E.+Coli+Genes">
             E. Coli Genes
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">Data giving characteristics of each ORF (potential gene) in the E. coli genome. Sequence, homology (similarity to other genes) and structural information, and function (if known) are provided.  </p></td> -->
      <td>
       <p class="normal">
        Relational
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
        2001
       </p>
      </td>
      <!-- <td><p class="normal">Life </p></td> -->
     </tr>
     <tr>
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/EEG+Database">
           <img border="1" src="assets/MLimages/SmallLarge121.jpg"/>
```

```
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/EEG+Database">
             EEG Database
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">This data arises from a large study to examine EEG correlates of genetic predi
sposition to alcoholism. It contains measurements from 64 electrodes placed on the scalp sampled at 256 Hz&n
bsp;</p></td> -->
      <td>
       <p class="normal">
        Multivariate, Time-Series
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
        Categorical, Integer, Real
       </p>
      </td>
      <td>
       <p class="normal">
        122
       </p>
      </td>
      <td>
       <p class="normal">
        4
       </p>
      </td>
      <td>
       <p class="normal">
        1999
       </p>
      </td>
      <!-- <td><p class="normal">Life </p></td> -->
     </tr>
     <tr bgcolor="DDEEFF">
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/El+Nino">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/El+Nino">
             El Nino
```

```
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">The data set contains oceanographic and surface meteorological readings take
n from a series of buoys positioned throughout the equatorial Pacific. </p></td> -->
    <td>
     <p class="normal">
      Spatio-temporal
     </p>
    </td>
    <td>
     <p class="normal">
     </p>
    </td>
    <td>
     <p class="normal">
      Integer, Real
     </p>
    </td>
    <td>
     <p class="normal">
      178080
     </p>
    </td>
    <td>
     <p class="normal">
      12
     </p>
    </td>
    <td>
     <p class="normal">
      1999
     </p>
    </td>
    <!-- <td><p class="normal">Physical </p></td> -->
   </tr>
   <tr>
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/Entree+Chicago+Recommendation+Data">
         <img border="1" src="assets/MLimages/SmallLarge123.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Entree+Chicago+Recommendation+Data">
           Entree Chicago Recommendation Data
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">This data contains a record of user interactions with the Entree Chicago restau
```

<!-- <td> This data contains a record of user interactions with the Entree Chicago restaurant recommendation system. </td> -->
<td>

Transactional, Sequential

</td>
<td>

Recommender-Systems

</td>
<td>

Categorical

</td>
<td>

50672

</td>
<td>

</td>
<td>

2000

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/CMU+Face+Images">

</a>
</td>
<td>

<a href="datasets/CMU+Face+Images">
CMU Face Images
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This data consists of 640 black and white face images of people taken with varying pose (straight, left, right, up), expression (neutral, happy, sad, angry), eyes (wearing sunglasses or not), and size </td> -->
<td>

Image

</td>
<td>
</td>

<td>

Classification

</td>
<td>

Integer

</td>
<td>

640

</td>
<td>

</td>
<td>

1999

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Insurance+Company+Benchmark+%28COIL+2000%29">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Insurance+Company+Benchmark+%28COIL+2000%29">
Insurance Company Benchmark (COIL 2000)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This data set used in the CoIL 2000 Challenge contains information on custom
ers of an insurance company. The data consists of 86 variables and includes product usage data and socio-de
mographic data </td> -->
<td>

Multivariate

</td>
<td>

Regression, Description

</td>
<td>

Categorical, Integer

</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

```
      Categorical, Integer
     </p>
    </td>
    <td>
     <p class="normal">
      9000
     </p>
    </td>
    <td>
     <p class="normal">
      86
     </p>
    </td>
    <td>
     <p class="normal">
      2000
     </p>
    </td>
    <!-- <td><p class="normal">Social&amp;nbsp;</p></td> -->
   </tr>
   <tr bgcolor="DDEEFF">
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/Internet+Usage+Data">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Internet+Usage+Data">
           Internet Usage Data
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">This data contains general demographic information on internet users in 1997.
&amp;nbsp;</p></td> -->
    <td>
     <p class="normal">
      Multivariate
     </p>
    </td>
    <td>
     <p class="normal">
     </p>
    </td>
    <td>
     <p class="normal">
      Categorical, Integer
     </p>
    </td>
    <td>
     <p class="normal">
      10104
     </p>
    </td>
```

<td>

72

</td>
<td>

1999

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/IPUMS+Census+Database">
<img border="1" src="assets/MLimages/SmallLarge2.jpg"/>
</a>
</td>
<td>

<a href="datasets/IPUMS+Census+Database">
IPUMS Census Database
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This data set contains unweighted PUMS census data from the Los Angeles an
d Long Beach areas for the years 1970, 1980, and 1990. </td> -->
<td>

Multivariate

</td>
<td>

</td>
<td>

Categorical, Integer

</td>
<td>

256932

</td>
<td>

61

</td>
<td>

1999

</td>
</tr>
</table>
</body>
</html>

```
      </p>
     </td>
     <!-- <td><p class="normal">Social&amp;nbsp;</p></td> -->
    </tr>
    <tr bgcolor="DDEEFF">
     <td>
      <table>
       <tr>
        <td>
         <a href="datasets/Japanese+Vowels">
          <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
         </a>
        </td>
        <td>
         <p class="normal">
          <b>
           <a href="datasets/Japanese+Vowels">
            Japanese Vowels
           </a>
          </b>
         </p>
        </td>
       </tr>
      </table>
     </td>
     <!-- <td><p class="normal">This dataset records 640 time series of 12 LPC cepstrum coefficients taken fro
m nine male speakers.&amp;nbsp;</p></td> -->
     <td>
      <p class="normal">
       Multivariate, Time-Series
      </p>
     </td>
     <td>
      <p class="normal">
       Classification
      </p>
     </td>
     <td>
      <p class="normal">
       Real
      </p>
     </td>
     <td>
      <p class="normal">
       640
      </p>
     </td>
     <td>
      <p class="normal">
       12
      </p>
     </td>
     <td>
      <p class="normal">
      </p>
     </td>
     <!-- <td><p class="normal">Other&amp;nbsp;</p></td> -->
    </tr>
    <tr>
     <td>
      <table>
       <tr>
```

```
        <td>
         <a href="datasets/KDD+Cup+1998+Data">
          <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
         </a>
        </td>
        <td>
         <p class="normal">
          <b>
           <a href="datasets/KDD+Cup+1998+Data">
            KDD Cup 1998 Data
           </a>
          </b>
         </p>
        </td>
       </tr>
      </table>
     </td>
     <!-- <td><p class="normal">This is the data set used for The Second International Knowledge Discovery an
d Data Mining Tools Competition, which was held in conjunction with KDD-98&amp;nbsp;</p></td> -->
     <td>
      <p class="normal">
       Multivariate
      </p>
     </td>
     <td>
      <p class="normal">
       Regression
      </p>
     </td>
     <td>
      <p class="normal">
       Categorical, Integer
      </p>
     </td>
     <td>
      <p class="normal">
       191779
      </p>
     </td>
     <td>
      <p class="normal">
       481
      </p>
     </td>
     <td>
      <p class="normal">
       1998
      </p>
     </td>
     <!-- <td><p class="normal">Other&amp;nbsp;</p></td> -->
    </tr>
    <tr bgcolor="DDEEFF">
     <td>
      <table>
       <tr>
        <td>
         <a href="datasets/KDD+Cup+1999+Data">
          <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
         </a>
        </td>
        <td>
         <p class="normal">
```

<a href="datasets/KDD+Cup+1999+Data">
KDD Cup 1999 Data
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This is the data set used for The Third International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-99 </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer

</td>
<td>

4000000

</td>
<td>

42

</td>
<td>

1999

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/M.+Tuberculosis+Genes">

</a>
</td>
<td>

<a href="datasets/M.+Tuberculosis+Genes">
M. Tuberculosis Genes
</a>

</td>

```
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal"> Data giving characteristics of each ORF (potential gene) in the M. tuberculosis bacterium. Sequence, homology (similarity to other genes) and structural information, and function (if known) ar e provided </p></td> -->
    <td>
     <p class="normal">
      Relational
     </p>
    </td>
    <td>
     <p class="normal">
     </p>
    </td>
    <td>
     <p class="normal">
     </p>
    </td>
    <td>
     <p class="normal">
     </p>
    </td>
    <td>
     <p class="normal">
     </p>
    </td>
    <td>
     <p class="normal">
      2001
     </p>
    </td>
    <!-- <td><p class="normal">Life </p></td> -->
   </tr>
   <tr bgcolor="DDEEFF">
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/Movie">
         <img border="1" src="assets/MLimages/SmallLarge132.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Movie">
           Movie
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">This data set contains a list of over 10000 films including many older, odd, and cult films. There is information on actors, casts, directors, producers, studios, etc. </p></td> -->
    <td>
     <p class="normal">
      Multivariate, Relational
     </p>
    </td>
```

<td>

</td>
<td>

</td>
<td>

10000

</td>
<td>

</td>
<td>

1999

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/MSNBC.com+Anonymous+Web+Data">

</a>
</td>
<td>

<a href="datasets/MSNBC.com+Anonymous+Web+Data">
MSNBC.com Anonymous Web Data
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This data describes the page visits of users who visited msnbc.com on Septem
ber 28, 1999. Visits are recorded at the level of URL category (see description) and are recorded in time order.
  </td> -->
<td>

Sequential

</td>
<td>

</td>
<td>

Categorical

</td>

<td>

989818

</td>
<td>

</td>
<td>

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/NSF+Research+Award+Abstracts+1990-2003">

</a>
</td>
<td>

<a href="datasets/NSF+Research+Award+Abstracts+1990-2003">
NSF Research Award Abstracts 1990-2003
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This data set consists of (a) 129,000 abstracts describing NSF awards for basi
c research, (b) bag-of-word data files extracted from the abstracts, (c) a list of words used for indexing the bag-
of-word </td> -->
<td>

Text

</td>
<td>

</td>
<td>

</td>
<td>

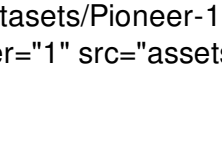
129000

</td>
<td>

</td>
</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

<td>

2003

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Pioneer-1+Mobile+Robot+Data">

</a>
</td>
<td>

<a href="datasets/Pioneer-1+Mobile+Robot+Data">
Pioneer-1 Mobile Robot Data
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset contains time series sensor readings of the Pioneer-1 mobile robo
t. The data is broken into "experiences" in which the robot takes action for some period of time and experiences
a control </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

</td>
<td>

Categorical, Real

</td>
<td>

</td>
<td>

</td>
<td>

</td>
<td>

1999

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

```
<table>
 <tr>
  <td>
   <a href="datasets/Pseudo+Periodic+Synthetic+Time+Series">
    <img border="1" src="assets/MLimages/SmallLarge136.jpg"/>
   </a>
  </td>
  <td>
   <p class="normal">
    <b>
     <a href="datasets/Pseudo+Periodic+Synthetic+Time+Series">
      Pseudo Periodic Synthetic Time Series
     </a>
    </b>
   </p>
  </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">This data set is designed for testing indexing schemes in time series databases
. The data appears highly periodic, but never exactly repeats itself. </p></td> -->
<td>
 <p class="normal">
  Univariate, Time-Series
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
  100000
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
  1999
 </p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Reuters-21578+Text+Categorization+Collection">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
```

<a href="datasets/Reuters-21578+Text+Categorization+Collection">
Reuters-21578 Text Categorization Collection
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This is a collection of documents that appeared on Reuters newswire in 1987. The documents were assembled and indexed with categories. </td> -->
<td>

Text

</td>
<td>

Classification

</td>
<td>

Categorical

</td>
<td>

21578

</td>
<td>

5

</td>
<td>

1997

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Robot+Execution+Failures">

</a>
</td>
<td>

<a href="datasets/Robot+Execution+Failures">
Robot Execution Failures
</a>

</td>
</tr>

```
 </table>
</td>
<!-- <td><p class="normal">This dataset contains force and torque measurements on a robot after failure d
etection. Each failure is characterized by 15 force/torque samples collected at regular time intervals&amp;nbsp;</p>
</td> -->
<td>
 <p class="normal">
  Multivariate, Time-Series
 </p>
</td>
<td>
 <p class="normal">
  Classification
 </p>
</td>
<td>
 <p class="normal">
  Integer
 </p>
</td>
<td>
 <p class="normal">
  463
 </p>
</td>
<td>
 <p class="normal">
  90
 </p>
</td>
<td>
 <p class="normal">
  1999
 </p>
</td>
<!-- <td><p class="normal">Physical&amp;nbsp;</p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Synthetic+Control+Chart+Time+Series">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Synthetic+Control+Chart+Time+Series">
        Synthetic Control Chart Time Series
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">This data consists of synthetically generated control charts.&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Time-Series
```

```
 </p>
</td>
<td>
 <p class="normal">
  Classification, Clustering
 </p>
</td>
<td>
 <p class="normal">
  Real
 </p>
</td>
<td>
 <p class="normal">
  600
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
  1999
 </p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Syskill+and+Webert+Web+Page+Ratings">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Syskill+and+Webert+Web+Page+Ratings">
        Syskill and Webert Web Page Ratings
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">This database contains HTML source of web pages plus the ratings of a single user on these web pages. Web pages are on four seperate subjects (Bands- recording artists; Goats; Sheep; a nd BioMedical) </p></td> -->
 <td>
  <p class="normal">
   Multivariate, Text
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
```

<td>

Categorical

</td>
<td>

332

</td>
<td>

5

</td>
<td>

1998

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/UNIX+User+Data">

</a>
</td>
<td>

<a href="datasets/UNIX+User+Data">
UNIX User Data
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This file contains 9 sets of sanitized user data drawn from the command histori
es of 8 UNIX computer users at Purdue over the course of up to 2 years. </td> -->
<td>

Text, Sequential

</td>
<td>

</td>
<td>

</td>
<td>

</td>
</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

<td>

</td>
<td>

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Volcanoes+on+Venus+-+JARtool+experiment">

</a>
</td>
<td>

<a href="datasets/Volcanoes+on+Venus+-+JARtool+experiment">
Volcanoes on Venus - JARtool experiment
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The JARtool project was a pioneering effort to develop an automatic system for cataloging small volcanoes in the large set of Venus images returned by the Magellan spacecraft. </td> -->
<td>

Image

</td>
<td>

Classification

</td>
<td>

</td>
<td>

</td>
<td>

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>

<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Statlog+%28Australian+Credit+Approval%29">

</a>
</td>
<td>

<a href="datasets/Statlog+%28Australian+Credit+Approval%29">
Statlog (Australian Credit Approval)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This file concerns credit card applications. This database exists elsewhere in th
e repository (Credit Screening Database) in a slightly different form </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer, Real

</td>
<td>

690

</td>
<td>

14

</td>
<td>

</td>
<!-- <td> Financial </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Statlog+%28German+Credit+Data%29">

</a>
</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

```
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Statlog+%28German+Credit+Data%29">
           Statlog (German Credit Data)
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">This dataset classifies people described by a set of attributes as good or bad cr
edit risks. Comes in two formats (one all numeric). Also comes with a cost matrix&amp;nbsp;</p></td> -->
    <td>
     <p class="normal">
      Multivariate
     </p>
    </td>
    <td>
     <p class="normal">
      Classification
     </p>
    </td>
    <td>
     <p class="normal">
      Categorical, Integer
     </p>
    </td>
    <td>
     <p class="normal">
      1000
     </p>
    </td>
    <td>
     <p class="normal">
      20
     </p>
    </td>
    <td>
     <p class="normal">
      1994
     </p>
    </td>
    <!-- <td><p class="normal">Financial&amp;nbsp;</p></td> -->
   </tr>
   <tr>
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/Statlog+%28Heart%29">
         <img border="1" src="assets/MLimages/SmallLarge45.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Statlog+%28Heart%29">
           Statlog (Heart)
```

```
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">This dataset is a heart disease database similar to a database already present in the repository (Heart Disease databases) but in a slightly different form&amp;nbsp;</p></td> -->
    <td>
     <p class="normal">
      Multivariate
     </p>
    </td>
    <td>
     <p class="normal">
      Classification
     </p>
    </td>
    <td>
     <p class="normal">
      Categorical, Real
     </p>
    </td>
    <td>
     <p class="normal">
      270
     </p>
    </td>
    <td>
     <p class="normal">
      13
     </p>
    </td>
    <td>
     <p class="normal">
     </p>
    </td>
    <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
   </tr>
   <tr bgcolor="DDEEFF">
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/Statlog+%28Landsat+Satellite%29">
         <img border="1" src="assets/MLimages/SmallLarge146.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Statlog+%28Landsat+Satellite%29">
           Statlog (Landsat Satellite)
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">Multi-spectral values of pixels in 3x3 neighbourhoods in a satellite image, and t
```

he classification associated with the central pixel in each neighbourhood </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

6435

</td>
<td>

36

</td>
<td>

1993

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Statlog+%28Image+Segmentation%29">

</a>
</td>
<td>

<a href="datasets/Statlog+%28Image+Segmentation%29">
Statlog (Image Segmentation)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset is an image segmentation database similar to a database already present in the repository (Image segmentation database) but in a slightly different form. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

2310

</td>
<td>

19

</td>
<td>

1990

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Statlog+%28Shuttle%29">

</a>
</td>
<td>

<a href="datasets/Statlog+%28Shuttle%29">
Statlog (Shuttle)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The shuttle dataset contains 9 attributes all of which are numerical. Approximat
ely 80% of the data belongs to class 1 </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

9

</td>
<td>

43500

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Statlog+%28Vehicle+Silhouettes%29">

</a>
</td>
<td>

<a href="datasets/Statlog+%28Vehicle+Silhouettes%29">
Statlog (Vehicle Silhouettes)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The purpose is to classify a given silhouette as one of four types of vehicle, u
sing a set of features extracted from the silhouette. The vehicle may be viewed from one of many different angles. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

18

</td>
<td>

846

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Statlog+%28Vowel+Recognition+-+Deterding+Data%29">

</a>
</td>
<td>

<a href="datasets/Statlog+%28Vowel+Recognition+-+Deterding+Data%29">
Statlog (Vowel Recognition - Deterding Data)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Speaker independent recognition of the eleven steady state vowels of British Engl
ish using a specified training set of lpc derived log area ratios. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

11

</td>
<td>

528

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Statlog+%28Vowel+Recognition+-+Deterding+Data%29">

</a>
</td>
<td>

<a href="datasets/Statlog+%28Vowel+Recognition+-+Deterding+Data%29">
Statlog (Vowel Recognition - Deterding Data)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Speaker independent recognition of the eleven steady state vowels of British Engl
ish using a specified training set of lpc derived log area ratios. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

11

</td>
<td>

990

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Statlog+%28Heart%29">

</a>
</td>
<td>

<a href="datasets/Statlog+%28Heart%29">
Statlog (Heart)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This database contains 13 attributes which were extracted from a larger set of 7
5. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

13

</td>
<td>

270

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Statlog+%28German+Credit+Data%29">

</a>
</td>
<td>

<a href="datasets/Statlog+%28German+Credit+Data%29">
Statlog (German Credit Data)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset classifies people described by a set of attributes as good or bad cr
edit risks. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

24

</td>
<td>

1000

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Statlog+%28Australian+Credit+Approval%29">

</a>
</td>
<td>

<a href="datasets/Statlog+%28Australian+Credit+Approval%29">
Statlog (Australian Credit Approval)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This file concerns credit card applications. All attribute names and values have
been changed to meaningless symbols to protect confidentiality of the data. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer, Real

</td>
<td>

14

</td>
<td>

690

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Statlog+%28Landsat+Satellite%29">

</a>
</td>
<td>

<a href="datasets/Statlog+%28Landsat+Satellite%29">
Statlog (Landsat Satellite)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The database consists of the multi-spectral values of pixels in 3x3 neighbourhood
s in a satellite image, and the classification associated with the central pixel in each neighbourhood. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

36

</td>
<td>

6435

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Statlog+%28Image+Segmentation%29">

</a>
</td>
<td>

<a href="datasets/Statlog+%28Image+Segmentation%29">
Statlog (Image Segmentation)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The instances were drawn randomly from a database of 7 outdoor images. The images
were hand-segmented to create a classification for every pixel. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

19

</td>
<td>

2310

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Statlog+%28Shuttle%29">

</a>
</td>
<td>

<a href="datasets/Statlog+%28Shuttle%29">
Statlog (Shuttle)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The shuttle dataset contains 9 attributes all of which are numerical. Approximat
ely 80% of the data belongs to class 1 </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

9

</td>
<td>

43500

</td>
<!-- <td> Other </td> -->
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>
</table>
</td>

```
 </p>
</td>
<td>
 <p class="normal">
  58000
 </p>
</td>
<td>
 <p class="normal">
  9
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<!-- <td><p class="normal">Physical&amp;nbsp;</p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Statlog+%28Vehicle+Silhouettes%29">
      <img border="1" src="assets/MLimages/SmallLarge149.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Statlog+%28Vehicle+Silhouettes%29">
        Statlog (Vehicle Silhouettes)
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">3D objects within a 2D image by application of an ensemble of shape feature e
xtractors to the 2D silhouettes of the objects.&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Integer
  </p>
 </td>
 <td>
  <p class="normal">
   946
  </p>
 </td>
 <td>
```

18

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Connectionist+Bench+%28Nettalk+Corpus%29">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Connectionist+Bench+%28Nettalk+Corpus%29">
Connectionist Bench (Nettalk Corpus)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The file "nettalk.data" contains a list of 20,008 English words, along with a phon
etic transcription for each word. The task is to train a network to produce the proper phonemes </td>
-->
<td>

Multivariate

</td>
<td>

</td>
<td>

Categorical

</td>
<td>

20008

</td>
<td>

4

</td>
<td>

</td>

<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29">
Connectionist Bench (Sonar, Mines vs. Rocks)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The task is to train a network to discriminate between sonar signals bounced of
f a metal cylinder and those bounced off a roughly cylindrical rock. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

208

</td>
<td>

60

</td>
<td>

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29">

<a href="datasets/Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29">
Connectionist Bench (Vowel Recognition - Deterding Data)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Speaker independent recognition of the eleven steady state vowels of British E
nglish using a specified training set of lpc derived log area ratios. </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

528

</td>
<td>

10

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Economic+Sanctions">
<img border="1" src="assets/MLimages/SmallLarge153.jpg"/>
</a>
</td>
<td>

<a href="datasets/Economic+Sanctions">
Economic Sanctions
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Speaker independent recognition of the eleven steady state vowels of British E
nglish using a specified training set of lpc derived log area ratios. </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

1000

</td>
<td>

4

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Forest+Fires">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Forest+Fires">
Forest Fires
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This is a difficult regression task, where the aim is to predict the burned area
of forest fires, in the northeast region of Portugal, by using meteorological and other data. </td> -->
<td>

</td>
<td>

Regression

</td>
<td>

Real

</td>
<td>

517

</td>
<td>

12

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Glass+Identification">

</a>
</td>
<td>

<a href="datasets/Glass+Identification">
Glass Identification
</a>

</td>
</tr>
</table>
</td>
<!-- <td> A comprehensive database for examining the task of forensic glass identification
. </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

214

</td>
<td>

10

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Ionosphere">

</a>
</td>
<td>

<a href="datasets/Ionosphere">
Ionosphere
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Classification of radar returns from the ionosphere. </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

351

</td>
<td>

34

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Iris">

</a>
</td>
<td>

<a href="datasets/Iris">
Iris
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Famous database; from Fisher, 1936. </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

150

</td>
<td>

4

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Iris+Plants+Database">

</a>
</td>
<td>

<a href="datasets/Iris+Plants+Database">
Iris Plants Database
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This is perhaps the best known database to be found in the pattern recognition l
iterature. Fisher's paper is a classic in the field and is referenced frequently to this day. (See Duda & Hart, for example.) The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2; the latter are NOT linearly separable from each other. </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

150

</td>
<td>

4

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Isolet">

</a>
</td>
<td>

<a href="datasets/Isolet">
Isolet
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Isolet Spoken Letter Recognition </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

7797

</td>
<td>

617

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Letter+Recognition">

</a>
</td>
<td>

<a href="datasets/Letter+Recognition">
Letter Recognition
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Letter Image Recognition Data </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

20000

</td>
<td>

16

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Libras+Movement">

</a>
</td>
<td>

<a href="datasets/Libras+Movement">
Libras Movement
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Libras Movement Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

360

</td>
<td>

90

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Low+Resolution+Spectrometer">

</a>
</td>
<td>

<a href="datasets/Low+Resolution+Spectrometer">
Low Resolution Spectrometer
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Low Resolution Spectrometer </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

531

</td>
<td>

100

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/MAGIC+Gamma+Telescope">

</a>
</td>
<td>

<a href="datasets/MAGIC+Gamma+Telescope">
MAGIC Gamma Telescope
</a>

</td>
</tr>
</table>
</td>
<!-- <td> MAGIC Gamma Telescope Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

19020

</td>
<td>

10

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/MAGIC+Gamma+Telescope+Data+Set">

</a>
</td>
<td>

<a href="datasets/MAGIC+Gamma+Telescope+Data+Set">
MAGIC Gamma Telescope Data Set
</a>

</td>
</tr>
</table>
</td>
<!-- <td> MAGIC Gamma Telescope Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

19020

</td>
<td>

10

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Musk+%28Version+1%29">

</a>
</td>
<td>

<a href="datasets/Musk+%28Version+1%29">
Musk (Version 1)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The Musk (Version 1) data set. </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

476

</td>
<td>

166

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Musk+%28Version+2%29">

</a>
</td>
<td>

<a href="datasets/Musk+%28Version+2%29">
Musk (Version 2)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The Musk (Version 2) data set. </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

6598

</td>
<td>

166

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Optical+Recognition+of+Handwritten+Digits">

</a>
</td>
<td>

<a href="datasets/Optical+Recognition+of+Handwritten+Digits">
Optical Recognition of Handwritten Digits
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Optical Recognition of Handwritten Digits Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

5620

</td>
<td>

64

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Page+Blocks+Classification">

</a>
</td>
<td>

<a href="datasets/Page+Blocks+Classification">
Page Blocks Classification
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Page Blocks Classification Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

5473

</td>
<td>

10

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits">

</a>
</td>
<td>

<a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits">
Pen-Based Recognition of Handwritten Digits
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Pen-Based Recognition of Handwritten Digits Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

10992

</td>
<td>

16

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Pima+Indians+Diabetes">

</a>
</td>
<td>

<a href="datasets/Pima+Indians+Diabetes">
Pima Indians Diabetes
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Pima Indians Diabetes Database </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

768

</td>
<td>

8

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Poker+Hand">

</a>
</td>
<td>

<a href="datasets/Poker+Hand">
Poker Hand
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Poker Hand Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

1025010

</td>
<td>

10

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Post-Operative+Patient">

</a>
</td>
<td>

<a href="datasets/Post-Operative+Patient">
Post-Operative Patient
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Post-Operative Patient Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

90

</td>
<td>

8

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Qualitative_Bankruptcy">

</a>
</td>
<td>

<a href="datasets/Qualitative_Bankruptcy">
Qualitative_Bankruptcy
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Qualitative Bankruptcy Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Categorical

</td>
<td>

250

</td>
<td>

6

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Seeds">

</a>
</td>
<td>

<a href="datasets/Seeds">
Seeds
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Seeds Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

210

</td>
<td>

7

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Spambase">

</a>
</td>
<td>

<a href="datasets/Spambase">
Spambase
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Spambase Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

4601

</td>
<td>

57

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Spect+Heart">

</a>
</td>
<td>

<a href="datasets/Spect+Heart">
Spect Heart
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Spect Heart Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

267

</td>
<td>

44

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Statlog+%28Heart%29">

</a>
</td>
<td>

<a href="datasets/Statlog+%28Heart%29">
Statlog (Heart)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Statlog (Heart) Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

270

</td>
<td>

13

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Statlog+%28Image+Segmentation%29">

</a>
</td>
<td>

<a href="datasets/Statlog+%28Image+Segmentation%29">
Statlog (Image Segmentation)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Statlog (Image Segmentation) Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

2310

</td>
<td>

19

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Statlog+%28Landsat+Satellite%29">

</a>
</td>
<td>

<a href="datasets/Statlog+%28Landsat+Satellite%29">
Statlog (Landsat Satellite)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Statlog (Landsat Satellite) Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

6435

</td>
<td>

36

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Statlog+%28Shuttle%29">

</a>
</td>
<td>

<a href="datasets/Statlog+%28Shuttle%29">
Statlog (Shuttle)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Statlog (Shuttle) Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

58000

</td>
<td>

9

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Statlog+%28Vehicle+Silhouettes%29">

</a>
</td>
<td>

<a href="datasets/Statlog+%28Vehicle+Silhouettes%29">
Statlog (Vehicle Silhouettes)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Statlog (Vehicle Silhouettes) Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

946

</td>
<td>

18

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Steel+Plates+Faults">

</a>
</td>
<td>

<a href="datasets/Steel+Plates+Faults">
Steel Plates Faults
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Steel Plates Faults Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

1941

</td>
<td>

27

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Synthetic+Control+Chart+Time+Series">

</a>
</td>
<td>

<a href="datasets/Synthetic+Control+Chart+Time+Series">
Synthetic Control Chart Time Series
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Synthetic Control Chart Time Series Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

600

</td>
<td>

60

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Thyroid+Disease">

</a>
</td>
<td>

<a href="datasets/Thyroid+Disease">
Thyroid Disease
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Thyroid Disease Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

9172

</td>
<td>

29

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Wall-Following+Robot+Navigation+Data">

</a>
</td>
<td>

<a href="datasets/Wall-Following+Robot+Navigation+Data">
Wall-Following Robot Navigation Data
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Wall-Following Robot Navigation Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

5456

</td>
<td>

24

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Waveform+Database+Generator+%28Version+1%29">

</a>
</td>
<td>

<a href="datasets/Waveform+Database+Generator+%28Version+1%29">
Waveform Database Generator (Version 1)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Waveform Database Generator (Version 1) Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

5000

</td>
<td>

21

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Waveform+Database+Generator+%28Version+2%29">

</a>
</td>
<td>

<a href="datasets/Waveform+Database+Generator+%28Version+2%29">
Waveform Database Generator (Version 2)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Waveform Database Generator (Version 2) Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

5000

</td>
<td>

40

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Wine">

</a>
</td>
<td>

<a href="datasets/Wine">
Wine
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Wine Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

178

</td>
<td>

13

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Wine+Quality">

</a>
</td>
<td>

<a href="datasets/Wine+Quality">
Wine Quality
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Wine Quality Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

4898

</td>
<td>

11

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Yeast">

</a>
</td>
<td>

<a href="datasets/Yeast">
Yeast
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Yeast Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

1484

</td>
<td>

8

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Zoo">

</a>
</td>
<td>

<a href="datasets/Zoo">
Zoo
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Zoo Data Set </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Categorical

</td>
<td>

101

</td>
<td>

16

</td>
<td>

</td>
<!-- <td> Other </td> -->
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>
</table>
</td>

```
      </a>
     </b>
    </p>
   </td>
  </tr>
 </table>
</td>
<!-- <td><p class="normal">Domain Theory on Economic Sanctions; Undocumented </p></td> -->
<td>
 <p class="normal">
  Domain-Theory
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<!-- <td><p class="normal">Financial </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Protein+Data">
      <img border="1" src="assets/MLimages/SmallLarge154.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Protein+Data">
        Protein Data
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Undocumented </p></td> -->
 <td>
  <p class="normal">
  </p>
 </td>
 <td>
```

</td>
<td>

</td>
<td>

</td>
<td>

</td>
<td>

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Cloud">

</a>
</td>
<td>

<a href="datasets/Cloud">
Cloud
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Little Documentation </td> -->
<td>

Multivariate

</td>
<td>

</td>
<td>

Real

</td>
<td>

1024

</td>

```
      <td>
       <p class="normal">
        10
       </p>
      </td>
      <td>
       <p class="normal">
        1989
       </p>
      </td>
      <!-- <td><p class="normal">Physical </p></td> -->
     </tr>
     <tr bgcolor="DDEEFF">
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/CalIt2+Building+People+Counts">
           <img border="1" src="assets/MLimages/SmallLarge156.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/CalIt2+Building+People+Counts">
             CalIt2 Building People Counts
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">This data comes from the main door of the CalIt2 building at UCI. </p></t
d> -->
      <td>
       <p class="normal">
        Multivariate, Time-Series
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
        Categorical, Integer
       </p>
      </td>
      <td>
       <p class="normal">
        10080
       </p>
      </td>
      <td>
       <p class="normal">
        4
       </p>
      </td>
      <td>
       <p class="normal">
        2006
```

```
      </p>
     </td>
     <!-- <td><p class="normal">Other </p></td> -->
    </tr>
    <tr>
     <td>
      <table>
       <tr>
        <td>
         <a href="datasets/Dodgers+Loop+Sensor">
          <img border="1" src="assets/MLimages/SmallLarge157.jpg"/>
         </a>
        </td>
        <td>
         <p class="normal">
          <b>
           <a href="datasets/Dodgers+Loop+Sensor">
            Dodgers Loop Sensor
           </a>
          </b>
         </p>
        </td>
       </tr>
      </table>
     </td>
     <!-- <td><p class="normal">Loop sensor data was collected for the Glendale on ramp for the 101 North fre
eway in Los Angeles </p></td> -->
     <td>
      <p class="normal">
       Multivariate, Time-Series
      </p>
     </td>
     <td>
      <p class="normal">
      </p>
     </td>
     <td>
      <p class="normal">
       Categorical, Integer
      </p>
     </td>
     <td>
      <p class="normal">
       50400
      </p>
     </td>
     <td>
      <p class="normal">
       3
      </p>
     </td>
     <td>
      <p class="normal">
       2006
      </p>
     </td>
     <!-- <td><p class="normal">Other </p></td> -->
    </tr>
    <tr bgcolor="DDEEFF">
     <td>
      <table>
       <tr>
```

<td>
<a href="datasets/Poker+Hand">

</a>
</td>
<td>

<a href="datasets/Poker+Hand">
Poker Hand
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Purpose is to predict poker hands </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer

</td>
<td>

1025010

</td>
<td>

11

</td>
<td>

2007

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/MAGIC+Gamma+Telescope">

</a>
</td>
<td>

<a href="datasets/MAGIC+Gamma+Telescope">
MAGIC Gamma Telescope
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Data are MC generated to simulate registration of high energy gamma particle
s in an atmospheric Cherenkov telescope </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

19020

</td>
<td>

11

</td>
<td>

2007

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/UJI+Pen+Characters">

</a>
</td>
<td>

<a href="datasets/UJI+Pen+Characters">
UJI Pen Characters
</a>

</td>
</tr>

```
 </table>
</td>
<!-- <td><p class="normal">Data consists of written characters in a UNIPEN-like format </p></td> -->
<td>
 <p class="normal">
  Multivariate, Sequential
 </p>
</td>
<td>
 <p class="normal">
  Classification
 </p>
</td>
<td>
 <p class="normal">
  Integer
 </p>
</td>
<td>
 <p class="normal">
  1364
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
  2007
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
 <table>
  <tr>
   <td>
    <a href="datasets/Mammographic+Mass">
     <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   </td>
   <td>
    <p class="normal">
     <b>
      <a href="datasets/Mammographic+Mass">
       Mammographic Mass
      </a>
     </b>
    </p>
   </td>
  </tr>
 </table>
</td>
<!-- <td><p class="normal">Discrimination of benign and malignant mammographic masses based on BI-R
ADS attributes and the patient's age. </p></td> -->
<td>
 <p class="normal">
  Multivariate
 </p>
</td>
```

<td>

Classification

</td>
<td>

Integer

</td>
<td>

961

</td>
<td>

6

</td>
<td>

2007

</td>
<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Forest+Fires">

</a>
</td>
<td>

<a href="datasets/Forest+Fires">
Forest Fires
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This is a difficult regression task, where the aim is to predict the burned area of forest fires, in the northeast region of Portugal, by using meteorological and other data (see details at: http://www.dsi.uminho.pt/~pcortez/forestfires). </td> -->
<td>

Multivariate

</td>
<td>

Regression

</td>
<td>

<td>

Real

</td>
<td>

517

</td>
<td>

13

</td>
<td>

2008

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Reuters+Transcribed+Subset">

</a>
</td>
<td>

<a href="datasets/Reuters+Transcribed+Subset">
Reuters Transcribed Subset
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset is created by reading out 200 files from the 10 largest Reuters classes and using an Automatic Speech Recognition system to create corresponding transcriptions. </td> -->
<td>

Text

</td>
<td>

Classification

</td>
<td>

</td>
<td>

200

</td>
</tr>
</table>
</body>
</html>

```
 200
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
  2008
 </p>
</td>
<!-- <td><p class="normal">Business </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Bag+of+Words">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Bag+of+Words">
        Bag of Words
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">This data set contains five text collections in the form of bags-of-words. <
/p></td> -->
 <td>
  <p class="normal">
   Text
  </p>
 </td>
 <td>
  <p class="normal">
   Clustering
  </p>
 </td>
 <td>
  <p class="normal">
   Integer
  </p>
 </td>
 <td>
  <p class="normal">
   8000000
  </p>
 </td>
 <td>
  <p class="normal">
   100000
  </p>
 </td>
```

```
 <td>
  <p class="normal">
   2008
  </p>
 </td>
 <!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Concrete+Compressive+Strength">
      <img border="1" src="assets/MLimages/SmallLarge165.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Concrete+Compressive+Strength">
        Concrete Compressive Strength
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Concrete is the most important material in civil engineering. The concrete com
pressive strength is a highly nonlinear function of age and ingredients.  </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Regression
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
 </td>
 <td>
  <p class="normal">
   1030
  </p>
 </td>
 <td>
  <p class="normal">
   9
  </p>
 </td>
 <td>
  <p class="normal">
   2007
  </p>
 </td>
 <!-- <td><p class="normal">Physical </p></td> -->
</tr>
```

<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Hill-Valley">

</a>
</td>
<td>

<a href="datasets/Hill-Valley">
Hill-Valley
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Each record represents 100 points on a two-dimensional graph. When plotted i
n order (from 1 through 100) as the Y co-ordinate, the points will create either a Hill (a &#65533;bump&#65533; in the terrain)
or a Valley (a &#65533;dip&#65533; in the terrain). </td> -->
<td>

Sequential

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

606

</td>
<td>

101

</td>
<td>

2008

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Arcene">

```
          <img border="1" src="assets/MLimages/SmallLarge167.jpg"/>
         </a>
        </td>
        <td>
         <p class="normal">
          <b>
           <a href="datasets/Arcene">
            Arcene
           </a>
          </b>
         </p>
        </td>
       </tr>
      </table>
     </td>
     <!-- <td><p class="normal">ARCENE's task is to distinguish cancer versus normal patterns from mass-spec
trometric data. This is a two-class classification problem with continuous input variables. This dataset is one of
5 datasets of the NIPS 2003 feature selection challenge.&amp;nbsp;</p></td> -->
     <td>
      <p class="normal">
       Multivariate
      </p>
     </td>
     <td>
      <p class="normal">
       Classification
      </p>
     </td>
     <td>
      <p class="normal">
       Real
      </p>
     </td>
     <td>
      <p class="normal">
       900
      </p>
     </td>
     <td>
      <p class="normal">
       10000
      </p>
     </td>
     <td>
      <p class="normal">
       2008
      </p>
     </td>
     <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
    </tr>
    <tr bgcolor="DDEEFF">
     <td>
      <table>
       <tr>
        <td>
         <a href="datasets/Dexter">
          <img border="1" src="assets/MLimages/SmallLarge168.jpg"/>
         </a>
        </td>
        <td>
         <p class="normal">
          <b>
```

<a href="datasets/Dexter">
Dexter
</a>

</td>
</tr>
</table>
</td>
<!-- <td> DEXTER is a text classification problem in a bag-of-word representation. This i
s a two-class classification problem with sparse continuous input variables. This dataset is one of five datasets
of the NIPS 2003 feature selection challenge.
  </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

2600

</td>
<td>

20000

</td>
<td>

2008

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Dorothea">

</a>
</td>
<td>

<a href="datasets/Dorothea">
Dorothea
</a>

</td>
</tr>
</table>
</td>
<!-- <td> DOROTHEA is a drug discovery dataset. Chemical compounds represented by structural molecular features must be classified as active (binding to thrombin) or inactive. This is one of 5 data sets of the NIPS 2003 feature selection challenge. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

1950

</td>
<td>

100000

</td>
<td>

2008

</td>
<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Gisette">

</a>
</td>
<td>

<a href="datasets/Gisette">
Gisette
</a>

</td>
</tr>
</table>
</td>
<!-- <td> GISETTE is a handwritten digit recognition problem. The problem is to separate the highly confusible digits '4' and '9'. This dataset is one of five datasets of the NIPS 2003 feature selection c

hallenge.
  </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

13500

</td>
<td>

5000

</td>
<td>

2008

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Madelon">

</a>
</td>
<td>

<a href="datasets/Madelon">
Madelon
</a>

</td>
</tr>
</table>
</td>
<!-- <td> MADELON is an artificial dataset, which was part of the NIPS 2003 feature sele
ction challenge. This is a two-class classification problem with continuous input variables. The difficulty is that th
e problem is multivariate and highly non-linear.   </td> -->
<td>

Multivariate

</td>

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

4400

</td>
<td>

500

</td>
<td>

2008

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Ozone+Level+Detection">

</a>
</td>
<td>

<a href="datasets/Ozone+Level+Detection">
Ozone Level Detection
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Two ground ozone level data sets are included in this collection. One is the eig
ht hour peak set (eighthr.data), the other is the one hour peak set (onehr.data). Those data were collected fro
m 1998 to 2004 at the Houston, Galveston and Brazoria area. </td> -->
<td>

Multivariate, Sequential, Time-Series

</td>
<td>

Classification

</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

<td>

Real

</td>
<td>

2536

</td>
<td>

73

</td>
<td>

2008

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Abscisic+Acid+Signaling+Network">

</a>
</td>
<td>

<a href="datasets/Abscisic+Acid+Signaling+Network">
Abscisic Acid Signaling Network
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The objective is to determine the set of boolean rules that describe the interacti
ons of the nodes within this plant signaling network. The dataset includes 300 separate boolean pseudodynami
c simulations using an asynchronous update scheme.   </td> -->
<td>

Multivariate

</td>
<td>

Causal-Discovery

</td>
<td>

Integer

</td>
<td>

<td>

300

</td>
<td>

43

</td>
<td>

2008

</td>
<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Parkinsons">

</a>
</td>
<td>

<a href="datasets/Parkinsons">
Parkinsons
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Oxford Parkinson's Disease Detection Dataset </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

197

</td>
<td>

23

</td>

```
 </p>
</td>
<td>
 <p class="normal">
  2008
 </p>
</td>
<!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Character+Trajectories">
      <img border="1" src="assets/MLimages/SmallLarge175.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Character+Trajectories">
        Character Trajectories
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Multiple, labelled samples of pen tip trajectories recorded whilst writing individu
al characters. All samples are from the same writer, for the purposes of primitive extraction. Only characters wit
h a single pen-down segment were considered.&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Time-Series
  </p>
 </td>
 <td>
  <p class="normal">
   Classification, Clustering
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
 </td>
 <td>
  <p class="normal">
   2858
  </p>
 </td>
 <td>
  <p class="normal">
   3
  </p>
 </td>
 <td>
  <p class="normal">
   2008
  </p>
```

```
 </td>
 <!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Blood+Transfusion+Service+Center">
      <img border="1" src="assets/MLimages/SmallLarge176.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Blood+Transfusion+Service+Center">
        Blood Transfusion Service Center
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Data taken from the Blood Transfusion Service Center in Hsin-Chu City in Taiw
an -- this is a classification problem.  </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
 </td>
 <td>
  <p class="normal">
   748
  </p>
 </td>
 <td>
  <p class="normal">
   5
  </p>
 </td>
 <td>
  <p class="normal">
   2008
  </p>
 </td>
 <!-- <td><p class="normal">Business </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
```

<td>
<a href="datasets/UJI+Pen+Characters+%28Version+2%29">

</a>
</td>
<td>

<a href="datasets/UJI+Pen+Characters+%28Version+2%29">
UJI Pen Characters (Version 2)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> A pen-based database with more than 11k isolated handwritten characters&nb
sp; </td> -->
<td>

Multivariate, Sequential

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

11640

</td>
<td>

</td>
<td>

2009

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Semeion+Handwritten+Digit">

</a>
</td>
<td>

<a href="datasets/Semeion+Handwritten+Digit">
Semeion Handwritten Digit
</a>

</td>
</tr>
</table>
</td>
<!-- <td> 1593 handwritten digits from around 80 persons were scanned, stretched in a r
ectangular box 16x16 in a gray scale of 256 values. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

1593

</td>
<td>

256

</td>
<td>

2008

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/SECOM">

</a>
</td>
<td>

<a href="datasets/SECOM">
SECOM
</a>

</td>
</tr>

```
 </table>
</td>
<!-- <td><p class="normal">Data from a semi-conductor manufacturing process </p></td> -->
<td>
 <p class="normal">
  Multivariate
 </p>
</td>
<td>
 <p class="normal">
  Classification, Causal-Discovery
 </p>
</td>
<td>
 <p class="normal">
  Real
 </p>
</td>
<td>
 <p class="normal">
  1567
 </p>
</td>
<td>
 <p class="normal">
  591
 </p>
</td>
<td>
 <p class="normal">
  2008
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Plants">
      <img border="1" src="assets/MLimages/SmallLarge180.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Plants">
        Plants
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Data has been extracted from the USDA plants database. It contains all plants (species and genera) in the database and the states of USA and Canada where they occur. </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
```

</td>
<td>

Clustering

</td>
<td>

Categorical

</td>
<td>

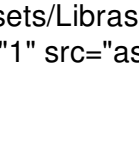
22632

</td>
<td>

70

</td>
<td>

2008

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Libras+Movement">

</a>
</td>
<td>

<a href="datasets/Libras+Movement">
Libras Movement
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The data set contains 15 classes of 24 instances each. Each class references t
o a hand movement type in LIBRAS (Portuguese
name 'L&#65533;ngua BRAsileira de Sinais', oficial brazilian signal language). </td> -->
<td>

Multivariate, Sequential

</td>
<td>

Classification, Clustering

</td>

<td>

Real

</td>
<td>

360

</td>
<td>

91

</td>
<td>

2009

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Concrete+Slump+Test">

</a>
</td>
<td>

<a href="datasets/Concrete+Slump+Test">
Concrete Slump Test
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Concrete is a highly complex material. The slump flow of concrete is not only d
etermined by the water content, but that is also influenced by other concrete ingredients. </td> -->
<td>

Multivariate

</td>
<td>

Regression

</td>
<td>

Real

</td>
<td>

103

</td>
<td>

10

</td>
<td>

2009

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Communities+and+Crime">
<img border="1" src="assets/MLimages/SmallLarge183.jpg"/>
</a>
</td>
<td>

<a href="datasets/Communities+and+Crime">
Communities and Crime
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Communities within the United States. The data combines socio-economic data from the 1990 US Census, law enforcement data from the 1990 US LEMAS survey, and crime data from the 19 95 FBI UCR. </td> -->
<td>

Multivariate

</td>
<td>

Regression

</td>
<td>

Real

</td>
<td>

1994

</td>
<td>

```
      128
     </p>
    </td>
    <td>
     <p class="normal">
      2009
     </p>
    </td>
    <!-- <td><p class="normal">Social </p></td> -->
   </tr>
   <tr bgcolor="DDEEFF">
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/Acute+Inflammations">
         <img border="1" src="assets/MLimages/SmallLarge184.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Acute+Inflammations">
           Acute Inflammations
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">The data was created by a medical expert as a data set to test the expert syste
m,
which will perform the presumptive diagnosis of two diseases of the urinary system.
 </p></td> -->
    <td>
     <p class="normal">
      Multivariate
     </p>
    </td>
    <td>
     <p class="normal">
      Classification
     </p>
    </td>
    <td>
     <p class="normal">
      Categorical, Integer
     </p>
    </td>
    <td>
     <p class="normal">
      120
     </p>
    </td>
    <td>
     <p class="normal">
      6
     </p>
    </td>
    <td>
     <p class="normal">
```

```
    2009
   </p>
  </td>
  <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
 </tr>
 <tr>
  <td>
   <table>
    <tr>
     <td>
      <a href="datasets/Wine+Quality">
       <img border="1" src="assets/MLimages/SmallLarge186.jpg"/>
      </a>
     </td>
     <td>
      <p class="normal">
       <b>
        <a href="datasets/Wine+Quality">
         Wine Quality
        </a>
       </b>
      </p>
     </td>
    </tr>
   </table>
  </td>
  <!-- <td><p class="normal">Two datasets are included, related to red and white vinho verde wine samples, from the north of Portugal. The goal is to model wine quality based on physicochemical tests (see [Cortez et al., 2009], http://www3.dsi.uminho.pt/pcortez/wine/).&amp;nbsp;</p></td> -->
  <td>
   <p class="normal">
    Multivariate
   </p>
  </td>
  <td>
   <p class="normal">
    Classification, Regression
   </p>
  </td>
  <td>
   <p class="normal">
    Real
   </p>
  </td>
  <td>
   <p class="normal">
    4898
   </p>
  </td>
  <td>
   <p class="normal">
    12
   </p>
  </td>
  <td>
   <p class="normal">
    2009
   </p>
  </td>
  <!-- <td><p class="normal">Business&amp;nbsp;</p></td> -->
 </tr>
 <tr bgcolor="DDEEFF">
```

<td>
<table>
<tr>
<td>
<a href="datasets/URL+Reputation">

</a>
</td>
<td>

<a href="datasets/URL+Reputation">
URL Reputation
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Anonymized 120-day subset of the ICML-09 URL data containing 2.4 million ex
amples and 3.2 million features. </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

Classification

</td>
<td>

Integer, Real

</td>
<td>

2396130

</td>
<td>

3231961

</td>
<td>

2009

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/p53+Mutants">

</a>
</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

```
          </td>
          <td>
           <p class="normal">
            <b>
             <a href="datasets/p53+Mutants">
              p53 Mutants
             </a>
            </b>
           </p>
          </td>
         </tr>
        </table>
       </td>
       <!-- <td><p class="normal">The goal is to model mutant p53 transcriptional activity (active vs inactive) base
d on data extracted from biophysical simulations.
 </p></td> -->
       <td>
        <p class="normal">
         Multivariate
        </p>
       </td>
       <td>
        <p class="normal">
         Classification
        </p>
       </td>
       <td>
        <p class="normal">
         Real
        </p>
       </td>
       <td>
        <p class="normal">
         16772
        </p>
       </td>
       <td>
        <p class="normal">
         5409
        </p>
       </td>
       <td>
        <p class="normal">
         2010
        </p>
       </td>
       <!-- <td><p class="normal">Life </p></td> -->
      </tr>
      <tr bgcolor="DDEEFF">
       <td>
        <table>
         <tr>
          <td>
           <a href="datasets/Parkinsons+Telemonitoring">
            <img border="1" src="assets/MLimages/SmallLarge174.jpg"/>
           </a>
          </td>
          <td>
           <p class="normal">
            <b>
             <a href="datasets/Parkinsons+Telemonitoring">
              Parkinsons Telemonitoring
```

```
      </a>
     </b>
    </p>
   </td>
  </tr>
 </table>
</td>
<!-- <td><p class="normal">Oxford Parkinson's Disease Telemonitoring Dataset </p></td> -->
<td>
 <p class="normal">
  Multivariate
 </p>
</td>
<td>
 <p class="normal">
  Regression
 </p>
</td>
<td>
 <p class="normal">
  Integer, Real
 </p>
</td>
<td>
 <p class="normal">
  5875
 </p>
</td>
<td>
 <p class="normal">
  26
 </p>
</td>
<td>
 <p class="normal">
  2009
 </p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Demospongiae">
      <img border="1" src="assets/MLimages/SmallLarge190.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Demospongiae">
        Demospongiae
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Marine sponges of the Demospongiae class classification domain. </p></
```

td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

503

</td>
<td>

</td>
<td>

2010

</td>
<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Opinosis+Opinion+%26frasl%3B+Review">

</a>
</td>
<td>

<a href="datasets/Opinosis+Opinion+%26frasl%3B+Review">
Opinosis Opinion／Review
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset contains sentences extracted from user reviews on a given topic. Example topics are “performance of Toyota Camry” and “sound quality of ipod nano”.   </td> -->
<td>

Text

</td>
<td>

</td>
<td>

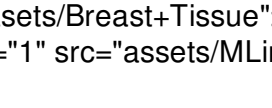
</td>
<td>

51

</td>
<td>

</td>
<td>

2010

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Breast+Tissue">

</a>
</td>
<td>

<a href="datasets/Breast+Tissue">
Breast Tissue
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Dataset with electrical impedance measurements of freshly excised tissue sam
ples from the breast. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

106

</td>
<td>

10

</td>
<td>

2010

</td>
<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Cardiotocography">

</a>
</td>
<td>

<a href="datasets/Cardiotocography">
Cardiotocography
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The dataset consists of measurements of fetal heart rate (FHR) and uterine co
ntraction (UC) features on cardiotocograms classified by expert obstetricians. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

2126

</td>
<td>

23

</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

```
</p>
</td>
<td>
<p class="normal">
2010
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Wall-Following+Robot+Navigation+Data">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Wall-Following+Robot+Navigation+Data">
Wall-Following Robot Navigation Data
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The data were collected as the SCITOS G5 robot navigates through the room f
ollowing the wall in a clockwise direction, for 4 rounds, using 24 ultrasound sensors arranged circularly around i
ts 'waist'. </p></td> -->
<td>
<p class="normal">
Multivariate, Sequential
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Real
</p>
</td>
<td>
<p class="normal">
5456
</p>
</td>
<td>
<p class="normal">
24
</p>
</td>
<td>
<p class="normal">
2010
</p>
```

```
 </td>
 <!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Spoken+Arabic+Digit">
      <img border="1" src="assets/MLimages/SmallLarge195.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Spoken+Arabic+Digit">
        Spoken Arabic Digit
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">This dataset contains timeseries of mel-frequency cepstrum coefficients (MFC
Cs) corresponding to spoken Arabic digits.  Includes data from 44 male and 44 female native Arabic speakers.&
nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate, Time-Series
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
 </td>
 <td>
  <p class="normal">
   8800
  </p>
 </td>
 <td>
  <p class="normal">
   13
  </p>
 </td>
 <td>
  <p class="normal">
   2010
  </p>
 </td>
 <!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
 <td>
  <table>
```

<tr>
<td>
<a href="datasets/Localization+Data+for+Person+Activity">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Localization+Data+for+Person+Activity">
Localization Data for Person Activity
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Data contains recordings of five people performing different activities. Each per
son wore four sensors (tags) while performing the same scenario five times.   </td> -->
<td>

Univariate, Sequential, Time-Series

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

164860

</td>
<td>

8

</td>
<td>

2010

</td>
<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/AutoUniv">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/AutoUniv">
AutoUniv
</a>

</td>
</tr>
</table>
</td>
<!-- <td> AutoUniv is an advanced data generator for classifications tasks. The aim is to reflect the nuances and heterogeneity of real data. Data can be generated in .csv, ARFF or C4.5 formats. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Categorical, Integer, Real

</td>
<td>

</td>
<td>

</td>
<td>

2010

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Steel+Plates+Faults">

</a>
</td>
<td>

<a href="datasets/Steel+Plates+Faults">
Steel Plates Faults
</a>

</td>

```
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">A dataset of steel plates' faults, classified into 7 different types.
The goal was to train machine learning for automatic pattern recognition.
&amp;nbsp;</p></td> -->
    <td>
     <p class="normal">
      Multivariate
     </p>
    </td>
    <td>
     <p class="normal">
      Classification
     </p>
    </td>
    <td>
     <p class="normal">
      Integer, Real
     </p>
    </td>
    <td>
     <p class="normal">
      1941
     </p>
    </td>
    <td>
     <p class="normal">
      27
     </p>
    </td>
    <td>
     <p class="normal">
      2010
     </p>
    </td>
    <!-- <td><p class="normal">Physical&amp;nbsp;</p></td> -->
   </tr>
   <tr bgcolor="DDEEFF">
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/MiniBooNE+particle+identification">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/MiniBooNE+particle+identification">
           MiniBooNE particle identification
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">This dataset is taken from the MiniBooNE experiment and is used to distinguish
electron neutrinos (signal) from muon neutrinos (background).&amp;nbsp;</p></td> -->
    <td>
```

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

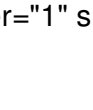
130065

</td>
<td>

50

</td>
<td>

2010

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/YearPredictionMSD">

</a>
</td>
<td>

<a href="datasets/YearPredictionMSD">
YearPredictionMSD
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Prediction of the release year of a song from audio features. Songs are mostly western, commercial tracks ranging from 1922 to 2011, with a peak in the year 2000s. </td> -->
<td>

Multivariate

</td>
<td>

Regression

</td>

```
 </p>
</td>
<td>
 <p class="normal">
  Real
 </p>
</td>
<td>
 <p class="normal">
  515345
 </p>
</td>
<td>
 <p class="normal">
  90
 </p>
</td>
<td>
 <p class="normal">
  2011
 </p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/PEMS-SF">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/PEMS-SF">
        PEMS-SF
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">15 months worth of daily data (440 daily records) that describes the occupancy rate, between 0 and 1, of different car lanes of the San Francisco bay area freeways across time. </p></td> -->
 <td>
  <p class="normal">
   Multivariate, Time-Series
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
```

```
</td>
<td>
 <p class="normal">
  440
 </p>
</td>
<td>
 <p class="normal">
  138672
 </p>
</td>
<td>
 <p class="normal">
  2011
 </p>
</td>
<!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/OpinRank+Review+Dataset">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/OpinRank+Review+Dataset">
        OpinRank Review Dataset
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">This data set contains user reviews of cars and and hotels collected from Tripa
dvisor (~259,000
reviews) and Edmunds (~42,230 reviews).  &amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Text
  </p>
 </td>
 <td>
  <p class="normal">
  </p>
 </td>
 <td>
  <p class="normal">
  </p>
 </td>
 <td>
  <p class="normal">
  </p>
 </td>
 <td>
  <p class="normal">
```

</td>
<td>

2011

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Relative+location+of+CT+slices+on+axial+axis">

</a>
</td>
<td>

<a href="datasets/Relative+location+of+CT+slices+on+axial+axis">
Relative location of CT slices on axial axis
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The dataset consists of 384 features extracted from CT images. The class variable is numeric and denotes the relative location of the CT slice on the axial axis of the human body. </td> -->
<td>

Domain-Theory

</td>
<td>

Regression

</td>
<td>

Real

</td>
<td>

53500

</td>
<td>

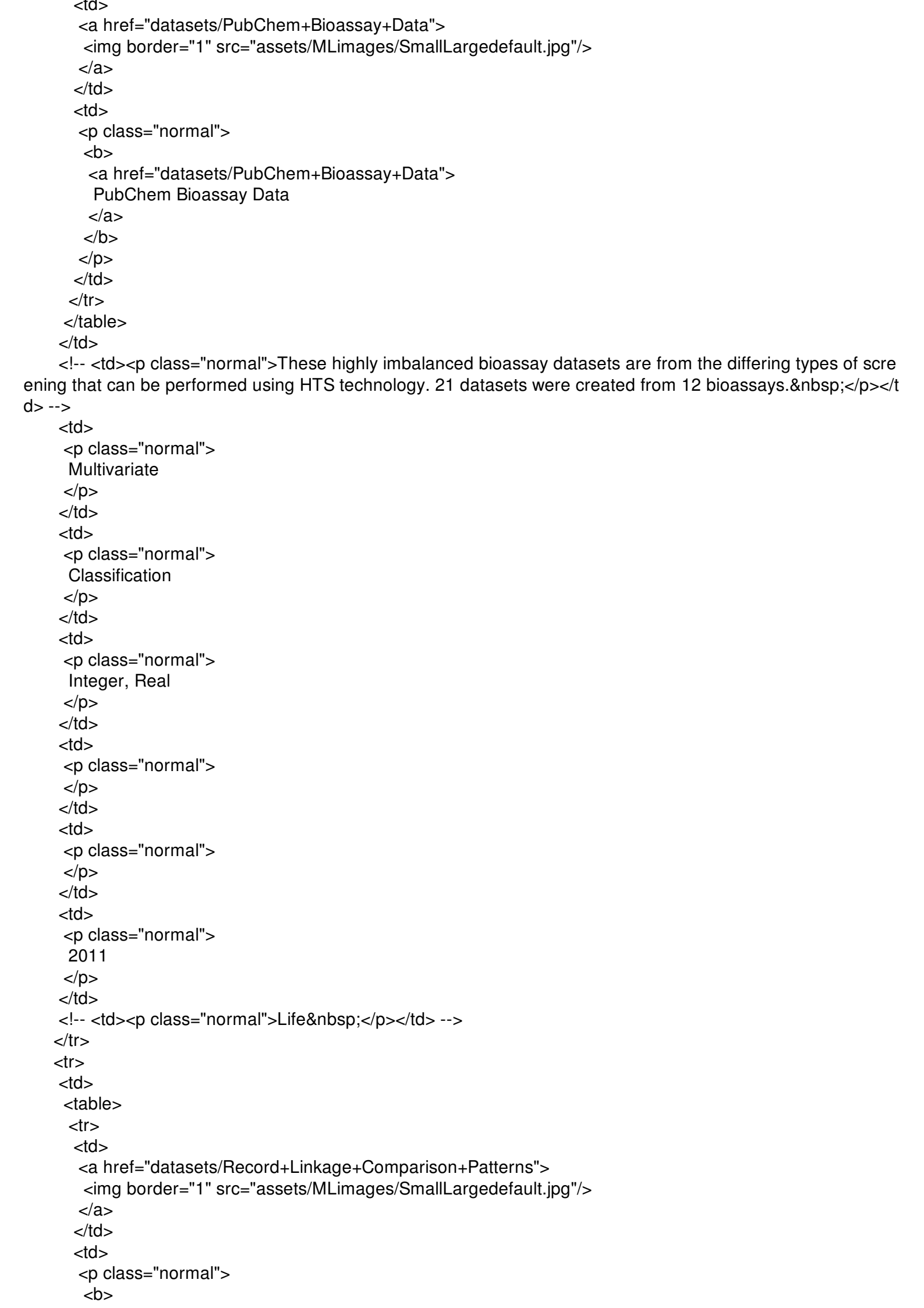
386

</td>
<td>

2011

</td>

```
 </td>
 <!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Online+Handwritten+Assamese+Characters+Dataset">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Online+Handwritten+Assamese+Characters+Dataset">
        Online Handwritten Assamese Characters Dataset
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">This is a dataset of 8235 online handwritten assamese characters. The “online” process involves capturing of data as text is written on a digitizing tablet with an electronic pen. </p></td> -->
 <td>
  <p class="normal">
   Multivariate, Sequential
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Integer
  </p>
 </td>
 <td>
  <p class="normal">
   8235
  </p>
 </td>
 <td>
  <p class="normal">
  </p>
 </td>
 <td>
  <p class="normal">
   2011
  </p>
 </td>
 <!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
```

<td>
<a href="datasets/PubChem+Bioassay+Data">

</a>
</td>
<td>

<a href="datasets/PubChem+Bioassay+Data">
PubChem Bioassay Data
</a>

</td>
</tr>
</table>
</td>
<!-- <td> These highly imbalanced bioassay datasets are from the differing types of screening that can be performed using HTS technology. 21 datasets were created from 12 bioassays. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer, Real

</td>
<td>

</td>
<td>

</td>
<td>

2011

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Record+Linkage+Comparison+Patterns">

</a>
</td>
<td>

<a href="datasets/Record+Linkage+Comparison+Patterns">
Record Linkage Comparison Patterns
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Element-wise comparison of records with personal data from a record linkage s
etting. The task is to decide from a comparison pattern whether the underlying records belong to one person.&
nbsp; </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

5749132

</td>
<td>

12

</td>
<td>

2011

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Communities+and+Crime+Unnormalized">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Communities+and+Crime+Unnormalized">
Communities and Crime Unnormalized
</a>

</td>

</tr>
</table>
</td>
<!-- <td> Communities in the US. Data combines socio-economic data from the '90 Cens us, law enforcement data from the 1990 Law Enforcement Management and Admin Stats survey, and crime dat a from the 1995 FBI UCR </td> -->
<td>

Multivariate

</td>
<td>

Regression

</td>
<td>

Real

</td>
<td>

2215

</td>
<td>

147

</td>
<td>

2011

</td>
<!-- <td> Social </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Vertebral+Column">

</a>
</td>
<td>

<a href="datasets/Vertebral+Column">
Vertebral Column
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Data set containing values for six biomechanical features used to classify ortho paedic patients into 3 classes (normal, disk hernia or spondilolysthesis) or 2 classes (normal or abnormal).&nbs p; </td> -->

<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

310

</td>
<td>

6

</td>
<td>

2011

</td>
<!-- <td>   </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/EMG+Physical+Action+Data+Set">

</a>
</td>
<td>

<a href="datasets/EMG+Physical+Action+Data+Set">
EMG Physical Action Data Set
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The Physical Action Data Set includes 10 normal and 10 aggressive physical a
ctions that measure the human activity. The data have been collected by 4 subjects using the Delsys EMG wirel
ess apparatus. </td> -->
<td>

Time-Series

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

10000

</td>
<td>

8

</td>
<td>

2011

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Vicon+Physical+Action+Data+Set">

</a>
</td>
<td>

<a href="datasets/Vicon+Physical+Action+Data+Set">
Vicon Physical Action Data Set
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The Physical Action Data Set includes 10 normal and 10 aggressive physical a
ctions that measure the human activity. The data have been collected by 10 subjects using the Vicon 3D tracke
r. </td> -->
<td>

Time-Series

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

3000

</td>
<td>

27

</td>
<td>

2011

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Amazon+Commerce+reviews+set">

</a>
</td>
<td>

<a href="datasets/Amazon+Commerce+reviews+set">
Amazon Commerce reviews set
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The dataset is used for authorship identification in online Writeprint which is a n
ew research field of pattern recognition.   </td> -->
<td>

Multivariate, Text, Domain-Theory

</td>
<td>

Classification

</td>
<td>

Real

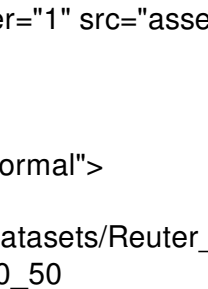
</td>
<td>

1500

</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

```
</p>
</td>
<td>
<p class="normal">
10000
</p>
</td>
<td>
<p class="normal">
2011
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Amazon+Access+Samples">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Amazon+Access+Samples">
Amazon Access Samples
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Amazon's InfoSec is getting smarter about the way Access data is leveraged. T
his is an anonymized sample of access provisioned within the company. </p></td> -->
<td>
<p class="normal">
Time-Series, Domain-Theory
</p>
</td>
<td>
<p class="normal">
Regression, Clustering, Causal-Discovery
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
30000
</p>
</td>
<td>
<p class="normal">
20000
</p>
</td>
<td>
<p class="normal">
```

2011

</td>
<!-- <td> Business </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Reuter_50_50">

</a>
</td>
<td>

<a href="datasets/Reuter_50_50">
Reuter_50_50
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The dataset is used for authorship identification in online Writeprint which is a n
ew research field of pattern recognition.   </td> -->
<td>

Multivariate, Text, Domain-Theory

</td>
<td>

Classification, Clustering

</td>
<td>

Real

</td>
<td>

2500

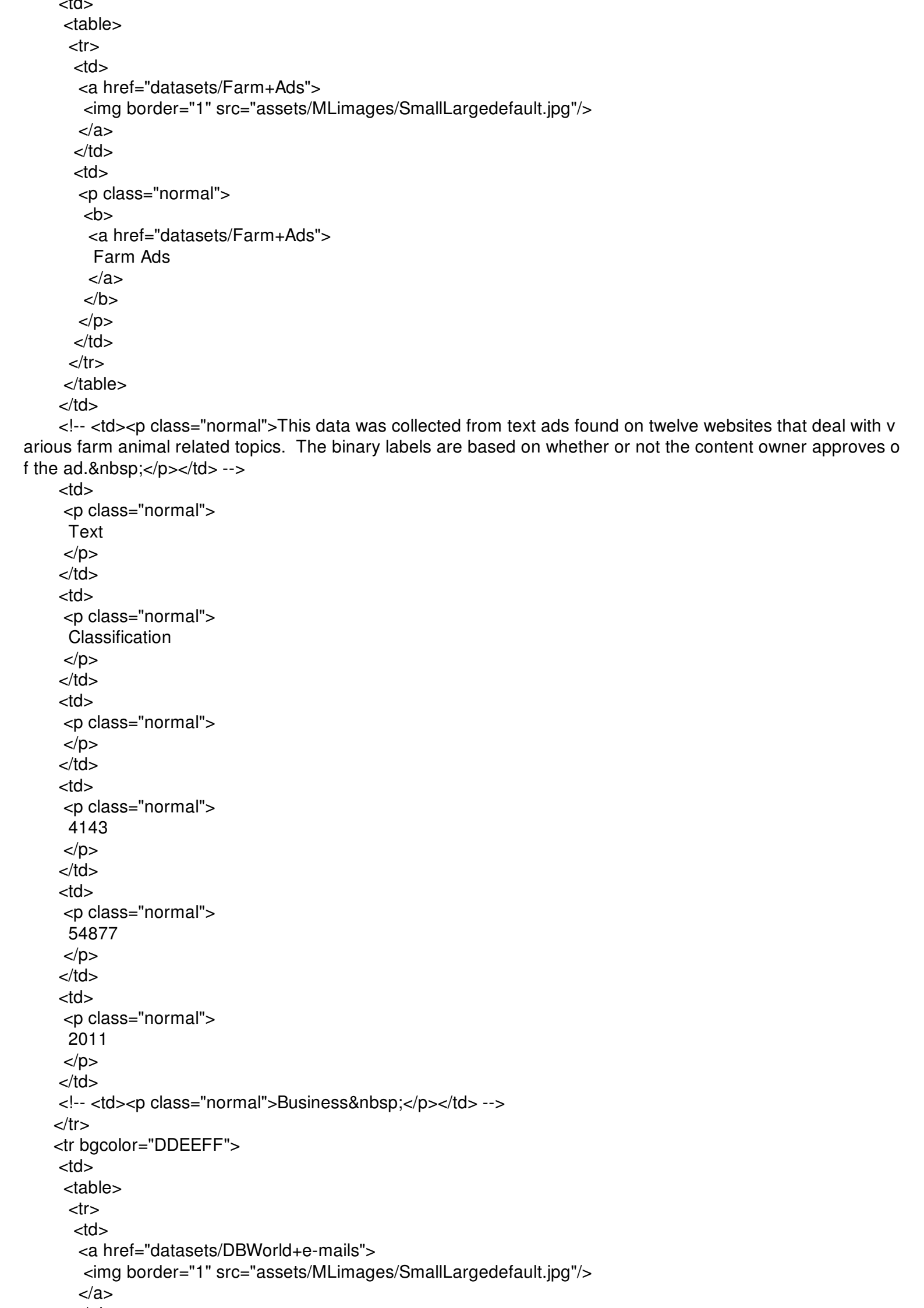
</td>
<td>

10000

</td>
<td>

2011

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>

<td>
<table>
<tr>
<td>
<a href="datasets/Farm+Ads">

</a>
</td>
<td>

<a href="datasets/Farm+Ads">
Farm Ads
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This data was collected from text ads found on twelve websites that deal with v
arious farm animal related topics. The binary labels are based on whether or not the content owner approves o
f the ad. </td> -->
<td>

Text

</td>
<td>

Classification

</td>
<td>

</td>
<td>

4143

</td>
<td>

54877

</td>
<td>

2011

</td>
<!-- <td> Business </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/DBWorld+e-mails">

</a>
</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

```
        </td>
        <td>
         <p class="normal">
          <b>
           <a href="datasets/DBWorld+e-mails">
            DBWorld e-mails
           </a>
          </b>
         </p>
        </td>
       </tr>
      </table>
     </td>
     <!-- <td><p class="normal">It contains 64 e-mails which I have manually collected from DBWorld mailing lis
t. They are classified in: 'announces of conferences' and 'everything else'.&amp;nbsp;</p></td> -->
     <td>
      <p class="normal">
       Text
      </p>
     </td>
     <td>
      <p class="normal">
       Classification
      </p>
     </td>
     <td>
      <p class="normal">
      </p>
     </td>
     <td>
      <p class="normal">
       64
      </p>
     </td>
     <td>
      <p class="normal">
       4702
      </p>
     </td>
     <td>
      <p class="normal">
       2011
      </p>
     </td>
     <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
    </tr>
    <tr>
     <td>
      <table>
       <tr>
        <td>
         <a href="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29">
          <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
         </a>
        </td>
        <td>
         <p class="normal">
          <b>
           <a href="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29">
            KEGG Metabolic Relation Network (Directed)
           </a>
          </b>
```

```
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">KEGG Metabolic pathways modeled as directed relation network. Variety of gra
phical features presented.&amp;nbsp;</p></td> -->
      <td>
       <p class="normal">
        Multivariate, Univariate, Text
       </p>
      </td>
      <td>
       <p class="normal">
        Classification, Regression, Clustering
       </p>
      </td>
      <td>
       <p class="normal">
        Integer, Real
       </p>
      </td>
      <td>
       <p class="normal">
        53414
       </p>
      </td>
      <td>
       <p class="normal">
        24
       </p>
      </td>
      <td>
       <p class="normal">
        2011
       </p>
      </td>
      <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
     </tr>
     <tr bgcolor="DDEEFF">
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29">
             KEGG Metabolic Reaction Network (Undirected)
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">KEGG Metabolic pathways modeled as un-directed reaction network. Variety of
graphical features presented.&amp;nbsp;</p></td> -->
```

<td>

Multivariate, Univariate, Text

</td>
<td>

Classification, Regression, Clustering

</td>
<td>

Integer, Real

</td>
<td>

65554

</td>
<td>

29

</td>
<td>

2011

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Bank+Marketing">

</a>
</td>
<td>

<a href="datasets/Bank+Marketing">
Bank Marketing
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The data is related with direct marketing campaigns (phone calls) of a Portugu
ese banking institution. The classification goal is to predict if the client will subscribe a term deposit (variable y).
  </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

45211

</td>
<td>

17

</td>
<td>

2012

</td>
<!-- <td> Business </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/YouTube+Comedy+Slam+Preference+Data">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/YouTube+Comedy+Slam+Preference+Data">
YouTube Comedy Slam Preference Data
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset provides user vote data on which video from a pair of videos is fun
nier collected on YouTube Comedy Slam. The task is to automatically predict this preference based on video m
etadata. </td> -->
<td>

Text

</td>
<td>

Classification

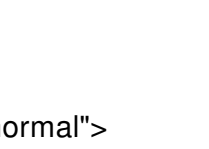
</td>
<td>

```
 </p>
</td>
<td>
 <p class="normal">
  1138562
 </p>
</td>
<td>
 <p class="normal">
  3
 </p>
</td>
<td>
 <p class="normal">
  2012
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Gas+Sensor+Array+Drift+Dataset">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Gas+Sensor+Array+Drift+Dataset">
        Gas Sensor Array Drift Dataset
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">This archive contains 13910 measurements from 16 chemical sensors utilized i
n simulations for drift compensation in a discrimination task of 6 gases at various levels of concentrations.&nbs
p;</p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
 </td>
 <td>
  <p class="normal">
   13910
  </p>
```

```
</td>
<td>
<p class="normal">
128
</p>
</td>
<td>
<p class="normal">
2012
</p>
</td>
<!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/ILPD+%28Indian+Liver+Patient+Dataset%29">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/ILPD+%28Indian+Liver+Patient+Dataset%29">
ILPD (Indian Liver Patient Dataset)
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This data set contains 10 variables that are age, gender, total Bilirubin, direct B
ilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT and Alkphos.&amp;nbsp;</p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Integer, Real
</p>
</td>
<td>
<p class="normal">
583
</p>
</td>
<td>
<p class="normal">
10
</p>
</td>
<td>
```

<td>

2012

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/OPPORTUNITY+Activity+Recognition">

</a>
</td>
<td>

<a href="datasets/OPPORTUNITY+Activity+Recognition">
OPPORTUNITY Activity Recognition
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The OPPORTUNITY Dataset for Human Activity Recognition from Wearable, O
bject, and Ambient Sensors is a dataset devised to benchmark human activity recognition algorithms (classificat
ion, automatic data segmentation, sensor fusion, feature extraction, etc). </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

2551

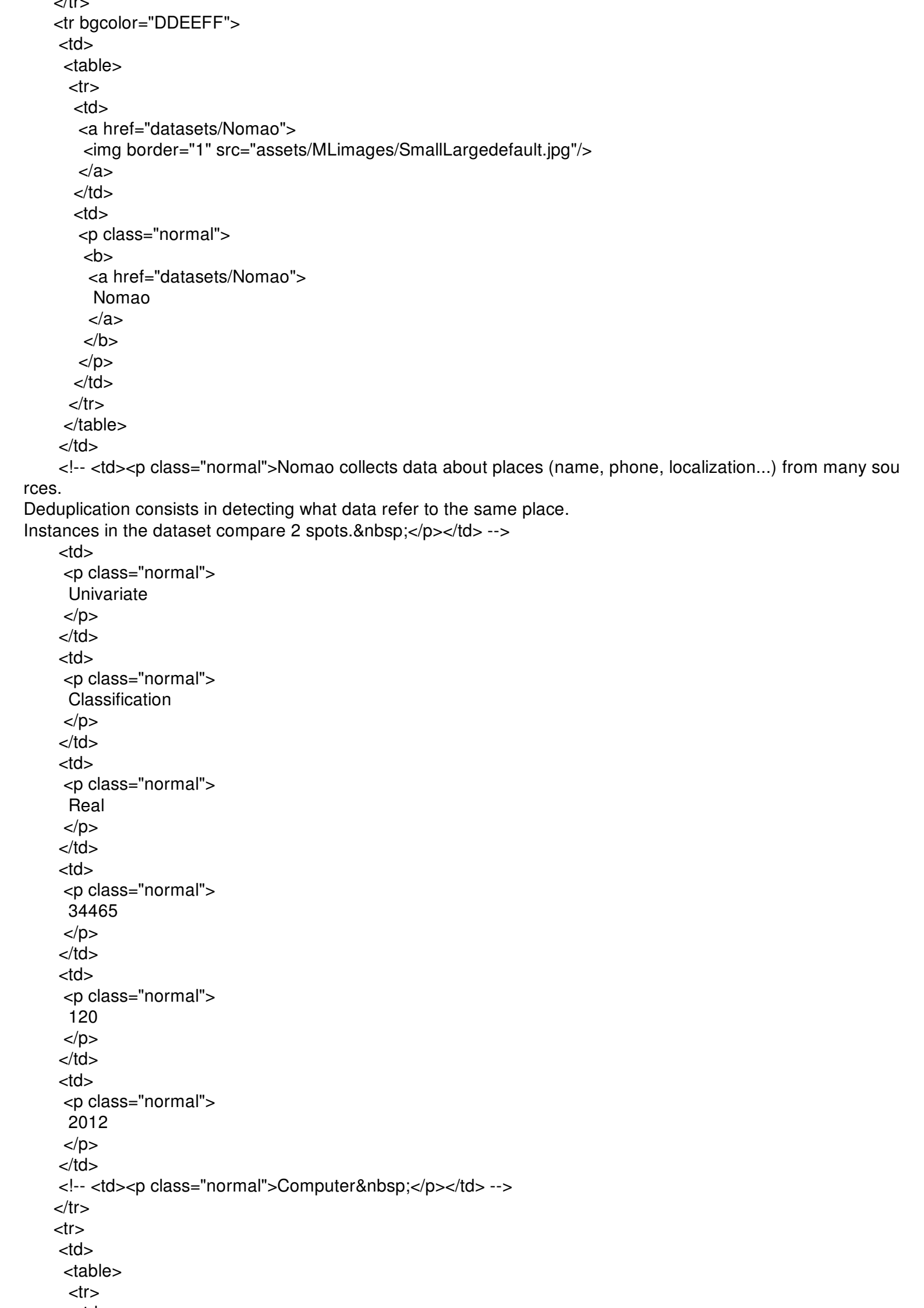
</td>
<td>

242

</td>
<td>

2012

</td>
<!-- <td> Computer </td> -->
</tr>

</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Nomao">

</a>
</td>
<td>

<a href="datasets/Nomao">
Nomao
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Nomao collects data about places (name, phone, localization...) from many sou
rces.
Deduplication consists in detecting what data refer to the same place.
Instances in the dataset compare 2 spots. </td> -->
<td>

Univariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

34465

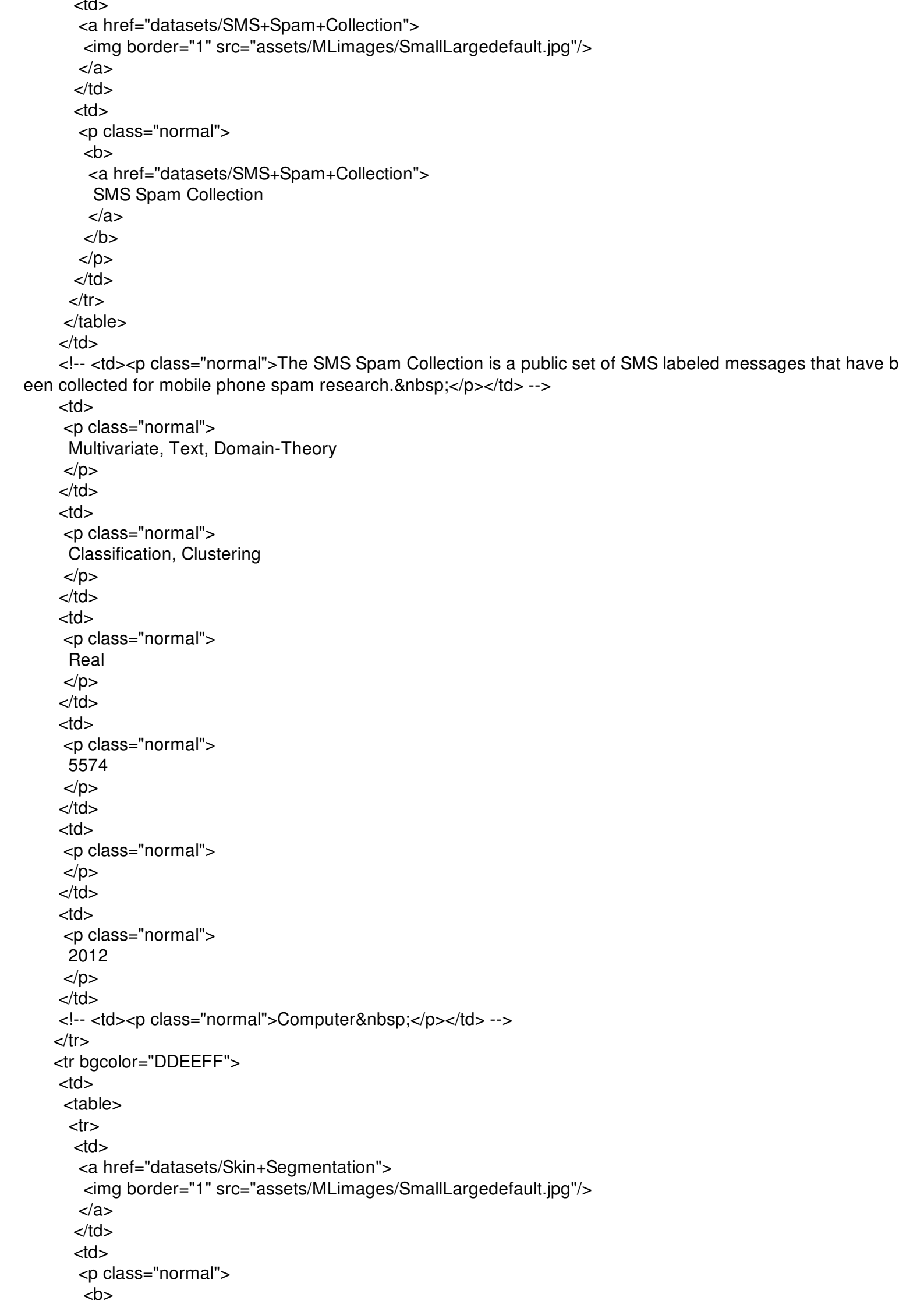
</td>
<td>

120

</td>
<td>

2012

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>

<td>
<a href="datasets/SMS+Spam+Collection">

</a>
</td>
<td>

<a href="datasets/SMS+Spam+Collection">
SMS Spam Collection
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The SMS Spam Collection is a public set of SMS labeled messages that have b
een collected for mobile phone spam research. </td> -->
<td>

Multivariate, Text, Domain-Theory

</td>
<td>

Classification, Clustering

</td>
<td>

Real

</td>
<td>

5574

</td>
<td>

</td>
<td>

2012

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Skin+Segmentation">

</a>
</td>
<td>

<a href="datasets/Skin+Segmentation">
Skin Segmentation
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The Skin Segmentation dataset is constructed over B, G, R color space. Skin a
nd Nonskin dataset is generated using skin textures from face images of diversity of age, gender, and race peo
ple. </td> -->
<td>

Univariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

245057

</td>
<td>

4

</td>
<td>

2012

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Planning+Relax">

</a>
</td>
<td>

<a href="datasets/Planning+Relax">
Planning Relax
</a>

</td>

```
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">The dataset concerns with the classification of two mental stages from recorde
d EEG signals: Planning (during imagination of motor act) and Relax state. &amp;nbsp;</p></td> -->
    <td>
     <p class="normal">
      Univariate
     </p>
    </td>
    <td>
     <p class="normal">
      Classification
     </p>
    </td>
    <td>
     <p class="normal">
      Real
     </p>
    </td>
    <td>
     <p class="normal">
      182
     </p>
    </td>
    <td>
     <p class="normal">
      13
     </p>
    </td>
    <td>
     <p class="normal">
      2012
     </p>
    </td>
    <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
   </tr>
   <tr bgcolor="DDEEFF">
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/PAMAP2+Physical+Activity+Monitoring">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/PAMAP2+Physical+Activity+Monitoring">
           PAMAP2 Physical Activity Monitoring
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">The PAMAP2 Physical Activity Monitoring dataset contains data of 18 different
physical activities, performed by 9 subjects wearing 3 inertial measurement units and a heart rate monitor.&nbs
p;</p></td> -->
    <td>
```

Multivariate, Time-Series

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

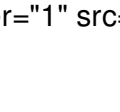
3850505

</td>
<td>

52

</td>
<td>

2012

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Restaurant+%26+consumer+data">

</a>
</td>
<td>

<a href="datasets/Restaurant+%26+consumer+data">
Restaurant & consumer data
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The dataset was obtained from a recommender system prototype. The task wa
s to generate a top-n list of restaurants according to the consumer preferences.   </td> -->
<td>

Multivariate

</td>
<td>

</td>

```
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
  138
 </p>
</td>
<td>
 <p class="normal">
  47
 </p>
</td>
<td>
 <p class="normal">
  2012
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/CNAE-9">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/CNAE-9">
        CNAE-9
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">This is a data set containing 1080 documents of free text business descriptions
of Brazilian companies categorized into a
subset of 9 categories </p></td> -->
 <td>
  <p class="normal">
   Multivariate, Text
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Integer
  </p>
 </td>
 <td>
```

1080

</td>
<td>

857

</td>
<td>

2012

</td>
<!-- <td> Business </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Individual+household+electric+power+consumption">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Individual+household+electric+power+consumption">
Individual household electric power consumption
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Measurements of electric power consumption in one household with a one-min
ute sampling rate over a period of almost 4 years. Different electrical quantities and some sub-metering values
are available. </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

Regression, Clustering

</td>
<td>

Real

</td>
<td>

2075259

</td>
<td>

9

</td>
<td>

2012

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/seeds">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/seeds">
seeds
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Measurements of geometrical properties of kernels belonging to three different varieties of wheat. A soft X-ray technique and GRAINS package were used to construct all seven, real-valued at tributes. </td> -->
<td>

Multivariate

</td>
<td>

Classification, Clustering

</td>
<td>

Real

</td>
<td>

210

</td>
<td>

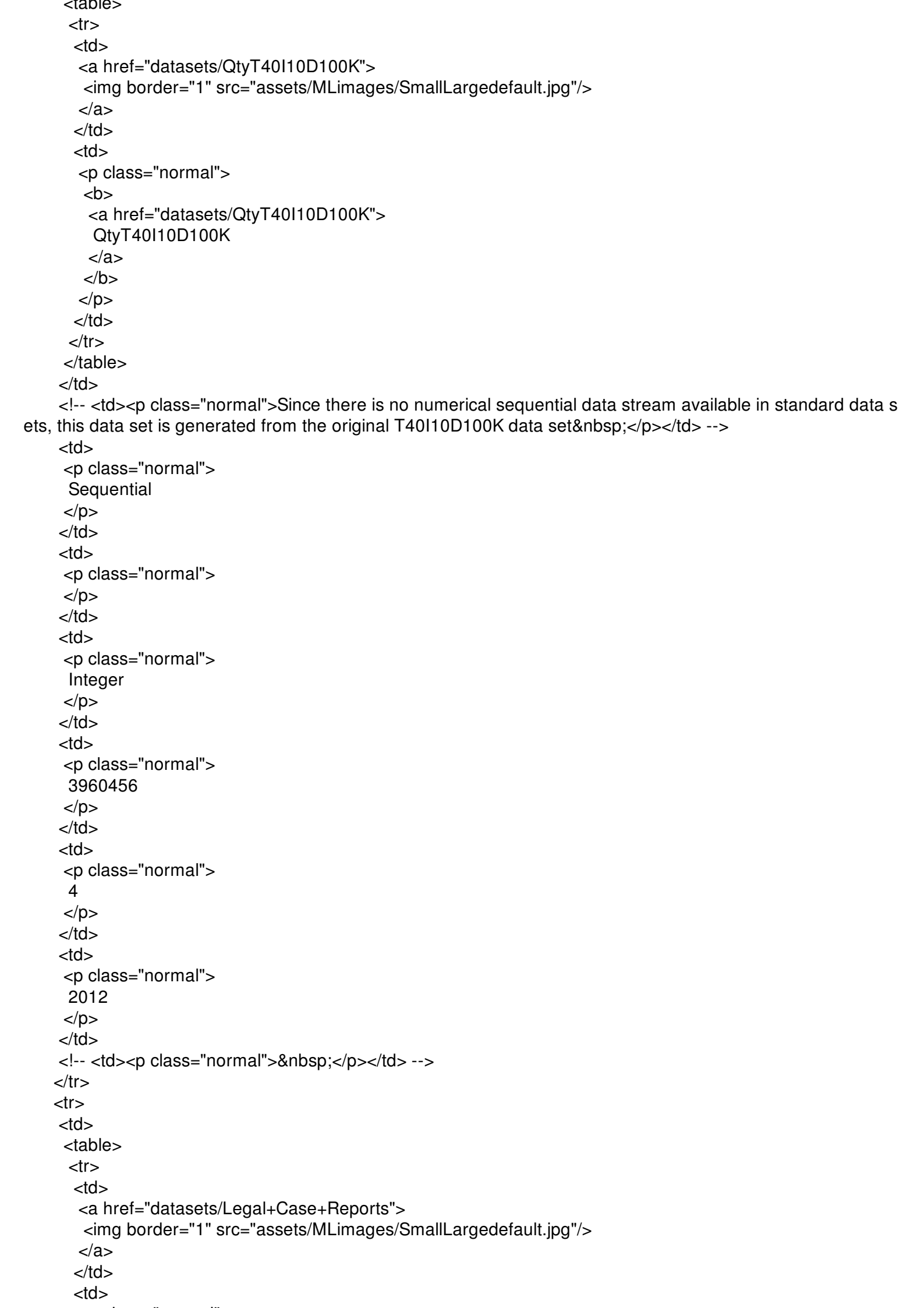
7

</td>
<td>

2012

</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

```
      2012
     </p>
    </td>
    <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
   </tr>
   <tr>
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/Northix">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Northix">
           Northix
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">Northix is designed to be a schema matching benchmark problem for data inte
gration of two entity relationship databases. &amp;nbsp;</p></td> -->
    <td>
     <p class="normal">
      Multivariate, Univariate, Text
     </p>
    </td>
    <td>
     <p class="normal">
      Classification
     </p>
    </td>
    <td>
     <p class="normal">
      Integer, Real
     </p>
    </td>
    <td>
     <p class="normal">
      115
     </p>
    </td>
    <td>
     <p class="normal">
      200
     </p>
    </td>
    <td>
     <p class="normal">
      2012
     </p>
    </td>
    <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
   </tr>
   <tr bgcolor="DDEEFF">
    <td>
```

<table>
<tr>
<td>
<a href="datasets/QtyT40I10D100K">

</a>
</td>
<td>

<a href="datasets/QtyT40I10D100K">
QtyT40I10D100K
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Since there is no numerical sequential data stream available in standard data s
ets, this data set is generated from the original T40I10D100K data set </td> -->
<td>

Sequential

</td>
<td>

</td>
<td>

Integer

</td>
<td>

3960456

</td>
<td>

4

</td>
<td>

2012

</td>
<!-- <td>   </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Legal+Case+Reports">

</a>
</td>
<td>

<a href="datasets/Legal+Case+Reports">
Legal Case Reports
</a>

</td>
</tr>
</table>
</td>
<!-- <td> A textual corpus of 4000 legal cases for automatic summarization and citation analysis. For each document we collect catchphrases, citations sentences, citation catchphrases and citation classes. </td> -->
<td>

Text

</td>
<td>

Classification

</td>
<td>

</td>
<td>

</td>
<td>

</td>
<td>

2012

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Human+Activity+Recognition+Using+Smartphones">

</a>
</td>
<td>

<a href="datasets/Human+Activity+Recognition+Using+Smartphones">
Human Activity Recognition Using Smartphones
</a>

</td>
</tr>

</table>
</td>
<!-- <td> Human Activity Recognition database built from the recordings of 30 subjects p
erforming activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sen
sors. </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

Classification, Clustering

</td>
<td>

</td>
<td>

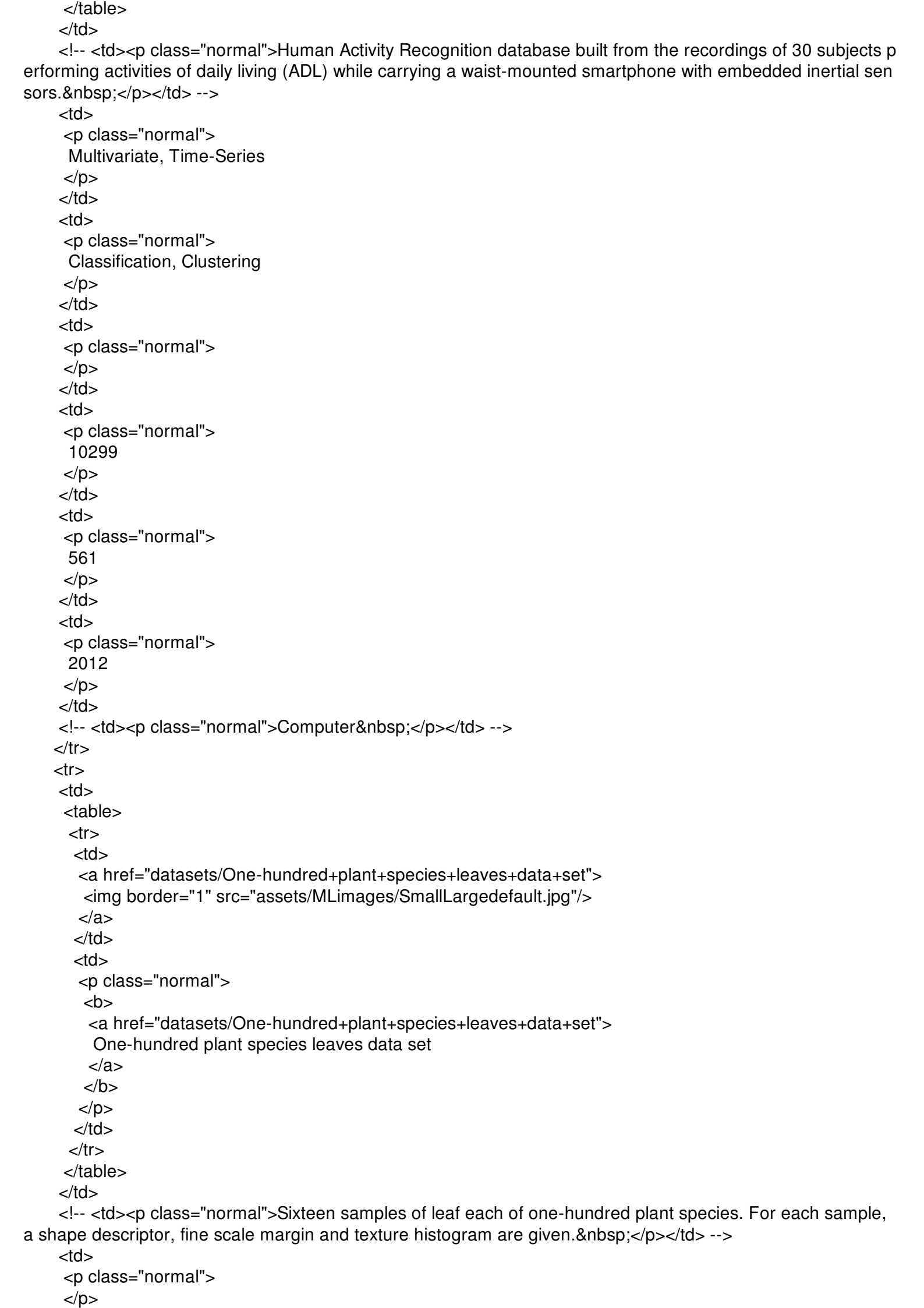
10299

</td>
<td>

561

</td>
<td>

2012

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/One-hundred+plant+species+leaves+data+set">

</a>
</td>
<td>

<a href="datasets/One-hundred+plant+species+leaves+data+set">
One-hundred plant species leaves data set
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Sixteen samples of leaf each of one-hundred plant species. For each sample,
a shape descriptor, fine scale margin and texture histogram are given. </td> -->
<td>

```
</td>
<td>
 <p class="normal">
  Classification
 </p>
</td>
<td>
 <p class="normal">
  Real
 </p>
</td>
<td>
 <p class="normal">
  1600
 </p>
</td>
<td>
 <p class="normal">
  64
 </p>
</td>
<td>
 <p class="normal">
  2012
 </p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Energy+efficiency">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Energy+efficiency">
        Energy efficiency
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">This study looked into assessing the heating load and cooling load requirement
s of buildings (that is, energy efficiency) as a function of building parameters. </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification, Regression
  </p>
 </td>
 <td>
```

Integer, Real

</td>
<td>

768

</td>
<td>

8

</td>
<td>

2012

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Yacht+Hydrodynamics">

</a>
</td>
<td>

<a href="datasets/Yacht+Hydrodynamics">
Yacht Hydrodynamics
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Delft data set, used to predict the hydodynamic performance of sailing yachts fr
om dimensions and velocity. </td> -->
<td>

Multivariate

</td>
<td>

Regression

</td>
<td>

Real

</td>
<td>

308

</td>
</tr>
</table>
</body>
</html>

```
 </p>
</td>
<td>
 <p class="normal">
  7
 </p>
</td>
<td>
 <p class="normal">
  2013
 </p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
 <table>
  <tr>
   <td>
    <a href="datasets/Fertility">
     <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   </td>
   <td>
    <p class="normal">
     <b>
      <a href="datasets/Fertility">
       Fertility
      </a>
     </b>
    </p>
   </td>
  </tr>
 </table>
</td>
<!-- <td><p class="normal">100 volunteers provide a semen sample analyzed according to the WHO 2010 criteria. Sperm concentration are related to socio-demographic data, environmental factors, health status, and li fe habits </p></td> -->
<td>
 <p class="normal">
  Multivariate
 </p>
</td>
<td>
 <p class="normal">
  Classification, Regression
 </p>
</td>
<td>
 <p class="normal">
  Real
 </p>
</td>
<td>
 <p class="normal">
  100
 </p>
</td>
<td>
 <p class="normal">
  10
 </p>
```

```
</td>
<td>
<p class="normal">
2013
</p>
</td>
<!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Daphnet+Freezing+of+Gait">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Daphnet+Freezing+of+Gait">
Daphnet Freezing of Gait
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset contains the annotated readings of 3 acceleration sensors at the h
ip and leg of Parkinson's disease patients that experience freezing of gait (FoG) during walking tasks.
&amp;nbsp;</p></td> -->
<td>
<p class="normal">
Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Real
</p>
</td>
<td>
<p class="normal">
237
</p>
</td>
<td>
<p class="normal">
9
</p>
</td>
<td>
<p class="normal">
2013
</p>
</td>
```

<td>
<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29">
3D Road Network (North Jutland, Denmark)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> 3D road network with highly accurate elevation information (+-20cm) from Den
mark used in eco-routing and fuel/Co2-estimation routing algorithms. </td> -->
<td>

Sequential, Text

</td>
<td>

Regression, Clustering

</td>
<td>

Real

</td>
<td>

434874

</td>
<td>

4

</td>
<td>

2013

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>
</table>

<td>
<a href="datasets/ISTANBUL+STOCK+EXCHANGE">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/ISTANBUL+STOCK+EXCHANGE">
ISTANBUL STOCK EXCHANGE
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Data sets includes returns of Istanbul Stock Exchange with seven other internat
ional index; SP, DAX, FTSE, NIKKEI, BOVESPA, MSCE_EU, MSCI_EM from Jun 5, 2009 to Feb 22, 2011.&nbs
p; </td> -->
<td>

Multivariate, Univariate, Time-Series

</td>
<td>

Classification, Regression

</td>
<td>

Real

</td>
<td>

536

</td>
<td>

8

</td>
<td>

2013

</td>
<!-- <td> Business </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Buzz+in+social+media+">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Buzz+in+social+media+">
Buzz in social media
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This data-set contains examples of buzz events from two different social networks: Twitter, and Tom's Hardware, a forum network focusing on new technology with more conservative dynamics. </td> -->
<td>

Time-Series, Multivariate

</td>
<td>

Regression, Classification

</td>
<td>

Integer, Real

</td>
<td>

140000

</td>
<td>

77

</td>
<td>

2013

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/First-order+theorem+proving">

</a>
</td>
<td>

<a href="datasets/First-order+theorem+proving">
First-order theorem proving
</a>

```
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">Given a theorem, predict which of five heuristics will give the fastest proof when used by a first-order prover. A sixth prediction declines to attempt a proof, should the theorem be too difficult. </p></td> -->
      <td>
       <p class="normal">
        Multivariate
       </p>
      </td>
      <td>
       <p class="normal">
        Classification
       </p>
      </td>
      <td>
       <p class="normal">
        Real
       </p>
      </td>
      <td>
       <p class="normal">
        6118
       </p>
      </td>
      <td>
       <p class="normal">
        51
       </p>
      </td>
      <td>
       <p class="normal">
        2013
       </p>
      </td>
      <!-- <td><p class="normal">Computer </p></td> -->
     </tr>
     <tr bgcolor="DDEEFF">
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29">
             Wearable Computing: Classification of Body Postures and Movements (PUC-Rio)
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
```

</td>
<!-- <td> A dataset with 5 classes (sitting-down, standing-up, standing, walking, and sitting) collected on 8 hours of activities of 4 healthy subjects. We also established a baseline performance index. </td> -->
<td>

Sequential

</td>
<td>

Classification

</td>
<td>

Integer, Real

</td>
<td>

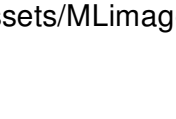
165632

</td>
<td>

18

</td>
<td>

2013

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Gas+sensor+arrays+in+open+sampling+settings">

</a>
</td>
<td>

<a href="datasets/Gas+sensor+arrays+in+open+sampling+settings">
Gas sensor arrays in open sampling settings
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The dataset contains 18000 time-series recordings from a chemical detection platform at six different locations in a wind tunnel facility in response to ten high-priority chemical gaseous substances </td> -->
<td>

```
 Multivariate, Time-Series
 </p>
</td>
<td>
 <p class="normal">
  Classification
 </p>
</td>
<td>
 <p class="normal">
  Real
 </p>
</td>
<td>
 <p class="normal">
  18000
 </p>
</td>
<td>
 <p class="normal">
  1950000
 </p>
</td>
<td>
 <p class="normal">
  2013
 </p>
</td>
<!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Climate+Model+Simulation+Crashes">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Climate+Model+Simulation+Crashes">
        Climate Model Simulation Crashes
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Given Latin hypercube samples of 18 climate model input parameter values, pr
edict climate model simulation crashes and determine the parameter value combinations that cause the failures
.&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
```

```
 </p>
</td>
<td>
 <p class="normal">
  Real
 </p>
</td>
<td>
 <p class="normal">
  540
 </p>
</td>
<td>
 <p class="normal">
  18
 </p>
</td>
<td>
 <p class="normal">
  2013
 </p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/MicroMass">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/MicroMass">
        MicroMass
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">A dataset to explore machine learning approaches for the identification of micr
oorganisms from mass-spectrometry data. </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
 </td>
```

<td>

931

</td>
<td>

1300

</td>
<td>

2013

</td>
<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/QSAR+biodegradation">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/QSAR+biodegradation">
QSAR biodegradation
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Data set containing values for 41 attributes (molecular descriptors) used to clas
sify 1055 chemicals into 2 classes (ready and not ready biodegradable). </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer, Real

</td>
<td>

1055

</td>
<td>

41

</td>
<td>

2013

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/BLOGGER">

</a>
</td>
<td>

<a href="datasets/BLOGGER">
BLOGGER
</a>

</td>
</tr>
</table>
</td>
<!-- <td> In this paper, we look for to recognize the causes of users tend to cyber space in Kohkiloye and Boyer Ahmad Province in Iran </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

</td>
<td>

100

</td>
<td>

6

</td>
<td>

2013

</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

```
 </p>
 </td>
 <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Daily+and+Sports+Activities">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Daily+and+Sports+Activities">
        Daily and Sports Activities
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">The dataset comprises motion sensor data of 19 daily and sports activities eac
h performed by 8 subjects in their own style for 5 minutes. Five Xsens MTx units are used on the torso, arms, a
nd legs.
&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate, Time-Series
  </p>
 </td>
 <td>
  <p class="normal">
   Classification, Clustering
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
 </td>
 <td>
  <p class="normal">
   9120
  </p>
 </td>
 <td>
  <p class="normal">
   5625
  </p>
 </td>
 <td>
  <p class="normal">
   2013
  </p>
 </td>
 <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
</tr>
<tr>
```

<td>
<table>
<tr>
<td>
<a href="datasets/User+Knowledge+Modeling">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/User+Knowledge+Modeling">
User Knowledge Modeling
</a>

</td>
</tr>
</table>
</td>
<!-- <td> It is the real dataset about the students' knowledge status about the subject of Electrical DC Machines. The dataset had been obtained from Ph.D. Thesis. </td> -->
<td>

Multivariate

</td>
<td>

Classification, Clustering

</td>
<td>

Integer

</td>
<td>

403

</td>
<td>

5

</td>
<td>

2013

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection">
Reuters RCV1 RCV2 Multilingual, Multiview Text Categorization Test collection
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This is a test collection for multilingual, multiview text categorization. The original data comes from the Reuters RCV1/RCV2 collections. The collection contains documents written in five languages (English, French, German, Italian, Spanish) and labeled according to a four-class category system (corporate/industrial, economics, government/social, markets). For each language version, documents are represented in two views (i.e., two different sets of attributes). </td> -->
<td>

Multivariate, Multilingual, Multiview

</td>
<td>

Classification, Clustering

</td>
<td>

Integer, Real

</td>
<td>

111740

</td>
<td>

21531

</td>
<td>

2009

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection">
Reuters RCV1 RCV2 Multilingual, Multiview Text Categorization Test collection
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This is a test collection for multilingual, multiview text categorization. The original data comes from the Reuters RCV1/RCV2 collections. The collection contains documents written in five languages (English, French, German, Italian, Spanish) and labeled according to a four-class category system (corporate/industrial, economics, government/social, markets). For each language version, documents are represented in two views (i.e., two different sets of attributes). </td> -->
<td>

Multivariate, Multilingual, Multiview

</td>
<td>

Classification, Clustering

</td>
<td>

Integer, Real

</td>
<td>

111740

</td>
<td>

21531

</td>
<td>

2009

</td>
<!-- <td> Computer </td> -->
</tr>

```
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collect
ion">
             Reuters RCV1 RCV2 Multilingual, Multiview Text Categorization Test collection
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">This test collection contains feature characteristics of documents originally writt
en in five different languages and their translations, over a common set of 6 categories. &amp;nbsp;</p></td> -->
      <td>
       <p class="normal">
        Multivariate
       </p>
      </td>
      <td>
       <p class="normal">
        Classification
       </p>
      </td>
      <td>
       <p class="normal">
        Real
       </p>
      </td>
      <td>
       <p class="normal">
        111740
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
        2013
       </p>
      </td>
      <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
     </tr>
     <tr>
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/NYSK">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/NYSK">
             NYSK
```

```
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">NYSK (New York v. Strauss-Kahn) is a collection of English news articles about the case relating to allegations of sexual assault against the former IMF director Dominique Strauss-Kahn (May 2011). </p></td> -->
    <td>
     <p class="normal">
      Multivariate, Sequential, Text
     </p>
    </td>
    <td>
     <p class="normal">
      Clustering
     </p>
    </td>
    <td>
     <p class="normal">
     </p>
    </td>
    <td>
     <p class="normal">
      10421
     </p>
    </td>
    <td>
     <p class="normal">
      7
     </p>
    </td>
    <td>
     <p class="normal">
      2013
     </p>
    </td>
    <!-- <td><p class="normal">Social </p></td> -->
   </tr>
   <tr bgcolor="DDEEFF">
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/Turkiye+Student+Evaluation">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Turkiye+Student+Evaluation">
           Turkiye Student Evaluation
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
```

<!-- <td> This data set contains a total 5820 evaluation scores provided by students fro
m Gazi University in Ankara (Turkey). There is a total of 28 course specific questions and additional 5 attributes
. </td> -->
<td>

Multivariate

</td>
<td>

Classification, Clustering

</td>
<td>

</td>
<td>

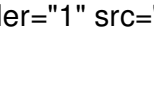
5820

</td>
<td>

33

</td>
<td>

2013

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electr
ical+Machines%29">

</a>
</td>
<td>

<a href="datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Elec
trical+Machines%29">
ser Knowledge Modeling Data (Students' Knowledge Levels on DC Electrical Machines)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The dataset is about the users' learning activities and knowledge levels on subj
ects of DC Electrical Machines. The dataset had been obtained from online web-courses and reported in my P
h.D. Thesis. </td> -->
<td>

```
      Multivariate
     </p>
    </td>
    <td>
     <p class="normal">
      Classification
     </p>
    </td>
    <td>
     <p class="normal">
      Real
     </p>
    </td>
    <td>
     <p class="normal">
      403
     </p>
    </td>
    <td>
     <p class="normal">
      5
     </p>
    </td>
    <td>
     <p class="normal">
      2013
     </p>
    </td>
    <!-- <td><p class="normal">Computer </p></td> -->
   </tr>
   <tr bgcolor="DDEEFF">
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/EEG+Eye+State">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/EEG+Eye+State">
           EEG Eye State
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">The data set consists of 14 EEG values and a value indicating the eye state.&n
bsp;</p></td> -->
    <td>
     <p class="normal">
      Multivariate, Sequential, Time-Series
     </p>
    </td>
    <td>
     <p class="normal">
      Classification
     </p>
```

```
</td>
<td>
 <p class="normal">
  Integer, Real
 </p>
</td>
<td>
 <p class="normal">
  14980
 </p>
</td>
<td>
 <p class="normal">
  15
 </p>
</td>
<td>
 <p class="normal">
  2013
 </p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure">
        Physicochemical Properties of Protein Tertiary Structure
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">This is a data set of Physicochemical Properties of Protein Tertiary Structure. T
he data set is taken from CASP 5-9. There are 45730 decoys and size varying from 0 to 21 armstrong. </
p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Regression
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
 </td>
```

<td>

45730

</td>
<td>

9

</td>
<td>

2013

</td>
<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/seismic-bumps">

</a>
</td>
<td>

<a href="datasets/seismic-bumps">
seismic-bumps
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The data describe the problem of high energy (higher than 10^4 J) seismic bu
mps forecasting in a coal
mine. Data come from two of longwalls located in a Polish coal mine. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

2584

</td>
<td>

<td>

19

</td>
<td>

2013

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/banknote+authentication">

</a>
</td>
<td>

<a href="datasets/banknote+authentication">
banknote authentication
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Data were extracted from images that were taken for the evaluation of an auth
entication procedure for banknotes. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

1372

</td>
<td>

5

</td>
<td>

```
      2013
     </p>
    </td>
    <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
   </tr>
   <tr bgcolor="DDEEFF">
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+T
opCoder++++Problem%3A+Pat">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+
TopCoder++++Problem%3A+Pat">
           USPTO Algorithm Challenge, run by NASA-Harvard Tournament Lab and TopCoder     Problem: Pat
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">Data used for USPTO Algorithm Competition. Contains drawing pages from US
patents with manually labeled figure and part labels.&amp;nbsp;</p></td> -->
    <td>
     <p class="normal">
      Domain-Theory
     </p>
    </td>
    <td>
     <p class="normal">
      Classification
     </p>
    </td>
    <td>
     <p class="normal">
      Integer
     </p>
    </td>
    <td>
     <p class="normal">
      306
     </p>
    </td>
    <td>
     <p class="normal">
      5
     </p>
    </td>
    <td>
     <p class="normal">
      2013
     </p>
    </td>
    <!-- <td><p class="normal">Other&amp;nbsp;</p></td> -->
   </tr>
```

<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/YouTube+Multiview+Video+Games+Dataset">

</a>
</td>
<td>

<a href="datasets/YouTube+Multiview+Video+Games+Dataset">
YouTube Multiview Video Games Dataset
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset contains about 120k instances, each described by 13 feature type
s, with class information, specially useful for exploring multiview topics (cotraining, ensembles, clustering,..).&nb
sp; </td> -->
<td>

Multivariate, Text

</td>
<td>

Classification, Clustering

</td>
<td>

Integer, Real

</td>
<td>

120000

</td>
<td>

1000000

</td>
<td>

2013

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations">

```
          <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
         </a>
        </td>
        <td>
         <p class="normal">
          <b>
           <a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations">
            Gas Sensor Array Drift Dataset at Different Concentrations
           </a>
          </b>
         </p>
        </td>
       </tr>
      </table>
     </td>
     <!-- <td><p class="normal">This archive contains 13910 measurements from 16 chemical sensors expose
d to 6 different gases at various concentration levels.&amp;nbsp;</p></td> -->
     <td>
      <p class="normal">
       Multivariate, Time-Series
      </p>
     </td>
     <td>
      <p class="normal">
       Classification, Regression, Clustering, Causa
      </p>
     </td>
     <td>
      <p class="normal">
       Real
      </p>
     </td>
     <td>
      <p class="normal">
       13910
      </p>
     </td>
     <td>
      <p class="normal">
       129
      </p>
     </td>
     <td>
      <p class="normal">
       2013
      </p>
     </td>
     <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
    </tr>
    <tr>
     <td>
      <table>
       <tr>
        <td>
         <a href="datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors">
          <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
         </a>
        </td>
        <td>
         <p class="normal">
          <b>
           <a href="datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors">
```

```
          Activities of Daily Living (ADLs) Recognition Using Binary Sensors
         </a>
        </b>
       </p>
      </td>
     </tr>
    </table>
   </td>
   <!-- <td><p class="normal">This dataset comprises information regarding the ADLs performed by two user
s on a daily basis in their
own homes.  &amp;nbsp;</p></td> -->
   <td>
    <p class="normal">
     Multivariate, Sequential, Time-Series
    </p>
   </td>
   <td>
    <p class="normal">
     Classification, Clustering
    </p>
   </td>
   <td>
    <p class="normal">
    </p>
   </td>
   <td>
    <p class="normal">
     2747
    </p>
   </td>
   <td>
    <p class="normal">
    </p>
   </td>
   <td>
    <p class="normal">
     2013
    </p>
   </td>
   <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
  </tr>
  <tr bgcolor="DDEEFF">
   <td>
    <table>
     <tr>
      <td>
       <a href="datasets/SkillCraft1+Master+Table+Dataset">
        <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
       </a>
      </td>
      <td>
       <p class="normal">
        <b>
         <a href="datasets/SkillCraft1+Master+Table+Dataset">
          SkillCraft1 Master Table Dataset
         </a>
        </b>
       </p>
      </td>
     </tr>
    </table>
   </td>
```

<!-- <td> This data was used in Thompson et al. (2013). A list of possible game actions i
s discussed in Thompson, Blair, Chen, & Henrey (2013). </td> -->
<td>

Multivariate

</td>
<td>

Regression

</td>
<td>

Integer, Real

</td>
<td>

3395

</td>
<td>

20

</td>
<td>

2013

</td>
<!-- <td> Game </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units">

</a>
</td>
<td>

<a href="datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units">
Weight Lifting Exercises monitored with Inertial Measurement Units
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Six young health subjects were asked to perform 5 variations of the biceps curl
weight lifting exercise. One of the variations is the one predicted by the health professional. </td> -->
<td>

Multivariate

</td>

<td>

Classification

</td>
<td>

Real

</td>
<td>

39242

</td>
<td>

152

</td>
<td>

2013

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/SML2010">

</a>
</td>
<td>

<a href="datasets/SML2010">
SML2010
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset is collected from a monitor system mounted in a domotic house. It corresponds to approximately 40 days of monitoring data. </td> -->
<td>

Multivariate, Sequential, Time-Series, Text

</td>
<td>

Regression

</td>
<td>

Real

</td>
<td>

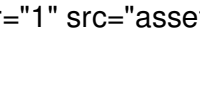
4137

</td>
<td>

24

</td>
<td>

2014

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Bike+Sharing+Dataset">

</a>
</td>
<td>

<a href="datasets/Bike+Sharing+Dataset">
Bike Sharing Dataset
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset contains the hourly and daily count of rental bikes between years 2011 and 2012 in Capital bikeshare system with the corresponding weather and seasonal information. </td> -->
<td>

Univariate

</td>
<td>

Regression

</td>
<td>

Integer, Real

</td>
<td>

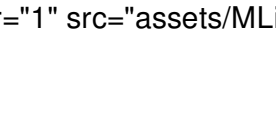
17389

</td>

```
  17389
  </p>
 </td>
 <td>
  <p class="normal">
   16
  </p>
 </td>
 <td>
  <p class="normal">
   2013
  </p>
 </td>
 <!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Predict+keywords+activities+in+a+online+social+media">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Predict+keywords+activities+in+a+online+social+media">
        Predict keywords activities in a online social media
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">The data from Twitter was collected during 360 consecutive days. It was done
by querying 1497 English keywords sampled from Wikipedia. This dataset is proposed in a Learning to rank set
ting. </p></td> -->
 <td>
  <p class="normal">
   Multivariate, Sequential, Time-Series
  </p>
 </td>
 <td>
  <p class="normal">
  </p>
 </td>
 <td>
  <p class="normal">
   Integer, Real
  </p>
 </td>
 <td>
  <p class="normal">
   51
  </p>
 </td>
 <td>
  <p class="normal">
   35
  </p>
 </td>
```

</td>
<td>

2013

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Thoracic+Surgery+Data">

</a>
</td>
<td>

<a href="datasets/Thoracic+Surgery+Data">
Thoracic Surgery Data
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The data is dedicated to classification problem related to the post-operative life expectancy in the lung cancer patients: class 1 - death within one year after surgery, class 2 - survival. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer, Real

</td>
<td>

470

</td>
<td>

17

</td>
<td>

2013

</td>
</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/EMG+dataset+in+Lower+Limb">

</a>
</td>
<td>

<a href="datasets/EMG+dataset+in+Lower+Limb">
EMG dataset in Lower Limb
</a>

</td>
</tr>
</table>
</td>
<!-- <td> 3 different exercises: sitting, standing and walking in the muscles: biceps femor
is, vastus medialis, rectus femoris and semitendinosus addition to goniometry in the exercises. </td>
-->
<td>

Multivariate, Time-Series

</td>
<td>

</td>
<td>

Real

</td>
<td>

132

</td>
<td>

5

</td>
<td>

2014

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

<a href="datasets/SUSY">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/SUSY">
SUSY
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This is a classification problem to distinguish between a signal process which p
roduces supersymmetric particles and a background process which does not. </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

5000000

</td>
<td>

18

</td>
<td>

2014

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/HIGGS">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/HIGGS">

```
          HIGGS
         </a>
        </b>
       </p>
      </td>
     </tr>
    </table>
   </td>
   <!-- <td><p class="normal">This is a classification problem to distinguish between a signal process which p
roduces Higgs bosons and a background process which does not. &amp;nbsp;</p></td> -->
   <td>
    <p class="normal">
    </p>
   </td>
   <td>
    <p class="normal">
     Classification
    </p>
   </td>
   <td>
    <p class="normal">
     Real
    </p>
   </td>
   <td>
    <p class="normal">
     11000000
    </p>
   </td>
   <td>
    <p class="normal">
     28
    </p>
   </td>
   <td>
    <p class="normal">
     2014
    </p>
   </td>
   <!-- <td><p class="normal">Physical&amp;nbsp;</p></td> -->
  </tr>
  <tr>
   <td>
    <table>
     <tr>
      <td>
       <a href="datasets/Qualitative_Bankruptcy">
        <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
       </a>
      </td>
      <td>
       <p class="normal">
        <b>
         <a href="datasets/Qualitative_Bankruptcy">
          Qualitative_Bankruptcy
         </a>
        </b>
       </p>
      </td>
     </tr>
    </table>
   </td>
```

<!-- <td> Predict the Bankruptcy from Qualitative parameters from experts. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

</td>
<td>

250

</td>
<td>

7

</td>
<td>

2014

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/LSVT+Voice+Rehabilitation">

</a>
</td>
<td>

<a href="datasets/LSVT+Voice+Rehabilitation">
LSVT Voice Rehabilitation
</a>

</td>
</tr>
</table>
</td>
<!-- <td> 126 samples from 14 participants, 309 features. Aim: assess whether voice rehabilitation treatment lead to phonations considered 'acceptable' or 'unacceptable' (binary class classification problem). </td> -->
<td>

Multivariate

</td>

<td>

Classification

</td>
<td>

Real

</td>
<td>

126

</td>
<td>

309

</td>
<td>

2014

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer">

</a>
</td>
<td>

<a href="datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer">
Dataset for ADL Recognition with Wrist-worn Accelerometer
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Recordings of 16 volunteers performing 14 Activities of Daily Living (ADL) while carrying a single wrist-worn tri-axial accelerometer. </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

Classification, Clustering

</td>
<td>

```
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
  3
 </p>
</td>
<td>
 <p class="normal">
  2014
 </p>
</td>
<!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Wilt">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Wilt">
        Wilt
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">High-resolution Remote Sensing data set (Quickbird). Small number of training samples of diseased trees, large number for other land cover. Testing data set from stratified random sample o f image.&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
  </p>
 </td>
 <td>
  <p class="normal">
   4889
  </p>
 </td>
 <td>
```

<td>

6

</td>
<td>

2014

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/User+Identification+From+Walking+Activity">

</a>
</td>
<td>

<a href="datasets/User+Identification+From+Walking+Activity">
User Identification From Walking Activity
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The dataset collects data from an Android smartphone positioned in the chest pocket from 22 participants walking in the wild over a predefined path.
  </td> -->
<td>

Univariate, Sequential, Time-Series

</td>
<td>

Classification, Clustering

</td>
<td>

Real

</td>
<td>

</td>
<td>

</td>
<td>

2014

</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

```
 </p>
 </td>
 <!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer">
        Activity Recognition from Single Chest-Mounted Accelerometer
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">The dataset collects data from a wearable accelerometer mounted on the ches
t. The dataset is intended for Activity Recognition research purposes. </p></td> -->
 <td>
  <p class="normal">
   Univariate, Sequential, Time-Series
  </p>
 </td>
 <td>
  <p class="normal">
   Classification, Clustering
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
 </td>
 <td>
  <p class="normal">
  </p>
 </td>
 <td>
  <p class="normal">
  </p>
 </td>
 <td>
  <p class="normal">
   2014
  </p>
 </td>
 <!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Leaf"
```

<a href="datasets/Leaf">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Leaf">
Leaf
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset consists in a collection of shape and texture features extracted fro
m digital images of leaf specimens originating from a total of 40 different plant species. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

340

</td>
<td>

16

</td>
<td>

2014

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Dresses_Attribute_Sales">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

```
          <a href="datasets/Dresses_Attribute_Sales">
           Dresses_Attribute_Sales
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">This dataset contain Attributes of dresses and their recommendations accordin
g to their sales.Sales are monitor on the basis of alternate days. &amp;nbsp;</p></td> -->
    <td>
     <p class="normal">
      Text
     </p>
    </td>
    <td>
     <p class="normal">
      Classification, Clustering
     </p>
    </td>
    <td>
     <p class="normal">
     </p>
    </td>
    <td>
     <p class="normal">
      501
     </p>
    </td>
    <td>
     <p class="normal">
      13
     </p>
    </td>
    <td>
     <p class="normal">
      2014
     </p>
    </td>
    <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
   </tr>
   <tr>
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/Tamilnadu+Electricity+Board+Hourly+Readings">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Tamilnadu+Electricity+Board+Hourly+Readings">
           Tamilnadu Electricity Board Hourly Readings
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
```

```
      </td>
      <!-- <td><p class="normal">This data can be effectively produced the result to fewer parameter of the Load profile can be reduced in the Database  &amp;nbsp;</p></td> -->
      <td>
       <p class="normal">
        Multivariate
       </p>
      </td>
      <td>
       <p class="normal">
        Classification, Regression, Clustering
       </p>
      </td>
      <td>
       <p class="normal">
        Real
       </p>
      </td>
      <td>
       <p class="normal">
        45781
       </p>
      </td>
      <td>
       <p class="normal">
        5
       </p>
      </td>
      <td>
       <p class="normal">
        2013
       </p>
      </td>
      <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
     </tr>
     <tr bgcolor="DDEEFF">
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/Airfoil+Self-Noise">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Airfoil+Self-Noise">
             Airfoil Self-Noise
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">NASA data set, obtained from a series of aerodynamic and acoustic tests of two and three-dimensional airfoil blade sections conducted in an anechoic wind tunnel.&amp;nbsp;</p></td> -->
      <td>
       <p class="normal">
        Multivariate
       </p>
```

```
      </td>
      <td>
       <p class="normal">
        Regression
       </p>
      </td>
      <td>
       <p class="normal">
        Real
       </p>
      </td>
      <td>
       <p class="normal">
        1503
       </p>
      </td>
      <td>
       <p class="normal">
        6
       </p>
      </td>
      <td>
       <p class="normal">
        2014
       </p>
      </td>
      <!-- <td><p class="normal">Physical&amp;nbsp;</p></td> -->
     </tr>
     <tr>
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/Wholesale+customers">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Wholesale+customers">
             Wholesale customers
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">The data set refers to clients of a wholesale distributor. It includes the annual s
pending in monetary units (m.u.) on diverse product categories&amp;nbsp;</p></td> -->
      <td>
       <p class="normal">
        Multivariate
       </p>
      </td>
      <td>
       <p class="normal">
        Classification, Clustering
       </p>
      </td>
      <td>
```

Integer

</td>
<td>

440

</td>
<td>

8

</td>
<td>

2014

</td>
<!-- <td> Business </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis">
Twitter Data set for Arabic Sentiment Analysis
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This problem of Sentiment Analysis (SA) has been studied well on the English l
anguage but not Arabic one. Two main approaches have been devised: corpus-based and lexicon-based. &nbs
p; </td> -->
<td>

Text

</td>
<td>

Classification

</td>
<td>

</td>
<td>

2000

</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

```
 </p>
</td>
<td>
 <p class="normal">
  2
 </p>
</td>
<td>
 <p class="normal">
  2014
 </p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Combined+Cycle+Power+Plant">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Combined+Cycle+Power+Plant">
        Combined Cycle Power Plant
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">The dataset contains 9568 data points collected from a Combined Cycle Power Plant over 6 years (2006-2011), when the plant was set to work with full load.  </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Regression
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
 </td>
 <td>
  <p class="normal">
   9568
  </p>
 </td>
 <td>
  <p class="normal">
   4
  </p>
 </td>
```

<td>

2014

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Urban+Land+Cover">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Urban+Land+Cover">
Urban Land Cover
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Classification of urban land cover using high resolution aerial imagery. Intende
d to assist sustainable urban planning efforts. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

</td>
<td>

168

</td>
<td>

148

</td>
<td>

2014

</td>
<!-- <td> Physical </td> -->
</tr>
<tr>

<td>
<table>
<tr>
<td>
<a href="datasets/Diabetes+130-US+hospitals+for+years+1999-2008">

</a>
</td>
<td>

<a href="datasets/Diabetes+130-US+hospitals+for+years+1999-2008">
Diabetes 130-US hospitals for years 1999-2008
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This data has been prepared to analyze factors related to readmission as well as other

outcomes pertaining to patients with diabetes. </td> -->
<td>

Multivariate

</td>
<td>

Classification, Clustering

</td>
<td>

Integer

</td>
<td>

100000

</td>
<td>

55

</td>
<td>

2014

</td>
<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Bach+Choral+Harmony">

<a href="datasets/Bach+Choral+Harmony">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Bach+Choral+Harmony">
Bach Choral Harmony
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The data set is composed of 60 chorales (5665 events) by J.S. Bach (1675-17
50).
Each event of each chorale is labelled using 1 among 101 chord labels and described
through 14 features. </td> -->
<td>

Sequential

</td>
<td>

Classification

</td>
<td>

</td>
<td>

5665

</td>
<td>

17

</td>
<td>

2014

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/StoneFlakes">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/StoneFlakes">
StoneFlakes
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Stone flakes are waste products of the stone tool production in
the prehistoric era. The variables are means of geometric and
stylistic features of the flakes contained in different inventories. </td> -->
<td>

Multivariate

</td>
<td>

Classification, Clustering, Causal-Discovery

</td>
<td>

Real

</td>
<td>

79

</td>
<td>

8

</td>
<td>

2014

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Tennis+Major+Tournament+Match+Statistics">

</a>
</td>
<td>

<a href="datasets/Tennis+Major+Tournament+Match+Statistics">
Tennis Major Tournament Match Statistics
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This is a collection of 8 files containing the match statistics for both women and men at the four major tennis tournaments of the year 2013. Each file has 42 columns and a minimum of 76 rows. </td> -->
<td>

Multivariate

</td>
<td>

Classification, Regression, Clustering

</td>
<td>

Integer, Real

</td>
<td>

127

</td>
<td>

42

</td>
<td>

2014

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings">

</a>
</td>
<td>

<a href="datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings">
Parkinson Speech Dataset with Multiple Types of Sound Recordings
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The training data belongs to 20 Parkinson's Disease (PD) patients and 20 healthy subjects. From all subjects, multiple types of sound recordings (26) are taken. </td> -->

<td>

Multivariate

</td>
<td>

Classification, Regression

</td>
<td>

Integer, Real

</td>
<td>

1040

</td>
<td>

26

</td>
<td>

2014

</td>
<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Gesture+Phase+Segmentation">

</a>
</td>
<td>

<a href="datasets/Gesture+Phase+Segmentation">
Gesture Phase Segmentation
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The dataset is composed by features extracted from 7 videos with people gesti
culating, aiming at studying Gesture Phase Segmentation. It contains 50 attributes divided into two files for eac
h video. </td> -->
<td>

Multivariate, Sequential, Time-Series

</td>
<td>

Classification, Clustering

</td>
<td>

Real

</td>
<td>

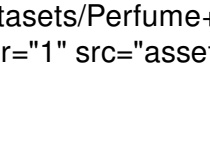
9900

</td>
<td>

50

</td>
<td>

2014

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Perfume+Data">

</a>
</td>
<td>

<a href="datasets/Perfume+Data">
Perfume Data
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This data consists of odors of 20 different perfumes. Data was obtained by usin
g a handheld odor meter (OMX-GR sensor) per second for 28 seconds period. </td> -->
<td>

Univariate, Domain-Theory

</td>
<td>

Classification, Clustering

</td>
<td>

Integer

</td>
</tr>
</table>
</body>
</html>

```
 </p>
</td>
<td>
 <p class="normal">
  560
 </p>
</td>
<td>
 <p class="normal">
  2
 </p>
</td>
<td>
 <p class="normal">
  2014
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/BlogFeedback">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/BlogFeedback">
        BlogFeedback
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Instances in this dataset contain features extracted from blog posts. The task a
ssociated with the data is to predict how many comments the post will receive. </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Regression
  </p>
 </td>
 <td>
  <p class="normal">
   Integer, Real
  </p>
 </td>
 <td>
  <p class="normal">
   60021
  </p>
 </td>
```

<td>

281

</td>
<td>

2014

</td>
<!-- <td> Social </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/REALDISP+Activity+Recognition+Dataset">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/REALDISP+Activity+Recognition+Dataset">
REALDISP Activity Recognition Dataset
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The REALDISP dataset is devised to evaluate techniques dealing with the effec
ts of sensor displacement in wearable activity recognition as well as to benchmark general activity recognition al
gorithms   </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

1419

</td>
<td>

120

</td>
<td>

<td>

2014

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Newspaper+and+magazine+images+segmentation+dataset">

</a>
</td>
<td>

<a href="datasets/Newspaper+and+magazine+images+segmentation+dataset">
Newspaper and magazine images segmentation dataset
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Dataset is well suited for segmentation tasks. It contains 101 scanned pages fr
om different newspapers and magazines in Russian with ground truth pixel-based masks. </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

</td>
<td>

101

</td>
<td>

</td>
<td>

2014

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>
</table>

<tr>
<td>
<a href="datasets/AAAI+2014+Accepted+Papers">

</a>
</td>
<td>

<a href="datasets/AAAI+2014+Accepted+Papers">
AAAI 2014 Accepted Papers
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This data set compromises the metadata for the 2014 AAAI conference's accepted papers, including paper titles, authors, abstracts, and keywords of varying granularity. </td> -->
<td>

Multivariate

</td>
<td>

Clustering

</td>
<td>

</td>
<td>

399

</td>
<td>

6

</td>
<td>

2014

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Gas+sensor+array+under+flow+modulation">

</a>
</td>
<td>

<a href="datasets/Gas+sensor+array+under+flow+modulation">
Gas sensor array under flow modulation
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The data set contains 58 time series acquired from 16 chemical sensors under gas flow modulation conditions. The sensors were exposed to different gaseous binary mixtures of acetone and ethanol. </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

Classification, Regression

</td>
<td>

Real

</td>
<td>

58

</td>
<td>

120432

</td>
<td>

2014

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures">

</a>
</td>
<td>

<a href="datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures">
Gas sensor array exposed to turbulent gas mixtures
</a>

</td>
</tr>
</table>
</td>
<!-- <td> A chemical detection platform composed of 8 chemoresistive gas sensors was exposed to turbulent gas mixtures generated naturally in a wind tunnel. The acquired time series of the sensors are provided. </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

Classification, Regression

</td>
<td>

Real

</td>
<td>

180

</td>
<td>

150000

</td>
<td>

2014

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/UJIIndoorLoc">

</a>
</td>
<td>

<a href="datasets/UJIIndoorLoc">
UJIIndoorLoc
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The UJIIndoorLoc is a Multi-Building Multi-Floor indoor localization database to test Indoor Positioning System that rely on WLAN/WiFi fingerprint. </td> -->

<td>

Multivariate

</td>
<td>

Classification, Regression

</td>
<td>

Integer, Real

</td>
<td>

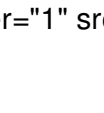
21048

</td>
<td>

529

</td>
<td>

2014

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Sentence+Classification">

</a>
</td>
<td>

<a href="datasets/Sentence+Classification">
Sentence Classification
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Contains sentences from the abstract and introduction of 30 articles annotated with a modified Argumentative Zones annotation scheme. These articles come from biology, machine learning and psychology. </td> -->
<td>

Text

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

</td>
<td>

</td>
<td>

2014

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Dow+Jones+Index">

</a>
</td>
<td>

<a href="datasets/Dow+Jones+Index">
Dow Jones Index
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset contains weekly data for the Dow Jones Industrial Index. It has be
en used in computational investing research. </td> -->
<td>

Time-Series

</td>
<td>

Classification, Clustering

</td>
<td>

Integer, Real

</td>

<td>

750

</td>
<td>

16

</td>
<td>

2014

</td>
<!-- <td> Business </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/sEMG+for+Basic+Hand+movements">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/sEMG+for+Basic+Hand+movements">
sEMG for Basic Hand movements
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The “sEMG for Basic Hand movements” includes 2 databases of surface electr
omyographic signals of 6 hand movements using Delsys' EMG System. Healthy subjects conducted six daily life
grasps. </td> -->
<td>

Time-Series

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

3000

</td>
<td>

2500

</td>
<td>

2014

</td>
<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/AAAI+2013+Accepted+Papers">

</a>
</td>
<td>

<a href="datasets/AAAI+2013+Accepted+Papers">
AAAI 2013 Accepted Papers
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This data set compromises the metadata for the 2013 AAAI conference's acce
pted papers (main track only), including paper titles, abstracts, and keywords of varying granularity.
</td> -->
<td>

Multivariate

</td>
<td>

Clustering

</td>
<td>

</td>
<td>

150

</td>
<td>

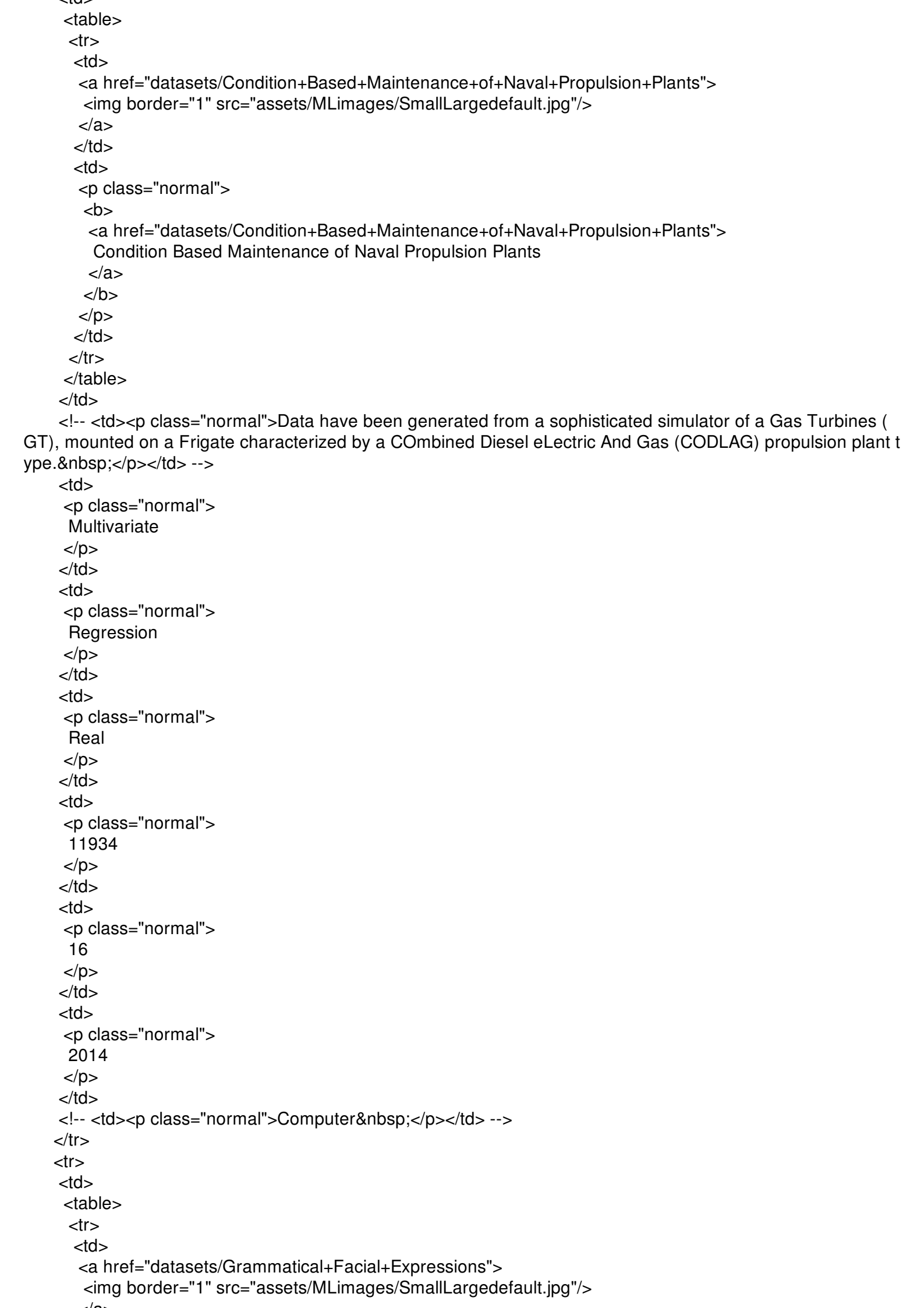
5

</td>
<td>

2014

</td>
</tr>
</table>
</body>
</html>

```
      </p>
     </td>
     <!-- <td><p class="normal">Computer </p></td> -->
    </tr>
    <tr>
     <td>
      <table>
       <tr>
        <td>
         <a href="datasets/Geographical+Original+of+Music">
          <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
         </a>
        </td>
        <td>
         <p class="normal">
          <b>
           <a href="datasets/Geographical+Original+of+Music">
            Geographical Original of Music
           </a>
          </b>
         </p>
        </td>
       </tr>
      </table>
     </td>
     <!-- <td><p class="normal">Instances in this dataset contain audio features extracted from 1059 wave files.
The task associated with the data is to predict the geographical origin of music.
 </p></td> -->
     <td>
      <p class="normal">
       Multivariate
      </p>
     </td>
     <td>
      <p class="normal">
       Classification, Regression
      </p>
     </td>
     <td>
      <p class="normal">
       Real
      </p>
     </td>
     <td>
      <p class="normal">
       1059
      </p>
     </td>
     <td>
      <p class="normal">
       68
      </p>
     </td>
     <td>
      <p class="normal">
       2014
      </p>
     </td>
     <!-- <td><p class="normal">Other </p></td> -->
    </tr>
    <tr bgcolor="DDEEFF">
     <td>
```

<td>
<table>
<tr>
<td>
<a href="datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants">

</a>
</td>
<td>

<a href="datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants">
Condition Based Maintenance of Naval Propulsion Plants
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Data have been generated from a sophisticated simulator of a Gas Turbines (GT), mounted on a Frigate characterized by a COmbined Diesel eLectric And Gas (CODLAG) propulsion plant t ype. </td> -->
<td>

Multivariate

</td>
<td>

Regression

</td>
<td>

Real

</td>
<td>

11934

</td>
<td>

16

</td>
<td>

2014

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Grammatical+Facial+Expressions">

</a>
</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

```
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Grammatical+Facial+Expressions">
           Grammatical Facial Expressions
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">This dataset supports the development of models that make possible to interpr
et Grammatical Facial Expressions from Brazilian Sign Language (Libras).&amp;nbsp;</p></td> -->
    <td>
     <p class="normal">
      Multivariate, Sequential
     </p>
    </td>
    <td>
     <p class="normal">
      Classification, Clustering
     </p>
    </td>
    <td>
     <p class="normal">
      Real
     </p>
    </td>
    <td>
     <p class="normal">
      27965
     </p>
    </td>
    <td>
     <p class="normal">
      100
     </p>
    </td>
    <td>
     <p class="normal">
      2014
     </p>
    </td>
    <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
   </tr>
   <tr bgcolor="DDEEFF">
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/NoisyOffice">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/NoisyOffice">
           NoisyOffice
```

```
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">Corpus intended to do cleaning (or binarization) and enhancement of noisy gra
yscale printed text images using supervised learning methods. Noisy images and their corresponding ground tr
uth provided. </p></td> -->
    <td>
     <p class="normal">
      Multivariate
     </p>
    </td>
    <td>
     <p class="normal">
      Classification, Regression
     </p>
    </td>
    <td>
     <p class="normal">
      Real
     </p>
    </td>
    <td>
     <p class="normal">
      216
     </p>
    </td>
    <td>
     <p class="normal">
      216
     </p>
    </td>
    <td>
     <p class="normal">
      2015
     </p>
    </td>
    <!-- <td><p class="normal">Computer </p></td> -->
   </tr>
   <tr>
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/MHEALTH+Dataset">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/MHEALTH+Dataset">
           MHEALTH Dataset
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
```

```
      </td>
      <!-- <td><p class="normal">The MHEALTH (Mobile Health) dataset is devised to benchmark techniques de
aling with human behavior analysis based on multimodal body sensing.&amp;nbsp;</p></td> -->
      <td>
       <p class="normal">
        Multivariate, Time-Series
       </p>
      </td>
      <td>
       <p class="normal">
        Classification
       </p>
      </td>
      <td>
       <p class="normal">
        Real
       </p>
      </td>
      <td>
       <p class="normal">
        120
       </p>
      </td>
      <td>
       <p class="normal">
        23
       </p>
      </td>
      <td>
       <p class="normal">
        2014
       </p>
      </td>
      <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
     </tr>
     <tr bgcolor="DDEEFF">
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/Student+Performance">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Student+Performance">
             Student Performance
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">Predict student performance in secondary education (high school). &amp;nbsp;</p>
</td> -->
      <td>
       <p class="normal">
        Multivariate
       </p>
```

```
      </td>
      <td>
       <p class="normal">
        Classification, Regression
       </p>
      </td>
      <td>
       <p class="normal">
        Integer
       </p>
      </td>
      <td>
       <p class="normal">
        649
       </p>
      </td>
      <td>
       <p class="normal">
        33
       </p>
      </td>
      <td>
       <p class="normal">
        2014
       </p>
      </td>
      <!-- <td><p class="normal">Social&amp;nbsp;</p></td> -->
     </tr>
     <tr>
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/ElectricityLoadDiagrams20112014">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/ElectricityLoadDiagrams20112014">
             ElectricityLoadDiagrams20112014
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">This data set contains electricity consumption of 370 points/clients.
&amp;nbsp;</p></td> -->
      <td>
       <p class="normal">
        Time-Series
       </p>
      </td>
      <td>
       <p class="normal">
        Regression, Clustering
       </p>
      </td>
      <td>
```

Real

</td>
<td>

370

</td>
<td>

140256

</td>
<td>

2015

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Gas+sensor+array+under+dynamic+gas+mixtures">

</a>
</td>
<td>

<a href="datasets/Gas+sensor+array+under+dynamic+gas+mixtures">
Gas sensor array under dynamic gas mixtures
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The data set contains the recordings of 16 chemical sensors exposed to two dy
namic gas mixtures at varying concentrations. For each mixture, signals were acquired continuously during 12
hours. </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

Classification, Regression

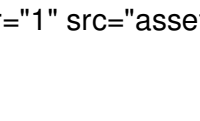
</td>
<td>

Real

</td>
<td>

```
        4178504
       </p>
      </td>
      <td>
       <p class="normal">
        19
       </p>
      </td>
      <td>
       <p class="normal">
        2015
       </p>
      </td>
      <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
     </tr>
     <tr>
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/microblogPCU">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/microblogPCU">
             microblogPCU
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">MicroblogPCU data is crawled from sina weibo microblog[http://weibo.com/]. Th
is data can be used to study machine learning methods  as well as do some social network research. &amp;nbsp;</p
></td> -->
      <td>
       <p class="normal">
        Multivariate, Univariate, Sequential, Text
       </p>
      </td>
      <td>
       <p class="normal">
        Classification, Causal-Discovery
       </p>
      </td>
      <td>
       <p class="normal">
        Integer, Real
       </p>
      </td>
      <td>
       <p class="normal">
        221579
       </p>
      </td>
      <td>
       <p class="normal">
        20
```

```
 </p>
</td>
<td>
 <p class="normal">
  2015
 </p>
</td>
<!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Firm-Teacher_Clave-Direction_Classification">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Firm-Teacher_Clave-Direction_Classification">
        Firm-Teacher_Clave-Direction_Classification
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">The data are binary attack-point vectors and their clave-direction class(es) acc
ording to the partido-alto-based paradigm.&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
  </p>
 </td>
 <td>
  <p class="normal">
   10800
  </p>
 </td>
 <td>
  <p class="normal">
   20
  </p>
 </td>
 <td>
  <p class="normal">
   2015
  </p>
 </td>
 <!-- <td><p class="normal">Other&amp;nbsp;</p></td> -->
```

<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Dataset+for+Sensorless+Drive+Diagnosis">

</a>
</td>
<td>

<a href="datasets/Dataset+for+Sensorless+Drive+Diagnosis">
Dataset for Sensorless Drive Diagnosis
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Features are extracted from motor current. The motor has intact and defective components. This results in 11 different classes with different conditions.   </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

58509

</td>
<td>

49

</td>
<td>

2015

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/TV+News+Channel+Commercial+Detection+Dataset">

<a href="datasets/TV+News+Channel+Commercial+Detection+Dataset">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/TV+News+Channel+Commercial+Detection+Dataset">
TV News Channel Commercial Detection Dataset
</a>

</td>
</tr>
</table>
</td>
<!-- <td> TV Commercials data set consists of standard audio-visual features of video s
hots extracted from 150 hours of TV news broadcast of 3 Indian and 2 international news channels (30 Hours
each).   </td> -->
<td>

Multivariate

</td>
<td>

Classification, Clustering

</td>
<td>

Real

</td>
<td>

129685

</td>
<td>

12

</td>
<td>

2015

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Phishing+Websites">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Phishing+Websites">
Phishing Websites
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset collected mainly from: PhishTank archive, MillerSmiles archive, G
oogle’s searching operators. </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

2456

</td>
<td>

30

</td>
<td>

2015

</td>
<!-- <td> Computer Security </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Greenhouse+Gas+Observing+Network">

</a>
</td>
<td>

<a href="datasets/Greenhouse+Gas+Observing+Network">
Greenhouse Gas Observing Network
</a>

</td>
</tr>

```
      </table>
     </td>
     <!-- <td><p class="normal">Design an observing network to monitor emissions of a greenhouse gas (GHG) in California given time series of synthetic observations and tracers from weather model simulations. &amp;nbsp;</p></td> -->
     <td>
      <p class="normal">
       Multivariate, Time-Series
      </p>
     </td>
     <td>
      <p class="normal">
       Regression
      </p>
     </td>
     <td>
      <p class="normal">
       Real
      </p>
     </td>
     <td>
      <p class="normal">
       2921
      </p>
     </td>
     <td>
      <p class="normal">
       5232
      </p>
     </td>
     <td>
      <p class="normal">
       2015
      </p>
     </td>
     <!-- <td><p class="normal">Physical&amp;nbsp;</p></td> -->
    </tr>
    <tr>
     <td>
      <table>
       <tr>
        <td>
         <a href="datasets/Diabetic+Retinopathy+Debrecen+Data+Set">
          <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
         </a>
        </td>
        <td>
         <p class="normal">
          <b>
           <a href="datasets/Diabetic+Retinopathy+Debrecen+Data+Set">
            Diabetic Retinopathy Debrecen Data Set
           </a>
          </b>
         </p>
        </td>
       </tr>
      </table>
     </td>
     <!-- <td><p class="normal">This dataset contains features extracted from the Messidor image set to predic t whether an image contains signs of diabetic retinopathy or not. &amp;nbsp;</p></td> -->
     <td>
      <p class="normal">
```

```
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
  Classification
 </p>
</td>
<td>
 <p class="normal">
  Integer, Real
 </p>
</td>
<td>
 <p class="normal">
  1151
 </p>
</td>
<td>
 <p class="normal">
  20
 </p>
</td>
<td>
 <p class="normal">
  2014
 </p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/HIV-1+protease+cleavage">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/HIV-1+protease+cleavage">
        HIV-1 protease cleavage
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">The data contains lists of octamers (8 amino acids) and a flag (-1 or 1) dependi
ng on whether HIV-1 protease will cleave in the central position (between amino acids 4 and 5). </p></td
> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
```

```
 </p>
</td>
<td>
 <p class="normal">
  Categorical
 </p>
</td>
<td>
 <p class="normal">
  6590
 </p>
</td>
<td>
 <p class="normal">
  1
 </p>
</td>
<td>
 <p class="normal">
  2015
 </p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Sentiment+Labelled+Sentences">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Sentiment+Labelled+Sentences">
        Sentiment Labelled Sentences
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">The dataset contains sentences labelled with positive or negative sentiment.&n
bsp;</p></td> -->
 <td>
  <p class="normal">
   Text
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
  </p>
 </td>
 <td>
```

3000

</td>
<td>

</td>
<td>

2015

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Online+News+Popularity">

</a>
</td>
<td>

<a href="datasets/Online+News+Popularity">
Online News Popularity
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset summarizes a heterogeneous set of features about articles publis
hed by Mashable in a period of two years. The goal is to predict the number of shares in social networks (popul
arity). </td> -->
<td>

Multivariate

</td>
<td>

Classification, Regression

</td>
<td>

Integer, Real

</td>
<td>

39797

</td>
<td>

61

</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

```
 </p>
</td>
<td>
 <p class="normal">
  2015
 </p>
</td>
<!-- <td><p class="normal">Business </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Forest+type+mapping">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Forest+type+mapping">
        Forest type mapping
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Multi-temporal remote sensing data of a forested area in Japan. The goal is to map different forest types using spectral data. </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
  </p>
 </td>
 <td>
  <p class="normal">
   326
  </p>
 </td>
 <td>
  <p class="normal">
   27
  </p>
 </td>
 <td>
  <p class="normal">
   2015
  </p>
 </td>
 <!-- <td><p class="normal">Life </p></td> -->
```

</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/wiki4HE">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/wiki4HE">
wiki4HE
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Survey of faculty members from two Spanish universities on teaching uses of Wikipedia </td> -->
<td>

Multivariate

</td>
<td>

Regression, Clustering, Causal-Discovery

</td>
<td>

</td>
<td>

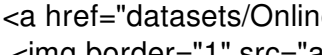
913

</td>
<td>

53

</td>
<td>

2015

</td>
<!-- <td> Social </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset">


```
<img border="1" src="assets/MLimages/SmallLargedefault.jpg" />
</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset">
Online Video Characteristics and Transcoding Time Dataset
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The dataset contains a million randomly sampled video instances listing 10 fun
damental video characteristics along with the YouTube video ID. &amp;nbsp;</p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Regression
</p>
</td>
<td>
<p class="normal">
Integer, Real
</p>
</td>
<td>
<p class="normal">
168286
</p>
</td>
<td>
<p class="normal">
11
</p>
</td>
<td>
<p class="normal">
2015
</p>
</td>
<!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Chronic_Kidney_Disease">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Chronic_Kidney_Disease">
Chronic_Kidney_Disease
```

Chronic_Kidney_Disease
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset can be used to predict the chronic kidney disease and it can be co
llected from the hospital nearly 2 months of period. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

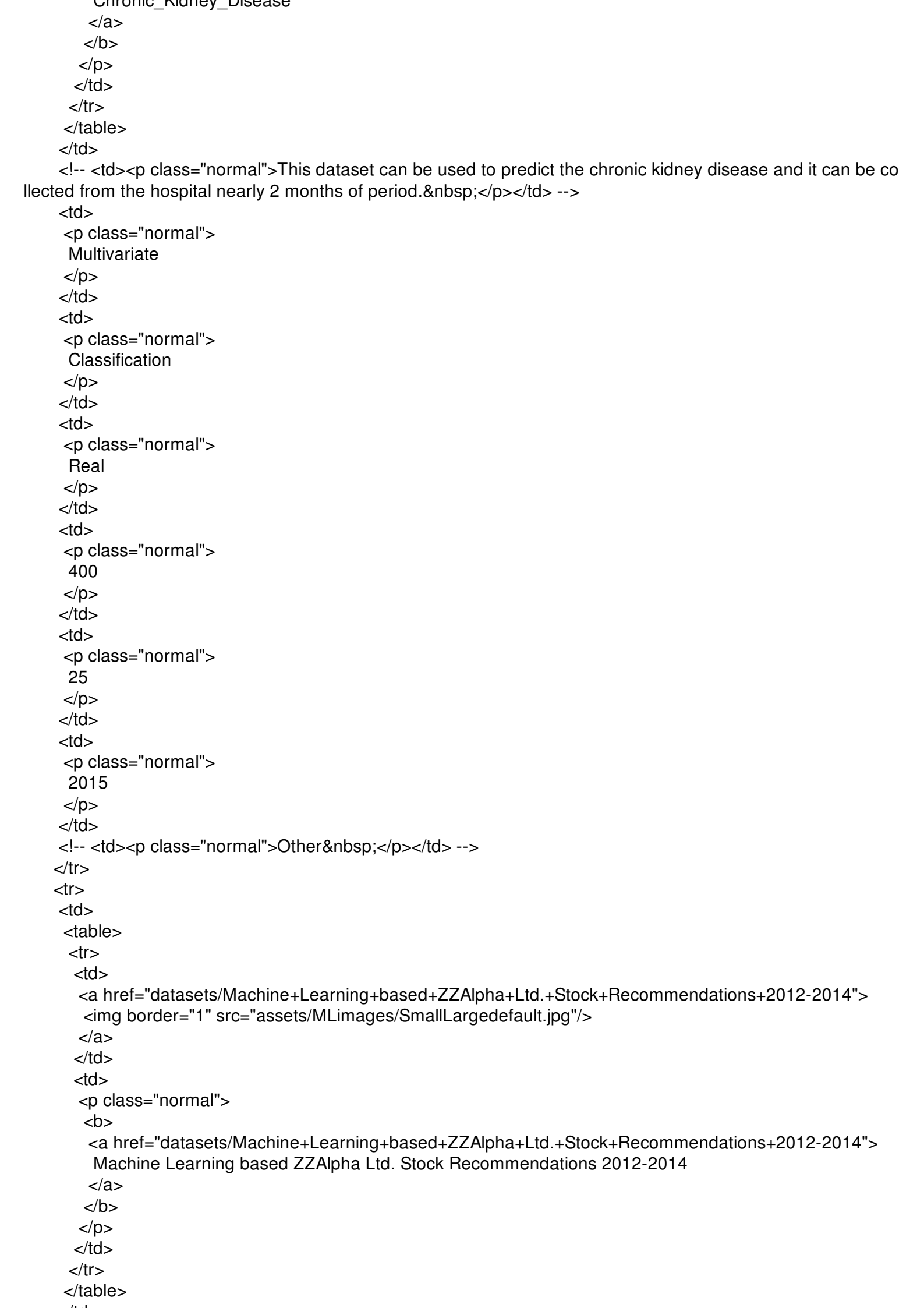
400

</td>
<td>

25

</td>
<td>

2015

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014">

</a>
</td>
<td>

<a href="datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014">
Machine Learning based ZZAlpha Ltd. Stock Recommendations 2012-2014
</a>

</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

```
      </td>
      <!-- <td><p class="normal">The data here are the ZZAlpha® machine learning recommendations made for various US traded stock portfolios the morning of each day during the 3 year period Jan 1, 2012 - Dec 31, 2014 .  &amp;nbsp;</p></td> -->
      <td>
       <p class="normal">
        Sequential, Time-Series
       </p>
      </td>
      <td>
       <p class="normal">
        Classification
       </p>
      </td>
      <td>
       <p class="normal">
        Real
       </p>
      </td>
      <td>
       <p class="normal">
        314080
       </p>
      </td>
      <td>
       <p class="normal">
        0
       </p>
      </td>
      <td>
       <p class="normal">
        2015
       </p>
      </td>
      <!-- <td><p class="normal">Business&amp;nbsp;</p></td> -->
     </tr>
     <tr bgcolor="DDEEFF">
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/Folio">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Folio">
             Folio
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">20 photos of leaves for each of 32 different species.&amp;nbsp;</p></td> -->
      <td>
       <p class="normal">
        Multivariate
       </p>
```

</td>
<td>

Classification, Clustering

</td>
<td>

</td>
<td>

637

</td>
<td>

20

</td>
<td>

2015

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015">

</a>
</td>
<td>

<a href="datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015">
Taxi Service Trajectory - Prediction Challenge, ECML PKDD 2015
</a>

</td>
</tr>
</table>
</td>
<!-- <td> An accurate dataset describing trajectories performed by all the 442 taxis runni
ng in the city of Porto, in Portugal.
  </td> -->
<td>

Multivariate, Sequential, Time-Series, Domain-Theory

</td>
<td>

Clustering, Causal-Discovery

</td>
<td>

Real

</td>
<td>

1710671

</td>
<td>

9

</td>
<td>

2015

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Cuff-Less+Blood+Pressure+Estimation">

</a>
</td>
<td>

<a href="datasets/Cuff-Less+Blood+Pressure+Estimation">
Cuff-Less Blood Pressure Estimation
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This Data set provides preprocessed and cleaned vital signals which can be us
ed in designing algorithms for cuff-less estimation of the blood pressure. </td> -->
<td>

Multivariate

</td>
<td>

Classification, Regression

</td>
<td>

Real

</td>
<td>

12000

</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

```
 </p>
</td>
<td>
 <p class="normal">
  3
 </p>
</td>
<td>
 <p class="normal">
  2015
 </p>
</td>
<!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions">
        Smartphone-Based Recognition of Human Activities and Postural Transitions
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Activity recognition data set built from the recordings of 30 subjects performing basic activities and postural transitions while carrying a waist-mounted smartphone with embedded inertial sens ors.
&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate, Time-Series
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
 </td>
 <td>
  <p class="normal">
   10929
  </p>
 </td>
 <td>
  <p class="normal">
   561
```

```
 </p>
</td>
<td>
 <p class="normal">
  2015
 </p>
</td>
<!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
 <table>
  <tr>
   <td>
    <a href="datasets/Mice+Protein+Expression">
     <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   </td>
   <td>
    <p class="normal">
     <b>
      <a href="datasets/Mice+Protein+Expression">
       Mice Protein Expression
      </a>
     </b>
    </p>
   </td>
  </tr>
 </table>
</td>
<!-- <td><p class="normal">Expression levels of 77 proteins measured in the cerebral cortex of 8 classes o
f control and Down syndrome mice exposed to context fear conditioning, a task used to assess associative lear
ning.&amp;nbsp;</p></td> -->
<td>
 <p class="normal">
  Multivariate
 </p>
</td>
<td>
 <p class="normal">
  Classification, Clustering
 </p>
</td>
<td>
 <p class="normal">
  Real
 </p>
</td>
<td>
 <p class="normal">
  1080
 </p>
</td>
<td>
 <p class="normal">
  82
 </p>
</td>
<td>
 <p class="normal">
  2015
 </p>
```

```
 </td>
 <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/UJIIndoorLoc-Mag">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/UJIIndoorLoc-Mag">
        UJIIndoorLoc-Mag
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">The UJIIndoorLoc-Mag is an indoor localization database to test Indoor Positio
ning System that rely on Earth's magnetic field variations.&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate, Sequential, Time-Series
  </p>
 </td>
 <td>
  <p class="normal">
   Classification, Regression, Clustering
  </p>
 </td>
 <td>
  <p class="normal">
   Integer, Real
  </p>
 </td>
 <td>
  <p class="normal">
   40000
  </p>
 </td>
 <td>
  <p class="normal">
   13
  </p>
 </td>
 <td>
  <p class="normal">
   2015
  </p>
 </td>
 <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
```

<td>
<a href="datasets/Heterogeneity+Activity+Recognition">

</a>
</td>
<td>

<a href="datasets/Heterogeneity+Activity+Recognition">
Heterogeneity Activity Recognition
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The Heterogeneity Human Activity Recognition (HHAR) dataset from Smartpho
nes and Smartwatches is a dataset devised to benchmark human activity recognition algorithms (classification,
automatic data segmentation, sensor fusion, feature extraction, etc.) in real-world contexts; specifically, the dat
aset is gathered with a variety of different device models and use-scenarios, in order to reflect sensing heterog
eneities to be expected in real deployments. </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

Classification, Clustering

</td>
<td>

Real

</td>
<td>

43930257

</td>
<td>

16

</td>
<td>

2015

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set">

</a>
</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

</a>
</td>
<td>

<a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set">
Educational Process Mining (EPM): A Learning Analytics Data Set
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Educational Process Mining data set is built from the recordings of 115 subject
s' activities through a logging application while learning with an educational simulator. </td> -->
<td>

Multivariate, Sequential, Time-Series

</td>
<td>

Classification, Regression, Clustering

</td>
<td>

Integer

</td>
<td>

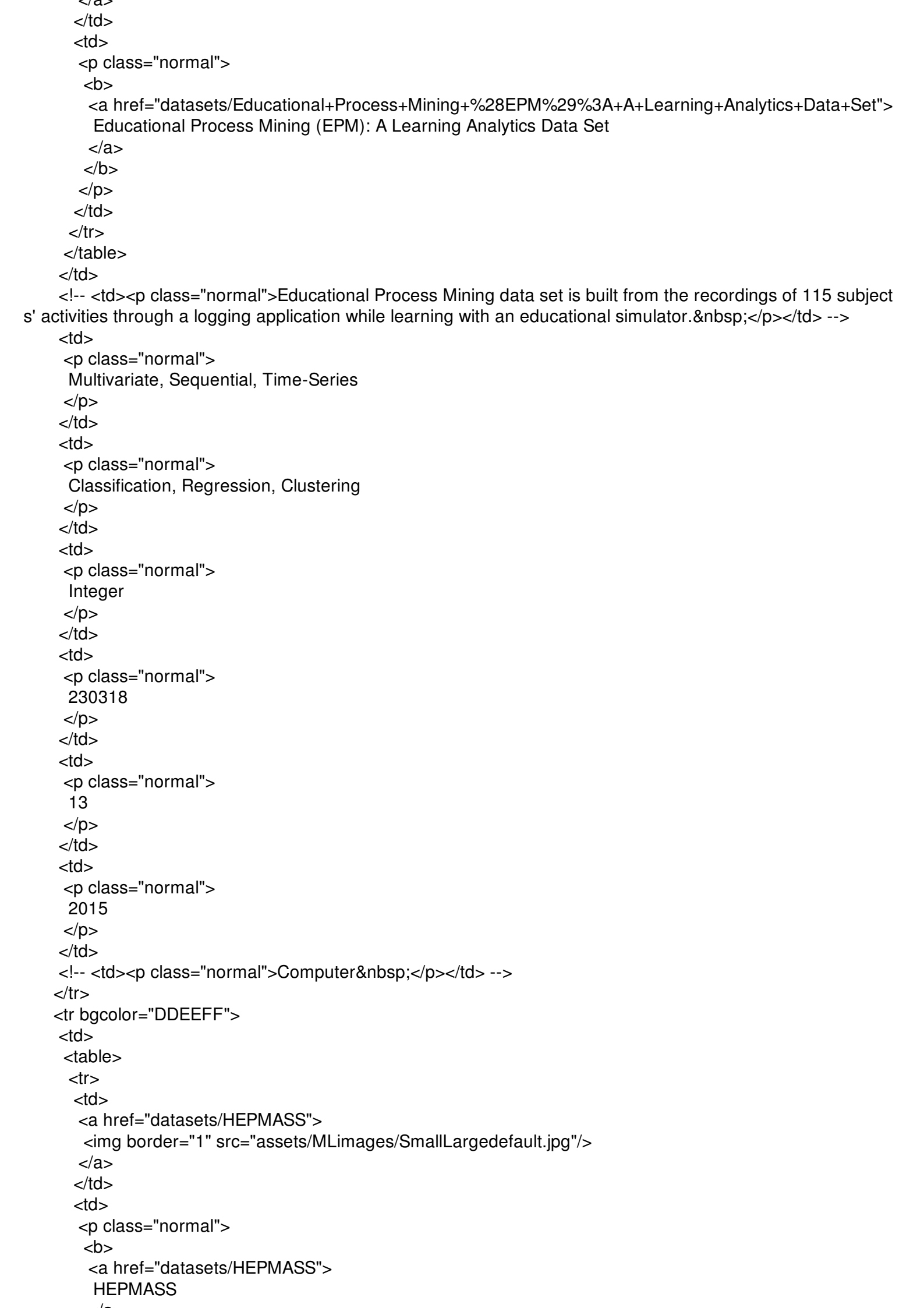
230318

</td>
<td>

13

</td>
<td>

2015

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/HEPMASS">

</a>
</td>
<td>

<a href="datasets/HEPMASS">
HEPMASS
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Particle Physics data set in which the task is to distinguish between a signal
process which produces exotic particles and a background process which does not. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

10500000

</td>
<td>

28

</td>
<td>

2016

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations">

</a>
</td>
<td>

<a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations">
Gas Sensor Array Drift Dataset at Different Concentrations
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Gas sensor array drift dataset for 6 gases at different concentrations. </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

Classification, Regression, Clustering

</td>
<td>

Real

</td>
<td>

13910

</td>
<td>

128

</td>
<td>

2012

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Gas+Sensor+Array+Drift+Dataset">

</a>
</td>
<td>

<a href="datasets/Gas+Sensor+Array+Drift+Dataset">
Gas Sensor Array Drift Dataset
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Gas sensor array drift dataset for 6 gases at different concentrations. </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

Classification, Regression, Clustering

</td>
<td>

Real

</td>
<td>

13910

</td>
<td>

128

</td>
<td>

2012

</td>
<!-- <td> Computer

```
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">The search for exotic particles requires sorting through a large number of collis
ions to find the events of interest. This data set challenges one to detect a new particle of unknown mass.&nbs
p;</p></td> -->
    <td>
     <p class="normal">
      Multivariate
     </p>
    </td>
    <td>
     <p class="normal">
      Classification
     </p>
    </td>
    <td>
     <p class="normal">
      Real
     </p>
    </td>
    <td>
     <p class="normal">
      10500000
     </p>
    </td>
    <td>
     <p class="normal">
      28
     </p>
    </td>
    <td>
     <p class="normal">
      2016
     </p>
    </td>
    <!-- <td><p class="normal">Physical </p></td> -->
   </tr>
   <tr>
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/Indoor+User+Movement+Prediction+from+RSS+data">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Indoor+User+Movement+Prediction+from+RSS+data">
           Indoor User Movement Prediction from RSS data
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
```

</td>
<!-- <td> This dataset contains temporal data from a Wireless Sensor Network deployed in real-world office environments. The task is intended as real-life benchmark in the area of Ambient Assisted Living. </td> -->
<td>

Multivariate, Sequential, Time-Series

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

13197

</td>
<td>

4

</td>
<td>

2016

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Open+University+Learning+Analytics+dataset">

</a>
</td>
<td>

<a href="datasets/Open+University+Learning+Analytics+dataset">
Open University Learning Analytics dataset
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Open University Learning Analytics Dataset contains data about courses, students and their interactions with Virtual Learning Environment for seven selected courses and more than 30000 students. </td> -->
<td>

```
 Multivariate, Sequential, Time-Series
 </p>
</td>
<td>
 <p class="normal">
  Classification, Regression, Clustering
 </p>
</td>
<td>
 <p class="normal">
  Integer
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
  2015
 </p>
</td>
<!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/default+of+credit+card+clients">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/default+of+credit+card+clients">
        default of credit card clients
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">This research aimed at the case of customers' default payments in Taiwan and compares the predictive accuracy of probability of default among six data mining methods.&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
```

Integer, Real

</td>
<td>

30000

</td>
<td>

24

</td>
<td>

2016

</td>
<!-- <td> Business </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Mesothelioma%E2%80%99s+disease+data+set+">

</a>
</td>
<td>

<a href="datasets/Mesothelioma%E2%80%99s+disease+data+set+">
Mesothelioma’s disease data set
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Mesothelioma’s disease data set were prepared at Dicle University Faculty of Medicine in Turkey.
Three hundred and twenty-four Mesothelioma patient data. In the dataset, all samples have 34 features.
</td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

324

</td>
<td>

34

</td>
<td>

2016

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Online+Retail">

</a>
</td>
<td>

<a href="datasets/Online+Retail">
Online Retail
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This is a transnational data set which contains all the transactions occurring bet
ween 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail. </td> -->
<td>

Multivariate, Sequential, Time-Series

</td>
<td>

Classification, Clustering

</td>
<td>

Integer, Real

</td>
<td>

541909

</td>
<td>

8

</td>
</tr>
</table>
</body>
</html>

```
 </p>
</td>
<td>
 <p class="normal">
  2015
 </p>
</td>
<!-- <td><p class="normal">Business&amp;nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/SIFT10M">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/SIFT10M">
        SIFT10M
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">In SIFT10M, each data point is a SIFT feature which is extracted from Caltech-256 by the open source VLFeat library. The corresponding patches of the SIFT features are provided.&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Causal-Discovery
  </p>
 </td>
 <td>
  <p class="normal">
   Integer
  </p>
 </td>
 <td>
  <p class="normal">
   11164866
  </p>
 </td>
 <td>
  <p class="normal">
   128
  </p>
 </td>
 <td>
  <p class="normal">
   2016
  </p>
```

```
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/GPS+Trajectories">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/GPS+Trajectories">
GPS Trajectories
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The dataset has been feed by Android app called Go!Track. It is available at G
oolge Play Store(https://play.google.com/store/apps/details?id=com.go.router).  </p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Classification, Regression
</p>
</td>
<td>
<p class="normal">
Real
</p>
</td>
<td>
<p class="normal">
163
</p>
</td>
<td>
<p class="normal">
15
</p>
</td>
<td>
<p class="normal">
2016
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
```

<td>
<a href="datasets/Detect+Malacious+Executable%28AntiVirus%29">

</a>
</td>
<td>

<a href="datasets/Detect+Malacious+Executable%28AntiVirus%29">
Detect Malacious Executable(AntiVirus)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> I extract features from malacious and non-malacious and create and training d
ataset to teach svm classifier.Dataset made of unknown executable to detect if it is virus or normal safe executa
ble. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

373

</td>
<td>

513

</td>
<td>

2016

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Occupancy+Detection+">

</a>
</td>
<td>

<td>

<a href="datasets/Occupancy+Detection+">
Occupancy Detection
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Experimental data used for binary classification (room occupancy) from Tempe
rature,Humidity,Light and CO2. Ground-truth occupancy was obtained from time stamped pictures that were ta
ken every minute. </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

20560

</td>
<td>

7

</td>
<td>

2016

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Parkinson%
E2%80%99s+Disease">

</a>
</td>
<td>

<a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Parkinson
%E2%80%99s+Disease">

Improved Spiral Test Using Digitized Graphics Tablet for Monitoring Parkinson's Disease
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Handwriting database consists of 25 PWP(People with Parkinson) and 15 healt
hy individuals.Three types of recordings (Static Spiral Test, Dynamic Spiral Test and Stability Test) are taken.&
nbsp; </td> -->
<td>

Multivariate

</td>
<td>

Classification, Regression, Clustering

</td>
<td>

Real

</td>
<td>

40

</td>
<td>

7

</td>
<td>

2016

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/News+Aggregator">

</a>
</td>
<td>

<a href="datasets/News+Aggregator">
News Aggregator
</a>

</td>
</tr>

</table>
</td>
<!-- <td> References to news pages collected from an web aggregator in the period fro
m 10-March-2014 to 10-August-2014. The resources are grouped into clusters that represent pages discussing
the same story. </td> -->
<td>

Multivariate

</td>
<td>

Classification, Clustering

</td>
<td>

</td>
<td>

422937

</td>
<td>

5

</td>
<td>

2016

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Air+Quality">

</a>
</td>
<td>

<a href="datasets/Air+Quality">
Air Quality
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Contains the responses of a gas multisensor device deployed on the field in an
Italian city. Hourly responses averages are recorded along with gas concentrations references from a certified
analyzer.   </td> -->
<td>

```
 Multivariate, Time-Series
 </p>
</td>
<td>
 <p class="normal">
  Regression
 </p>
</td>
<td>
 <p class="normal">
  Real
 </p>
</td>
<td>
 <p class="normal">
  9358
 </p>
</td>
<td>
 <p class="normal">
  15
 </p>
</td>
<td>
 <p class="normal">
  2016
 </p>
</td>
<!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Twin+gas+sensor+arrays">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Twin+gas+sensor+arrays">
        Twin gas sensor arrays
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">5 replicates of an 8-MOX gas sensor array were exposed to different gas condi
tions (4 volatiles at 10 concentration levels each).&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate, Time-Series, Domain-Theory
  </p>
 </td>
 <td>
  <p class="normal">
   Classification, Regression
  </p>
```

```
      </td>
      <td>
       <p class="normal">
        Real
       </p>
      </td>
      <td>
       <p class="normal">
        640
       </p>
      </td>
      <td>
       <p class="normal">
        480000
       </p>
      </td>
      <td>
       <p class="normal">
        2016
       </p>
      </td>
      <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
     </tr>
     <tr bgcolor="DDEEFF">
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/Gas+sensors+for+home+activity+monitoring">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Gas+sensors+for+home+activity+monitoring">
             Gas sensors for home activity monitoring
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">100 recordings of a sensor array under different conditions in a home setting: b
ackground, wine and banana presentations. The array includes 8 MOX gas sensors, and humidity and tempera
ture sensors.
&amp;nbsp;</p></td> -->
      <td>
       <p class="normal">
        Multivariate, Time-Series
       </p>
      </td>
      <td>
       <p class="normal">
        Classification
       </p>
      </td>
      <td>
       <p class="normal">
        Real
       </p>
```

</td>
<td>

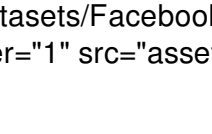
919438

</td>
<td>

11

</td>
<td>

2016

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Facebook+Comment+Volume+Dataset">

</a>
</td>
<td>

<a href="datasets/Facebook+Comment+Volume+Dataset">
Facebook Comment Volume Dataset
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Instances in this dataset contain features extracted from facebook posts. The t
ask associated with the data is to predict how many comments the post will receive. </td> -->
<td>

Multivariate

</td>
<td>

Regression

</td>
<td>

Integer, Real

</td>
<td>

40949

</td>
<td>

54

</td>
<td>

2016

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Living+%28AAL%29">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Living+%28AAL%29">
Smartphone Dataset for Human Activity Recognition (HAR) in Ambient Assisted Living (AAL)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This data is an addition to an existing dataset on UCI. We collected more data to improve the accuracy of our human activity recognition algorithms applied in the domain of Ambient Assisted Living.   </td> -->
<td>

Time-Series

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

5744

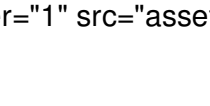
</td>
<td>

561

</td>

<td>

2016

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Polish+companies+bankruptcy+data">

</a>
</td>
<td>

<a href="datasets/Polish+companies+bankruptcy+data">
Polish companies bankruptcy data
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The dataset is about bankruptcy prediction of Polish companies.The bankrupt c
ompanies were analyzed in the period 2000-2012, while the still operating companies were evaluated from 200
7 to 2013. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

10503

</td>
<td>

64

</td>
<td>

2016

</td>
<!-- <td> Business </td> -->

<!-- <td> Business </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29">

</a>
</td>
<td>

<a href="datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29">
Activity Recognition system based on Multisensor data fusion (AReM)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset contains temporal data from a Wireless Sensor Network worn by a
n actor performing the activities: bending, cycling, lying down, sitting, standing, walking. </td> -->
<td>

Multivariate, Sequential, Time-Series

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

42240

</td>
<td>

6

</td>
<td>

2016

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Data2+Games+Results"

<a href="datasets/Dota2+Games+Results">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Dota2+Games+Results">
Dota2 Games Results
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Dota 2 is a popular computer game with two teams of 5 players. At the start of t
he game each player chooses a unique hero with different strengths and weaknesses. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

</td>
<td>

102944

</td>
<td>

116

</td>
<td>

2016

</td>
<!-- <td> Game </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Facebook+metrics">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Facebook+metrics">

```
          Facebook metrics
         </a>
        </b>
       </p>
      </td>
     </tr>
    </table>
   </td>
   <!-- <td><p class="normal">Facebook performance metrics of a renowned cosmetic's brand Facebook pag
e.&amp;nbsp;</p></td> -->
   <td>
    <p class="normal">
     Multivariate
    </p>
   </td>
   <td>
    <p class="normal">
     Regression
    </p>
   </td>
   <td>
    <p class="normal">
     Integer
    </p>
   </td>
   <td>
    <p class="normal">
     500
    </p>
   </td>
   <td>
    <p class="normal">
     19
    </p>
   </td>
   <td>
    <p class="normal">
     2016
    </p>
   </td>
   <!-- <td><p class="normal">Business&amp;nbsp;</p></td> -->
  </tr>
  <tr>
   <td>
    <table>
     <tr>
      <td>
       <a href="datasets/UbiqLog+%28smartphone+lifelogging%29">
        <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
       </a>
      </td>
      <td>
       <p class="normal">
        <b>
         <a href="datasets/UbiqLog+%28smartphone+lifelogging%29">
          UbiqLog (smartphone lifelogging)
         </a>
        </b>
       </p>
      </td>
     </tr>
    </table>
```

```
      </td>
      <!-- <td><p class="normal">UbiqLog is the smartphone lifelogging tool that runs on the smartphone of 35 u
sers for about 2 months.
&amp;nbsp;</p></td> -->
      <td>
       <p class="normal">
        Multivariate
       </p>
      </td>
      <td>
       <p class="normal">
        Causal-Discovery
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
        9782222
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
        2016
       </p>
      </td>
      <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
     </tr>
     <tr bgcolor="DDEEFF">
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/NIPS+Conference+Papers+1987-2015">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/NIPS+Conference+Papers+1987-2015">
             NIPS Conference Papers 1987-2015
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">This data set contains the distribution of words in the full text of the NIPS confe
rence papers published from 1987 to 2015.&amp;nbsp;</p></td> -->
      <td>
       <p class="normal">
        Text
       </p>
      </td>
```

<td>

Clustering

</td>
<td>

Integer

</td>
<td>

11463

</td>
<td>

5812

</td>
<td>

2016

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/HTRU2">

</a>
</td>
<td>

<a href="datasets/HTRU2">
HTRU2
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Pulsar candidates collected during the HTRU survey. Pulsars are a type of star, of considerable scientific interest. Candidates must be classified in to pulsar and non-pulsar classes to aid disco very. </td> -->
<td>

Multivariate

</td>
<td>

Classification, Clustering

</td>
<td>

```
 <p class="normal">
  Real
 </p>
</td>
<td>
 <p class="normal">
  17898
 </p>
</td>
<td>
 <p class="normal">
  9
 </p>
</td>
<td>
 <p class="normal">
  2017
 </p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
 <table>
  <tr>
   <td>
    <a href="datasets/Drug+consumption+%28quantified%29">
     <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   </td>
   <td>
    <p class="normal">
     <b>
      <a href="datasets/Drug+consumption+%28quantified%29">
       Drug consumption (quantified)
      </a>
     </b>
    </p>
   </td>
  </tr>
 </table>
</td>
<!-- <td><p class="normal">Classify type of drug consumer by personality data </p></td> -->
<td>
 <p class="normal">
  Multivariate
 </p>
</td>
<td>
 <p class="normal">
  Classification
 </p>
</td>
<td>
 <p class="normal">
  Real
 </p>
</td>
<td>
 <p class="normal">
  1885
 </p>
```

```
</td>
<td>
 <p class="normal">
  32
 </p>
</td>
<td>
 <p class="normal">
  2016
 </p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Appliances+energy+prediction">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Appliances+energy+prediction">
        Appliances energy prediction
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Experimental data used to create regression models of appliances energy use i
n a low energy building. </p></td> -->
 <td>
  <p class="normal">
   Multivariate, Time-Series
  </p>
 </td>
 <td>
  <p class="normal">
   Regression
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
 </td>
 <td>
  <p class="normal">
   19735
  </p>
 </td>
 <td>
  <p class="normal">
   29
  </p>
 </td>
 <td>
```

2017

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Miskolc+IIS+Hybrid+IPS">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Miskolc+IIS+Hybrid+IPS">
Miskolc IIS Hybrid IPS
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The dataset was created for the comparison and evaluation of hybrid indoor po
sitioning methods. The dataset presented contains data from W-LAN and Bluetooth interfaces, and Magnetome
ter.   </td> -->
<td>

Text

</td>
<td>

Classification, Clustering, Causal-Discovery

</td>
<td>

Integer

</td>
<td>

1540

</td>
<td>

67

</td>
<td>

2016

</td>
<!-- <td> Computer </td> -->
</tr>

```
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/KDC-4007+dataset+Collection">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/KDC-4007+dataset+Collection">
        KDC-4007 dataset Collection
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">KDC-4007 dataset Collection is the Kurdish Documents Classification text used in categories regarding Kurdish Sorani news and articles. </p></td> -->
 <td>
  <p class="normal">
   Multivariate, Text
  </p>
 </td>
 <td>
  <p class="normal">
   Classification, Regression
  </p>
 </td>
 <td>
  <p class="normal">
   Integer
  </p>
 </td>
 <td>
  <p class="normal">
   4007
  </p>
 </td>
 <td>
  <p class="normal">
  </p>
 </td>
 <td>
  <p class="normal">
   2017
  </p>
 </td>
 <!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Geo-Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
```

<img border="1" src="assets/MLimages/SmallLargedefault.jpg" />
</a>
</td>
<td>

<a href="datasets/Geo-Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone">
Geo-Magnetic field and WLAN dataset for indoor localisation from wristband and smartphone
</a>

</td>
</tr>
</table>
</td>
<!-- <td> A multisource and multivariate dataset for indoor localisation methods based on WLAN and Geo-Magnetic field fingerprinting </td> -->
<td>

Multivariate, Sequential, Time-Series

</td>
<td>

Classification, Regression, Clustering

</td>
<td>

Integer, Real

</td>
<td>

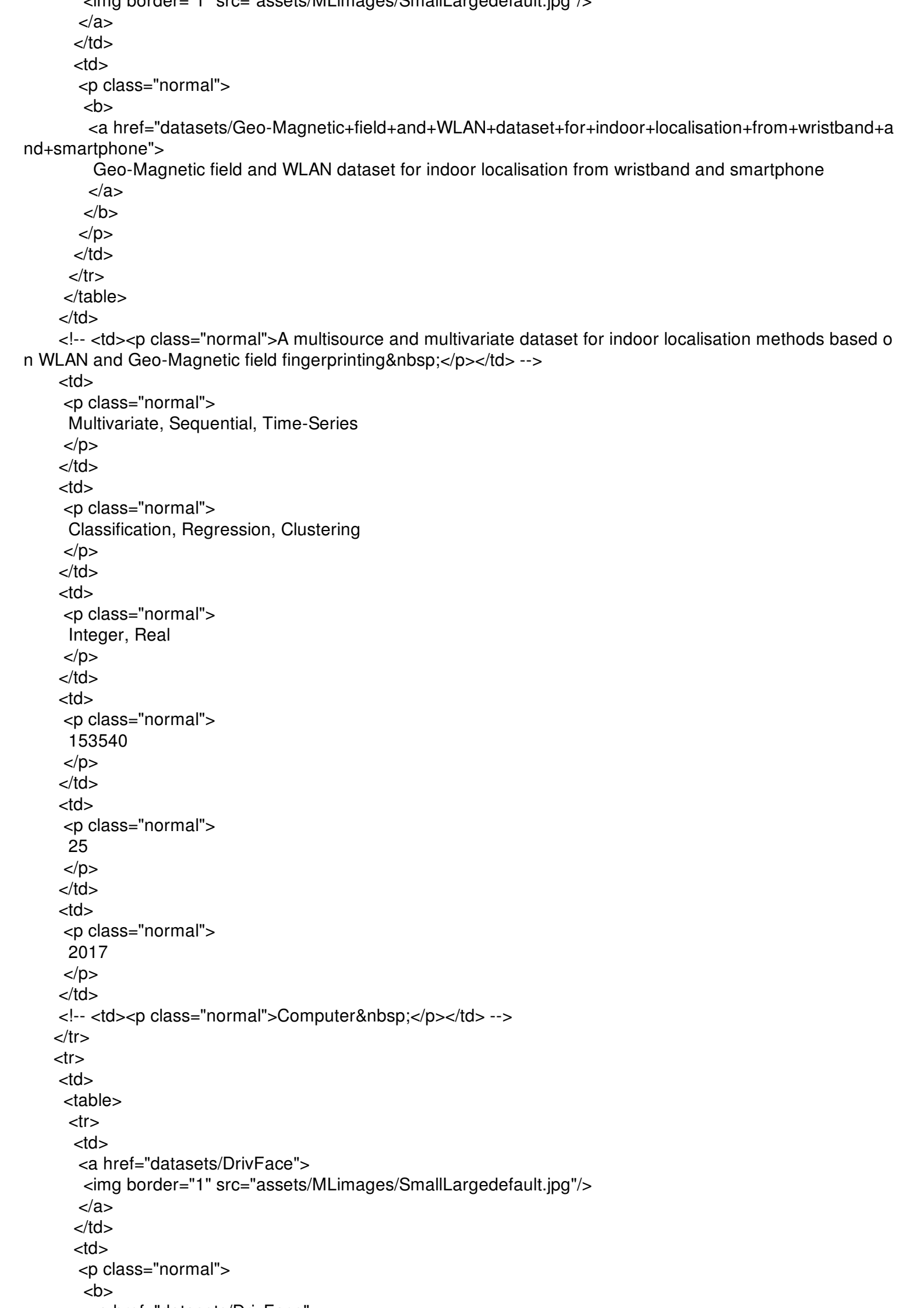
153540

</td>
<td>

25

</td>
<td>

2017

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/DrivFace">

</a>
</td>
<td>

<a href="datasets/DrivFace">
DrivFace
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The DrivFace contains images sequences of subjects while driving in real scen
arios. It is composed of 606 samples of 640×480, acquired over different days from 4 drivers with several facial
features. </td> -->
<td>

Multivariate

</td>
<td>

Classification, Regression, Clustering

</td>
<td>

Real

</td>
<td>

606

</td>
<td>

6400

</td>
<td>

2016

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Website+Phishing">

</a>
</td>
<td>

<a href="datasets/Website+Phishing">
Website Phishing
</a>

</td>

```
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">

&amp;nbsp;</p></td> -->
    <td>
     <p class="normal">
      Multivariate
     </p>
    </td>
    <td>
     <p class="normal">
      Classification
     </p>
    </td>
    <td>
     <p class="normal">
      Integer
     </p>
    </td>
    <td>
     <p class="normal">
      1353
     </p>
    </td>
    <td>
     <p class="normal">
      10
     </p>
    </td>
    <td>
     <p class="normal">
      2016
     </p>
    </td>
    <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
   </tr>
   <tr>
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/YouTube+Spam+Collection">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/YouTube+Spam+Collection">
           YouTube Spam Collection
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">It is a public set of comments collected for spam research. It has five datasets composed by 1,956 real messages extracted from five videos that were among the 10 most viewed on the colle ction period.&amp;nbsp;</p></td> -->
```

<td>

Text

</td>
<td>

Classification

</td>
<td>

</td>
<td>

1956

</td>
<td>

5

</td>
<td>

2017

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Beijing+PM2.5+Data">

</a>
</td>
<td>

<a href="datasets/Beijing+PM2.5+Data">
Beijing PM2.5 Data
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This hourly data set contains the PM2.5 data of US Embassy in Beijing. Meanw
hile, meteorological data from Beijing Capital International Airport are also included.   </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

Regression

</td>
</tr>
</table>
</body>
</html>

```
 </p>
</td>
<td>
 <p class="normal">
  Integer, Real
 </p>
</td>
<td>
 <p class="normal">
  43824
 </p>
</td>
<td>
 <p class="normal">
  13
 </p>
</td>
<td>
 <p class="normal">
  2017
 </p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Cargo+2000+Freight+Tracking+and+Tracing">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Cargo+2000+Freight+Tracking+and+Tracing">
        Cargo 2000 Freight Tracking and Tracing
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Sanitized and anonymized Cargo 2000 (C2K) airfreight tracking and tracing ev
ents, covering five months of business execution (3,942 process instances, 7,932 transport legs, 56,082 activiti
es).  </p></td> -->
 <td>
  <p class="normal">
   Multivariate, Sequential
  </p>
 </td>
 <td>
  <p class="normal">
   Classification, Regression
  </p>
 </td>
 <td>
  <p class="normal">
   Integer
  </p>
```

```
</td>
<td>
 <p class="normal">
  3942
 </p>
</td>
<td>
 <p class="normal">
  98
 </p>
</td>
<td>
 <p class="normal">
  2016
 </p>
</td>
<!-- <td><p class="normal">Business&amp;nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Cervical+cancer+%28Risk+Factors%29">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Cervical+cancer+%28Risk+Factors%29">
        Cervical cancer (Risk Factors)
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">This dataset focuses on the prediction of indicators/diagnosis of cervical cancer
. The features cover demographic information, habits, and historic medical records.&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Integer, Real
  </p>
 </td>
 <td>
  <p class="normal">
   858
  </p>
 </td>
 <td>
```

36

</td>
<td>

2017

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Quality+Assessment+of+Digital+Colposcopies">

</a>
</td>
<td>

<a href="datasets/Quality+Assessment+of+Digital+Colposcopies">
Quality Assessment of Digital Colposcopies
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset explores the subjective quality assessment of digital colposcopies.
  </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

287

</td>
<td>

69

</td>
<td>

2017

</td>
</tr>
</table>
</body>
</html>

```
    2017
   </p>
  </td>
  <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
 </tr>
 <tr bgcolor="DDEEFF">
  <td>
   <table>
    <tr>
     <td>
      <a href="datasets/KASANDR">
       <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
      </a>
     </td>
     <td>
      <p class="normal">
       <b>
        <a href="datasets/KASANDR">
         KASANDR
        </a>
       </b>
      </p>
     </td>
    </tr>
   </table>
  </td>
  <!-- <td><p class="normal">KASANDR is a novel, publicly available collection for recommendation systems that records the behavior of customers of the European leader in e-Commerce advertising, Kelkoo. &amp;nbsp;</p></td> -->
  <td>
   <p class="normal">
    Multivariate
   </p>
  </td>
  <td>
   <p class="normal">
    Causal-Discovery
   </p>
  </td>
  <td>
   <p class="normal">
    Integer
   </p>
  </td>
  <td>
   <p class="normal">
    17764280
   </p>
  </td>
  <td>
   <p class="normal">
    2158859
   </p>
  </td>
  <td>
   <p class="normal">
    2017
   </p>
  </td>
  <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
 </tr>
 <tr>
  <td>
```

<td>
<table>
<tr>
<td>
<a href="datasets/FMA%3A+A+Dataset+For+Music+Analysis">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/FMA%3A+A+Dataset+For+Music+Analysis">
FMA: A Dataset For Music Analysis
</a>

</td>
</tr>
</table>
</td>
<!-- <td> FMA features 106,574 tracks and includes song title, album, artist, genres; play counts, favorites, comments; description, biography, tags; together with audio (343 days, 917 GiB) and features . </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

Classification, Clustering

</td>
<td>

Real

</td>
<td>

106574

</td>
<td>

518

</td>
<td>

2017

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Air+quality">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

```
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Air+quality">
             Air quality
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal"> Contains the responses of a gas multisensor device deployed on the field in a
n Italian city.  </p></td> -->
      <td>
       <p class="normal">
        Multivariate, Time-Series
       </p>
      </td>
      <td>
       <p class="normal">
        Regression
       </p>
      </td>
      <td>
       <p class="normal">
        Real
       </p>
      </td>
      <td>
       <p class="normal">
        9358
       </p>
      </td>
      <td>
       <p class="normal">
        15
       </p>
      </td>
      <td>
       <p class="normal">
        2016
       </p>
      </td>
      <!-- <td><p class="normal">Other </p></td> -->
     </tr>
     <tr>
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/Epileptic+Seizure+Recognition">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Epileptic+Seizure+Recognition">
             Epileptic Seizure Recognition
```

```
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">This dataset is a pre-processed and re-structured/reshaped version of a very c
ommonly used dataset featuring epileptic seizure detection. &amp;nbsp;</p></td> -->
    <td>
     <p class="normal">
      Multivariate, Time-Series
     </p>
    </td>
    <td>
     <p class="normal">
      Classification, Clustering
     </p>
    </td>
    <td>
     <p class="normal">
      Integer, Real
     </p>
    </td>
    <td>
     <p class="normal">
      11500
     </p>
    </td>
    <td>
     <p class="normal">
      179
     </p>
    </td>
    <td>
     <p class="normal">
      2017
     </p>
    </td>
    <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
   </tr>
   <tr bgcolor="DDEEFF">
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/Devanagari+Handwritten+Character+Dataset">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Devanagari+Handwritten+Character+Dataset">
           Devanagari Handwritten Character Dataset
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
```

<!-- <td> This is an image database of Handwritten Devanagari characters. There are 4
6 classes of characters with 2000 examples each. The dataset is split into training set(85%) and testing set(15
%).   </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer

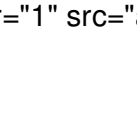
</td>
<td>

92000

</td>
<td>

</td>
<td>

2016

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Stock+portfolio+performance">

</a>
</td>
<td>

<a href="datasets/Stock+portfolio+performance">
Stock portfolio performance
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The data set of performances of weighted scoring stock portfolios are obtained
with mixture design from the US stock market historical database. </td> -->
<td>

Multivariate

</td>
<td>

Regression

</td>
<td>

Real

</td>
<td>

315

</td>
<td>

12

</td>
<td>

2016

</td>
<!-- <td> Business </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/MoCap+Hand+Postures">

</a>
</td>
<td>

<a href="datasets/MoCap+Hand+Postures">
MoCap Hand Postures
</a>

</td>
</tr>
</table>
</td>
<!-- <td> 5 types of hand postures from 12 users were recorded using unlabeled markers attached to fingers of a glove in a motion capture environment. Due to resolution and occlusion, missing values are common. </td> -->
<td>

Multivariate

</td>
<td>

Classification, Clustering

</td>
<td>

```
 Integer, Real
 </p>
</td>
<td>
 <p class="normal">
  78095
 </p>
</td>
<td>
 <p class="normal">
  38
 </p>
</td>
<td>
 <p class="normal">
  2016
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech
">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speec
h">
        Early biomarkers of Parkinson's disease based on natural connected speech
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Predict a pattern of neurodegeneration in the dataset of speech features obtain
ed from patients with early untreated Parkinson's disease and patients at high risk developing Parkinson's disea
se. </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification, Regression
  </p>
 </td>
 <td>
  <p class="normal">
   Integer, Real
  </p>
 </td>
 <td>
```

130

</td>
<td>

65

</td>
<td>

2017

</td>
<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting">
Data for Software Engineering Teamwork Assessment in Education Setting
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Data include over 100 Team Activity Measures and outcomes (ML classes) obt
ained from activities of 74 student teams during the creation of final class project in SW Eng. classes at SFSU,
Fulda, FAU </td> -->
<td>

Sequential, Time-Series

</td>
<td>

Classification

</td>
<td>

Integer, Real

</td>
<td>

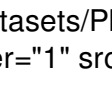
74

</td>
<td>

102

</td>
<td>

2017

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/PM2.5+Data+of+Five+Chinese+Cities">

</a>
</td>
<td>

<a href="datasets/PM2.5+Data+of+Five+Chinese+Cities">
PM2.5 Data of Five Chinese Cities
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This hourly data set contains the PM2.5 data in Beijing, Shanghai, Guangzhou, Chengdu and Shenyang. Meanwhile, meteorological data for each city are also included. </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

Regression

</td>
<td>

Integer, Real

</td>
<td>

52854

</td>
<td>

86

</td>
<td>

2017

</td>

```
 </td>
 <!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet">
        Parkinson Disease Spiral Drawings Using Digitized Graphics Tablet
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Handwriting database consists of 62 PWP(People with Parkinson) and 15 healt
hy individuals. Three types of recordings (Static Spiral Test, Dynamic Spiral Test and Stability Test) are taken.&
nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification, Regression, Clustering
  </p>
 </td>
 <td>
  <p class="normal">
   Integer
  </p>
 </td>
 <td>
  <p class="normal">
   77
  </p>
 </td>
 <td>
  <p class="normal">
   7
  </p>
 </td>
 <td>
  <p class="normal">
   2017
  </p>
 </td>
 <!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
 <td>
```

<table>
<tr>
<td>
<a href="datasets/Sales_Transactions_Dataset_Weekly">

</a>
</td>
<td>

<a href="datasets/Sales_Transactions_Dataset_Weekly">
Sales_Transactions_Dataset_Weekly
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Contains weekly purchased quantities of 800 over products over 52 weeks. Nor
malised values are provided too. </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

Clustering

</td>
<td>

Integer, Real

</td>
<td>

811

</td>
<td>

53

</td>
<td>

2017

</td>
<!-- <td>   </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Las+Vegas+Strip">

</a>
</td>

<td>

<a href="datasets/Las+Vegas+Strip">
Las Vegas Strip
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset includes quantitative and categorical features from online reviews from 21 hotels located in Las Vegas Strip, extracted from TripAdvisor (http://www.tripadvisor.com). </td> -->
<td>

</td>
<td>

Classification, Regression

</td>
<td>

Integer

</td>
<td>

504

</td>
<td>

20

</td>
<td>

2017

</td>
<!-- <td> Business </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Eco-hotel">

</a>
</td>
<td>

<a href="datasets/Eco-hotel">
Eco-hotel
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset includes quantitative and categorical features from online reviews from 20 hotels located in Las Vegas Strip, extracted from TripAdvisor (http://www.tripadvisor.com). </td> -->
<td>

</td>
<td>

Classification, Regression

</td>
<td>

Integer

</td>
<td>

504

</td>
<td>

20

</td>
<td>

2017

</td>
<!-- <td> Business </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Energy-consumption-generation-prediction">

</a>
</td>
<td>

<a href="datasets/Energy-consumption-generation-prediction">
Energy consumption generation prediction
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset contains 1 year (2017) of electricity consumption and solar generation data for 4 UK households. The dataset contains measurements recorded every minute. </td> -->
<td>

</td>
<td>

Regression

</td>
<td>

Integer

</td>
<td>

251274

</td>
<td>

9

</td>
<td>

2017

</td>
<!-- <td> Business </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Energy-consumption-generation-prediction-using-LSTM">

</a>
</td>
<td>

<a href="datasets/Energy-consumption-generation-prediction-using-LSTM">
Energy consumption generation prediction using LSTM
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset contains 1 year (2017) of electricity consumption and solar generation data for 4 UK households. The dataset contains measurements recorded every minute. </td> -->
<td>

</td>
<td>

Regression

</td>
<td>

Integer

</td>
<td>

251274

</td>
<td>

9

</td>
<td>

2017

</td>
<!-- <td> Business </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Energy-consumption-generation-prediction-using-statistical-learning-methods">

</a>
</td>
<td>

<a href="datasets/Energy-consumption-generation-prediction-using-statistical-learning-methods">
Energy consumption generation prediction using statistical learning methods
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset contains 1 year (2017) of electricity consumption and solar generation data for 4 UK households. The dataset contains measurements recorded every minute. </td> -->
<td>

</td>
<td>

Regression

</td>
<td>

Integer

</td>
<td>

251274

</td>
<td>

9

</td>
<td>

2017

</td>
<!-- <td> Business </td> -->
</tr>
</table>
</body>
</html>

```
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">This dataset includes Online Textual Reviews from both online (e.g., TripAdvis
or) and offline (e.g., Guests' book) sources from the Areias do Seixo Eco-Resort.&amp;nbsp;</p></td> -->
    <td>
     <p class="normal">
      Text
     </p>
    </td>
    <td>
     <p class="normal">
     </p>
    </td>
    <td>
     <p class="normal">
     </p>
    </td>
    <td>
     <p class="normal">
      401
     </p>
    </td>
    <td>
     <p class="normal">
      1
     </p>
    </td>
    <td>
     <p class="normal">
      2017
     </p>
    </td>
    <!-- <td><p class="normal">Business&amp;nbsp;</p></td> -->
   </tr>
   <tr bgcolor="DDEEFF">
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/MEU-Mobile+KSD">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/MEU-Mobile+KSD">
           MEU-Mobile KSD
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">This dataset contains keystroke dynamics data collected on a touch mobile dev
ice (Nexus 7). The dataset contains 2856 records, 51 records per subject for 56 subjects. &amp;nbsp;</p></td> -->
    <td>
     <p class="normal">
```

```
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
  Classification
 </p>
</td>
<td>
 <p class="normal">
  Integer, Real
 </p>
</td>
<td>
 <p class="normal">
  2856
 </p>
</td>
<td>
 <p class="normal">
  71
 </p>
</td>
<td>
 <p class="normal">
  2016
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Crowdsourced+Mapping">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Crowdsourced+Mapping">
        Crowdsourced Mapping
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Crowdsourced data from OpenStreetMap is used to automate the classification of satellite images into different land cover classes (impervious, farm, forest, grass, orchard, water).  </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
```

```
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
  10546
 </p>
</td>
<td>
 <p class="normal">
  29
 </p>
</td>
<td>
 <p class="normal">
  2016
 </p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/gene+expression+cancer+RNA-Seq">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/gene+expression+cancer+RNA-Seq">
        gene expression cancer RNA-Seq
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">This collection of data is part of the RNA-Seq (HiSeq) PANCAN data set, it is a random extraction of gene expressions of patients having different types of tumor: BRCA, KIRC, COAD, LUAD and PRAD. </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification, Clustering
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
 </td>
```

<td>

801

</td>
<td>

20531

</td>
<td>

2016

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetom
eter">

</a>
</td>
<td>

<a href="datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magneto
meter">
Hybrid Indoor Positioning Dataset from WiFi RSSI, Bluetooth and magnetometer
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The dataset was created for the comparison and evaluation of hybrid indoor po
sitioning methods. The dataset presented contains data from W-LAN and Bluetooth interfaces, and Magnetome
ter.   </td> -->
<td>

Multivariate, Sequential, Time-Series

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

1540

</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

```
</td>
<td>
 <p class="normal">
  65
 </p>
</td>
<td>
 <p class="normal">
  2016
 </p>
</td>
<!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
 <table>
  <tr>
   <td>
    <a href="datasets/chestnut+%E2%80%93+LARVIC">
     <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   </td>
   <td>
    <p class="normal">
     <b>
      <a href="datasets/chestnut+%E2%80%93+LARVIC">
       chestnut – LARVIC
      </a>
     </b>
    </p>
   </td>
  </tr>
 </table>
</td>
<!-- <td><p class="normal">The research project presents this database, shows the images of chestnuts th
at will be processed to determine the presence or absence of defects&amp;nbsp;</p></td> -->
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
  Classification, Clustering
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
  1451
 </p>
</td>
<td>
 <p class="normal">
  3
 </p>
</td>
<td>
 <p class="normal">
  2017
```

```
      2017
     </p>
    </td>
    <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
   </tr>
   <tr>
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%
28OBS%29+Network">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+
%28OBS%29+Network">
           Burst Header Packet (BHP) flooding attack on Optical Burst Switching (OBS) Network
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">One of the primary challenges in identifying the risks of the Burst Header Pack
et (BHP) flood attacks in Optical Burst Switching networks (OBS) is the scarcity of reliable historical data.  &nbs
p;</p></td> -->
    <td>
     <p class="normal">
      Text
     </p>
    </td>
    <td>
     <p class="normal">
      Classification
     </p>
    </td>
    <td>
     <p class="normal">
      Integer
     </p>
    </td>
    <td>
     <p class="normal">
      1075
     </p>
    </td>
    <td>
     <p class="normal">
      22
     </p>
    </td>
    <td>
     <p class="normal">
      2017
     </p>
    </td>
    <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
   </tr>
```

<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Motion+Capture+Hand+Postures">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Motion+Capture+Hand+Postures">
Motion Capture Hand Postures
</a>

</td>
</tr>
</table>
</td>
<!-- <td> 5 types of hand postures from 12 users were recorded using unlabeled markers on fingers of a glove in a motion capture environment. Due to resolution and occlusion, missing values are common. </td> -->
<td>

Multivariate

</td>
<td>

Classification, Clustering

</td>
<td>

Real

</td>
<td>

78095

</td>
<td>

38

</td>
<td>

2017

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Anuran+Calls+%28MFCCs%29

<a href="datasets/Anuran+Calls+%28MFCCs%29">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Anuran+Calls+%28MFCCs%29">
Anuran Calls (MFCCs)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Acoustic features extracted from syllables of anuran (frogs) calls, including the family, the genus, and the species labels (multilabel).   </td> -->
<td>

Multivariate

</td>
<td>

Classification, Clustering

</td>
<td>

Real

</td>
<td>

7195

</td>
<td>

22

</td>
<td>

2017

</td>
<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization">
TTC-3600: Benchmark dataset for Turkish text categorization
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The TTC-3600 data set is a collection of Turkish news and articles including ca
tegorized 3,600 documents from 6 well-known portals in Turkey. It has 4 different forms in ARFF Weka format.
  </td> -->
<td>

Text

</td>
<td>

Classification, Clustering

</td>
<td>

Integer

</td>
<td>

3600

</td>
<td>

4814

</td>
<td>

2017

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy">

</a>
</td>
<td>

<a href="datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy">
Gastrointestinal Lesions in Regular Colonoscopy
</a>

</td>

```
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">This dataset contains features extracted from colonoscopy videos used to dete
ct gastrointestinal lesions. It contains 76 lesions: 15 serrated adenomas, 21 hyperplastic lesions and 40 adeno
ma. &amp;nbsp;</p></td> -->
    <td>
     <p class="normal">
      Multivariate
     </p>
    </td>
    <td>
     <p class="normal">
      Classification
     </p>
    </td>
    <td>
     <p class="normal">
      Real
     </p>
    </td>
    <td>
     <p class="normal">
      76
     </p>
    </td>
    <td>
     <p class="normal">
      698
     </p>
    </td>
    <td>
     <p class="normal">
      2016
     </p>
    </td>
    <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
   </tr>
   <tr bgcolor="DDEEFF">
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/Daily+Demand+Forecasting+Orders">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Daily+Demand+Forecasting+Orders">
           Daily Demand Forecasting Orders
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">The dataset was collected during 60 days, this is a real database of a brazilian
logistics company.&amp;nbsp;</p></td> -->
    <td>
```

Time-Series

</td>
<td>

Regression

</td>
<td>

Integer

</td>
<td>

60

</td>
<td>

13

</td>
<td>

2017

</td>
<!-- <td> Business </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Paper+Reviews">

</a>
</td>
<td>

<a href="datasets/Paper+Reviews">
Paper Reviews
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This sentiment analysis data set contains scientific paper reviews from an inter
national conference on computing and informatics. The task is to predict the orientation or the evaluation of a r
eview. </td> -->
<td>

Text

</td>
<td>

```
  Classification, Regression
 </p>
</td>
<td>
 <p class="normal">
  Integer
 </p>
</td>
<td>
 <p class="normal">
  405
 </p>
</td>
<td>
 <p class="normal">
  10
 </p>
</td>
<td>
 <p class="normal">
  2017
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/extention+of+Z-Alizadeh+sani+dataset">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/extention+of+Z-Alizadeh+sani+dataset">
        extention of Z-Alizadeh sani dataset
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">It was collected for CAD diagnosis. </p></td> -->
 <td>
  <p class="normal">
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Integer, Real
  </p>
 </td>
 <td>
```

303

</td>
<td>

59

</td>
<td>

2017

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Z-Alizadeh+Sani">

</a>
</td>
<td>

<a href="datasets/Z-Alizadeh+Sani">
Z-Alizadeh Sani
</a>

</td>
</tr>
</table>
</td>
<!-- <td> It was collected for CAD diagnosis. </td> -->
<td>

</td>
<td>

Classification

</td>
<td>

Integer, Real

</td>
<td>

303

</td>
<td>

56

</td>

<td>

2017

</td>
<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Dynamic+Features+of+VirusShare+Executables">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Dynamic+Features+of+VirusShare+Executables">
Dynamic Features of VirusShare Executables
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset contains the dynamic features of 107,888 executables, collected b
y VirusShare from Nov/2010 to Jul/2014. </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

Classification, Regression

</td>
<td>

Integer

</td>
<td>

107888

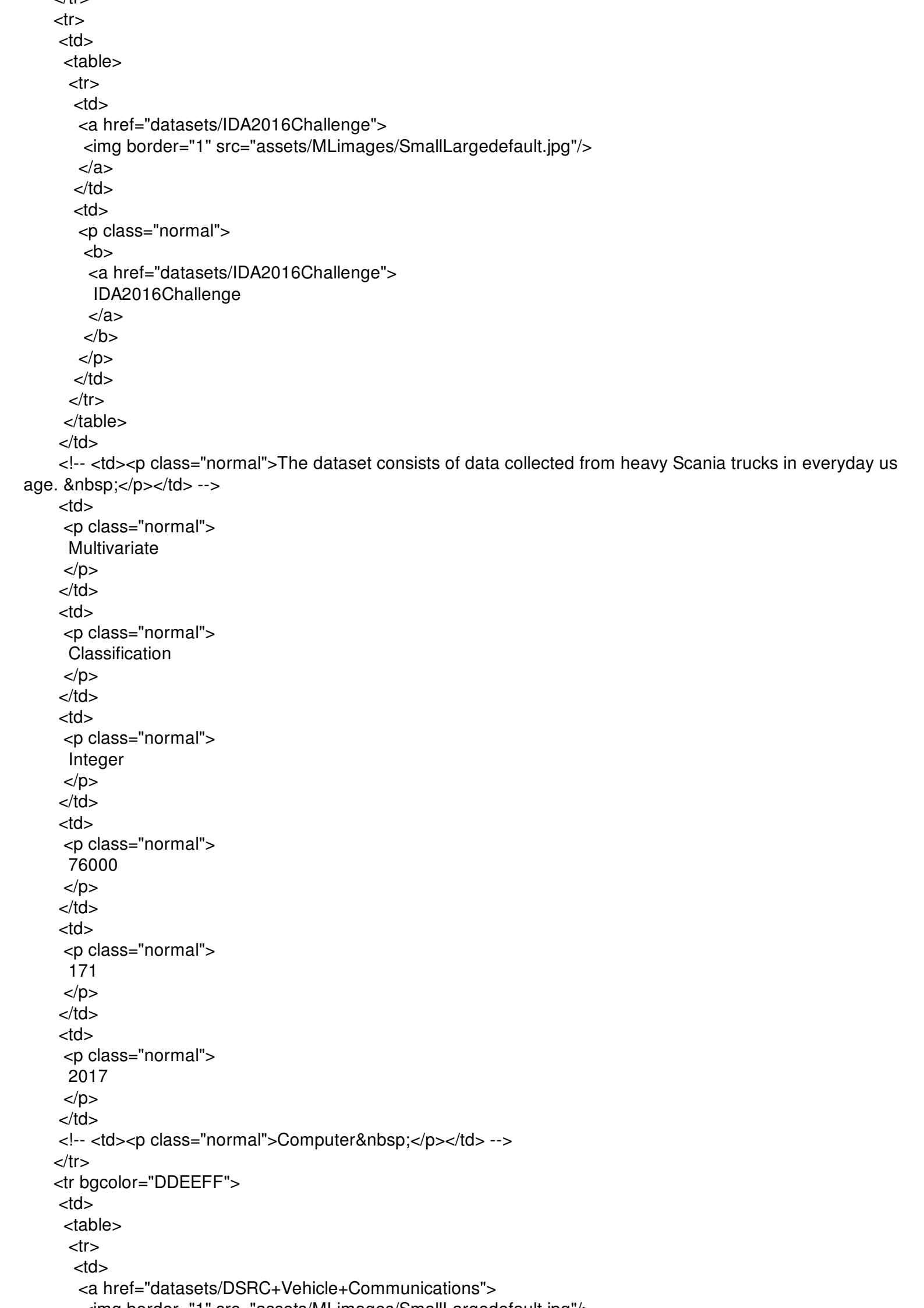
</td>
<td>

482

</td>
<td>

2017

</td>
<!-- <td> Computer </td> -->
</tr>

<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/IDA2016Challenge">

</a>
</td>
<td>

<a href="datasets/IDA2016Challenge">
IDA2016Challenge
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The dataset consists of data collected from heavy Scania trucks in everyday us
age.   </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

76000

</td>
<td>

171

</td>
<td>

2017

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/DSRC+Vehicle+Communications">

</a>
</td>
<td>

<a href="datasets/DSRC+Vehicle+Communications">
DSRC Vehicle Communications
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The dataset consists of data collected from heavy Scania trucks in everyday us
age.   </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

10000

</td>
<td>

100

</td>
<td>

2017

</td>
<!-- <td> Computer </td> -->
</tr>
</table>
</body>
</html>

<img border="1" src="assets/MLimages/SmallLargedefault.jpg" />
</a>
</td>
<td>

<a href="datasets/DSRC+Vehicle+Communications">
DSRC Vehicle Communications
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This set Provides data regarding wireless communications between vehicles and road side units. two separate data sets are provided (normal scenario) and in the presence of attacker (jammer). </td> -->
<td>

Sequential, Text

</td>
<td>

Clustering

</td>
<td>

Real

</td>
<td>

10000

</td>
<td>

5

</td>
<td>

2017

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Mturk+User-Perceived+Clusters+over+Images">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Mturk+User-Perceived+Clusters+over+Images">
Mturk User-Perceived Clusters over Images
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset was collected by Shan-Hung Wu and DataLab members at NTHU, Taiwan. There're 325 user-perceived clusters from 100 users and their corresponding descriptions. </td> -->
<td>

Multivariate, Text

</td>
<td>

Clustering

</td>
<td>

Integer

</td>
<td>

180

</td>
<td>

500

</td>
<td>

2016

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Character+Font+Images">

</a>
</td>
<td>

<a href="datasets/Character+Font+Images">
Character Font Images
</a>

</td>

```
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">Character images from scanned and computer generated fonts. </p></td
> -->
    <td>
     <p class="normal">
      Multivariate
     </p>
    </td>
    <td>
     <p class="normal">
      Classification
     </p>
    </td>
    <td>
     <p class="normal">
      Integer, Real
     </p>
    </td>
    <td>
     <p class="normal">
      745000
     </p>
    </td>
    <td>
     <p class="normal">
      411
     </p>
    </td>
    <td>
     <p class="normal">
      2016
     </p>
    </td>
    <!-- <td><p class="normal">Computer </p></td> -->
   </tr>
   <tr>
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Instanc
e+Labels">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Insta
nce+Labels">
           DeliciousMIL: A Data Set for Multi-Label Multi-Instance Learning with Instance Labels
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">This dataset includes 1) 12234 documents (8251 training, 3983 test) extracted
from DeliciousT140 dataset, 2) class labels for all documents, 3) labels for a subset of sentences of the test doc
```

```
uments. </p></td> -->
      <td>
       <p class="normal">
        Text
       </p>
      </td>
      <td>
       <p class="normal">
        Classification
       </p>
      </td>
      <td>
       <p class="normal">
        Integer
       </p>
      </td>
      <td>
       <p class="normal">
        12234
       </p>
      </td>
      <td>
       <p class="normal">
        8519
       </p>
      </td>
      <td>
       <p class="normal">
        2016
       </p>
      </td>
      <!-- <td><p class="normal">Computer </p></td> -->
     </tr>
     <tr bgcolor="DDEEFF">
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++">
             Autistic Spectrum Disorder Screening Data for Children
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">Children screening data for autism suitable for classification and predictive task
s  </p></td> -->
      <td>
       <p class="normal">
        Multivariate
       </p>
      </td>
      <td>
```

Classification

</td>
<td>

Integer

</td>
<td>

292

</td>
<td>

21

</td>
<td>

2017

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++">

</a>
</td>
<td>

<a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++">
Autistic Spectrum Disorder Screening Data for Adolescent
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Autistic Spectrum Disorder Screening Data for Adolescent. This dataset is relat
ed to classification and predictive tasks. </td> -->
<td>

Multivariate

</td>
<td>

```
 </p>
</td>
<td>
 <p class="normal">
  104
 </p>
</td>
<td>
 <p class="normal">
  21
 </p>
</td>
<td>
 <p class="normal">
  2017
 </p>
</td>
<!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/APS+Failure+at+Scania+Trucks">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/APS+Failure+at+Scania+Trucks">
        APS Failure at Scania Trucks
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">The datasets' positive class consists of component failures for a specific compo
nent of the APS system. The negative class consists of trucks with failures for components not related to the AP
S.&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Integer, Real
  </p>
 </td>
 <td>
  <p class="normal">
   60000
  </p>
```

```
</td>
<td>
 <p class="normal">
  171
 </p>
</td>
<td>
 <p class="normal">
  2017
 </p>
</td>
<!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Wireless+Indoor+Localization">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Wireless+Indoor+Localization">
        Wireless Indoor Localization
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Collected in indoor space by observing signal strengths of seven WiFi signals vi
sible on a smartphone. The decision variable is one of the four rooms. &amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
 </td>
 <td>
  <p class="normal">
   2000
  </p>
 </td>
 <td>
  <p class="normal">
   7
  </p>
 </td>
 <td>
```

<td>

2017

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/HCC+Survival">

</a>
</td>
<td>

<a href="datasets/HCC+Survival">
HCC Survival
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Hepatocellular Carcinoma dataset (HCC dataset) was collected at a University Hospital in Portugal. It contains real clinical data of 165 patients diagnosed with HCC. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer, Real

</td>
<td>

165

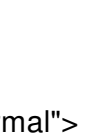
</td>
<td>

49

</td>
<td>

2017

</td>
<!-- <td> Life </td> -->
</tr>
<tr>

<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015">

</a>
</td>
<td>

<a href="datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015"
>
CSM (Conventional and Social Media Movies) Dataset 2014 and 2015
</a>

</td>
</tr>
</table>
</td>
<!-- <td> 12 features categorized as conventional and social media features. Both conve
ntional features, collected from movies databases on Web as well as social media features(YouTube,Twitter).&
nbsp; </td> -->
<td>

Multivariate

</td>
<td>

Classification, Regression

</td>
<td>

Integer

</td>
<td>

217

</td>
<td>

12

</td>
<td>

2017

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD+2016%29"

<a href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29">

<a href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29">

<a href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29">
University of Tehran Question Dataset 2016 (UTQD.2016)
Persian questions gathered from a jeopardy game broadcasted on Iranian national television.
Text
Classification
1175
3
2017
Other
<a href="datasets/Autism+Screening+Adult">

<a href="datasets/Autism+Screening+Adult">
Autism Screening Adult
</table>
</td>
<!-- <td> Persian questions gathered from a jeopardy game broadcasted on Iranian nati
onal television.   </td> -->
<td>

Text

</td>
<td>

Classification

</td>
<td>

</td>
<td>

1175

</td>
<td>

3

</td>
<td>

2017

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Autism+Screening+Adult">

```
          Autism Screening Adult
         </a>
        </b>
       </p>
      </td>
     </tr>
    </table>
   </td>
   <!-- <td><p class="normal">Autistic Spectrum Disorder Screening Data for Adult. This dataset is related to classification and predictive tasks.&amp;nbsp;</p></td> -->
   <td>
    <p class="normal">
    </p>
   </td>
   <td>
    <p class="normal">
     Classification
    </p>
   </td>
   <td>
    <p class="normal">
     Integer
    </p>
   </td>
   <td>
    <p class="normal">
     704
    </p>
   </td>
   <td>
    <p class="normal">
     21
    </p>
   </td>
   <td>
    <p class="normal">
     2017
    </p>
   </td>
   <!-- <td><p class="normal">Social&amp;nbsp;</p></td> -->
  </tr>
  <tr bgcolor="DDEEFF">
   <td>
    <table>
     <tr>
      <td>
       <a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor">
        <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
       </a>
      </td>
      <td>
       <p class="normal">
        <b>
         <a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor">
          Activity recognition with healthy older people using a batteryless wearable sensor
         </a>
        </b>
       </p>
      </td>
     </tr>
```

</table>
</td>
<!-- <td> Sequential motion data from 14 healthy older people aged 66 to 86 years old using a batteryless, wearable sensor on top of their clothing for the recognition of activities in clinical environments. </td> -->
<td>

Sequential

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

75128

</td>
<td>

9

</td>
<td>

2016

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Immunotherapy+Dataset">

</a>
</td>
<td>

<a href="datasets/Immunotherapy+Dataset">
Immunotherapy Dataset
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset contains information about wart treatment results of 90 patients using immunotherapy. </td> -->
<td>

```
 Univariate
 </p>
</td>
<td>
 <p class="normal">
  Classification
 </p>
</td>
<td>
 <p class="normal">
  Integer, Real
 </p>
</td>
<td>
 <p class="normal">
  90
 </p>
</td>
<td>
 <p class="normal">
  8
 </p>
</td>
<td>
 <p class="normal">
  2018
 </p>
</td>
<!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Cryotherapy+Dataset+">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Cryotherapy+Dataset+">
        Cryotherapy Dataset
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">This dataset contains information about wart treatment results of 90 patients us
ing cryotherapy.&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Univariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
```

</td>
<td>

Integer, Real

</td>
<td>

90

</td>
<td>

7

</td>
<td>

2018

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes">

</a>
</td>
<td>

<a href="datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes">
OCT data & Color Fundus Images of Left & Right Eyes
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset contains OCT data (in mat format) and color fundus data (in jpg fo
rmat) of left & right eyes of 50 healthy persons. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

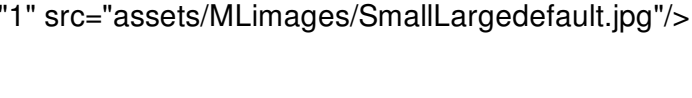
50

</td>
<td>

2

</td>
<td>

2016

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Discrete+Tone+Image+Dataset">

</a>
</td>
<td>

<a href="datasets/Discrete+Tone+Image+Dataset">
Discrete Tone Image Dataset
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Discrete Tone Images(DTI)are available which needs to be analyzed in detail. Here, we created this dataset for those who do research in DTI.
  </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

</td>
<td>

71

</td>
<td>

11

</td>
</tr>
</table>
</body>
</html>

```
 </p>
</td>
<td>
 <p class="normal">
  2018
 </p>
</td>
<!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/News+Popularity+in+Multiple+Social+Media+Platforms">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/News+Popularity+in+Multiple+Social+Media+Platforms">
        News Popularity in Multiple Social Media Platforms
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Large data set of news items and their respective social feedback on multiple p
latforms: Facebook, Google+ and LinkedIn.&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate, Time-Series, Text
  </p>
 </td>
 <td>
  <p class="normal">
   Regression
  </p>
 </td>
 <td>
  <p class="normal">
   Integer, Real
  </p>
 </td>
 <td>
  <p class="normal">
   93239
  </p>
 </td>
 <td>
  <p class="normal">
   11
  </p>
 </td>
 <td>
  <p class="normal">
   2018
  </p>
 </td>
```

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Ultrasonic+flowmeter+diagnostics">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Ultrasonic+flowmeter+diagnostics">
Ultrasonic flowmeter diagnostics
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Fault diagnosis of four liquid ultrasonic flowmeters </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

540

</td>
<td>

173

</td>
<td>

2018

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/ICMLA+2014+Accepted+Papers+Data+Set">

<a href="datasets/ICMLA+2014+Accepted+Papers+Data+Set">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/ICMLA+2014+Accepted+Papers+Data+Set">
ICMLA 2014 Accepted Papers Data Set
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This data set compromises the metadata for the 2014 ICMLA conference's acc
epted papers, including ID, paper titles, author's keywords, abstracts and sessions in which they were exposed.
  </td> -->
<td>

Multivariate

</td>
<td>

Classification, Clustering

</td>
<td>

</td>
<td>

105

</td>
<td>

5

</td>
<td>

2018

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation">
BLE RSSI Dataset for Indoor localization and Navigation
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset contains RSSI readings gathered from an array of Bluetooth Low Energy (BLE) iBeacons in a real-world and operational indoor environment for localization and navigation purposes. </td> -->
<td>

Multivariate, Sequential, Time-Series

</td>
<td>

Classification, Clustering

</td>
<td>

Integer

</td>
<td>

6611

</td>
<td>

15

</td>
<td>

2018

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Container+Crane+Controller+Data+Set">

</a>
</td>
<td>

<a href="datasets/Container+Crane+Controller+Data+Set">
Container Crane Controller Data Set
</a>

</td>

```
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">A container crane has the function of transporting containers from one point to
another point.&amp;nbsp;</p></td> -->
    <td>
     <p class="normal">
      Univariate, Domain-Theory
     </p>
    </td>
    <td>
     <p class="normal">
      Classification, Regression
     </p>
    </td>
    <td>
     <p class="normal">
      Real
     </p>
    </td>
    <td>
     <p class="normal">
      15
     </p>
    </td>
    <td>
     <p class="normal">
      3
     </p>
    </td>
    <td>
     <p class="normal">
      2018
     </p>
    </td>
    <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
   </tr>
   <tr bgcolor="DDEEFF">
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/Residential+Building+Data+Set">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Residential+Building+Data+Set">
           Residential Building Data Set
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">Data set includes construction cost, sale prices, project variables, and economi
c variables corresponding to real estate single-family residential apartments in Tehran, Iran. &amp;nbsp;</p></td> --
>
    <td>
```

Multivariate

</td>
<td>

Regression

</td>
<td>

Real

</td>
<td>

372

</td>
<td>

105

</td>
<td>

2018

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Health+News+in+Twitter">

</a>
</td>
<td>

<a href="datasets/Health+News+in+Twitter">
Health News in Twitter
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The data was collected in 2015 using Twitter API. This dataset contains health news from more than 15 major health news agencies such as BBC, CNN, and NYT.   </td> -->
<td>

Text

</td>
<td>

Clustering

</td>

```
 </p>
</td>
<td>
 <p class="normal">
  Real
 </p>
</td>
<td>
 <p class="normal">
  58000
 </p>
</td>
<td>
 <p class="normal">
  25000
 </p>
</td>
<td>
 <p class="normal">
  2018
 </p>
</td>
<!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/chipseq">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/chipseq">
        chipseq
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">ChIP-seq experiments characterize protein modifications or binding at
specific genomic locations in specific samples. The machine learning
problem in these data is structured binary classification.&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Sequential
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Integer
  </p>
```

```
      </td>
      <td>
       <p class="normal">
        4960
       </p>
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
        2018
       </p>
      </td>
      <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
     </tr>
     <tr>
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/SGEMM+GPU+kernel+performance">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/SGEMM+GPU+kernel+performance">
             SGEMM GPU kernel performance
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">Running times for multiplying two 2048 x 2048 matrices using a GPU OpenCL
SGEMM kernel with varying parameters (using the library 'CLTune').&amp;nbsp;</p></td> -->
      <td>
       <p class="normal">
        Multivariate
       </p>
      </td>
      <td>
       <p class="normal">
        Regression
       </p>
      </td>
      <td>
       <p class="normal">
        Integer
       </p>
      </td>
      <td>
       <p class="normal">
        241600
       </p>
      </td>
      <td>
       <p class="normal">
```

```
      18
     </p>
    </td>
    <td>
     <p class="normal">
      2018
     </p>
    </td>
    <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
   </tr>
   <tr bgcolor="DDEEFF">
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/Repeat+Consumption+Matrices">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Repeat+Consumption+Matrices">
           Repeat Consumption Matrices
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">The dataset contains 7 datasets of User - Item matrices, where each entry repr
esents how many times a user consumed an item. Item is used as an umbrella term for various categories.&nb
sp;</p></td> -->
    <td>
     <p class="normal">
      Multivariate
     </p>
    </td>
    <td>
     <p class="normal">
      Clustering
     </p>
    </td>
    <td>
     <p class="normal">
      Real
     </p>
    </td>
    <td>
     <p class="normal">
      130000
     </p>
    </td>
    <td>
     <p class="normal">
      21000
     </p>
    </td>
    <td>
     <p class="normal">
      2018
```

```
  2018
  </p>
 </td>
 <!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/detection_of_IoT_botnet_attacks_N_BaIoT">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/detection_of_IoT_botnet_attacks_N_BaIoT">
        detection_of_IoT_botnet_attacks_N_BaIoT
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">This dataset addresses the lack of public botnet datasets, especially for the IoT
. It suggests *real* traffic data, gathered from 9 commercial IoT devices authentically infected by Mirai and BAS
HLITE. </p></td> -->
 <td>
  <p class="normal">
   Multivariate, Sequential
  </p>
 </td>
 <td>
  <p class="normal">
   Classification, Clustering
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
 </td>
 <td>
  <p class="normal">
   7062606
  </p>
 </td>
 <td>
  <p class="normal">
   115
  </p>
 </td>
 <td>
  <p class="normal">
   2018
  </p>
 </td>
 <!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
```

<td>
<table>
<tr>
<td>
<a href="datasets/Absenteeism+at+work">

</a>
</td>
<td>

<a href="datasets/Absenteeism+at+work">
Absenteeism at work
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The database was created with records of absenteeism at work from July 2007 to July 2010 at a courier company in Brazil. </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

Classification, Clustering

</td>
<td>

Integer, Real

</td>
<td>

740

</td>
<td>

21

</td>
<td>

2018

</td>
<!-- <td> Business </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/SCADI">

</a>
</td>

```
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/SCADI">
        SCADI
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">First self-care activities dataset based on ICF-CY. </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification, Clustering
  </p>
 </td>
 <td>
  <p class="normal">
  </p>
 </td>
 <td>
  <p class="normal">
   70
  </p>
 </td>
 <td>
  <p class="normal">
   206
  </p>
 </td>
 <td>
  <p class="normal">
   2018
  </p>
 </td>
 <!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Condition+monitoring+of+hydraulic+systems">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Condition+monitoring+of+hydraulic+systems">
        Condition monitoring of hydraulic systems
       </a>
      </b>
     </p>
```

```
      </td>
     </tr>
    </table>
   </td>
   <!-- <td><p class="normal">The data set addresses the condition assessment of a hydraulic test rig based on multi sensor data. Four fault types are superimposed with several severity grades impeding selective quantification. </p></td> -->
   <td>
    <p class="normal">
     Multivariate, Time-Series
    </p>
   </td>
   <td>
    <p class="normal">
     Classification, Regression
    </p>
   </td>
   <td>
    <p class="normal">
     Real
    </p>
   </td>
   <td>
    <p class="normal">
     2205
    </p>
   </td>
   <td>
    <p class="normal">
     43680
    </p>
   </td>
   <td>
    <p class="normal">
     2018
    </p>
   </td>
   <!-- <td><p class="normal">Computer </p></td> -->
  </tr>
  <tr>
   <td>
    <table>
     <tr>
      <td>
       <a href="datasets/Carbon+Nanotubes">
        <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
       </a>
      </td>
      <td>
       <p class="normal">
        <b>
         <a href="datasets/Carbon+Nanotubes">
          Carbon Nanotubes
         </a>
        </b>
       </p>
      </td>
     </tr>
    </table>
   </td>
   <!-- <td><p class="normal">This dataset contains 10721 initial and calculated atomic coordinates of carbon nanotubes. </p></td> -->
```

<td>

Univariate

</td>
<td>

Regression

</td>
<td>

Real

</td>
<td>

10721

</td>
<td>

8

</td>
<td>

2018

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Optical+Interconnection+Network+">

</a>
</td>
<td>

<a href="datasets/Optical+Interconnection+Network+">
Optical Interconnection Network
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset contains 640 performance measurements from a simulation of 2-Dimensional Multiprocessor Optical Interconnection Network.   </td> -->
<td>

Multivariate

</td>
<td>

```
 Classification, Regression
 </p>
</td>
<td>
 <p class="normal">
  Integer, Real
 </p>
</td>
<td>
 <p class="normal">
  640
 </p>
</td>
<td>
 <p class="normal">
  10
 </p>
</td>
<td>
 <p class="normal">
  2018
 </p>
</td>
<!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Sports+articles+for+objectivity+analysis">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Sports+articles+for+objectivity+analysis">
        Sports articles for objectivity analysis
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">1000 sports articles were labeled using Amazon Mechanical Turk as objective or subjective. The raw texts, extracted features, and the URLs from which the articles were retrieved are provid ed.&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate, Text
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Integer
```

```
 </p>
</td>
<td>
 <p class="normal">
  1000
 </p>
</td>
<td>
 <p class="normal">
  59
 </p>
</td>
<td>
 <p class="normal">
  2018
 </p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Breast+Cancer+Coimbra">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Breast+Cancer+Coimbra">
        Breast Cancer Coimbra
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Clinical features were observed or measured for 64 patients with breast cancer
and 52 healthy controls.  </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Integer
  </p>
 </td>
 <td>
  <p class="normal">
   116
  </p>
 </td>
```

<td>

10

</td>
<td>

2018

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data">
GNFUV Unmanned Surface Vehicles Sensor Data
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The data-set contains four (4) sets of mobile sensor readings data (humidity, te
mperature) corresponding to a swarm of four (4) Unmanned Surface Vehicles (USVs) in a test-bed in Athens (
Greece).   </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

Regression

</td>
<td>

Real

</td>
<td>

1672

</td>
<td>

5

</td>
<td>

<td>

2018

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Dishonest+Internet+users+Dataset">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Dishonest+Internet+users+Dataset">
Dishonest Internet users Dataset
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The dataset was used to test an architecture based on a trust model capable t
o cope with the evaluation of the trustworthiness of users interacting in pervasive environments. </td
> -->
<td>

Multivariate

</td>
<td>

Classification, Clustering

</td>
<td>

</td>
<td>

322

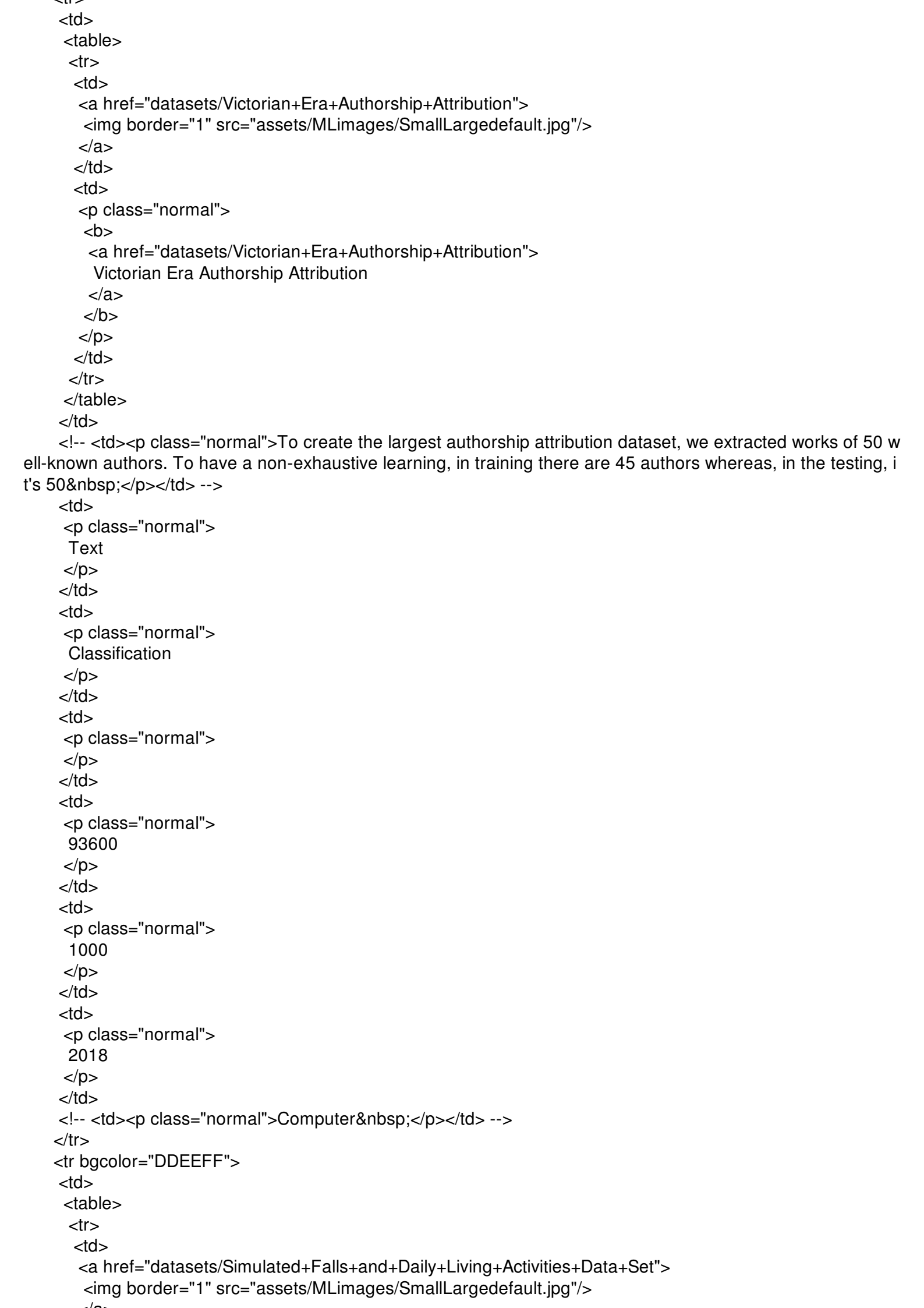
</td>
<td>

5

</td>
<td>

2018

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>

<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Victorian+Era+Authorship+Attribution">

</a>
</td>
<td>

<a href="datasets/Victorian+Era+Authorship+Attribution">
Victorian Era Authorship Attribution
</a>

</td>
</tr>
</table>
</td>
<!-- <td> To create the largest authorship attribution dataset, we extracted works of 50 well-known authors. To have a non-exhaustive learning, in training there are 45 authors whereas, in the testing, it's 50 </td> -->
<td>

Text

</td>
<td>

Classification

</td>
<td>

</td>
<td>

93600

</td>
<td>

1000

</td>
<td>

2018

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set">

</a>
</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

```
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set">
           Simulated Falls and Daily Living Activities Data Set
          </a>
         </b>
        </p>
       </td>
      </tr>
     </table>
    </td>
    <!-- <td><p class="normal">20 falls and 16 daily living activities were performed by 17 volunteers with 5 rep
etitions while wearing 6 sensors (3.060 instances) that attached to their head, chest, waist, wrist, thigh and ankl
e.&amp;nbsp;</p></td> -->
    <td>
     <p class="normal">
      Time-Series
     </p>
    </td>
    <td>
     <p class="normal">
      Classification
     </p>
    </td>
    <td>
     <p class="normal">
      Integer
     </p>
    </td>
    <td>
     <p class="normal">
      3060
     </p>
    </td>
    <td>
     <p class="normal">
      138
     </p>
    </td>
    <td>
     <p class="normal">
      2018
     </p>
    </td>
    <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
   </tr>
   <tr>
    <td>
     <table>
      <tr>
       <td>
        <a href="datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing">
         <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
        </a>
       </td>
       <td>
        <p class="normal">
         <b>
          <a href="datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing">
```

Multimodal Damage Identification for Humanitarian Computing
</a>

</td>
</tr>
</table>
</td>
<!-- <td> 5879 captioned images (image and text) from social media related to damage during natural disasters/wars, and belong to 6 classes: Fires, Floods, Natural landscape, Infrastructural, Human , Non-damage. </td> -->
<td>

Multivariate, Text

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

5879

</td>
<td>

</td>
<td>

2018

</td>
<!-- <td> Social </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals">

</a>
</td>
<td>

<a href="datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals">
EEG Steady-State Visual Evoked Potential Signals
</a>

</td>
</tr>
</table>

```
      </td>
      <!-- <td><p class="normal">This database consists on 30 subjects performing Brain Computer Interface for Steady State Visual Evoked Potentials (BCI-SSVEP). &amp;nbsp;</p></td> -->
      <td>
       <p class="normal">
        Multivariate, Time-Series
       </p>
      </td>
      <td>
       <p class="normal">
        Classification, Regression
       </p>
      </td>
      <td>
       <p class="normal">
        Integer
       </p>
      </td>
      <td>
       <p class="normal">
        9200
       </p>
      </td>
      <td>
       <p class="normal">
        16
       </p>
      </td>
      <td>
       <p class="normal">
        2018
       </p>
      </td>
      <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
     </tr>
     <tr>
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/Roman+Urdu+Data+Set">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Roman+Urdu+Data+Set">
             Roman Urdu Data Set
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">Roman Urdu (the scripting style for Urdu language) is one of the limited resource languages.A data corpus comprising of more than 20000 records was collected.&amp;nbsp;</p></td> -->
      <td>
       <p class="normal">
        Text
       </p>
```

```
</td>
<td>
 <p class="normal">
  Classification
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
  20000
 </p>
</td>
<td>
 <p class="normal">
  2
 </p>
</td>
<td>
 <p class="normal">
  2018
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Avila">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Avila">
        Avila
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">The Avila data set has been extracted from 800 images of the 'Avila Bible', an XII century giant Latin copy of the Bible.  The prediction task consists in associating each pattern to a copyist. </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
```

Real

</td>
<td>

20867

</td>
<td>

10

</td>
<td>

2018

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/PANDOR">

</a>
</td>
<td>

<a href="datasets/PANDOR">
PANDOR
</a>

</td>
</tr>
</table>
</td>
<!-- <td> PANDOR is a novel and publicly available dataset for online recommendation p
rovided by Purch (http://www.purch.com/).   </td> -->
<td>

Multivariate

</td>
<td>

Recommendation

</td>
<td>

Categorical

</td>
<td>

```
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
  2018
 </p>
</td>
<!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Drug+Review+Dataset+%28Druglib.com%29">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Drug+Review+Dataset+%28Druglib.com%29">
        Drug Review Dataset (Druglib.com)
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">The dataset provides patient reviews on specific drugs along with related condi
tions. Reviews and ratings are grouped into reports on the three aspects benefits, side effects and overall com
ment.&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate, Text
  </p>
 </td>
 <td>
  <p class="normal">
   Classification, Regression, Clustering
  </p>
 </td>
 <td>
  <p class="normal">
   Integer
  </p>
 </td>
 <td>
  <p class="normal">
   4143
  </p>
 </td>
 <td>
  <p class="normal">
   8
  </p>
 </td>
 <td>
```

2018

</td>
<!-- <td>   </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Drug+Review+Dataset+%28Drugs.com%29">

</a>
</td>
<td>

<a href="datasets/Drug+Review+Dataset+%28Drugs.com%29">
Drug Review Dataset (Drugs.com)
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The dataset provides patient reviews on specific drugs along with related condi
tions and a 10 star patient rating reflecting overall patient satisfaction. </td> -->
<td>

Multivariate, Text

</td>
<td>

Classification, Regression, Clustering

</td>
<td>

Integer

</td>
<td>

215063

</td>
<td>

6

</td>
<td>

2018

</td>
<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEEE">

<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Physical+Unclonable+Functions">

</a>
</td>
<td>

<a href="datasets/Physical+Unclonable+Functions">
Physical Unclonable Functions
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The dataset is generated from Physical Unclonable Functions (PUFs) simulatio
n, specifically XOR Arbiter PUFs. PUFs are used for authentication purposes. For more info, refer to our paper
below. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

6000000

</td>
<td>

129

</td>
<td>

2018

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Superconductivty+Data">

</a>
</td>
<td>

<a href="datasets/Superconductivty+Data">
Superconductivty Data
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The dataset is generated from Physical Unclonable Functions (PUFs) simulatio
n, specifically XOR Arbiter PUFs. PUFs are used for authentication purposes. For more info, refer to our paper
below. </td> -->
<td>

Multivariate

</td>
<td>

Regression

</td>
<td>

Integer

</td>
<td>

21263

</td>
<td>

81

</td>
<td>

2018

</td>
<!-- <td> Computer </td> -->
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>
</table>
</td>

<img border="1" src="assets/MLimages/SmallLargedefault.jpg" />
</a>
</td>
<td>

<a href="datasets/Superconductivty+Data">
Superconductivty Data
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Two file s contain data on 21263 superconductors and their relevant features.&
nbsp; </td> -->
<td>

Multivariate

</td>
<td>

Regression

</td>
<td>

Real

</td>
<td>

21263

</td>
<td>

81

</td>
<td>

2018

</td>
<!-- <td> Physical </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29">

```
          WESAD (Wearable Stress and Affect Detection)
         </a>
        </b>
       </p>
      </td>
     </tr>
    </table>
   </td>
   <!-- <td><p class="normal">WESAD (Wearable Stress and Affect Detection) contains data of 15 subjects d
uring a stress-affect lab study, while wearing physiological and motion sensors. </p></td> -->
   <td>
    <p class="normal">
     Multivariate, Time-Series
    </p>
   </td>
   <td>
    <p class="normal">
     Classification, Regression
    </p>
   </td>
   <td>
    <p class="normal">
     Real
    </p>
   </td>
   <td>
    <p class="normal">
     63000000
    </p>
   </td>
   <td>
    <p class="normal">
     12
    </p>
   </td>
   <td>
    <p class="normal">
     2018
    </p>
   </td>
   <!-- <td><p class="normal">Computer </p></td> -->
  </tr>
  <tr>
   <td>
    <table>
     <tr>
      <td>
       <a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2">
        <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
       </a>
      </td>
      <td>
       <p class="normal">
        <b>
         <a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2">
          GNFUV Unmanned Surface Vehicles Sensor Data Set 2
         </a>
        </b>
       </p>
      </td>
     </tr>
    </table>
```

```
</td>
<!-- <td><p class="normal">The data-set contains eight (2x4) data-sets of mobile sensor readings data (humidity, temperature) corresponding to a swarm of four Unmanned Surface Vehicles (USVs) in a test-bed, Athens, Greece. </p></td> -->
<td>
 <p class="normal">
  Multivariate, Sequential, Time-Series
 </p>
</td>
<td>
 <p class="normal">
  Regression
 </p>
</td>
<td>
 <p class="normal">
  Real
 </p>
</td>
<td>
 <p class="normal">
  10190
 </p>
</td>
<td>
 <p class="normal">
  6
 </p>
</td>
<td>
 <p class="normal">
  2018
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Student+Academics+Performance">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Student+Academics+Performance">
        Student Academics Performance
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
<!-- <td><p class="normal">The dataset tried to find the end semester percentage prediction based on different social, economic and academic attributes.  </p></td> -->
<td>
 <p class="normal">
  Multivariate
```

```
 </p>
</td>
<td>
 <p class="normal">
  Classification
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
  300
 </p>
</td>
<td>
 <p class="normal">
  22
 </p>
</td>
<td>
 <p class="normal">
  2018
 </p>
</td>
<!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Online+Shoppers+Purchasing+Intention+Dataset">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Online+Shoppers+Purchasing+Intention+Dataset">
        Online Shoppers Purchasing Intention Dataset
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Of the 12,330 sessions in the dataset,
84.5% (10,422) were negative class samples that did not
end with shopping, and the rest (1908) were positive class
samples ending with shopping.&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification, Clustering
  </p>
```

</td>
<td>

Integer, Real

</td>
<td>

12330

</td>
<td>

18

</td>
<td>

2018

</td>
<!-- <td> Business </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/PMU-UD">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/PMU-UD">
PMU-UD
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The handwritten dataset was collected from 170 participants with a total of 5,18
0 numeral patterns. The dataset is named Prince Mohammad Bin Fahd University - Urdu/Arabic Database (PM
U-UD).   </td> -->
<td>

Univariate

</td>
<td>

Classification

</td>
<td>

</td>
<td>

5180

</td>
<td>

9

</td>
<td>

2018

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Parkinson%27s+Disease+Classification">

</a>
</td>
<td>

<a href="datasets/Parkinson%27s+Disease+Classification">
Parkinson's Disease Classification
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The data used in this study were gathered from 188 patients with PD (107 men and 81 women) with ages ranging from 33 to 87 (65.1±10.9). </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Integer, Real

</td>
<td>

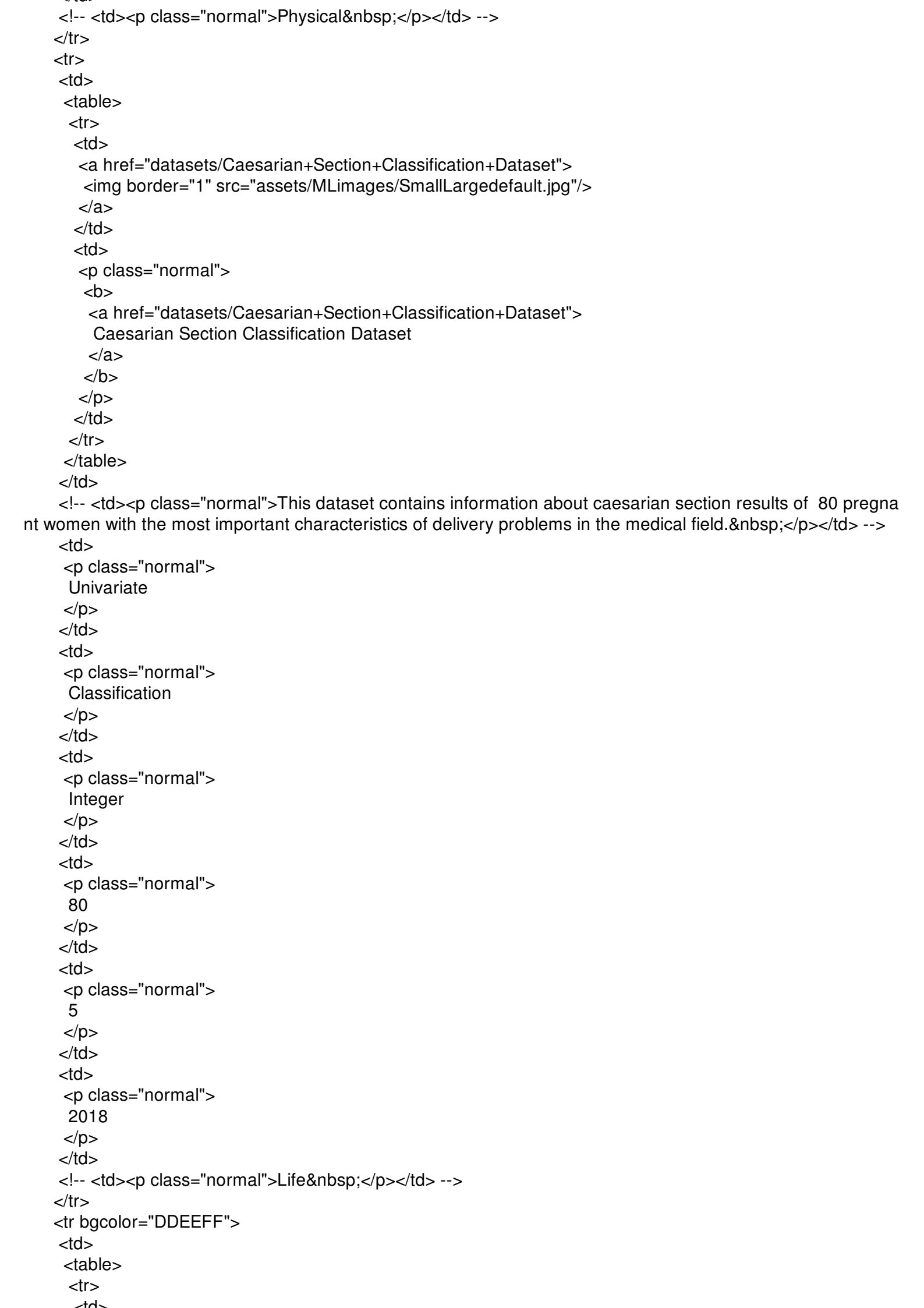
756

</td>
<td>

754

</td>
</tr>
</table>
</body>
</html>

```
 </p>
</td>
<td>
 <p class="normal">
  2018
 </p>
</td>
<!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Electrical+Grid+Stability+Simulated+Data+">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Electrical+Grid+Stability+Simulated+Data+">
        Electrical Grid Stability Simulated Data
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">The local stability analysis of the 4-node star system (electricity producer is in t
he center) implementing Decentral Smart Grid Control concept. &amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification, Regression
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
 </td>
 <td>
  <p class="normal">
   10000
  </p>
 </td>
 <td>
  <p class="normal">
   14
  </p>
 </td>
 <td>
  <p class="normal">
   2018
  </p>
 </td>
```

<!-- <td> Physical </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Caesarian+Section+Classification+Dataset">

</a>
</td>
<td>

<a href="datasets/Caesarian+Section+Classification+Dataset">
Caesarian Section Classification Dataset
</a>

</td>
</tr>
</table>
</td>
<!-- <td> This dataset contains information about caesarian section results of 80 pregna
nt women with the most important characteristics of delivery problems in the medical field. </td> -->
<td>

Univariate

</td>
<td>

Classification

</td>
<td>

Integer

</td>
<td>

80

</td>
<td>

5

</td>
<td>

2018

</td>
<!-- <td> Life </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

<td>
<a href="datasets/BAUM-1">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/BAUM-1">
BAUM-1
</a>

</td>
</tr>
</table>
</td>
<!-- <td> BAUM-1 dataset contains 1184 multimodal facial video clips collected from 31 s
ubjects. The 1184 video clips contain spontaneous facial expressions and speech of 13 emotional and mental s
tates. </td> -->
<td>

Time-Series

</td>
<td>

Classification

</td>
<td>

</td>
<td>

1184

</td>
<td>

</td>
<td>

2018

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/BAUM-2">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

</td>
</tr>
</table>
</td>
</tr>
</table>
</td>
</tr>

<a href= datasets/BAUM-2 >
BAUM-2
</a>

</td>
</tr>
</table>
</td>
<!-- <td> A multilingual audio-visual affective face database consisting of 1047 video clip
s of 286 subjects.   </td> -->
<td>

Time-Series

</td>
<td>

Classification

</td>
<td>

</td>
<td>

1047

</td>
<td>

</td>
<td>

2018

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Audit+Data">
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>
</td>
<td>

<a href="datasets/Audit+Data">
Audit Data
</a>

</td>
</tr>
</table>
</td>

<!-- <td> Exhaustive one year non-confidential data in the year 2015 to 2016 of firms is c
ollected from the Auditor Office of India to build a predictor for classifying suspicious firms. </td> -->
<td>

Multivariate

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

777

</td>
<td>

18

</td>
<td>

2018

</td>
<!-- <td> Other </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/BuddyMove+Data+Set">

</a>
</td>
<td>

<a href="datasets/BuddyMove+Data+Set">
BuddyMove Data Set
</a>

</td>
</tr>
</table>
</td>
<!-- <td> User interest information extracted from user reviews published in holidayiq.co
m about various types of point of interests in South India </td> -->
<td>

Multivariate, Text

</td>

<td>

Classification, Clustering

</td>
<td>

Real

</td>
<td>

249

</td>
<td>

7

</td>
<td>

2018

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Real+estate+valuation+data+set">

</a>
</td>
<td>

<a href="datasets/Real+estate+valuation+data+set">
Real estate valuation data set
</a>

</td>
</tr>
</table>
</td>
<!-- <td> The “real estate valuation” is a regression problem. The market historical data set of real estate valuation are collected from Sindian Dist., New Taipei City, Taiwan.   </td> -->
<td>

Multivariate

</td>
<td>

Regression

</td>
<td>

```
 Integer, Real
 </p>
</td>
<td>
 <p class="normal">
  414
 </p>
</td>
<td>
 <p class="normal">
  7
 </p>
</td>
<td>
 <p class="normal">
  2018
 </p>
</td>
<!-- <td><p class="normal">Business&amp;nbsp;</p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connecte
d+speech+Data+Set+">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connect
ed+speech+Data+Set+">
        Early biomarkers of Parkinson’s disease based on natural connected speech Data Set
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">.&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
 </td>
 <td>
  <p class="normal">
  </p>
```

```
      </td>
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
        2018
       </p>
      </td>
      <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
     </tr>
     <tr bgcolor="DDEEFF">
      <td>
       <table>
        <tr>
         <td>
          <a href="datasets/Somerville+Happiness+Survey">
           <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
          </a>
         </td>
         <td>
          <p class="normal">
           <b>
            <a href="datasets/Somerville+Happiness+Survey">
             Somerville Happiness Survey
            </a>
           </b>
          </p>
         </td>
        </tr>
       </table>
      </td>
      <!-- <td><p class="normal">A data extract of a non-federal dataset posted here https://catalog.data.gov/dat
aset/somerville-happiness-survey-responses-2011-2013-2015&amp;nbsp;</p></td> -->
      <td>
       <p class="normal">
       </p>
      </td>
      <td>
       <p class="normal">
        Classification
       </p>
      </td>
      <td>
       <p class="normal">
        Integer
       </p>
      </td>
      <td>
       <p class="normal">
        143
       </p>
      </td>
      <td>
       <p class="normal">
        7
       </p>
      </td>
      <td>
       <p class="normal">
        2018
```

```
 </p>
 </td>
 <!-- <td><p class="normal">Life&amp;nbsp;</p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/2.4+GHZ+Indoor+Channel+Measurements">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/2.4+GHZ+Indoor+Channel+Measurements">
        2.4 GHZ Indoor Channel Measurements
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">Measurement of the S21,consists of 10 sweeps, each sweep contains 601 freq
uency points with spacing of 0.167MHz to cover a 100MHz band centered at 2.4GHz.&amp;nbsp;</p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
  <p class="normal">
   Real
  </p>
 </td>
 <td>
  <p class="normal">
   7840
  </p>
 </td>
 <td>
  <p class="normal">
   5
  </p>
 </td>
 <td>
  <p class="normal">
   2018
  </p>
 </td>
 <!-- <td><p class="normal">Computer&amp;nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
  <table>
```

<tr>
<td>
<a href="datasets/EMG+data+for+gestures">

</a>
</td>
<td>

<a href="datasets/EMG+data+for+gestures">
EMG data for gestures
</a>

</td>
</tr>
</table>
</td>
<!-- <td> These are files of raw EMG data recorded by MYO Thalmic bracelet </p
></td> -->
<td>

Time-Series

</td>
<td>

Classification

</td>
<td>

Real

</td>
<td>

30000

</td>
<td>

6

</td>
<td>

2019

</td>
<!-- <td> Life </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Parking+Birmingham">

</a>
</td>
<td>

<td>

<a href="datasets/Parking+Birmingham">
Parking Birmingham
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Data collected from car parks in Birmingham that are operated by NCP from Birmingham City Council. UK Open Government Licence (OGL). https://data.birmingham.gov.uk/dataset/birmingham-parking </td> -->
<td>

Multivariate, Univariate, Sequential, Time-Series

</td>
<td>

Classification, Regression, Clustering

</td>
<td>

Real

</td>
<td>

35717

</td>
<td>

4

</td>
<td>

2019

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Behavior+of+the+urban+traffic+of+the+city+of+Sao+Paulo+in+Brazil">

</a>
</td>
<td>

<a href="datasets/Behavior+of+the+urban+traffic+of+the+city+of+Sao+Paulo+in+Brazil">
Behavior of the urban traffic of the city of Sao Paulo in Brazil
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate, Sequential, Time-Series

</td>
<td>

Classification, Regression, Clustering

</td>
<td>

Real

</td>
<td>

135

</td>
<td>

4

</td>
<td>

2019

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Bike+Sharing+Dataset">

</a>
</td>
<td>

<a href="datasets/Bike+Sharing+Dataset">
Bike Sharing Dataset
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate, Sequential, Time-Series

</td>
<td>

Regression

</td>
<td>

Integer

</td>
<td>

17389

</td>
<td>

16

</td>
<td>

2013

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Beijing+PM2.5+Data">

</a>
</td>
<td>

<a href="datasets/Beijing+PM2.5+Data">
Beijing PM2.5 Data
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate, Time-Series

</td>
<td>

Regression

</td>
<td>

Real

</td>
<td>

43824

</td>
<td>

11

</td>
<td>

2015

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Beijing+Multi-Site+Air-Quality+Data">

</a>
</td>
<td>

<a href="datasets/Beijing+Multi-Site+Air-Quality+Data">
Beijing Multi-Site Air-Quality Data
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate, Time-Series

</td>
<td>

Regression

</td>
<td>

Real

</td>
<td>

420768

</td>
<td>

18

</td>
<td>

2017

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Appliances+energy+prediction">

</a>
</td>
<td>

<a href="datasets/Appliances+energy+prediction">
Appliances energy prediction
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate, Time-Series

</td>
<td>

Regression

</td>
<td>

Real

</td>
<td>

19735

</td>
<td>

29

</td>
<td>

2017

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Metro+Interstate+Traffic+Volume">

</a>
</td>
<td>

<a href="datasets/Metro+Interstate+Traffic+Volume">
Metro Interstate Traffic Volume
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate, Time-Series

</td>
<td>

Regression

</td>
<td>

Integer, Real

</td>
<td>

48204

</td>
<td>

8

</td>
<td>

2019

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations">

</a>
</td>
<td>

<a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations">
Gas Sensor Array Drift Dataset at Different Concentrations
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate, Time-Series

</td>
<td>

Classification, Regression, Clustering

</td>
<td>

Real

</td>
<td>

13910

</td>
<td>

129

</td>
<td>

2015

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Gas+Sensor+Array+Drift+Dataset">

</a>
</td>
<td>

<a href="datasets/Gas+Sensor+Array+Drift+Dataset">
Gas Sensor Array Drift Dataset
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate, Time-Series

</td>
<td>

Classification, Regression, Clustering

</td>
<td>

Real

</td>
<td>

13910

</td>
<td>

128

</td>
<td>

2012

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Online+Retail+II">

</a>
</td>
<td>

<a href="datasets/Online+Retail+II">
Online Retail II
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate, Sequential, Time-Series

</td>
<td>

Clustering

</td>
<td>

Integer, Real

</td>
<td>

525461

</td>
<td>

8

</td>
<td>

2019

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Air+Quality">

</a>
</td>
<td>

<a href="datasets/Air+Quality">
Air Quality
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate, Time-Series

</td>
<td>

Regression

</td>
<td>

Real

</td>
<td>

9358

</td>
<td>

15

</td>
<td>

2016

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Beijing+Multi-Site+Air-Quality+Data">

</a>
</td>
<td>

<a href="datasets/Beijing+Multi-Site+Air-Quality+Data">
Beijing Multi-Site Air-Quality Data
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate, Time-Series

</td>
<td>

Regression

</td>
<td>

Real

</td>
<td>

420768

</td>
<td>

18

</td>
<td>

2017

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Appliances+energy+prediction">

</a>
</td>
<td>

<a href="datasets/Appliances+energy+prediction">
Appliances energy prediction
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate, Time-Series

</td>
<td>

Regression

</td>
<td>

Real

</td>
<td>

19735

</td>
<td>

29

</td>
<td>

2017

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Metro+Interstate+Traffic+Volume">

</a>
</td>
<td>

<a href="datasets/Metro+Interstate+Traffic+Volume">
Metro Interstate Traffic Volume
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate, Time-Series

</td>
<td>

Regression

</td>
<td>

Integer, Real

</td>
<td>

48204

</td>
<td>

8

</td>
<td>

2019

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations">

</a>
</td>
<td>

<a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations">
Gas Sensor Array Drift Dataset at Different Concentrations
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate, Time-Series

</td>
<td>

Classification, Regression, Clustering

</td>
<td>

Real

</td>
<td>

13910

</td>
<td>

129

</td>
<td>

2015

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Gas+Sensor+Array+Drift+Dataset">

</a>
</td>
<td>

<a href="datasets/Gas+Sensor+Array+Drift+Dataset">
Gas Sensor Array Drift Dataset
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate, Time-Series

</td>
<td>

Classification, Regression, Clustering

</td>
<td>

Real

</td>
<td>

13910

</td>
<td>

128

</td>
<td>

2012

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Online+Retail+II">

</a>
</td>
<td>

<a href="datasets/Online+Retail+II">
Online Retail II
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate, Sequential, Time-Series

</td>
<td>

Clustering

</td>
<td>

Integer, Real

</td>
<td>

525461

</td>
<td>

8

</td>
<td>

2019

</td>
<!-- <td> Computer </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Air+Quality">

</a>
</td>
<td>

<a href="datasets/Air+Quality">
Air Quality
</a>

</td>
</tr>
</table>
</td>
<td>

Multivariate, Time-Series

</td>
<td>

Regression

</td>
<td>

Real

</td>
<td>

9358

</td>
<td>

15

</td>
<td>

2016

</td>
<!-- <td> Computer </td> -->
</tr>
</table>
</body>
</html>

</td>
</tr>
</table>
</td>
<!-- <td> The database was created with records of behavior of the urban traffic of the ci
ty of Sao Paulo in Brazil. </td> -->
<td>

Multivariate, Time-Series

</td>
<td>

Classification, Regression

</td>
<td>

Integer, Real

</td>
<td>

135

</td>
<td>

18

</td>
<td>

2018

</td>
<!-- <td> Computer </td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Travel+Reviews">

</a>
</td>
<td>

<a href="datasets/Travel+Reviews">
Travel Reviews
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Reviews on destinations in 10 categories mentioned across East Asia. Each tra

veler rating is mapped as Excellent(4), Very Good(3), Average(2), Poor(1), and Terrible(0) and average rating i
s used. </td> -->
<td>

Multivariate, Text

</td>
<td>

Classification, Clustering

</td>
<td>

Real

</td>
<td>

980

</td>
<td>

11

</td>
<td>

2018

</td>
<!-- <td> Other </td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Tarvel+Review+Ratings">

</a>
</td>
<td>

<a href="datasets/Tarvel+Review+Ratings">
Tarvel Review Ratings
</a>

</td>
</tr>
</table>
</td>
<!-- <td> Google reviews on attractions from 24 categories across Europe are considere
d. Google user rating ranges from 1 to 5 and average user rating per category is calculated. </td> --
>
<td>

Multivariate, Text

</td>

```
</td>
<td>
 <p class="normal">
  Classification, Clustering
 </p>
</td>
<td>
 <p class="normal">
  Real
 </p>
</td>
<td>
 <p class="normal">
  5456
 </p>
</td>
<td>
 <p class="normal">
  25
 </p>
</td>
<td>
 <p class="normal">
  2018
 </p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
 <td>
  <table>
   <tr>
    <td>
     <a href="datasets/Rice+Leaf+Diseases">
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
    </td>
    <td>
     <p class="normal">
      <b>
       <a href="datasets/Rice+Leaf+Diseases">
        Rice Leaf Diseases
       </a>
      </b>
     </p>
    </td>
   </tr>
  </table>
 </td>
 <!-- <td><p class="normal">There are three classes/diseases: Bacterial leaf blight, Brown spot, and Leaf s
mut, each having 40 images. The format of all images is jpg.   </p></td> -->
 <td>
  <p class="normal">
   Multivariate
  </p>
 </td>
 <td>
  <p class="normal">
   Classification
  </p>
 </td>
 <td>
```

```
      <p class="normal">
       Integer
      </p>
     </td>
     <td>
      <p class="normal">
       120
      </p>
     </td>
     <td>
      <p class="normal">
      </p>
     </td>
     <td>
      <p class="normal">
       2019
      </p>
     </td>
     <!-- <td><p class="normal">Computer </p></td> -->
    </tr>
   </table>
  </td>
 </tr>
</table>
<table align="center" cellpadding="5">
 <tr valign="center">
  <td>
   <p class="normal">
    Supported By:
   </p>
  </td>
  <td>
   <img height="60" src="assets/nsfe.gif"/>
  </td>
  <td>
   <p class="normal">
    In Collaboration With:
   </p>
  </td>
  <td>
   <img src="assets/rexaSmall.jpg"/>
  </td>
 </tr>
</table>
<center>
 <span class="normal">
  <a href="about.html">
   About
  </a>
  ||
  <a href="citation_policy.html">
   Citation Policy
  </a>
  ||
  <a href="donation_policy.html">
   Donation Policy
  </a>
  ||
  <a href="contact.html">
   Contact
  </a>
  ||
```

```
    <a href="http://cml.ics.uci.edu">
     CML
    </a>
   </span>
  </center>
 </body>
</html>
```

In [25]:

```
links = soup.find_all("a")
links
```

Out[25]:

```
[<a alt="Home" href="index.html"><img border="0" src="assets/logo.gif"/></a>,
 <a href="http://cml.ics.uci.edu"><font color="FFDD33">Center for Machine Learning and Intelligent Systems</f
ont></a>,
 <a href="about.html">About</a>,
 <a href="citation_policy.html">Citation Policy</a>,
 <a href="donation_policy.html">Donate a Data Set</a>,
 <a href="contact.html">Contact</a>,
 <a href="http://www.google.com/search"><img align="middle" alt="Google" border="0" height="27" src="http://w
ww.google.com/logos/Logo_25blk.gif"/></a>,
 <a href="datasets.php"><font color="#FFDD33" size="3"><b>View ALL Data Sets</b></font></a>,
 <a href="datasets.php?format=&amp;task=cla&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=
&amp;sort=nameUp&amp;view=table">Classification</a>,
 <a href="datasets.php?format=&amp;task=reg&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type
=&amp;sort=nameUp&amp;view=table">Regression</a>,
 <a href="datasets.php?format=&amp;task=clu&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=
&amp;sort=nameUp&amp;view=table">Clustering</a>,
 <a href="datasets.php?format=&amp;task=other&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;typ
e=&amp;sort=nameUp&amp;view=table">Other</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=cat&amp;area=&amp;numAtt=&amp;numIns=&amp;type=
&amp;sort=nameUp&amp;view=table">Categorical</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=num&amp;area=&amp;numAtt=&amp;numIns=&amp;type
=&amp;sort=nameUp&amp;view=table">Numerical</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=mix&amp;area=&amp;numAtt=&amp;numIns=&amp;type
=&amp;sort=nameUp&amp;view=table">Mixed</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=mv
ar&amp;sort=nameUp&amp;view=table">Multivariate</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=uv
ar&amp;sort=nameUp&amp;view=table">Univariate</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=se
q&amp;sort=nameUp&amp;view=table">Sequential</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=ts&
amp;sort=nameUp&amp;view=table">Time-Series</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=tex
t&amp;sort=nameUp&amp;view=table">Text</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=dt
&amp;sort=nameUp&amp;view=table">Domain-Theory</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=oth
er&amp;sort=nameUp&amp;view=table">Other</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=life&amp;numAtt=&amp;numIns=&amp;type=
&amp;sort=nameUp&amp;view=table">Life Sciences</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=phys&amp;numAtt=&amp;numIns=&amp;typ
e=&amp;sort=nameUp&amp;view=table">Physical Sciences</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=comp&amp;numAtt=&amp;numIns=&amp;typ
e=&amp;sort=nameUp&amp;view=table">CS / Engineering</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=soc&amp;numAtt=&amp;numIns=&amp;type
=&amp;sort=nameUp&amp;view=table">Social Sciences</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=bus&amp;numAtt=&amp;numIns=&amp;type
=&amp;sort=nameUp&amp;view=table">Business</a>,
```

```
=&amp;sort=nameUp&amp;view=table">Business</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=game&amp;numAtt=&amp;numIns=&amp;type=&amp;sort=nameUp&amp;view=table">Game</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=other&amp;numAtt=&amp;numIns=&amp;type=&amp;sort=nameUp&amp;view=table">Other</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=less10&amp;numIns=&amp;type=&amp;sort=nameUp&amp;view=table">Less than 10</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=10to100&amp;numIns=&amp;type=&amp;sort=nameUp&amp;view=table">10 to 100</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=greater100&amp;numIns=&amp;type=&amp;sort=nameUp&amp;view=table">Greater than 100</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=less100&amp;type=&amp;sort=nameUp&amp;view=table">Less than 100</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=100to1000&amp;type=&amp;sort=nameUp&amp;view=table">100 to 1000</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=greater1000&amp;type=&amp;sort=nameUp&amp;view=table">Greater than 1000</a>,
 <a href="datasets.php?format=mat&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=&amp;sort=nameUp&amp;view=table">Matrix</a>,
 <a href="datasets.php?format=nonmat&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=&amp;sort=nameUp&amp;view=table">Non-Matrix</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=&amp;sort=nameUp&amp;view=list">List View</a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=&amp;sort=nameDown&amp;view=table"><b>Name</b></a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=&amp;sort=typeUp&amp;view=table"><b>Data Types</b></a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=&amp;sort=taskUp&amp;view=table"><b>Default Task</b></a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=&amp;sort=attTypeUp&amp;view=table"><b>Attribute Types</b></a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=&amp;sort=instUp&amp;view=table"><b># Instances</b></a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=&amp;sort=attUp&amp;view=table"><b># Attributes</b></a>,
 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=&amp;sort=dateUp&amp;view=table"><b>Year</b></a>,
 <a href="datasets/Abalone"><img border="1" src="assets/MLimages/SmallLarge1.jpg"/></a>,
 <a href="datasets/Abalone">Abalone</a>,
 <a href="datasets/Adult"><img border="1" src="assets/MLimages/SmallLarge2.jpg"/></a>,
 <a href="datasets/Adult">Adult</a>,
 <a href="datasets/Annealing"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/Annealing">Annealing</a>,
 <a href="datasets/Anonymous+Microsoft+Web+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/Anonymous+Microsoft+Web+Data">Anonymous Microsoft Web Data</a>,
 <a href="datasets/Arrhythmia"><img border="1" src="assets/MLimages/SmallLarge5.jpg"/></a>,
 <a href="datasets/Arrhythmia">Arrhythmia</a>,
 <a href="datasets/Artificial+Characters"><img border="1" src="assets/MLimages/SmallLarge6.jpg"/></a>,
 <a href="datasets/Artificial+Characters">Artificial Characters</a>,
 <a href="datasets/Audiology+%28Original%29"><img border="1" src="assets/MLimages/SmallLarge7.jpg"/></a>,
 <a href="datasets/Audiology+%28Original%29">Audiology (Original)</a>,
 <a href="datasets/Audiology+%28Standardized%29"><img border="1" src="assets/MLimages/SmallLarge7.jpg"/></a>,
 <a href="datasets/Audiology+%28Standardized%29">Audiology (Standardized)</a>,
 <a href="datasets/Auto+MPG"><img border="1" src="assets/MLimages/SmallLarge9.jpg"/></a>,
 <a href="datasets/Auto+MPG">Auto MPG</a>,
 <a href="datasets/Automobile"><img border="1" src="assets/MLimages/SmallLarge10.jpg"/></a>,
 <a href="datasets/Automobile">Automobile</a>,
 <a href="datasets/Badges"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/Badges">Badges</a>,
 <a href="datasets/Balance+Scale"><img border="1" src="assets/MLimages/SmallLarge12.jpg"/></a>,
```

<a href="datasets/Balance+Scale"><img border="1" src="assets/MLimages/SmallLarge12.jpg"/></a>,
<a href="datasets/Balance+Scale">Balance Scale</a>,
<a href="datasets/Balloons"><img border="1" src="assets/MLimages/SmallLarge13.jpg"/></a>,
<a href="datasets/Balloons">Balloons</a>,
<a href="datasets/Breast+Cancer"><img border="1" src="assets/MLimages/SmallLarge14.jpg"/></a>,
<a href="datasets/Breast+Cancer">Breast Cancer</a>,
<a href="datasets/Breast+Cancer+Wisconsin+%28Original%29"><img border="1" src="assets/MLimages/SmallLarge14.jpg"/></a>,
<a href="datasets/Breast+Cancer+Wisconsin+%28Original%29">Breast Cancer Wisconsin (Original)</a>,
<a href="datasets/Breast+Cancer+Wisconsin+%28Prognostic%29"><img border="1" src="assets/MLimages/SmallLarge14.jpg"/></a>,
<a href="datasets/Breast+Cancer+Wisconsin+%28Prognostic%29">Breast Cancer Wisconsin (Prognostic)</a>,
<a href="datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29"><img border="1" src="assets/MLimages/SmallLarge14.jpg"/></a>,
<a href="datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29">Breast Cancer Wisconsin (Diagnostic)</a>,
<a href="datasets/Pittsburgh+Bridges"><img border="1" src="assets/MLimages/SmallLarge18.jpg"/></a>,
<a href="datasets/Pittsburgh+Bridges">Pittsburgh Bridges</a>,
<a href="datasets/Car+Evaluation"><img border="1" src="assets/MLimages/SmallLarge19.jpg"/></a>,
<a href="datasets/Car+Evaluation">Car Evaluation</a>,
<a href="datasets/Census+Income"><img border="1" src="assets/MLimages/SmallLarge2.jpg"/></a>,
<a href="datasets/Census+Income">Census Income</a>,
<a href="datasets/Chess+%28King-Rook+vs.+King-Knight%29"><img border="1" src="assets/MLimages/SmallLarge24.jpg"/></a>,
<a href="datasets/Chess+%28King-Rook+vs.+King-Knight%29">Chess (King-Rook vs. King-Knight)</a>,
<a href="datasets/Chess+%28King-Rook+vs.+King-Pawn%29"><img border="1" src="assets/MLimages/SmallLarge24.jpg"/></a>,
<a href="datasets/Chess+%28King-Rook+vs.+King-Pawn%29">Chess (King-Rook vs. King-Pawn)</a>,
<a href="datasets/Chess+%28King-Rook+vs.+King%29"><img border="1" src="assets/MLimages/SmallLarge24.jpg"/></a>,
<a href="datasets/Chess+%28King-Rook+vs.+King%29">Chess (King-Rook vs. King)</a>,
<a href="datasets/Chess+%28Domain+Theories%29"><img border="1" src="assets/MLimages/SmallLarge24.jpg"/></a>,
<a href="datasets/Chess+%28Domain+Theories%29">Chess (Domain Theories)</a>,
<a href="datasets/Bach+Chorales"><img border="1" src="assets/MLimages/SmallLarge25.jpg"/></a>,
<a href="datasets/Bach+Chorales">Bach Chorales</a>,
<a href="datasets/Connect-4"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Connect-4">Connect-4</a>,
<a href="datasets/Credit+Approval"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Credit+Approval">Credit Approval</a>,
<a href="datasets/Japanese+Credit+Screening"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Japanese+Credit+Screening">Japanese Credit Screening</a>,
<a href="datasets/Computer+Hardware"><img border="1" src="assets/MLimages/SmallLarge29.jpg"/></a>,
<a href="datasets/Computer+Hardware">Computer Hardware</a>,
<a href="datasets/Contraceptive+Method+Choice"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Contraceptive+Method+Choice">Contraceptive Method Choice</a>,
<a href="datasets/Covertype"><img border="1" src="assets/MLimages/SmallLarge31.jpg"/></a>,
<a href="datasets/Covertype">Covertype</a>,
<a href="datasets/Cylinder+Bands"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Cylinder+Bands">Cylinder Bands</a>,
<a href="datasets/Dermatology"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Dermatology">Dermatology</a>,
<a href="datasets/Diabetes"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Diabetes">Diabetes</a>,
<a href="datasets/DGP2+-+The+Second+Data+Generation+Program"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/DGP2+-+The+Second+Data+Generation+Program">DGP2 - The Second Data Generation Program</a>,
<a href="datasets/Document+Understanding"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Document+Understanding">Document Understanding</a>,

<a href="datasets/EBL+Domain+Theories"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/EBL+Domain+Theories">EBL Domain Theories</a>,
<a href="datasets/Echocardiogram"><img border="1" src="assets/MLimages/SmallLarge38.jpg"/></a>,
<a href="datasets/Echocardiogram">Echocardiogram</a>,
<a href="datasets/Ecoli"><img border="1" src="assets/MLimages/SmallLarge120.jpg"/></a>,
<a href="datasets/Ecoli">Ecoli</a>,
<a href="datasets/Flags"><img border="1" src="assets/MLimages/SmallLarge40.jpg"/></a>,
<a href="datasets/Flags">Flags</a>,
<a href="datasets/Function+Finding"><img border="1" src="assets/MLimages/SmallLarge41.jpg"/></a>,
<a href="datasets/Function+Finding">Function Finding</a>,
<a href="datasets/Glass+Identification"><img border="1" src="assets/MLimages/SmallLarge42.jpg"/></a>,
<a href="datasets/Glass+Identification">Glass Identification</a>,
<a href="datasets/Haberman%27s+Survival"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Haberman%27s+Survival">Haberman's Survival</a>,
<a href="datasets/Hayes-Roth"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Hayes-Roth">Hayes-Roth</a>,
<a href="datasets/Heart+Disease"><img border="1" src="assets/MLimages/SmallLarge45.jpg"/></a>,
<a href="datasets/Heart+Disease">Heart Disease</a>,
<a href="datasets/Hepatitis"><img border="1" src="assets/MLimages/SmallLarge46.jpg"/></a>,
<a href="datasets/Hepatitis">Hepatitis</a>,
<a href="datasets/Horse+Colic"><img border="1" src="assets/MLimages/SmallLarge47.jpg"/></a>,
<a href="datasets/Horse+Colic">Horse Colic</a>,
<a href="datasets/ICU"><img border="1" src="assets/MLimages/SmallLarge49.jpg"/></a>,
<a href="datasets/ICU">ICU</a>,
<a href="datasets/Image+Segmentation"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Image+Segmentation">Image Segmentation</a>,
<a href="datasets/Internet+Advertisements"><img border="1" src="assets/MLimages/SmallLarge51.jpg"/></a>,
<a href="datasets/Internet+Advertisements">Internet Advertisements</a>,
<a href="datasets/Ionosphere"><img border="1" src="assets/MLimages/SmallLarge52.jpg"/></a>,
<a href="datasets/Ionosphere">Ionosphere</a>,
<a href="datasets/Iris"><img border="1" src="assets/MLimages/SmallLarge53.jpg"/></a>,
<a href="datasets/Iris">Iris</a>,
<a href="datasets/ISOLET"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/ISOLET">ISOLET</a>,
<a href="datasets/Kinship"><img border="1" src="assets/MLimages/SmallLarge55.jpg"/></a>,
<a href="datasets/Kinship">Kinship</a>,
<a href="datasets/Labor+Relations"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Labor+Relations">Labor Relations</a>,
<a href="datasets/LED+Display+Domain"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/LED+Display+Domain">LED Display Domain</a>,
<a href="datasets/Lenses"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Lenses">Lenses</a>,
<a href="datasets/Letter+Recognition"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Letter+Recognition">Letter Recognition</a>,
<a href="datasets/Liver+Disorders"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Liver+Disorders">Liver Disorders</a>,
<a href="datasets/Logic+Theorist"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Logic+Theorist">Logic Theorist</a>,
<a href="datasets/Lung+Cancer"><img border="1" src="assets/MLimages/SmallLarge62.jpg"/></a>,
<a href="datasets/Lung+Cancer">Lung Cancer</a>,
<a href="datasets/Lymphography"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Lymphography">Lymphography</a>,
<a href="datasets/Mechanical+Analysis"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Mechanical+Analysis">Mechanical Analysis</a>,
<a href="datasets/Meta-data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Meta-data">Meta-data</a>,
<a href="datasets/Mobile+Robots"><img border="1" src="assets/MLimages/SmallLarge66.jpg"/></a>,

<a href="datasets/Mobile+Robots">Mobile Robots</a>,
<a href="datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29"><img border="1" src="assets/MLimages/SmallLarge67.jpg"/></a>,
<a href="datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29">Molecular Biology (Promoter Gene Sequences)</a>,
<a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%29"><img border="1" src="assets/MLimages/SmallLarge67.jpg"/></a>,
<a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular Biology (Protein Secondary Structure)</a>,
<a href="datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29"><img border="1" src="assets/MLimages/SmallLarge67.jpg"/></a>,
<a href="datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29">Molecular Biology (Splice-junction Gene Sequences)</a>,
<a href="datasets/MONK%27s+Problems"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/MONK%27s+Problems">MONK's Problems</a>,
<a href="datasets/Moral+Reasoner"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Moral+Reasoner">Moral Reasoner</a>,
<a href="datasets/Multiple+Features"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Multiple+Features">Multiple Features</a>,
<a href="datasets/Mushroom"><img border="1" src="assets/MLimages/SmallLarge73.jpg"/></a>,
<a href="datasets/Mushroom">Mushroom</a>,
<a href="datasets/Musk+%28Version+1%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Musk+%28Version+1%29">Musk (Version 1)</a>,
<a href="datasets/Musk+%28Version+2%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Musk+%28Version+2%29">Musk (Version 2)</a>,
<a href="datasets/Nursery"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Nursery">Nursery</a>,
<a href="datasets/Othello+Domain+Theory"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Othello+Domain+Theory">Othello Domain Theory</a>,
<a href="datasets/Page+Blocks+Classification"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Page+Blocks+Classification">Page Blocks Classification</a>,
<a href="datasets/Optical+Recognition+of+Handwritten+Digits"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Optical+Recognition+of+Handwritten+Digits">Optical Recognition of Handwritten Digits</a>,
<a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits">Pen-Based Recognition of Handwritten Digits</a>,
<a href="datasets/Post-Operative+Patient"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Post-Operative+Patient">Post-Operative Patient</a>,
<a href="datasets/Primary+Tumor"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Primary+Tumor">Primary Tumor</a>,
<a href="datasets/Prodigy"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Prodigy">Prodigy</a>,
<a href="datasets/Qualitative+Structure+Activity+Relationships"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Qualitative+Structure+Activity+Relationships">Qualitative Structure Activity Relationships</a>,
<a href="datasets/Quadruped+Mammals"><img border="1" src="assets/MLimages/SmallLarge86.jpg"/></a>,
<a href="datasets/Quadruped+Mammals">Quadruped Mammals</a>,
<a href="datasets/Servo"><img border="1" src="assets/MLimages/SmallLarge87.jpg"/></a>,
<a href="datasets/Servo">Servo</a>,
<a href="datasets/Shuttle+Landing+Control"><img border="1" src="assets/MLimages/SmallLarge92.jpg"/></a>,
<a href="datasets/Shuttle+Landing+Control">Shuttle Landing Control</a>,
<a href="datasets/Solar+Flare"><img border="1" src="assets/MLimages/SmallLarge89.jpg"/></a>,
<a href="datasets/Solar+Flare">Solar Flare</a>,

<a href="datasets/Soybean+%28Large%29"><img border="1" src="assets/MLimages/SmallLarge90.jpg"/></a>,
<a href="datasets/Soybean+%28Large%29">Soybean (Large)</a>,
<a href="datasets/Soybean+%28Small%29"><img border="1" src="assets/MLimages/SmallLarge90.jpg"/></a>,
<a href="datasets/Soybean+%28Small%29">Soybean (Small)</a>,
<a href="datasets/Challenger+USA+Space+Shuttle+O-Ring"><img border="1" src="assets/MLimages/SmallLarge92.jpg"/></a>,
<a href="datasets/Challenger+USA+Space+Shuttle+O-Ring">Challenger USA Space Shuttle O-Ring</a>,
<a href="datasets/Low+Resolution+Spectrometer"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Low+Resolution+Spectrometer">Low Resolution Spectrometer</a>,
<a href="datasets/Spambase"><img border="1" src="assets/MLimages/SmallLarge94.jpg"/></a>,
<a href="datasets/Spambase">Spambase</a>,
<a href="datasets/SPECT+Heart"><img border="1" src="assets/MLimages/SmallLarge45.jpg"/></a>,
<a href="datasets/SPECT+Heart">SPECT Heart</a>,
<a href="datasets/SPECTF+Heart"><img border="1" src="assets/MLimages/SmallLarge45.jpg"/></a>,
<a href="datasets/SPECTF+Heart">SPECTF Heart</a>,
<a href="datasets/Sponge"><img border="1" src="assets/MLimages/SmallLarge97.jpg"/></a>,
<a href="datasets/Sponge">Sponge</a>,
<a href="datasets/Statlog+Project"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Statlog+Project">Statlog Project</a>,
<a href="datasets/Student+Loan+Relational"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Student+Loan+Relational">Student Loan Relational</a>,
<a href="datasets/Teaching+Assistant+Evaluation"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Teaching+Assistant+Evaluation">Teaching Assistant Evaluation</a>,
<a href="datasets/Tic-Tac-Toe+Endgame"><img border="1" src="assets/MLimages/SmallLarge101.jpg"/></a>,
<a href="datasets/Tic-Tac-Toe+Endgame">Tic-Tac-Toe Endgame</a>,
<a href="datasets/Thyroid+Disease"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Thyroid+Disease">Thyroid Disease</a>,
<a href="datasets/Trains"><img border="1" src="assets/MLimages/SmallLarge103.jpg"/></a>,
<a href="datasets/Trains">Trains</a>,
<a href="datasets/University"><img border="1" src="assets/MLimages/SmallLarge104.jpg"/></a>,
<a href="datasets/University">University</a>,
<a href="datasets/Congressional+Voting+Records"><img border="1" src="assets/MLimages/SmallLarge105.jpg"/></a>,
<a href="datasets/Congressional+Voting+Records">Congressional Voting Records</a>,
<a href="datasets/Water+Treatment+Plant"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Water+Treatment+Plant">Water Treatment Plant</a>,
<a href="datasets/Waveform+Database+Generator+%28Version+1%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Waveform+Database+Generator+%28Version+1%29">Waveform Database Generator (Version 1)</a>,
<a href="datasets/Waveform+Database+Generator+%28Version+2%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Waveform+Database+Generator+%28Version+2%29">Waveform Database Generator (Version 2)</a>,
<a href="datasets/Wine"><img border="1" src="assets/MLimages/SmallLarge109.jpg"/></a>,
<a href="datasets/Wine">Wine</a>,
<a href="datasets/Yeast"><img border="1" src="assets/MLimages/SmallLarge110.jpg"/></a>,
<a href="datasets/Yeast">Yeast</a>,
<a href="datasets/Zoo"><img border="1" src="assets/MLimages/SmallLarge111.jpg"/></a>,
<a href="datasets/Zoo">Zoo</a>,
<a href="datasets/Undocumented"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Undocumented">Undocumented</a>,
<a href="datasets/Twenty+Newsgroups"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Twenty+Newsgroups">Twenty Newsgroups</a>,
<a href="datasets/Australian+Sign+Language+signs"><img border="1" src="assets/MLimages/SmallLarge114.jpg"/></a>,
<a href="datasets/Australian+Sign+Language+signs">Australian Sign Language signs</a>,

<a href="datasets/Australian+Sign+Language+signs+%28High+Quality%29"><img border="1" src="assets/MLi
mages/SmallLarge114.jpg"/></a>,
<a href="datasets/Australian+Sign+Language+signs+%28High+Quality%29">Australian Sign Language signs (
High Quality)</a>,
<a href="datasets/US+Census+Data+%281990%29"><img border="1" src="assets/MLimages/SmallLarge2.jpg"
/></a>,
<a href="datasets/US+Census+Data+%281990%29">US Census Data (1990)</a>,
<a href="datasets/Census-Income+%28KDD%29"><img border="1" src="assets/MLimages/SmallLarge2.jpg"/>
</a>,
<a href="datasets/Census-Income+%28KDD%29">Census-Income (KDD)</a>,
<a href="datasets/Coil+1999+Competition+Data"><img border="1" src="assets/MLimages/SmallLarge118.jpg"/
></a>,
<a href="datasets/Coil+1999+Competition+Data">Coil 1999 Competition Data</a>,
<a href="datasets/Corel+Image+Features"><img border="1" src="assets/MLimages/SmallLarge119.jpg"/></a>,
<a href="datasets/Corel+Image+Features">Corel Image Features</a>,
<a href="datasets/E.+Coli+Genes"><img border="1" src="assets/MLimages/SmallLarge120.jpg"/></a>,
<a href="datasets/E.+Coli+Genes">E. Coli Genes</a>,
<a href="datasets/EEG+Database"><img border="1" src="assets/MLimages/SmallLarge121.jpg"/></a>,
<a href="datasets/EEG+Database">EEG Database</a>,
<a href="datasets/El+Nino"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/El+Nino">El Nino</a>,
<a href="datasets/Entree+Chicago+Recommendation+Data"><img border="1" src="assets/MLimages/SmallLar
ge123.jpg"/></a>,
<a href="datasets/Entree+Chicago+Recommendation+Data">Entree Chicago Recommendation Data</a>,
<a href="datasets/CMU+Face+Images"><img border="1" src="assets/MLimages/SmallLarge124.jpg"/></a>,
<a href="datasets/CMU+Face+Images">CMU Face Images</a>,
<a href="datasets/Insurance+Company+Benchmark+%28COIL+2000%29"><img border="1" src="assets/MLim
ages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Insurance+Company+Benchmark+%28COIL+2000%29">Insurance Company Benchmark (C
OIL 2000)</a>,
<a href="datasets/Internet+Usage+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a
>,
<a href="datasets/Internet+Usage+Data">Internet Usage Data</a>,
<a href="datasets/IPUMS+Census+Database"><img border="1" src="assets/MLimages/SmallLarge2.jpg"/></a>
,
<a href="datasets/IPUMS+Census+Database">IPUMS Census Database</a>,
<a href="datasets/Japanese+Vowels"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Japanese+Vowels">Japanese Vowels</a>,
<a href="datasets/KDD+Cup+1998+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a
>,
<a href="datasets/KDD+Cup+1998+Data">KDD Cup 1998 Data</a>,
<a href="datasets/KDD+Cup+1999+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a
>,
<a href="datasets/KDD+Cup+1999+Data">KDD Cup 1999 Data</a>,
<a href="datasets/M.+Tuberculosis+Genes"><img border="1" src="assets/MLimages/SmallLarge131.jpg"/></a
>,
<a href="datasets/M.+Tuberculosis+Genes">M. Tuberculosis Genes</a>,
<a href="datasets/Movie"><img border="1" src="assets/MLimages/SmallLarge132.jpg"/></a>,
<a href="datasets/Movie">Movie</a>,
<a href="datasets/MSNBC.com+Anonymous+Web+Data"><img border="1" src="assets/MLimages/SmallLarge
default.jpg"/></a>,
<a href="datasets/MSNBC.com+Anonymous+Web+Data">MSNBC.com Anonymous Web Data</a>,
<a href="datasets/NSF+Research+Award+Abstracts+1990-2003"><img border="1" src="assets/MLimages/Sma
llLarge134.jpg"/></a>,
<a href="datasets/NSF+Research+Award+Abstracts+1990-2003">NSF Research Award Abstracts 1990-2003<
/a>,
<a href="datasets/Pioneer-1+Mobile+Robot+Data"><img border="1" src="assets/MLimages/SmallLargedefault.j
pg"/></a>,
<a href="datasets/Pioneer-1+Mobile+Robot+Data">Pioneer-1 Mobile Robot Data</a>,
<a href="datasets/Pseudo+Periodic+Synthetic+Time+Series"><img border="1" src="assets/MLimages/SmallLar
ge136.jpg"/></a>,
<a href="datasets/Pseudo+Periodic+Synthetic+Time+Series">Pseudo Periodic Synthetic Time Series</a>,

<a href="datasets/Reuters-21578+Text+Categorization+Collection"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Reuters-21578+Text+Categorization+Collection">Reuters-21578 Text Categorization Collection</a>,
<a href="datasets/Robot+Execution+Failures"><img border="1" src="assets/MLimages/SmallLarge138.jpg"/></a>,
<a href="datasets/Robot+Execution+Failures">Robot Execution Failures</a>,
<a href="datasets/Synthetic+Control+Chart+Time+Series"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Synthetic+Control+Chart+Time+Series">Synthetic Control Chart Time Series</a>,
<a href="datasets/Syskill+and+Webert+Web+Page+Ratings"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Syskill+and+Webert+Web+Page+Ratings">Syskill and Webert Web Page Ratings</a>,
<a href="datasets/UNIX+User+Data"><img border="1" src="assets/MLimages/SmallLarge141.jpg"/></a>,
<a href="datasets/UNIX+User+Data">UNIX User Data</a>,
<a href="datasets/Volcanoes+on+Venus+-+JARtool+experiment"><img border="1" src="assets/MLimages/SmallLarge142.jpg"/></a>,
<a href="datasets/Volcanoes+on+Venus+-+JARtool+experiment">Volcanoes on Venus - JARtool experiment</a>,
<a href="datasets/Statlog+%28Australian+Credit+Approval%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Statlog+%28Australian+Credit+Approval%29">Statlog (Australian Credit Approval)</a>,
<a href="datasets/Statlog+%28German+Credit+Data%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Statlog+%28German+Credit+Data%29">Statlog (German Credit Data)</a>,
<a href="datasets/Statlog+%28Heart%29"><img border="1" src="assets/MLimages/SmallLarge45.jpg"/></a>,
<a href="datasets/Statlog+%28Heart%29">Statlog (Heart)</a>,
<a href="datasets/Statlog+%28Landsat+Satellite%29"><img border="1" src="assets/MLimages/SmallLarge146.jpg"/></a>,
<a href="datasets/Statlog+%28Landsat+Satellite%29">Statlog (Landsat Satellite)</a>,
<a href="datasets/Statlog+%28Image+Segmentation%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Statlog+%28Image+Segmentation%29">Statlog (Image Segmentation)</a>,
<a href="datasets/Statlog+%28Shuttle%29"><img border="1" src="assets/MLimages/SmallLarge92.jpg"/></a>,
<a href="datasets/Statlog+%28Shuttle%29">Statlog (Shuttle)</a>,
<a href="datasets/Statlog+%28Vehicle+Silhouettes%29"><img border="1" src="assets/MLimages/SmallLarge149.jpg"/></a>,
<a href="datasets/Statlog+%28Vehicle+Silhouettes%29">Statlog (Vehicle Silhouettes)</a>,
<a href="datasets/Connectionist+Bench+%28Nettalk+Corpus%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Connectionist+Bench+%28Nettalk+Corpus%29">Connectionist Bench (Nettalk Corpus)</a>,
<a href="datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29">Connectionist Bench (Sonar, Mines vs. Rocks)</a>,
<a href="datasets/Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29">Connectionist Bench (Vowel Recognition - Deterding Data)</a>,
<a href="datasets/Economic+Sanctions"><img border="1" src="assets/MLimages/SmallLarge153.jpg"/></a>,
<a href="datasets/Economic+Sanctions">Economic Sanctions</a>,
<a href="datasets/Protein+Data"><img border="1" src="assets/MLimages/SmallLarge154.jpg"/></a>,
<a href="datasets/Protein+Data">Protein Data</a>,
<a href="datasets/Cloud"><img border="1" src="assets/MLimages/SmallLarge155.jpg"/></a>,
<a href="datasets/Cloud">Cloud</a>,
<a href="datasets/CalIt2+Building+People+Counts"><img border="1" src="assets/MLimages/SmallLarge156.jpg"/></a>,
<a href="datasets/CalIt2+Building+People+Counts">CalIt2 Building People Counts</a>,
<a href="datasets/Dodgers+Loop+Sensor"><img border="1" src="assets/MLimages/SmallLarge157.jpg"/></a>,
<a href="datasets/Dodgers+Loop+Sensor">Dodgers Loop Sensor</a>,
<a href="datasets/Poker+Hand"><img border="1" src="assets/MLimages/SmallLarge158.jpg"/></a>,
<a href="datasets/Poker+Hand">Poker Hand</a>,

<a href="datasets/MAGIC+Gamma+Telescope"><img border="1" src="assets/MLimages/SmallLarge159.jpg"/></a>,
<a href="datasets/MAGIC+Gamma+Telescope">MAGIC Gamma Telescope</a>,
<a href="datasets/UJI+Pen+Characters"><img border="1" src="assets/MLimages/SmallLarge160.jpg"/></a>,
<a href="datasets/UJI+Pen+Characters">UJI Pen Characters</a>,
<a href="datasets/Mammographic+Mass"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Mammographic+Mass">Mammographic Mass</a>,
<a href="datasets/Forest+Fires"><img border="1" src="assets/MLimages/SmallLarge162.jpg"/></a>,
<a href="datasets/Forest+Fires">Forest Fires</a>,
<a href="datasets/Reuters+Transcribed+Subset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Reuters+Transcribed+Subset">Reuters Transcribed Subset</a>,
<a href="datasets/Bag+of+Words"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Bag+of+Words">Bag of Words</a>,
<a href="datasets/Concrete+Compressive+Strength"><img border="1" src="assets/MLimages/SmallLarge165.jpg"/></a>,
<a href="datasets/Concrete+Compressive+Strength">Concrete Compressive Strength</a>,
<a href="datasets/Hill-Valley"><img border="1" src="assets/MLimages/SmallLarge166.jpg"/></a>,
<a href="datasets/Hill-Valley">Hill-Valley</a>,
<a href="datasets/Arcene"><img border="1" src="assets/MLimages/SmallLarge167.jpg"/></a>,
<a href="datasets/Arcene">Arcene</a>,
<a href="datasets/Dexter"><img border="1" src="assets/MLimages/SmallLarge168.jpg"/></a>,
<a href="datasets/Dexter">Dexter</a>,
<a href="datasets/Dorothea"><img border="1" src="assets/MLimages/SmallLarge169.jpg"/></a>,
<a href="datasets/Dorothea">Dorothea</a>,
<a href="datasets/Gisette"><img border="1" src="assets/MLimages/SmallLarge170.jpg"/></a>,
<a href="datasets/Gisette">Gisette</a>,
<a href="datasets/Madelon"><img border="1" src="assets/MLimages/SmallLarge171.jpg"/></a>,
<a href="datasets/Madelon">Madelon</a>,
<a href="datasets/Ozone+Level+Detection"><img border="1" src="assets/MLimages/SmallLarge172.jpg"/></a>,
<a href="datasets/Ozone+Level+Detection">Ozone Level Detection</a>,
<a href="datasets/Abscisic+Acid+Signaling+Network"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Abscisic+Acid+Signaling+Network">Abscisic Acid Signaling Network</a>,
<a href="datasets/Parkinsons"><img border="1" src="assets/MLimages/SmallLarge174.jpg"/></a>,
<a href="datasets/Parkinsons">Parkinsons</a>,
<a href="datasets/Character+Trajectories"><img border="1" src="assets/MLimages/SmallLarge175.jpg"/></a>,
<a href="datasets/Character+Trajectories">Character Trajectories</a>,
<a href="datasets/Blood+Transfusion+Service+Center"><img border="1" src="assets/MLimages/SmallLarge176.jpg"/></a>,
<a href="datasets/Blood+Transfusion+Service+Center">Blood Transfusion Service Center</a>,
<a href="datasets/UJI+Pen+Characters+%28Version+2%29"><img border="1" src="assets/MLimages/SmallLarge160.jpg"/></a>,
<a href="datasets/UJI+Pen+Characters+%28Version+2%29">UJI Pen Characters (Version 2)</a>,
<a href="datasets/Semeion+Handwritten+Digit"><img border="1" src="assets/MLimages/SmallLarge178.jpg"/></a>,
<a href="datasets/Semeion+Handwritten+Digit">Semeion Handwritten Digit</a>,
<a href="datasets/SECOM"><img border="1" src="assets/MLimages/SmallLarge179.jpg"/></a>,
<a href="datasets/SECOM">SECOM</a>,
<a href="datasets/Plants"><img border="1" src="assets/MLimages/SmallLarge180.jpg"/></a>,
<a href="datasets/Plants">Plants</a>,
<a href="datasets/Libras+Movement"><img border="1" src="assets/MLimages/SmallLarge181.jpg"/></a>,
<a href="datasets/Libras+Movement">Libras Movement</a>,
<a href="datasets/Concrete+Slump+Test"><img border="1" src="assets/MLimages/SmallLarge165.jpg"/></a>,
<a href="datasets/Concrete+Slump+Test">Concrete Slump Test</a>,
<a href="datasets/Communities+and+Crime"><img border="1" src="assets/MLimages/SmallLarge183.jpg"/></a>,
<a href="datasets/Communities+and+Crime">Communities and Crime</a>,
<a href="datasets/Acute+Inflammations"><img border="1" src="assets/MLimages/SmallLarge184.jpg"/></a>,
<a href="datasets/Acute+Inflammations">Acute Inflammations</a>,

<a href="datasets/Acute+Inflammations">Acute Inflammations</a>,
<a href="datasets/Wine+Quality"><img border="1" src="assets/MLimages/SmallLarge186.jpg"/></a>,
<a href="datasets/Wine+Quality">Wine Quality</a>,
<a href="datasets/URL+Reputation"><img border="1" src="assets/MLimages/SmallLarge187.jpg"/></a>,
<a href="datasets/URL+Reputation">URL Reputation</a>,
<a href="datasets/p53+Mutants"><img border="1" src="assets/MLimages/SmallLarge188.jpg"/></a>,
<a href="datasets/p53+Mutants">p53 Mutants</a>,
<a href="datasets/Parkinsons+Telemonitoring"><img border="1" src="assets/MLimages/SmallLarge174.jpg"/></a>,
<a href="datasets/Parkinsons+Telemonitoring">Parkinsons Telemonitoring</a>,
<a href="datasets/Demospongiae"><img border="1" src="assets/MLimages/SmallLarge190.jpg"/></a>,
<a href="datasets/Demospongiae">Demospongiae</a>,
<a href="datasets/Opinosis+Opinion+%26frasl%3B+Review"><img border="1" src="assets/MLimages/SmallLarge191.jpg"/></a>,
<a href="datasets/Opinosis+Opinion+%26frasl%3B+Review">Opinosis Opinion ⁄ Review</a>,
<a href="datasets/Breast+Tissue"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Breast+Tissue">Breast Tissue</a>,
<a href="datasets/Cardiotocography"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Cardiotocography">Cardiotocography</a>,
<a href="datasets/Wall-Following+Robot+Navigation+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Wall-Following+Robot+Navigation+Data">Wall-Following Robot Navigation Data</a>,
<a href="datasets/Spoken+Arabic+Digit"><img border="1" src="assets/MLimages/SmallLarge195.jpg"/></a>,
<a href="datasets/Spoken+Arabic+Digit">Spoken Arabic Digit</a>,
<a href="datasets/Localization+Data+for+Person+Activity"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Localization+Data+for+Person+Activity">Localization Data for Person Activity</a>,
<a href="datasets/AutoUniv"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/AutoUniv">AutoUniv</a>,
<a href="datasets/Steel+Plates+Faults"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Steel+Plates+Faults">Steel Plates Faults</a>,
<a href="datasets/MiniBooNE+particle+identification"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/MiniBooNE+particle+identification">MiniBooNE particle identification</a>,
<a href="datasets/YearPredictionMSD"><img border="1" src="assets/MLimages/SmallLarge203.jpg"/></a>,
<a href="datasets/YearPredictionMSD">YearPredictionMSD</a>,
<a href="datasets/PEMS-SF"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/PEMS-SF">PEMS-SF</a>,
<a href="datasets/OpinRank+Review+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/OpinRank+Review+Dataset">OpinRank Review Dataset</a>,
<a href="datasets/Relative+location+of+CT+slices+on+axial+axis"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Relative+location+of+CT+slices+on+axial+axis">Relative location of CT slices on axial axis</a>,
<a href="datasets/Online+Handwritten+Assamese+Characters+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Online+Handwritten+Assamese+Characters+Dataset">Online Handwritten Assamese Characters Dataset</a>,
<a href="datasets/PubChem+Bioassay+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/PubChem+Bioassay+Data">PubChem Bioassay Data</a>,
<a href="datasets/Record+Linkage+Comparison+Patterns"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Record+Linkage+Comparison+Patterns">Record Linkage Comparison Patterns</a>,
<a href="datasets/Communities+and+Crime+Unnormalized"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Communities+and+Crime+Unnormalized">Communities and Crime Unnormalized</a>,
<a href="datasets/Vertebral+Column"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Vertebral+Column">Vertebral Column</a>,
<a href="datasets/EMG+Physical+Action+Data+Set"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/EMG+Physical+Action+Data+Set">EMG Physical Action Data Set</a>,

<a href="datasets/EMG+Physical+Action+Data+Set">EMG Physical Action Data Set</a>,
<a href="datasets/Vicon+Physical+Action+Data+Set"> </a>,
<a href="datasets/Vicon+Physical+Action+Data+Set">Vicon Physical Action Data Set</a>,
<a href="datasets/Amazon+Commerce+reviews+set"></a>,
<a href="datasets/Amazon+Commerce+reviews+set">Amazon Commerce reviews set</a>,
<a href="datasets/Amazon+Access+Samples"></a>,
<a href="datasets/Amazon+Access+Samples">Amazon Access Samples</a>,
<a href="datasets/Reuter_50_50"></a>,
<a href="datasets/Reuter_50_50">Reuter_50_50</a>,
<a href="datasets/Farm+Ads"></a>,
<a href="datasets/Farm+Ads">Farm Ads</a>,
<a href="datasets/DBWorld+e-mails"></a>,
<a href="datasets/DBWorld+e-mails">DBWorld e-mails</a>,
<a href="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29"></a>,
<a href="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29">KEGG Metabolic Relation Network (Directed)</a>,
<a href="datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29"></a>,
<a href="datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29">KEGG Metabolic Reaction Network (Undirected)</a>,
<a href="datasets/Bank+Marketing"></a>,
<a href="datasets/Bank+Marketing">Bank Marketing</a>,
<a href="datasets/YouTube+Comedy+Slam+Preference+Data"></a>,
<a href="datasets/YouTube+Comedy+Slam+Preference+Data">YouTube Comedy Slam Preference Data</a>,
<a href="datasets/Gas+Sensor+Array+Drift+Dataset"></a>,
<a href="datasets/Gas+Sensor+Array+Drift+Dataset">Gas Sensor Array Drift Dataset</a>,
<a href="datasets/ILPD+%28Indian+Liver+Patient+Dataset%29"></a>,
<a href="datasets/ILPD+%28Indian+Liver+Patient+Dataset%29">ILPD (Indian Liver Patient Dataset)</a>,
<a href="datasets/OPPORTUNITY+Activity+Recognition"></a>,
<a href="datasets/OPPORTUNITY+Activity+Recognition">OPPORTUNITY Activity Recognition</a>,
<a href="datasets/Nomao"></a>,
<a href="datasets/Nomao">Nomao</a>,
<a href="datasets/SMS+Spam+Collection"></a>,
<a href="datasets/SMS+Spam+Collection">SMS Spam Collection</a>,
<a href="datasets/Skin+Segmentation"></a>,
<a href="datasets/Skin+Segmentation">Skin Segmentation</a>,
<a href="datasets/Planning+Relax"></a>,
<a href="datasets/Planning+Relax">Planning Relax</a>,
<a href="datasets/PAMAP2+Physical+Activity+Monitoring"></a>,
<a href="datasets/PAMAP2+Physical+Activity+Monitoring">PAMAP2 Physical Activity Monitoring</a>,
<a href="datasets/Restaurant+%26+consumer+data"></a>,
<a href="datasets/Restaurant+%26+consumer+data">Restaurant & consumer data</a>,
<a href="datasets/CNAE-9"></a>,
<a href="datasets/CNAE-9">CNAE-9</a>,
<a href="datasets/Individual+household+electric+power+consumption"></a>,
<a href="datasets/Individual+household+electric+power+consumption">Individual household electric power consumption</a>,
<a href="datasets/seeds"></a>,
<a href="datasets/seeds">seeds</a>,
<a href="datasets/Northix"></a>,

<a href="datasets/Northrix">Northrix</a>,
<a href="datasets/QtyT40I10D100K"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/QtyT40I10D100K">QtyT40I10D100K</a>,
<a href="datasets/Legal+Case+Reports"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>
,
<a href="datasets/Legal+Case+Reports">Legal Case Reports</a>,
<a href="datasets/Human+Activity+Recognition+Using+Smartphones"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Human+Activity+Recognition+Using+Smartphones">Human Activity Recognition Using Smartphones</a>,
<a href="datasets/One-hundred+plant+species+leaves+data+set"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/One-hundred+plant+species+leaves+data+set">One-hundred plant species leaves data set</a>,
<a href="datasets/Energy+efficiency"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Energy+efficiency">Energy efficiency</a>,
<a href="datasets/Yacht+Hydrodynamics"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Yacht+Hydrodynamics">Yacht Hydrodynamics</a>,
<a href="datasets/Fertility"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Fertility">Fertility</a>,
<a href="datasets/Daphnet+Freezing+of+Gait"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Daphnet+Freezing+of+Gait">Daphnet Freezing of Gait</a>,
<a href="datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29">3D Road Network (North Jutland, Denmark)</a>,
<a href="datasets/ISTANBUL+STOCK+EXCHANGE"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/ISTANBUL+STOCK+EXCHANGE">ISTANBUL STOCK EXCHANGE</a>,
<a href="datasets/Buzz+in+social+media+"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Buzz+in+social+media+">Buzz in social media </a>,
<a href="datasets/First-order+theorem+proving"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/First-order+theorem+proving">First-order theorem proving</a>,
<a href="datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29">Wearable Computing: Classification of Body Postures and Movements (PUC-Rio)</a>,
<a href="datasets/Gas+sensor+arrays+in+open+sampling+settings"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Gas+sensor+arrays+in+open+sampling+settings">Gas sensor arrays in open sampling settings</a>,
<a href="datasets/Climate+Model+Simulation+Crashes"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Climate+Model+Simulation+Crashes">Climate Model Simulation Crashes</a>,
<a href="datasets/MicroMass"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/MicroMass">MicroMass</a>,
<a href="datasets/QSAR+biodegradation"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/QSAR+biodegradation">QSAR biodegradation</a>,
<a href="datasets/BLOGGER"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/BLOGGER">BLOGGER</a>,
<a href="datasets/Daily+and+Sports+Activities"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Daily+and+Sports+Activities">Daily and Sports Activities</a>,
<a href="datasets/User+Knowledge+Modeling"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/User+Knowledge+Modeling">User Knowledge Modeling</a>,
<a href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection"><i

mg border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection">Reuters RCV1 RCV2 Multilingual, Multiview Text Categorization Test collection</a>,
<a href="datasets/NYSK"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/NYSK">NYSK</a>,
<a href="datasets/Turkiye+Student+Evaluation"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Turkiye+Student+Evaluation">Turkiye Student Evaluation</a>,
<a href="datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machines%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machines%29">ser Knowledge Modeling Data (Students' Knowledge Levels on DC Electrical Machines)</a>,
<a href="datasets/EEG+Eye+State"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/EEG+Eye+State">EEG Eye State</a>,
<a href="datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure">Physicochemical Properties of Protein Tertiary Structure</a>,
<a href="datasets/seismic-bumps"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/seismic-bumps">seismic-bumps</a>,
<a href="datasets/banknote+authentication"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/banknote+authentication">banknote authentication</a>,
<a href="datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Problem%3A+Pat"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Problem%3A+Pat">USPTO Algorithm Challenge, run by NASA-Harvard Tournament Lab and TopCoder Problem: Pat</a>,
<a href="datasets/YouTube+Multiview+Video+Games+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/YouTube+Multiview+Video+Games+Dataset">YouTube Multiview Video Games Dataset</a>,
<a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations">Gas Sensor Array Drift Dataset at Different Concentrations</a>,
<a href="datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors">Activities of Daily Living (ADLs) Recognition Using Binary Sensors</a>,
<a href="datasets/SkillCraft1+Master+Table+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/SkillCraft1+Master+Table+Dataset">SkillCraft1 Master Table Dataset</a>,
<a href="datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units">Weight Lifting Exercises monitored with Inertial Measurement Units</a>,
<a href="datasets/SML2010"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/SML2010">SML2010</a>,
<a href="datasets/Bike+Sharing+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Bike+Sharing+Dataset">Bike Sharing Dataset</a>,
<a href="datasets/Predict+keywords+activities+in+a+online+social+media"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Predict+keywords+activities+in+a+online+social+media">Predict keywords activities in a online social media</a>,
<a href="datasets/Thoracic+Surgery+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Thoracic+Surgery+Data">Thoracic Surgery Data</a>,
<a href="datasets/EMG+dataset+in+Lower+Limb"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/EMG+dataset+in+Lower+Limb">EMG dataset in Lower Limb</a>,

<a href="datasets/SUSY"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/SUSY">SUSY</a>,
<a href="datasets/HIGGS"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/HIGGS">HIGGS</a>,
<a href="datasets/Qualitative_Bankruptcy"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Qualitative_Bankruptcy">Qualitative_Bankruptcy</a>,
<a href="datasets/LSVT+Voice+Rehabilitation"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/LSVT+Voice+Rehabilitation">LSVT Voice Rehabilitation</a>,
<a href="datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer">Dataset for ADL Recognition with Wrist-worn Accelerometer</a>,
<a href="datasets/Wilt"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Wilt">Wilt</a>,
<a href="datasets/User+Identification+From+Walking+Activity"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/User+Identification+From+Walking+Activity">User Identification From Walking Activity</a>,
<a href="datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer">Activity Recognition from Single Chest-Mounted Accelerometer</a>,
<a href="datasets/Leaf"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Leaf">Leaf</a>,
<a href="datasets/Dresses_Attribute_Sales"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Dresses_Attribute_Sales">Dresses_Attribute_Sales</a>,
<a href="datasets/Tamilnadu+Electricity+Board+Hourly+Readings"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Tamilnadu+Electricity+Board+Hourly+Readings">Tamilnadu Electricity Board Hourly Readings</a>,
<a href="datasets/Airfoil+Self-Noise"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Airfoil+Self-Noise">Airfoil Self-Noise</a>,
<a href="datasets/Wholesale+customers"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Wholesale+customers">Wholesale customers</a>,
<a href="datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis">Twitter Data set for Arabic Sentiment Analysis</a>,
<a href="datasets/Combined+Cycle+Power+Plant"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Combined+Cycle+Power+Plant">Combined Cycle Power Plant</a>,
<a href="datasets/Urban+Land+Cover"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Urban+Land+Cover">Urban Land Cover</a>,
<a href="datasets/Diabetes+130-US+hospitals+for+years+1999-2008"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Diabetes+130-US+hospitals+for+years+1999-2008">Diabetes 130-US hospitals for years 1999-2008</a>,
<a href="datasets/Bach+Choral+Harmony"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Bach+Choral+Harmony">Bach Choral Harmony</a>,
<a href="datasets/StoneFlakes"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/StoneFlakes">StoneFlakes</a>,
<a href="datasets/Tennis+Major+Tournament+Match+Statistics"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Tennis+Major+Tournament+Match+Statistics">Tennis Major Tournament Match Statistics</a>,
<a href="datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings">Parkinson Spee

ch Dataset with Multiple Types of Sound Recordings</a>,
 <a href="datasets/Gesture+Phase+Segmentation"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/Gesture+Phase+Segmentation">Gesture Phase Segmentation</a>,
 <a href="datasets/Perfume+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/Perfume+Data">Perfume Data</a>,
 <a href="datasets/BlogFeedback"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/BlogFeedback">BlogFeedback</a>,
 <a href="datasets/REALDISP+Activity+Recognition+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/REALDISP+Activity+Recognition+Dataset">REALDISP Activity Recognition Dataset</a>,
 <a href="datasets/Newspaper+and+magazine+images+segmentation+dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/Newspaper+and+magazine+images+segmentation+dataset">Newspaper and magazine images segmentation dataset</a>,
 <a href="datasets/AAAI+2014+Accepted+Papers"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/AAAI+2014+Accepted+Papers">AAAI 2014 Accepted Papers</a>,
 <a href="datasets/Gas+sensor+array+under+flow+modulation"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/Gas+sensor+array+under+flow+modulation">Gas sensor array under flow modulation</a>,
 <a href="datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures">Gas sensor array exposed to turbulent gas mixtures</a>,
 <a href="datasets/UJIIndoorLoc"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/UJIIndoorLoc">UJIIndoorLoc</a>,
 <a href="datasets/Sentence+Classification"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/Sentence+Classification">Sentence Classification</a>,
 <a href="datasets/Dow+Jones+Index"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/Dow+Jones+Index">Dow Jones Index</a>,
 <a href="datasets/sEMG+for+Basic+Hand+movements"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/sEMG+for+Basic+Hand+movements">sEMG for Basic Hand movements</a>,
 <a href="datasets/AAAI+2013+Accepted+Papers"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/AAAI+2013+Accepted+Papers">AAAI 2013 Accepted Papers</a>,
 <a href="datasets/Geographical+Original+of+Music"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/Geographical+Original+of+Music">Geographical Original of Music</a>,
 <a href="datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants">Condition Based Maintenance of Naval Propulsion Plants</a>,
 <a href="datasets/Grammatical+Facial+Expressions"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/Grammatical+Facial+Expressions">Grammatical Facial Expressions</a>,
 <a href="datasets/NoisyOffice"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/NoisyOffice">NoisyOffice</a>,
 <a href="datasets/MHEALTH+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/MHEALTH+Dataset">MHEALTH Dataset</a>,
 <a href="datasets/Student+Performance"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/Student+Performance">Student Performance</a>,
 <a href="datasets/ElectricityLoadDiagrams20112014"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/ElectricityLoadDiagrams20112014">ElectricityLoadDiagrams20112014</a>,
 <a href="datasets/Gas+sensor+array+under+dynamic+gas+mixtures"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
 <a href="datasets/Gas+sensor+array+under+dynamic+gas+mixtures">Gas sensor array under dynamic gas mixtures</a>,

<a href="datasets/microblogPCU"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/microblogPCU">microblogPCU</a>,
<a href="datasets/Firm-Teacher_Clave-Direction_Classification"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Firm-Teacher_Clave-Direction_Classification">Firm-Teacher_Clave-Direction_Classification</a>,
<a href="datasets/Dataset+for+Sensorless+Drive+Diagnosis"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Dataset+for+Sensorless+Drive+Diagnosis">Dataset for Sensorless Drive Diagnosis</a>,
<a href="datasets/TV+News+Channel+Commercial+Detection+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/TV+News+Channel+Commercial+Detection+Dataset">TV News Channel Commercial Detection Dataset</a>,
<a href="datasets/Phishing+Websites"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Phishing+Websites">Phishing Websites</a>,
<a href="datasets/Greenhouse+Gas+Observing+Network"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Greenhouse+Gas+Observing+Network">Greenhouse Gas Observing Network</a>,
<a href="datasets/Diabetic+Retinopathy+Debrecen+Data+Set"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Diabetic+Retinopathy+Debrecen+Data+Set">Diabetic Retinopathy Debrecen Data Set</a>,
<a href="datasets/HIV-1+protease+cleavage"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/HIV-1+protease+cleavage">HIV-1 protease cleavage</a>,
<a href="datasets/Sentiment+Labelled+Sentences"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Sentiment+Labelled+Sentences">Sentiment Labelled Sentences</a>,
<a href="datasets/Online+News+Popularity"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Online+News+Popularity">Online News Popularity</a>,
<a href="datasets/Forest+type+mapping"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Forest+type+mapping">Forest type mapping</a>,
<a href="datasets/wiki4HE"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/wiki4HE">wiki4HE</a>,
<a href="datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset">Online Video Characteristics and Transcoding Time Dataset</a>,
<a href="datasets/Chronic_Kidney_Disease"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Chronic_Kidney_Disease">Chronic_Kidney_Disease</a>,
<a href="datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014">Machine Learning based ZZAlpha Ltd. Stock Recommendations 2012-2014</a>,
<a href="datasets/Folio"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Folio">Folio</a>,
<a href="datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015">Taxi Service Trajectory - Prediction Challenge, ECML PKDD 2015</a>,
<a href="datasets/Cuff-Less+Blood+Pressure+Estimation"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Cuff-Less+Blood+Pressure+Estimation">Cuff-Less Blood Pressure Estimation</a>,
<a href="datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions">Smartphone-Based Recognition of Human Activities and Postural Transitions</a>,
<a href="datasets/Mice+Protein+Expression"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Mice+Protein+Expression">Mice Protein Expression</a>,

<a href="datasets/UJIIndoorLoc-Mag"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/UJIIndoorLoc-Mag">UJIIndoorLoc-Mag</a>,
<a href="datasets/Heterogeneity+Activity+Recognition"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Heterogeneity+Activity+Recognition">Heterogeneity Activity Recognition</a>,
<a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set">Educational Process Mining (EPM): A Learning Analytics Data Set</a>,
<a href="datasets/HEPMASS"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/HEPMASS">HEPMASS</a>,
<a href="datasets/Indoor+User+Movement+Prediction+from+RSS+data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Indoor+User+Movement+Prediction+from+RSS+data">Indoor User Movement Prediction from RSS data</a>,
<a href="datasets/Open+University+Learning+Analytics+dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Open+University+Learning+Analytics+dataset">Open University Learning Analytics dataset</a>,
<a href="datasets/default+of+credit+card+clients"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/default+of+credit+card+clients">default of credit card clients</a>,
<a href="datasets/Mesothelioma%E2%80%99s+disease+data+set+"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Mesothelioma%E2%80%99s+disease+data+set+">Mesothelioma’s disease data set </a>,
<a href="datasets/Online+Retail"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Online+Retail">Online Retail</a>,
<a href="datasets/SIFT10M"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/SIFT10M">SIFT10M</a>,
<a href="datasets/GPS+Trajectories"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/GPS+Trajectories">GPS Trajectories</a>,
<a href="datasets/Detect+Malacious+Executable%28AntiVirus%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Detect+Malacious+Executable%28AntiVirus%29">Detect Malacious Executable(AntiVirus)</a>,
<a href="datasets/Occupancy+Detection+"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Occupancy+Detection+">Occupancy Detection </a>,
<a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Parkinson%E2%80%99s+Disease"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Parkinson%E2%80%99s+Disease">Improved Spiral Test Using Digitized Graphics Tablet for Monitoring Parkinson’s Disease</a>,
<a href="datasets/News+Aggregator"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/News+Aggregator">News Aggregator</a>,
<a href="datasets/Air+Quality"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Air+Quality">Air Quality</a>,
<a href="datasets/Twin+gas+sensor+arrays"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Twin+gas+sensor+arrays">Twin gas sensor arrays</a>,
<a href="datasets/Gas+sensors+for+home+activity+monitoring"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Gas+sensors+for+home+activity+monitoring">Gas sensors for home activity monitoring</a>,
<a href="datasets/Facebook+Comment+Volume+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Facebook+Comment+Volume+Dataset">Facebook Comment Volume Dataset</a>,
<a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Living+%28AAL%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Living+%28AAL%29">Smartphone Dataset for Human Activity Recognition (HAR) in Ambient Assisted Living (AAL)</a>,
<a href="datasets/Polish+companies+bankruptcy+data"><img border="1" src="assets/MLimages/SmallLargede

fault.jpg"/></a>,
<a href="datasets/Polish+companies+bankruptcy+data">Polish companies bankruptcy data</a>,
<a href="datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29"><img bo
rder="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29">Activity
Recognition system based on Multisensor data fusion (AReM)</a>,
<a href="datasets/Dota2+Games+Results"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></
a>,
<a href="datasets/Dota2+Games+Results">Dota2 Games Results</a>,
<a href="datasets/Facebook+metrics"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Facebook+metrics">Facebook metrics</a>,
<a href="datasets/UbiqLog+%28smartphone+lifelogging%29"><img border="1" src="assets/MLimages/SmallLa
rgedefault.jpg"/></a>,
<a href="datasets/UbiqLog+%28smartphone+lifelogging%29">UbiqLog (smartphone lifelogging)</a>,
<a href="datasets/NIPS+Conference+Papers+1987-2015"><img border="1" src="assets/MLimages/SmallLarge
default.jpg"/></a>,
<a href="datasets/NIPS+Conference+Papers+1987-2015">NIPS Conference Papers 1987-2015</a>,
<a href="datasets/HTRU2"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/HTRU2">HTRU2</a>,
<a href="datasets/Drug+consumption+%28quantified%29"><img border="1" src="assets/MLimages/SmallLarge
default.jpg"/></a>,
<a href="datasets/Drug+consumption+%28quantified%29">Drug consumption (quantified)</a>,
<a href="datasets/Appliances+energy+prediction"><img border="1" src="assets/MLimages/SmallLargedefault.j
pg"/></a>,
<a href="datasets/Appliances+energy+prediction">Appliances energy prediction</a>,
<a href="datasets/Miskolc+IIS+Hybrid+IPS"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></
a>,
<a href="datasets/Miskolc+IIS+Hybrid+IPS">Miskolc IIS Hybrid IPS</a>,
<a href="datasets/KDC-4007+dataset+Collection"><img border="1" src="assets/MLimages/SmallLargedefault.j
pg"/></a>,
<a href="datasets/KDC-4007+dataset+Collection">KDC-4007 dataset Collection</a>,
<a href="datasets/Geo-Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+sma
rtphone"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Geo-Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+sma
rtphone">Geo-Magnetic field and WLAN dataset for indoor localisation from wristband and smartphone</a>,
<a href="datasets/DrivFace"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/DrivFace">DrivFace</a>,
<a href="datasets/Website+Phishing"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Website+Phishing">Website Phishing</a>,
<a href="datasets/YouTube+Spam+Collection"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"
/></a>,
<a href="datasets/YouTube+Spam+Collection">YouTube Spam Collection</a>,
<a href="datasets/Beijing+PM2.5+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Beijing+PM2.5+Data">Beijing PM2.5 Data</a>,
<a href="datasets/Cargo+2000+Freight+Tracking+and+Tracing"><img border="1" src="assets/MLimages/Small
Largedefault.jpg"/></a>,
<a href="datasets/Cargo+2000+Freight+Tracking+and+Tracing">Cargo 2000 Freight Tracking and Tracing</a
>,
<a href="datasets/Cervical+cancer+%28Risk+Factors%29"><img border="1" src="assets/MLimages/SmallLarg
edefault.jpg"/></a>,
<a href="datasets/Cervical+cancer+%28Risk+Factors%29">Cervical cancer (Risk Factors)</a>,
<a href="datasets/Quality+Assessment+of+Digital+Colposcopies"><img border="1" src="assets/MLimages/Sma
llLargedefault.jpg"/></a>,
<a href="datasets/Quality+Assessment+of+Digital+Colposcopies">Quality Assessment of Digital Colposcopies<
/a>,
<a href="datasets/KASANDR"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/KASANDR">KASANDR</a>,
<a href="datasets/FMA%3A+A+Dataset+For+Music+Analysis"><img border="1" src="assets/MLimages/SmallL
argedefault.jpg"/></a>,
<a href="datasets/FMA%3A+A+Dataset+For+Music+Analysis">FMA: A Dataset For Music Analysis</a>,
<a href="datasets/Air+quality"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Air+quality">Air quality</a>,

<a href="datasets/All+quality">All quality</a>,
<a href="datasets/Epileptic+Seizure+Recognition"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Epileptic+Seizure+Recognition">Epileptic Seizure Recognition</a>,
<a href="datasets/Devanagari+Handwritten+Character+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Devanagari+Handwritten+Character+Dataset">Devanagari Handwritten Character Dataset</a>,
<a href="datasets/Stock+portfolio+performance"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Stock+portfolio+performance">Stock portfolio performance</a>,
<a href="datasets/MoCap+Hand+Postures"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/MoCap+Hand+Postures">MoCap Hand Postures</a>,
<a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech">Early biomarkers of Parkinson's disease based on natural connected speech</a>,
<a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting">Data for Software Engineering Teamwork Assessment in Education Setting</a>,
<a href="datasets/PM2.5+Data+of+Five+Chinese+Cities"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/PM2.5+Data+of+Five+Chinese+Cities">PM2.5 Data of Five Chinese Cities</a>,
<a href="datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet">Parkinson Disease Spiral Drawings Using Digitized Graphics Tablet</a>,
<a href="datasets/Sales_Transactions_Dataset_Weekly"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly</a>,
<a href="datasets/Las+Vegas+Strip"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Las+Vegas+Strip">Las Vegas Strip</a>,
<a href="datasets/Eco-hotel"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Eco-hotel">Eco-hotel</a>,
<a href="datasets/MEU-Mobile+KSD"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/MEU-Mobile+KSD">MEU-Mobile KSD</a>,
<a href="datasets/Crowdsourced+Mapping"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Crowdsourced+Mapping">Crowdsourced Mapping</a>,
<a href="datasets/gene+expression+cancer+RNA-Seq"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/gene+expression+cancer+RNA-Seq">gene expression cancer RNA-Seq</a>,
<a href="datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer">Hybrid Indoor Positioning Dataset from WiFi RSSI, Bluetooth and magnetometer</a>,
<a href="datasets/chestnut+%E2%80%93+LARVIC"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/chestnut+%E2%80%93+LARVIC">chestnut – LARVIC</a>,
<a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%28OBS%29+Network"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%28OBS%29+Network">Burst Header Packet (BHP) flooding attack on Optical Burst Switching (OBS) Network</a>,
<a href="datasets/Motion+Capture+Hand+Postures"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Motion+Capture+Hand+Postures">Motion Capture Hand Postures</a>,
<a href="datasets/Anuran+Calls+%28MFCCs%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Anuran+Calls+%28MFCCs%29">Anuran Calls (MFCCs)</a>,
<a href="datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,

<a href="datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization">TTC-3600: Benchmark dataset for Turkish text categorization</a>,
<a href="datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy"> </a>,
<a href="datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy">Gastrointestinal Lesions in Regular Colonoscopy</a>,
<a href="datasets/Daily+Demand+Forecasting+Orders"></a>,
<a href="datasets/Daily+Demand+Forecasting+Orders">Daily Demand Forecasting Orders</a>,
<a href="datasets/Paper+Reviews"></a>,
<a href="datasets/Paper+Reviews">Paper Reviews</a>,
<a href="datasets/extention+of+Z-Alizadeh+sani+dataset"></a>,
<a href="datasets/extention+of+Z-Alizadeh+sani+dataset">extention of Z-Alizadeh sani dataset</a>,
<a href="datasets/Z-Alizadeh+Sani"></a>,
<a href="datasets/Z-Alizadeh+Sani">Z-Alizadeh Sani</a>,
<a href="datasets/Dynamic+Features+of+VirusShare+Executables"></a>,
<a href="datasets/Dynamic+Features+of+VirusShare+Executables">Dynamic Features of VirusShare Executables</a>,
<a href="datasets/IDA2016Challenge"></a>,
<a href="datasets/IDA2016Challenge">IDA2016Challenge</a>,
<a href="datasets/DSRC+Vehicle+Communications"></a>,
<a href="datasets/DSRC+Vehicle+Communications">DSRC Vehicle Communications</a>,
<a href="datasets/Mturk+User-Perceived+Clusters+over+Images"></a>,
<a href="datasets/Mturk+User-Perceived+Clusters+over+Images">Mturk User-Perceived Clusters over Images</a>,
<a href="datasets/Character+Font+Images"></a>,
<a href="datasets/Character+Font+Images">Character Font Images</a>,
<a href="datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Instance+Labels"></a>,
<a href="datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Instance+Labels">DeliciousMIL: A Data Set for Multi-Label Multi-Instance Learning with Instance Labels</a>,
<a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++"></a>,
<a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++">Autistic Spectrum Disorder Screening Data for Children </a>,
<a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++"></a>,
<a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++">Autistic Spectrum Disorder Screening Data for Adolescent </a>,
<a href="datasets/APS+Failure+at+Scania+Trucks"></a>,
<a href="datasets/APS+Failure+at+Scania+Trucks">APS Failure at Scania Trucks</a>,
<a href="datasets/Wireless+Indoor+Localization"></a>,
<a href="datasets/Wireless+Indoor+Localization">Wireless Indoor Localization</a>,
<a href="datasets/HCC+Survival"></a>,
<a href="datasets/HCC+Survival">HCC Survival</a>,
<a href="datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015"></a>,
<a href="datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015">CSM (Conventional and Social Media Movies) Dataset 2014 and 2015</a>,
<a href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29"></a>,
<a href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29">University of Tehran Question Dataset 2016 (UTQD.2016)</a>,
<a href="datasets/Autism+Screening+Adult"></a>,

/a>,
<a href="datasets/Autism+Screening+Adult">Autism Screening Adult</a>,
<a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor"><i
mg border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor">Act
ivity recognition with healthy older people using a batteryless wearable sensor</a>,
<a href="datasets/Immunotherapy+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a>,
<a href="datasets/Immunotherapy+Dataset">Immunotherapy Dataset</a>,
<a href="datasets/Cryotherapy+Dataset+"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a
>,
<a href="datasets/Cryotherapy+Dataset+">Cryotherapy Dataset </a>,
<a href="datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes"><img border="1" src="a
ssets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes">OCT data & Color
Fundus Images of Left & Right Eyes</a>,
<a href="datasets/Discrete+Tone+Image+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.
jpg"/></a>,
<a href="datasets/Discrete+Tone+Image+Dataset">Discrete Tone Image Dataset</a>,
<a href="datasets/News+Popularity+in+Multiple+Social+Media+Platforms"><img border="1" src="assets/MLima
ges/SmallLargedefault.jpg"/></a>,
<a href="datasets/News+Popularity+in+Multiple+Social+Media+Platforms">News Popularity in Multiple Social M
edia Platforms</a>,
<a href="datasets/Ultrasonic+flowmeter+diagnostics"><img border="1" src="assets/MLimages/SmallLargedefa
ult.jpg"/></a>,
<a href="datasets/Ultrasonic+flowmeter+diagnostics">Ultrasonic flowmeter diagnostics</a>,
<a href="datasets/ICMLA+2014+Accepted+Papers+Data+Set"><img border="1" src="assets/MLimages/SmallL
argedefault.jpg"/></a>,
<a href="datasets/ICMLA+2014+Accepted+Papers+Data+Set">ICMLA 2014 Accepted Papers Data Set</a>,
<a href="datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation"><img border="1" src="assets/
MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation">BLE RSSI Dataset for Indoor l
ocalization and Navigation</a>,
<a href="datasets/Container+Crane+Controller+Data+Set"><img border="1" src="assets/MLimages/SmallLarge
default.jpg"/></a>,
<a href="datasets/Container+Crane+Controller+Data+Set">Container Crane Controller Data Set</a>,
<a href="datasets/Residential+Building+Data+Set"><img border="1" src="assets/MLimages/SmallLargedefault.j
pg"/></a>,
<a href="datasets/Residential+Building+Data+Set">Residential Building Data Set</a>,
<a href="datasets/Health+News+in+Twitter"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/><
/a>,
<a href="datasets/Health+News+in+Twitter">Health News in Twitter</a>,
<a href="datasets/chipseq"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/chipseq">chipseq</a>,
<a href="datasets/SGEMM+GPU+kernel+performance"><img border="1" src="assets/MLimages/SmallLargede
fault.jpg"/></a>,
<a href="datasets/SGEMM+GPU+kernel+performance">SGEMM GPU kernel performance</a>,
<a href="datasets/Repeat+Consumption+Matrices"><img border="1" src="assets/MLimages/SmallLargedefault
.jpg"/></a>,
<a href="datasets/Repeat+Consumption+Matrices">Repeat Consumption Matrices</a>,
<a href="datasets/detection_of_IoT_botnet_attacks_N_BaIoT"><img border="1" src="assets/MLimages/SmallL
argedefault.jpg"/></a>,
<a href="datasets/detection_of_IoT_botnet_attacks_N_BaIoT">detection_of_IoT_botnet_attacks_N_BaIoT</a>
,
<a href="datasets/Absenteeism+at+work"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a
>,
<a href="datasets/Absenteeism+at+work">Absenteeism at work</a>,
<a href="datasets/SCADI"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/SCADI">SCADI</a>,
<a href="datasets/Condition+monitoring+of+hydraulic+systems"><img border="1" src="assets/MLimages/Small
Largedefault.jpg"/></a>,
<a href="datasets/Condition+monitoring+of+hydraulic+systems">Condition monitoring of hydraulic systems</a>

,
<a href="datasets/Carbon+Nanotubes"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Carbon+Nanotubes">Carbon Nanotubes</a>,
<a href="datasets/Optical+Interconnection+Network+"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Optical+Interconnection+Network+">Optical Interconnection Network </a>,
<a href="datasets/Sports+articles+for+objectivity+analysis"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Sports+articles+for+objectivity+analysis">Sports articles for objectivity analysis</a>,
<a href="datasets/Breast+Cancer+Coimbra"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Breast+Cancer+Coimbra">Breast Cancer Coimbra</a>,
<a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data">GNFUV Unmanned Surface Vehicles Sensor Data</a>,
<a href="datasets/Dishonest+Internet+users+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Dishonest+Internet+users+Dataset">Dishonest Internet users Dataset</a>,
<a href="datasets/Victorian+Era+Authorship+Attribution"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Victorian+Era+Authorship+Attribution">Victorian Era Authorship Attribution</a>,
<a href="datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set">Simulated Falls and Daily Living Activities Data Set</a>,
<a href="datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing">Multimodal Damage Identification for Humanitarian Computing</a>,
<a href="datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals">EEG Steady-State Visual Evoked Potential Signals</a>,
<a href="datasets/Roman+Urdu+Data+Set"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Roman+Urdu+Data+Set">Roman Urdu Data Set</a>,
<a href="datasets/Avila"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Avila">Avila</a>,
<a href="datasets/PANDOR"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/PANDOR">PANDOR</a>,
<a href="datasets/Drug+Review+Dataset+%28Druglib.com%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Drug+Review+Dataset+%28Druglib.com%29">Drug Review Dataset (Druglib.com)</a>,
<a href="datasets/Drug+Review+Dataset+%28Drugs.com%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Drug+Review+Dataset+%28Drugs.com%29">Drug Review Dataset (Drugs.com)</a>,
<a href="datasets/Physical+Unclonable+Functions"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Physical+Unclonable+Functions">Physical Unclonable Functions</a>,
<a href="datasets/Superconductivty+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Superconductivty+Data">Superconductivty Data</a>,
<a href="datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29">WESAD (Wearable Stress and Affect Detection)</a>,
<a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2">GNFUV Unmanned Surface Vehicles Sensor Data Set 2</a>,
<a href="datasets/Student+Academics+Performance"><img border="1" src="assets/MLimages/SmallLargedefa

ult.jpg"/></a>,
<a href="datasets/Student+Academics+Performance">Student Academics Performance</a>,
<a href="datasets/Online+Shoppers+Purchasing+Intention+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Online+Shoppers+Purchasing+Intention+Dataset">Online Shoppers Purchasing Intention Dataset</a>,
<a href="datasets/PMU-UD"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/PMU-UD">PMU-UD</a>,
<a href="datasets/Parkinson%27s+Disease+Classification"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Parkinson%27s+Disease+Classification">Parkinson's Disease Classification</a>,
<a href="datasets/Electrical+Grid+Stability+Simulated+Data+"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Electrical+Grid+Stability+Simulated+Data+">Electrical Grid Stability Simulated Data </a>,
<a href="datasets/Caesarian+Section+Classification+Dataset"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Caesarian+Section+Classification+Dataset">Caesarian Section Classification Dataset</a>,
<a href="datasets/BAUM-1"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/BAUM-1">BAUM-1</a>,
<a href="datasets/BAUM-2"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/BAUM-2">BAUM-2</a>,
<a href="datasets/Audit+Data"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Audit+Data">Audit Data</a>,
<a href="datasets/BuddyMove+Data+Set"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/BuddyMove+Data+Set">BuddyMove Data Set</a>,
<a href="datasets/Real+estate+valuation+data+set"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Real+estate+valuation+data+set">Real estate valuation data set</a>,
<a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connected+speech+Data+Set+"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connected+speech+Data+Set+">Early biomarkers of Parkinson’s disease based on natural connected speech Data Set </a>,
<a href="datasets/Somerville+Happiness+Survey"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Somerville+Happiness+Survey">Somerville Happiness Survey</a>,
<a href="datasets/2.4+GHZ+Indoor+Channel+Measurements"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/2.4+GHZ+Indoor+Channel+Measurements">2.4 GHZ Indoor Channel Measurements</a>,
<a href="datasets/EMG+data+for+gestures"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/EMG+data+for+gestures">EMG data for gestures</a>,
<a href="datasets/Parking+Birmingham"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Parking+Birmingham">Parking Birmingham</a>,
<a href="datasets/Behavior+of+the+urban+traffic+of+the+city+of+Sao+Paulo+in+Brazil"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Behavior+of+the+urban+traffic+of+the+city+of+Sao+Paulo+in+Brazil">Behavior of the urban traffic of the city of Sao Paulo in Brazil</a>,
<a href="datasets/Travel+Reviews"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Travel+Reviews">Travel Reviews</a>,
<a href="datasets/Tarvel+Review+Ratings"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Tarvel+Review+Ratings">Tarvel Review Ratings</a>,
<a href="datasets/Rice+Leaf+Diseases"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Rice+Leaf+Diseases">Rice Leaf Diseases</a>,
<a href="about.html">About</a>,
<a href="citation_policy.html">Citation Policy</a>,
<a href="donation_policy.html">Donation Policy</a>,
<a href="contact.html">Contact</a>,
<a href="http://cml.ics.uci.edu">CML</a>]

In [26]:

```
len(links)
```

Out[26]:

988

In [27]:

```
#buscar patrón
link_name = re.findall('"(datasets/.+?)"', str(links))
link_name
```

Out[27]:

['datasets/Abalone',
 'datasets/Abalone',
 'datasets/Adult',
 'datasets/Adult',
 'datasets/Annealing',
 'datasets/Annealing',
 'datasets/Anonymous+Microsoft+Web+Data',
 'datasets/Anonymous+Microsoft+Web+Data',
 'datasets/Arrhythmia',
 'datasets/Arrhythmia',
 'datasets/Artificial+Characters',
 'datasets/Artificial+Characters',
 'datasets/Audiology+%28Original%29',
 'datasets/Audiology+%28Original%29',
 'datasets/Audiology+%28Standardized%29',
 'datasets/Audiology+%28Standardized%29',
 'datasets/Auto+MPG',
 'datasets/Auto+MPG',
 'datasets/Automobile',
 'datasets/Automobile',
 'datasets/Badges',
 'datasets/Badges',
 'datasets/Balance+Scale',
 'datasets/Balance+Scale',
 'datasets/Balloons',
 'datasets/Balloons',
 'datasets/Breast+Cancer',
 'datasets/Breast+Cancer',
 'datasets/Breast+Cancer+Wisconsin+%28Original%29',
 'datasets/Breast+Cancer+Wisconsin+%28Original%29',
 'datasets/Breast+Cancer+Wisconsin+%28Prognostic%29',
 'datasets/Breast+Cancer+Wisconsin+%28Prognostic%29',
 'datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29',
 'datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29',
 'datasets/Pittsburgh+Bridges',
 'datasets/Pittsburgh+Bridges',
 'datasets/Car+Evaluation',
 'datasets/Car+Evaluation',
 'datasets/Census+Income',
 'datasets/Census+Income',
 'datasets/Chess+%28King-Rook+vs.+King-Knight%29',
 'datasets/Chess+%28King-Rook+vs.+King-Knight%29',
 'datasets/Chess+%28King-Rook+vs.+King-Pawn%29',
 'datasets/Chess+%28King-Rook+vs.+King-Pawn%29',
 'datasets/Chess+%28King-Rook+vs.+King%29',
 'datasets/Chess+%28King-Rook+vs.+King%29',
 'datasets/Chess+%28Domain+Theories%29',

'datasets/Chess+%28Domain+Theories%29',
'datasets/Bach+Chorales',
'datasets/Bach+Chorales',
'datasets/Connect-4',
'datasets/Connect-4',
'datasets/Credit+Approval',
'datasets/Credit+Approval',
'datasets/Japanese+Credit+Screening',
'datasets/Japanese+Credit+Screening',
'datasets/Computer+Hardware',
'datasets/Computer+Hardware',
'datasets/Contraceptive+Method+Choice',
'datasets/Contraceptive+Method+Choice',
'datasets/Covertype',
'datasets/Covertype',
'datasets/Cylinder+Bands',
'datasets/Cylinder+Bands',
'datasets/Dermatology',
'datasets/Dermatology',
'datasets/Diabetes',
'datasets/Diabetes',
'datasets/DGP2+-+The+Second+Data+Generation+Program',
'datasets/DGP2+-+The+Second+Data+Generation+Program',
'datasets/Document+Understanding',
'datasets/Document+Understanding',
'datasets/EBL+Domain+Theories',
'datasets/EBL+Domain+Theories',
'datasets/Echocardiogram',
'datasets/Echocardiogram',
'datasets/Ecoli',
'datasets/Ecoli',
'datasets/Flags',
'datasets/Flags',
'datasets/Function+Finding',
'datasets/Function+Finding',
'datasets/Glass+Identification',
'datasets/Glass+Identification',
'datasets/Haberman%27s+Survival',
'datasets/Haberman%27s+Survival',
'datasets/Hayes-Roth',
'datasets/Hayes-Roth',
'datasets/Heart+Disease',
'datasets/Heart+Disease',
'datasets/Hepatitis',
'datasets/Hepatitis',
'datasets/Horse+Colic',
'datasets/Horse+Colic',
'datasets/ICU',
'datasets/ICU',
'datasets/Image+Segmentation',
'datasets/Image+Segmentation',
'datasets/Internet+Advertisements',
'datasets/Internet+Advertisements',
'datasets/Ionosphere',
'datasets/Ionosphere',
'datasets/Iris',
'datasets/Iris',
'datasets/ISOLET',
'datasets/ISOLET',
'datasets/Kinship',
'datasets/Kinship',
'datasets/Labor+Relations',

'datasets/Labor+Relations',
'datasets/LED+Display+Domain',
'datasets/LED+Display+Domain',
'datasets/Lenses',
'datasets/Lenses',
'datasets/Letter+Recognition',
'datasets/Letter+Recognition',
'datasets/Liver+Disorders',
'datasets/Liver+Disorders',
'datasets/Logic+Theorist',
'datasets/Logic+Theorist',
'datasets/Lung+Cancer',
'datasets/Lung+Cancer',
'datasets/Lymphography',
'datasets/Lymphography',
'datasets/Mechanical+Analysis',
'datasets/Mechanical+Analysis',
'datasets/Meta-data',
'datasets/Meta-data',
'datasets/Mobile+Robots',
'datasets/Mobile+Robots',
'datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29',
'datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29',
'datasets/Molecular+Biology+%28Protein+Secondary+Structure%29',
'datasets/Molecular+Biology+%28Protein+Secondary+Structure%29',
'datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29',
'datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29',
'datasets/MONK%27s+Problems',
'datasets/MONK%27s+Problems',
'datasets/Moral+Reasoner',
'datasets/Moral+Reasoner',
'datasets/Multiple+Features',
'datasets/Multiple+Features',
'datasets/Mushroom',
'datasets/Mushroom',
'datasets/Musk+%28Version+1%29',
'datasets/Musk+%28Version+1%29',
'datasets/Musk+%28Version+2%29',
'datasets/Musk+%28Version+2%29',
'datasets/Nursery',
'datasets/Nursery',
'datasets/Othello+Domain+Theory',
'datasets/Othello+Domain+Theory',
'datasets/Page+Blocks+Classification',
'datasets/Page+Blocks+Classification',
'datasets/Optical+Recognition+of+Handwritten+Digits',
'datasets/Optical+Recognition+of+Handwritten+Digits',
'datasets/Pen-Based+Recognition+of+Handwritten+Digits',
'datasets/Pen-Based+Recognition+of+Handwritten+Digits',
'datasets/Post-Operative+Patient',
'datasets/Post-Operative+Patient',
'datasets/Primary+Tumor',
'datasets/Primary+Tumor',
'datasets/Prodigy',
'datasets/Prodigy',
'datasets/Qualitative+Structure+Activity+Relationships',
'datasets/Qualitative+Structure+Activity+Relationships',
'datasets/Quadruped+Mammals',
'datasets/Quadruped+Mammals',
'datasets/Servo',
'datasets/Servo',
'datasets/Shuttle+Landing+Control',

'datasets/Shuttle+Landing+Control',
'datasets/Solar+Flare',
'datasets/Solar+Flare',
'datasets/Soybean+%28Large%29',
'datasets/Soybean+%28Large%29',
'datasets/Soybean+%28Small%29',
'datasets/Soybean+%28Small%29',
'datasets/Challenger+USA+Space+Shuttle+O-Ring',
'datasets/Challenger+USA+Space+Shuttle+O-Ring',
'datasets/Low+Resolution+Spectrometer',
'datasets/Low+Resolution+Spectrometer',
'datasets/Spambase',
'datasets/Spambase',
'datasets/SPECT+Heart',
'datasets/SPECT+Heart',
'datasets/SPECTF+Heart',
'datasets/SPECTF+Heart',
'datasets/Sponge',
'datasets/Sponge',
'datasets/Statlog+Project',
'datasets/Statlog+Project',
'datasets/Student+Loan+Relational',
'datasets/Student+Loan+Relational',
'datasets/Teaching+Assistant+Evaluation',
'datasets/Teaching+Assistant+Evaluation',
'datasets/Tic-Tac-Toe+Endgame',
'datasets/Tic-Tac-Toe+Endgame',
'datasets/Thyroid+Disease',
'datasets/Thyroid+Disease',
'datasets/Trains',
'datasets/Trains',
'datasets/University',
'datasets/University',
'datasets/Congressional+Voting+Records',
'datasets/Congressional+Voting+Records',
'datasets/Water+Treatment+Plant',
'datasets/Water+Treatment+Plant',
'datasets/Waveform+Database+Generator+%28Version+1%29',
'datasets/Waveform+Database+Generator+%28Version+1%29',
'datasets/Waveform+Database+Generator+%28Version+2%29',
'datasets/Waveform+Database+Generator+%28Version+2%29',
'datasets/Wine',
'datasets/Wine',
'datasets/Yeast',
'datasets/Yeast',
'datasets/Zoo',
'datasets/Zoo',
'datasets/Undocumented',
'datasets/Undocumented',
'datasets/Twenty+Newsgroups',
'datasets/Twenty+Newsgroups',
'datasets/Australian+Sign+Language+signs',
'datasets/Australian+Sign+Language+signs',
'datasets/Australian+Sign+Language+signs+%28High+Quality%29',
'datasets/Australian+Sign+Language+signs+%28High+Quality%29',
'datasets/US+Census+Data+%281990%29',
'datasets/US+Census+Data+%281990%29',
'datasets/Census-Income+%28KDD%29',
'datasets/Census-Income+%28KDD%29',
'datasets/Coil+1999+Competition+Data',
'datasets/Coil+1999+Competition+Data',
'datasets/Corel+Image+Features',

'datasets/Corel+Image+Features',
'datasets/E.+Coli+Genes',
'datasets/E.+Coli+Genes',
'datasets/EEG+Database',
'datasets/EEG+Database',
'datasets/El+Nino',
'datasets/El+Nino',
'datasets/Entree+Chicago+Recommendation+Data',
'datasets/Entree+Chicago+Recommendation+Data',
'datasets/CMU+Face+Images',
'datasets/CMU+Face+Images',
'datasets/Insurance+Company+Benchmark+%28COIL+2000%29',
'datasets/Insurance+Company+Benchmark+%28COIL+2000%29',
'datasets/Internet+Usage+Data',
'datasets/Internet+Usage+Data',
'datasets/IPUMS+Census+Database',
'datasets/IPUMS+Census+Database',
'datasets/Japanese+Vowels',
'datasets/Japanese+Vowels',
'datasets/KDD+Cup+1998+Data',
'datasets/KDD+Cup+1998+Data',
'datasets/KDD+Cup+1999+Data',
'datasets/KDD+Cup+1999+Data',
'datasets/M.+Tuberculosis+Genes',
'datasets/M.+Tuberculosis+Genes',
'datasets/Movie',
'datasets/Movie',
'datasets/MSNBC.com+Anonymous+Web+Data',
'datasets/MSNBC.com+Anonymous+Web+Data',
'datasets/NSF+Research+Award+Abstracts+1990-2003',
'datasets/NSF+Research+Award+Abstracts+1990-2003',
'datasets/Pioneer-1+Mobile+Robot+Data',
'datasets/Pioneer-1+Mobile+Robot+Data',
'datasets/Pseudo+Periodic+Synthetic+Time+Series',
'datasets/Pseudo+Periodic+Synthetic+Time+Series',
'datasets/Reuters-21578+Text+Categorization+Collection',
'datasets/Reuters-21578+Text+Categorization+Collection',
'datasets/Robot+Execution+Failures',
'datasets/Robot+Execution+Failures',
'datasets/Synthetic+Control+Chart+Time+Series',
'datasets/Synthetic+Control+Chart+Time+Series',
'datasets/Syskill+and+Webert+Web+Page+Ratings',
'datasets/Syskill+and+Webert+Web+Page+Ratings',
'datasets/UNIX+User+Data',
'datasets/UNIX+User+Data',
'datasets/Volcanoes+on+Venus+-+JARtool+experiment',
'datasets/Volcanoes+on+Venus+-+JARtool+experiment',
'datasets/Statlog+%28Australian+Credit+Approval%29',
'datasets/Statlog+%28Australian+Credit+Approval%29',
'datasets/Statlog+%28German+Credit+Data%29',
'datasets/Statlog+%28German+Credit+Data%29',
'datasets/Statlog+%28Heart%29',
'datasets/Statlog+%28Heart%29',
'datasets/Statlog+%28Landsat+Satellite%29',
'datasets/Statlog+%28Landsat+Satellite%29',
'datasets/Statlog+%28Image+Segmentation%29',
'datasets/Statlog+%28Image+Segmentation%29',
'datasets/Statlog+%28Shuttle%29',
'datasets/Statlog+%28Shuttle%29',
'datasets/Statlog+%28Vehicle+Silhouettes%29',
'datasets/Statlog+%28Vehicle+Silhouettes%29',
'datasets/Connectionist+Bench+%28Nettalk+Corpus%29'

'datasets/Connectionist+Bench+%28Nettalk+Corpus%29',
'datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29',
'datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29',
'datasets/Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29',
'datasets/Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29',
'datasets/Economic+Sanctions',
'datasets/Economic+Sanctions',
'datasets/Protein+Data',
'datasets/Protein+Data',
'datasets/Cloud',
'datasets/Cloud',
'datasets/CalIt2+Building+People+Counts',
'datasets/CalIt2+Building+People+Counts',
'datasets/Dodgers+Loop+Sensor',
'datasets/Dodgers+Loop+Sensor',
'datasets/Poker+Hand',
'datasets/Poker+Hand',
'datasets/MAGIC+Gamma+Telescope',
'datasets/MAGIC+Gamma+Telescope',
'datasets/UJI+Pen+Characters',
'datasets/UJI+Pen+Characters',
'datasets/Mammographic+Mass',
'datasets/Mammographic+Mass',
'datasets/Forest+Fires',
'datasets/Forest+Fires',
'datasets/Reuters+Transcribed+Subset',
'datasets/Reuters+Transcribed+Subset',
'datasets/Bag+of+Words',
'datasets/Bag+of+Words',
'datasets/Concrete+Compressive+Strength',
'datasets/Concrete+Compressive+Strength',
'datasets/Hill-Valley',
'datasets/Hill-Valley',
'datasets/Arcene',
'datasets/Arcene',
'datasets/Dexter',
'datasets/Dexter',
'datasets/Dorothea',
'datasets/Dorothea',
'datasets/Gisette',
'datasets/Gisette',
'datasets/Madelon',
'datasets/Madelon',
'datasets/Ozone+Level+Detection',
'datasets/Ozone+Level+Detection',
'datasets/Abscisic+Acid+Signaling+Network',
'datasets/Abscisic+Acid+Signaling+Network',
'datasets/Parkinsons',
'datasets/Parkinsons',
'datasets/Character+Trajectories',
'datasets/Character+Trajectories',
'datasets/Blood+Transfusion+Service+Center',
'datasets/Blood+Transfusion+Service+Center',
'datasets/UJI+Pen+Characters+%28Version+2%29',
'datasets/UJI+Pen+Characters+%28Version+2%29',
'datasets/Semeion+Handwritten+Digit',
'datasets/Semeion+Handwritten+Digit',
'datasets/SECOM',
'datasets/SECOM',
'datasets/Plants',
'datasets/Plants',
'datasets/Libras+Movement',

'datasets/Libras+Movement',
'datasets/Concrete+Slump+Test',
'datasets/Concrete+Slump+Test',
'datasets/Communities+and+Crime',
'datasets/Communities+and+Crime',
'datasets/Acute+Inflammations',
'datasets/Acute+Inflammations',
'datasets/Wine+Quality',
'datasets/Wine+Quality',
'datasets/URL+Reputation',
'datasets/URL+Reputation',
'datasets/p53+Mutants',
'datasets/p53+Mutants',
'datasets/Parkinsons+Telemonitoring',
'datasets/Parkinsons+Telemonitoring',
'datasets/Demospongiae',
'datasets/Demospongiae',
'datasets/Opinosis+Opinion+%26frasl%3B+Review',
'datasets/Opinosis+Opinion+%26frasl%3B+Review',
'datasets/Breast+Tissue',
'datasets/Breast+Tissue',
'datasets/Cardiotocography',
'datasets/Cardiotocography',
'datasets/Wall-Following+Robot+Navigation+Data',
'datasets/Wall-Following+Robot+Navigation+Data',
'datasets/Spoken+Arabic+Digit',
'datasets/Spoken+Arabic+Digit',
'datasets/Localization+Data+for+Person+Activity',
'datasets/Localization+Data+for+Person+Activity',
'datasets/AutoUniv',
'datasets/AutoUniv',
'datasets/Steel+Plates+Faults',
'datasets/Steel+Plates+Faults',
'datasets/MiniBooNE+particle+identification',
'datasets/MiniBooNE+particle+identification',
'datasets/YearPredictionMSD',
'datasets/YearPredictionMSD',
'datasets/PEMS-SF',
'datasets/PEMS-SF',
'datasets/OpinRank+Review+Dataset',
'datasets/OpinRank+Review+Dataset',
'datasets/Relative+location+of+CT+slices+on+axial+axis',
'datasets/Relative+location+of+CT+slices+on+axial+axis',
'datasets/Online+Handwritten+Assamese+Characters+Dataset',
'datasets/Online+Handwritten+Assamese+Characters+Dataset',
'datasets/PubChem+Bioassay+Data',
'datasets/PubChem+Bioassay+Data',
'datasets/Record+Linkage+Comparison+Patterns',
'datasets/Record+Linkage+Comparison+Patterns',
'datasets/Communities+and+Crime+Unnormalized',
'datasets/Communities+and+Crime+Unnormalized',
'datasets/Vertebral+Column',
'datasets/Vertebral+Column',
'datasets/EMG+Physical+Action+Data+Set',
'datasets/EMG+Physical+Action+Data+Set',
'datasets/Vicon+Physical+Action+Data+Set',
'datasets/Vicon+Physical+Action+Data+Set',
'datasets/Amazon+Commerce+reviews+set',
'datasets/Amazon+Commerce+reviews+set',
'datasets/Amazon+Access+Samples',
'datasets/Amazon+Access+Samples',
'datasets/Reuter_50_50',

'datasets/Reuter_50_50',
'datasets/Reuter_50_50',
'datasets/Farm+Ads',
'datasets/Farm+Ads',
'datasets/DBWorld+e-mails',
'datasets/DBWorld+e-mails',
'datasets/KEGG+Metabolic+Relation+Network+%28Directed%29',
'datasets/KEGG+Metabolic+Relation+Network+%28Directed%29',
'datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29',
'datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29',
'datasets/Bank+Marketing',
'datasets/Bank+Marketing',
'datasets/YouTube+Comedy+Slam+Preference+Data',
'datasets/YouTube+Comedy+Slam+Preference+Data',
'datasets/Gas+Sensor+Array+Drift+Dataset',
'datasets/Gas+Sensor+Array+Drift+Dataset',
'datasets/ILPD+%28Indian+Liver+Patient+Dataset%29',
'datasets/ILPD+%28Indian+Liver+Patient+Dataset%29',
'datasets/OPPORTUNITY+Activity+Recognition',
'datasets/OPPORTUNITY+Activity+Recognition',
'datasets/Nomao',
'datasets/Nomao',
'datasets/SMS+Spam+Collection',
'datasets/SMS+Spam+Collection',
'datasets/Skin+Segmentation',
'datasets/Skin+Segmentation',
'datasets/Planning+Relax',
'datasets/Planning+Relax',
'datasets/PAMAP2+Physical+Activity+Monitoring',
'datasets/PAMAP2+Physical+Activity+Monitoring',
'datasets/Restaurant+%26+consumer+data',
'datasets/Restaurant+%26+consumer+data',
'datasets/CNAE-9',
'datasets/CNAE-9',
'datasets/Individual+household+electric+power+consumption',
'datasets/Individual+household+electric+power+consumption',
'datasets/seeds',
'datasets/seeds',
'datasets/Northix',
'datasets/Northix',
'datasets/QtyT40I10D100K',
'datasets/QtyT40I10D100K',
'datasets/Legal+Case+Reports',
'datasets/Legal+Case+Reports',
'datasets/Human+Activity+Recognition+Using+Smartphones',
'datasets/Human+Activity+Recognition+Using+Smartphones',
'datasets/One-hundred+plant+species+leaves+data+set',
'datasets/One-hundred+plant+species+leaves+data+set',
'datasets/Energy+efficiency',
'datasets/Energy+efficiency',
'datasets/Yacht+Hydrodynamics',
'datasets/Yacht+Hydrodynamics',
'datasets/Fertility',
'datasets/Fertility',
'datasets/Daphnet+Freezing+of+Gait',
'datasets/Daphnet+Freezing+of+Gait',
'datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29',
'datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29',
'datasets/ISTANBUL+STOCK+EXCHANGE',
'datasets/ISTANBUL+STOCK+EXCHANGE',
'datasets/Buzz+in+social+media+',
'datasets/Buzz+in+social+media+',

'datasets/First-order+theorem+proving',
'datasets/First-order+theorem+proving',
'datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29',
'datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29',
'datasets/Gas+sensor+arrays+in+open+sampling+settings',
'datasets/Gas+sensor+arrays+in+open+sampling+settings',
'datasets/Climate+Model+Simulation+Crashes',
'datasets/Climate+Model+Simulation+Crashes',
'datasets/MicroMass',
'datasets/MicroMass',
'datasets/QSAR+biodegradation',
'datasets/QSAR+biodegradation',
'datasets/BLOGGER',
'datasets/BLOGGER',
'datasets/Daily+and+Sports+Activities',
'datasets/Daily+and+Sports+Activities',
'datasets/User+Knowledge+Modeling',
'datasets/User+Knowledge+Modeling',
'datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection',
'datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection',
'datasets/NYSK',
'datasets/NYSK',
'datasets/Turkiye+Student+Evaluation',
'datasets/Turkiye+Student+Evaluation',
'datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machines
%29',
'datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machines
%29',
'datasets/EEG+Eye+State',
'datasets/EEG+Eye+State',
'datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure',
'datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure',
'datasets/seismic-bumps',
'datasets/seismic-bumps',
'datasets/banknote+authentication',
'datasets/banknote+authentication',
'datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Pr
oblem%3A+Pat',
'datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Pr
oblem%3A+Pat',
'datasets/YouTube+Multiview+Video+Games+Dataset',
'datasets/YouTube+Multiview+Video+Games+Dataset',
'datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations',
'datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations',
'datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors',
'datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors',
'datasets/SkillCraft1+Master+Table+Dataset',
'datasets/SkillCraft1+Master+Table+Dataset',
'datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units',
'datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units',
'datasets/SML2010',
'datasets/SML2010',
'datasets/Bike+Sharing+Dataset',
'datasets/Bike+Sharing+Dataset',
'datasets/Predict+keywords+activities+in+a+online+social+media',
'datasets/Predict+keywords+activities+in+a+online+social+media',
'datasets/Thoracic+Surgery+Data',
'datasets/Thoracic+Surgery+Data',
'datasets/EMG+dataset+in+Lower+Limb',
'datasets/EMG+dataset+in+Lower+Limb',
'datasets/SUSY',
'datasets/SUSY',

'datasets/HIGGS',
'datasets/HIGGS',
'datasets/Qualitative_Bankruptcy',
'datasets/Qualitative_Bankruptcy',
'datasets/LSVT+Voice+Rehabilitation',
'datasets/LSVT+Voice+Rehabilitation',
'datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer',
'datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer',
'datasets/Wilt',
'datasets/Wilt',
'datasets/User+Identification+From+Walking+Activity',
'datasets/User+Identification+From+Walking+Activity',
'datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer',
'datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer',
'datasets/Leaf',
'datasets/Leaf',
'datasets/Dresses_Attribute_Sales',
'datasets/Dresses_Attribute_Sales',
'datasets/Tamilnadu+Electricity+Board+Hourly+Readings',
'datasets/Tamilnadu+Electricity+Board+Hourly+Readings',
'datasets/Airfoil+Self-Noise',
'datasets/Airfoil+Self-Noise',
'datasets/Wholesale+customers',
'datasets/Wholesale+customers',
'datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis',
'datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis',
'datasets/Combined+Cycle+Power+Plant',
'datasets/Combined+Cycle+Power+Plant',
'datasets/Urban+Land+Cover',
'datasets/Urban+Land+Cover',
'datasets/Diabetes+130-US+hospitals+for+years+1999-2008',
'datasets/Diabetes+130-US+hospitals+for+years+1999-2008',
'datasets/Bach+Choral+Harmony',
'datasets/Bach+Choral+Harmony',
'datasets/StoneFlakes',
'datasets/StoneFlakes',
'datasets/Tennis+Major+Tournament+Match+Statistics',
'datasets/Tennis+Major+Tournament+Match+Statistics',
'datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings',
'datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings',
'datasets/Gesture+Phase+Segmentation',
'datasets/Gesture+Phase+Segmentation',
'datasets/Perfume+Data',
'datasets/Perfume+Data',
'datasets/BlogFeedback',
'datasets/BlogFeedback',
'datasets/REALDISP+Activity+Recognition+Dataset',
'datasets/REALDISP+Activity+Recognition+Dataset',
'datasets/Newspaper+and+magazine+images+segmentation+dataset',
'datasets/Newspaper+and+magazine+images+segmentation+dataset',
'datasets/AAAI+2014+Accepted+Papers',
'datasets/AAAI+2014+Accepted+Papers',
'datasets/Gas+sensor+array+under+flow+modulation',
'datasets/Gas+sensor+array+under+flow+modulation',
'datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures',
'datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures',
'datasets/UJIIndoorLoc',
'datasets/UJIIndoorLoc',
'datasets/Sentence+Classification',
'datasets/Sentence+Classification',
'datasets/Dow+Jones+Index',
'datasets/Dow+Jones+Index',

'datasets/sEMG+for+Basic+Hand+movements',
'datasets/sEMG+for+Basic+Hand+movements',
'datasets/AAAI+2013+Accepted+Papers',
'datasets/AAAI+2013+Accepted+Papers',
'datasets/Geographical+Original+of+Music',
'datasets/Geographical+Original+of+Music',
'datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants',
'datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants',
'datasets/Grammatical+Facial+Expressions',
'datasets/Grammatical+Facial+Expressions',
'datasets/NoisyOffice',
'datasets/NoisyOffice',
'datasets/MHEALTH+Dataset',
'datasets/MHEALTH+Dataset',
'datasets/Student+Performance',
'datasets/Student+Performance',
'datasets/ElectricityLoadDiagrams20112014',
'datasets/ElectricityLoadDiagrams20112014',
'datasets/Gas+sensor+array+under+dynamic+gas+mixtures',
'datasets/Gas+sensor+array+under+dynamic+gas+mixtures',
'datasets/microblogPCU',
'datasets/microblogPCU',
'datasets/Firm-Teacher_Clave-Direction_Classification',
'datasets/Firm-Teacher_Clave-Direction_Classification',
'datasets/Dataset+for+Sensorless+Drive+Diagnosis',
'datasets/Dataset+for+Sensorless+Drive+Diagnosis',
'datasets/TV+News+Channel+Commercial+Detection+Dataset',
'datasets/TV+News+Channel+Commercial+Detection+Dataset',
'datasets/Phishing+Websites',
'datasets/Phishing+Websites',
'datasets/Greenhouse+Gas+Observing+Network',
'datasets/Greenhouse+Gas+Observing+Network',
'datasets/Diabetic+Retinopathy+Debrecen+Data+Set',
'datasets/Diabetic+Retinopathy+Debrecen+Data+Set',
'datasets/HIV-1+protease+cleavage',
'datasets/HIV-1+protease+cleavage',
'datasets/Sentiment+Labelled+Sentences',
'datasets/Sentiment+Labelled+Sentences',
'datasets/Online+News+Popularity',
'datasets/Online+News+Popularity',
'datasets/Forest+type+mapping',
'datasets/Forest+type+mapping',
'datasets/wiki4HE',
'datasets/wiki4HE',
'datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset',
'datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset',
'datasets/Chronic_Kidney_Disease',
'datasets/Chronic_Kidney_Disease',
'datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014',
'datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014',
'datasets/Folio',
'datasets/Folio',
'datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015',
'datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015',
'datasets/Cuff-Less+Blood+Pressure+Estimation',
'datasets/Cuff-Less+Blood+Pressure+Estimation',
'datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions',
'datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions',
'datasets/Mice+Protein+Expression',
'datasets/Mice+Protein+Expression',
'datasets/UJIIndoorLoc-Mag',
'datasets/UJIIndoorLoc-Mag',

'datasets/Heterogeneity+Activity+Recognition',
'datasets/Heterogeneity+Activity+Recognition',
'datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set',
'datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set',
'datasets/HEPMASS',
'datasets/HEPMASS',
'datasets/Indoor+User+Movement+Prediction+from+RSS+data',
'datasets/Indoor+User+Movement+Prediction+from+RSS+data',
'datasets/Open+University+Learning+Analytics+dataset',
'datasets/Open+University+Learning+Analytics+dataset',
'datasets/default+of+credit+card+clients',
'datasets/default+of+credit+card+clients',
'datasets/Mesothelioma%E2%80%99s+disease+data+set+',
'datasets/Mesothelioma%E2%80%99s+disease+data+set+',
'datasets/Online+Retail',
'datasets/Online+Retail',
'datasets/SIFT10M',
'datasets/SIFT10M',
'datasets/GPS+Trajectories',
'datasets/GPS+Trajectories',
'datasets/Detect+Malacious+Executable%28AntiVirus%29',
'datasets/Detect+Malacious+Executable%28AntiVirus%29',
'datasets/Occupancy+Detection+',
'datasets/Occupancy+Detection+',
'datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Parkinson%E2%80%99s+D
isease',
'datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Parkinson%E2%80%99s+D
isease',
'datasets/News+Aggregator',
'datasets/News+Aggregator',
'datasets/Air+Quality',
'datasets/Air+Quality',
'datasets/Twin+gas+sensor+arrays',
'datasets/Twin+gas+sensor+arrays',
'datasets/Gas+sensors+for+home+activity+monitoring',
'datasets/Gas+sensors+for+home+activity+monitoring',
'datasets/Facebook+Comment+Volume+Dataset',
'datasets/Facebook+Comment+Volume+Dataset',
'datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Living+
%28AAL%29',
'datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Living+
%28AAL%29',
'datasets/Polish+companies+bankruptcy+data',
'datasets/Polish+companies+bankruptcy+data',
'datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29',
'datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29',
'datasets/Dota2+Games+Results',
'datasets/Dota2+Games+Results',
'datasets/Facebook+metrics',
'datasets/Facebook+metrics',
'datasets/UbiqLog+%28smartphone+lifelogging%29',
'datasets/UbiqLog+%28smartphone+lifelogging%29',
'datasets/NIPS+Conference+Papers+1987-2015',
'datasets/NIPS+Conference+Papers+1987-2015',
'datasets/HTRU2',
'datasets/HTRU2',
'datasets/Drug+consumption+%28quantified%29',
'datasets/Drug+consumption+%28quantified%29',
'datasets/Appliances+energy+prediction',
'datasets/Appliances+energy+prediction',
'datasets/Miskolc+IIS+Hybrid+IPS',
'datasets/Miskolc+IIS+Hybrid+IPS',

'datasets/KDC-4007+dataset+Collection',
'datasets/KDC-4007+dataset+Collection',
'datasets/Geo-Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone',
'datasets/Geo-Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone',
'datasets/DrivFace',
'datasets/DrivFace',
'datasets/Website+Phishing',
'datasets/Website+Phishing',
'datasets/YouTube+Spam+Collection',
'datasets/YouTube+Spam+Collection',
'datasets/Beijing+PM2.5+Data',
'datasets/Beijing+PM2.5+Data',
'datasets/Cargo+2000+Freight+Tracking+and+Tracing',
'datasets/Cargo+2000+Freight+Tracking+and+Tracing',
'datasets/Cervical+cancer+%28Risk+Factors%29',
'datasets/Cervical+cancer+%28Risk+Factors%29',
'datasets/Quality+Assessment+of+Digital+Colposcopies',
'datasets/Quality+Assessment+of+Digital+Colposcopies',
'datasets/KASANDR',
'datasets/KASANDR',
'datasets/FMA%3A+A+Dataset+For+Music+Analysis',
'datasets/FMA%3A+A+Dataset+For+Music+Analysis',
'datasets/Air+quality',
'datasets/Air+quality',
'datasets/Epileptic+Seizure+Recognition',
'datasets/Epileptic+Seizure+Recognition',
'datasets/Devanagari+Handwritten+Character+Dataset',
'datasets/Devanagari+Handwritten+Character+Dataset',
'datasets/Stock+portfolio+performance',
'datasets/Stock+portfolio+performance',
'datasets/MoCap+Hand+Postures',
'datasets/MoCap+Hand+Postures',
'datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech',
'datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech',
'datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting',
'datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting',
'datasets/PM2.5+Data+of+Five+Chinese+Cities',
'datasets/PM2.5+Data+of+Five+Chinese+Cities',
'datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet',
'datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet',
'datasets/Sales_Transactions_Dataset_Weekly',
'datasets/Sales_Transactions_Dataset_Weekly',
'datasets/Las+Vegas+Strip',
'datasets/Las+Vegas+Strip',
'datasets/Eco-hotel',
'datasets/Eco-hotel',
'datasets/MEU-Mobile+KSD',
'datasets/MEU-Mobile+KSD',
'datasets/Crowdsourced+Mapping',
'datasets/Crowdsourced+Mapping',
'datasets/gene+expression+cancer+RNA-Seq',
'datasets/gene+expression+cancer+RNA-Seq',
'datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer',
'datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer',
'datasets/chestnut+%E2%80%93+LARVIC',
'datasets/chestnut+%E2%80%93+LARVIC',
'datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%28OBS%29+Network',
'datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%28OBS%29+Network',
'datasets/Motion+Capture+Hand+Postures',
'datasets/Motion+Capture+Hand+Postures',

'datasets/Anuran+Calls+%28MFCCs%29',
'datasets/Anuran+Calls+%28MFCCs%29',
'datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization',
'datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization',
'datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy',
'datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy',
'datasets/Daily+Demand+Forecasting+Orders',
'datasets/Daily+Demand+Forecasting+Orders',
'datasets/Paper+Reviews',
'datasets/Paper+Reviews',
'datasets/extention+of+Z-Alizadeh+sani+dataset',
'datasets/extention+of+Z-Alizadeh+sani+dataset',
'datasets/Z-Alizadeh+Sani',
'datasets/Z-Alizadeh+Sani',
'datasets/Dynamic+Features+of+VirusShare+Executables',
'datasets/Dynamic+Features+of+VirusShare+Executables',
'datasets/IDA2016Challenge',
'datasets/IDA2016Challenge',
'datasets/DSRC+Vehicle+Communications',
'datasets/DSRC+Vehicle+Communications',
'datasets/Mturk+User-Perceived+Clusters+over+Images',
'datasets/Mturk+User-Perceived+Clusters+over+Images',
'datasets/Character+Font+Images',
'datasets/Character+Font+Images',
'datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Instance+Labels',
'datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Instance+Labels',
'datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++',
'datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++',
'datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++',
'datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++',
'datasets/APS+Failure+at+Scania+Trucks',
'datasets/APS+Failure+at+Scania+Trucks',
'datasets/Wireless+Indoor+Localization',
'datasets/Wireless+Indoor+Localization',
'datasets/HCC+Survival',
'datasets/HCC+Survival',
'datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015',
'datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015',
'datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29',
'datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29',
'datasets/Autism+Screening+Adult',
'datasets/Autism+Screening+Adult',
'datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor',
'datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor',
'datasets/Immunotherapy+Dataset',
'datasets/Immunotherapy+Dataset',
'datasets/Cryotherapy+Dataset+',
'datasets/Cryotherapy+Dataset+',
'datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes',
'datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes',
'datasets/Discrete+Tone+Image+Dataset',
'datasets/Discrete+Tone+Image+Dataset',
'datasets/News+Popularity+in+Multiple+Social+Media+Platforms',
'datasets/News+Popularity+in+Multiple+Social+Media+Platforms',
'datasets/Ultrasonic+flowmeter+diagnostics',
'datasets/Ultrasonic+flowmeter+diagnostics',
'datasets/ICMLA+2014+Accepted+Papers+Data+Set',
'datasets/ICMLA+2014+Accepted+Papers+Data+Set',
'datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation',
'datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation',
'datasets/Container+Crane+Controller+Data+Set',
'datasets/Container+Crane+Controller+Data+Set',

'datasets/Residential+Building+Data+Set',
'datasets/Residential+Building+Data+Set',
'datasets/Health+News+in+Twitter',
'datasets/Health+News+in+Twitter',
'datasets/chipseq',
'datasets/chipseq',
'datasets/SGEMM+GPU+kernel+performance',
'datasets/SGEMM+GPU+kernel+performance',
'datasets/Repeat+Consumption+Matrices',
'datasets/Repeat+Consumption+Matrices',
'datasets/detection_of_IoT_botnet_attacks_N_BaIoT',
'datasets/detection_of_IoT_botnet_attacks_N_BaIoT',
'datasets/Absenteeism+at+work',
'datasets/Absenteeism+at+work',
'datasets/SCADI',
'datasets/SCADI',
'datasets/Condition+monitoring+of+hydraulic+systems',
'datasets/Condition+monitoring+of+hydraulic+systems',
'datasets/Carbon+Nanotubes',
'datasets/Carbon+Nanotubes',
'datasets/Optical+Interconnection+Network+',
'datasets/Optical+Interconnection+Network+',
'datasets/Sports+articles+for+objectivity+analysis',
'datasets/Sports+articles+for+objectivity+analysis',
'datasets/Breast+Cancer+Coimbra',
'datasets/Breast+Cancer+Coimbra',
'datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data',
'datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data',
'datasets/Dishonest+Internet+users+Dataset',
'datasets/Dishonest+Internet+users+Dataset',
'datasets/Victorian+Era+Authorship+Attribution',
'datasets/Victorian+Era+Authorship+Attribution',
'datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set',
'datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set',
'datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing',
'datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing',
'datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals',
'datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals',
'datasets/Roman+Urdu+Data+Set',
'datasets/Roman+Urdu+Data+Set',
'datasets/Avila',
'datasets/Avila',
'datasets/PANDOR',
'datasets/PANDOR',
'datasets/Drug+Review+Dataset+%28Druglib.com%29',
'datasets/Drug+Review+Dataset+%28Druglib.com%29',
'datasets/Drug+Review+Dataset+%28Drugs.com%29',
'datasets/Drug+Review+Dataset+%28Drugs.com%29',
'datasets/Physical+Unclonable+Functions',
'datasets/Physical+Unclonable+Functions',
'datasets/Superconductivty+Data',
'datasets/Superconductivty+Data',
'datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29',
'datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29',
'datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2',
'datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2',
'datasets/Student+Academics+Performance',
'datasets/Student+Academics+Performance',
'datasets/Online+Shoppers+Purchasing+Intention+Dataset',
'datasets/Online+Shoppers+Purchasing+Intention+Dataset',
'datasets/PMU-UD',
'datasets/PMU-UD',

'datasets/Parkinson%27s+Disease+Classification',
'datasets/Parkinson%27s+Disease+Classification',
'datasets/Electrical+Grid+Stability+Simulated+Data+',
'datasets/Electrical+Grid+Stability+Simulated+Data+',
'datasets/Caesarian+Section+Classification+Dataset',
'datasets/Caesarian+Section+Classification+Dataset',
'datasets/BAUM-1',
'datasets/BAUM-1',
'datasets/BAUM-2',
'datasets/BAUM-2',
'datasets/Audit+Data',
'datasets/Audit+Data',
'datasets/BuddyMove+Data+Set',
'datasets/BuddyMove+Data+Set',
'datasets/Real+estate+valuation+data+set',
'datasets/Real+estate+valuation+data+set',
'datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connected+speech+Data+Set+',
'datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connected+speech+Data+Set+',
'datasets/Somerville+Happiness+Survey',
'datasets/Somerville+Happiness+Survey',
'datasets/2.4+GHZ+Indoor+Channel+Measurements',
'datasets/2.4+GHZ+Indoor+Channel+Measurements',
'datasets/EMG+data+for+gestures',
'datasets/EMG+data+for+gestures',
'datasets/Parking+Birmingham',
'datasets/Parking+Birmingham',
'datasets/Behavior+of+the+urban+traffic+of+the+city+of+Sao+Paulo+in+Brazil',
'datasets/Behavior+of+the+urban+traffic+of+the+city+of+Sao+Paulo+in+Brazil',
'datasets/Travel+Reviews',
'datasets/Travel+Reviews',
'datasets/Tarvel+Review+Ratings',
'datasets/Tarvel+Review+Ratings',
'datasets/Rice+Leaf+Diseases',
'datasets/Rice+Leaf+Diseases']

In [28]:

```
len(link_name)
#Se repiten los nombres
```

Out[28]:

938

In [29]:

```
#filtrar para que no se repitan
data_name = []
for i in link_name:
    if i not in data_name:
        data_name.append(i)
```

In [30]:

```
data_name
```

Out[30]:

['datasets/Abalone',
'datasets/Adult',
'datasets/Annealing'

'datasets/Annealing',
'datasets/Anonymous+Microsoft+Web+Data',
'datasets/Arrhythmia',
'datasets/Artificial+Characters',
'datasets/Audiology+%28Original%29',
'datasets/Audiology+%28Standardized%29',
'datasets/Auto+MPG',
'datasets/Automobile',
'datasets/Badges',
'datasets/Balance+Scale',
'datasets/Balloons',
'datasets/Breast+Cancer',
'datasets/Breast+Cancer+Wisconsin+%28Original%29',
'datasets/Breast+Cancer+Wisconsin+%28Prognostic%29',
'datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29',
'datasets/Pittsburgh+Bridges',
'datasets/Car+Evaluation',
'datasets/Census+Income',
'datasets/Chess+%28King-Rook+vs.+King-Knight%29',
'datasets/Chess+%28King-Rook+vs.+King-Pawn%29',
'datasets/Chess+%28King-Rook+vs.+King%29',
'datasets/Chess+%28Domain+Theories%29',
'datasets/Bach+Chorales',
'datasets/Connect-4',
'datasets/Credit+Approval',
'datasets/Japanese+Credit+Screening',
'datasets/Computer+Hardware',
'datasets/Contraceptive+Method+Choice',
'datasets/Covertype',
'datasets/Cylinder+Bands',
'datasets/Dermatology',
'datasets/Diabetes',
'datasets/DGP2+-+The+Second+Data+Generation+Program',
'datasets/Document+Understanding',
'datasets/EBL+Domain+Theories',
'datasets/Echocardiogram',
'datasets/Ecoli',
'datasets/Flags',
'datasets/Function+Finding',
'datasets/Glass+Identification',
'datasets/Haberman%27s+Survival',
'datasets/Hayes-Roth',
'datasets/Heart+Disease',
'datasets/Hepatitis',
'datasets/Horse+Colic',
'datasets/ICU',
'datasets/Image+Segmentation',
'datasets/Internet+Advertisements',
'datasets/Ionosphere',
'datasets/Iris',
'datasets/ISOLET',
'datasets/Kinship',
'datasets/Labor+Relations',
'datasets/LED+Display+Domain',
'datasets/Lenses',
'datasets/Letter+Recognition',
'datasets/Liver+Disorders',
'datasets/Logic+Theorist',
'datasets/Lung+Cancer',
'datasets/Lymphography',
'datasets/Mechanical+Analysis',
'datasets/Meta-data',
'datasets/Mobile+Robots',

'datasets/Mobile+Robots',
'datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29',
'datasets/Molecular+Biology+%28Protein+Secondary+Structure%29',
'datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29',
'datasets/MONK%27s+Problems',
'datasets/Moral+Reasoner',
'datasets/Multiple+Features',
'datasets/Mushroom',
'datasets/Musk+%28Version+1%29',
'datasets/Musk+%28Version+2%29',
'datasets/Nursery',
'datasets/Othello+Domain+Theory',
'datasets/Page+Blocks+Classification',
'datasets/Optical+Recognition+of+Handwritten+Digits',
'datasets/Pen-Based+Recognition+of+Handwritten+Digits',
'datasets/Post-Operative+Patient',
'datasets/Primary+Tumor',
'datasets/Prodigy',
'datasets/Qualitative+Structure+Activity+Relationships',
'datasets/Quadruped+Mammals',
'datasets/Servo',
'datasets/Shuttle+Landing+Control',
'datasets/Solar+Flare',
'datasets/Soybean+%28Large%29',
'datasets/Soybean+%28Small%29',
'datasets/Challenger+USA+Space+Shuttle+O-Ring',
'datasets/Low+Resolution+Spectrometer',
'datasets/Spambase',
'datasets/SPECT+Heart',
'datasets/SPECTF+Heart',
'datasets/Sponge',
'datasets/Statlog+Project',
'datasets/Student+Loan+Relational',
'datasets/Teaching+Assistant+Evaluation',
'datasets/Tic-Tac-Toe+Endgame',
'datasets/Thyroid+Disease',
'datasets/Trains',
'datasets/University',
'datasets/Congressional+Voting+Records',
'datasets/Water+Treatment+Plant',
'datasets/Waveform+Database+Generator+%28Version+1%29',
'datasets/Waveform+Database+Generator+%28Version+2%29',
'datasets/Wine',
'datasets/Yeast',
'datasets/Zoo',
'datasets/Undocumented',
'datasets/Twenty+Newsgroups',
'datasets/Australian+Sign+Language+signs',
'datasets/Australian+Sign+Language+signs+%28High+Quality%29',
'datasets/US+Census+Data+%281990%29',
'datasets/Census-Income+%28KDD%29',
'datasets/Coil+1999+Competition+Data',
'datasets/Corel+Image+Features',
'datasets/E.+Coli+Genes',
'datasets/EEG+Database',
'datasets/El+Nino',
'datasets/Entree+Chicago+Recommendation+Data',
'datasets/CMU+Face+Images',
'datasets/Insurance+Company+Benchmark+%28COIL+2000%29',
'datasets/Internet+Usage+Data',
'datasets/IPUMS+Census+Database',
'datasets/Japanese+Vowels',
'datasets/KDD+Cup+1998+Data',

'datasets/KDD+Cup+1998+Data',
'datasets/KDD+Cup+1999+Data',
'datasets/M.+Tuberculosis+Genes',
'datasets/Movie',
'datasets/MSNBC.com+Anonymous+Web+Data',
'datasets/NSF+Research+Award+Abstracts+1990-2003',
'datasets/Pioneer-1+Mobile+Robot+Data',
'datasets/Pseudo+Periodic+Synthetic+Time+Series',
'datasets/Reuters-21578+Text+Categorization+Collection',
'datasets/Robot+Execution+Failures',
'datasets/Synthetic+Control+Chart+Time+Series',
'datasets/Syskill+and+Webert+Web+Page+Ratings',
'datasets/UNIX+User+Data',
'datasets/Volcanoes+on+Venus+-+JARtool+experiment',
'datasets/Statlog+%28Australian+Credit+Approval%29',
'datasets/Statlog+%28German+Credit+Data%29',
'datasets/Statlog+%28Heart%29',
'datasets/Statlog+%28Landsat+Satellite%29',
'datasets/Statlog+%28Image+Segmentation%29',
'datasets/Statlog+%28Shuttle%29',
'datasets/Statlog+%28Vehicle+Silhouettes%29',
'datasets/Connectionist+Bench+%28Nettalk+Corpus%29',
'datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29',
'datasets/Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29',
'datasets/Economic+Sanctions',
'datasets/Protein+Data',
'datasets/Cloud',
'datasets/CalIt2+Building+People+Counts',
'datasets/Dodgers+Loop+Sensor',
'datasets/Poker+Hand',
'datasets/MAGIC+Gamma+Telescope',
'datasets/UJI+Pen+Characters',
'datasets/Mammographic+Mass',
'datasets/Forest+Fires',
'datasets/Reuters+Transcribed+Subset',
'datasets/Bag+of+Words',
'datasets/Concrete+Compressive+Strength',
'datasets/Hill-Valley',
'datasets/Arcene',
'datasets/Dexter',
'datasets/Dorothea',
'datasets/Gisette',
'datasets/Madelon',
'datasets/Ozone+Level+Detection',
'datasets/Abscisic+Acid+Signaling+Network',
'datasets/Parkinsons',
'datasets/Character+Trajectories',
'datasets/Blood+Transfusion+Service+Center',
'datasets/UJI+Pen+Characters+%28Version+2%29',
'datasets/Semeion+Handwritten+Digit',
'datasets/SECOM',
'datasets/Plants',
'datasets/Libras+Movement',
'datasets/Concrete+Slump+Test',
'datasets/Communities+and+Crime',
'datasets/Acute+Inflammations',
'datasets/Wine+Quality',
'datasets/URL+Reputation',
'datasets/p53+Mutants',
'datasets/Parkinsons+Telemonitoring',
'datasets/Demospongiae',
'datasets/Opinosis+Opinion+%26frasl%3B+Review',

'datasets/Breast+Tissue',
'datasets/Cardiotocography',
'datasets/Wall-Following+Robot+Navigation+Data',
'datasets/Spoken+Arabic+Digit',
'datasets/Localization+Data+for+Person+Activity',
'datasets/AutoUniv',
'datasets/Steel+Plates+Faults',
'datasets/MiniBooNE+particle+identification',
'datasets/YearPredictionMSD',
'datasets/PEMS-SF',
'datasets/OpinRank+Review+Dataset',
'datasets/Relative+location+of+CT+slices+on+axial+axis',
'datasets/Online+Handwritten+Assamese+Characters+Dataset',
'datasets/PubChem+Bioassay+Data',
'datasets/Record+Linkage+Comparison+Patterns',
'datasets/Communities+and+Crime+Unnormalized',
'datasets/Vertebral+Column',
'datasets/EMG+Physical+Action+Data+Set',
'datasets/Vicon+Physical+Action+Data+Set',
'datasets/Amazon+Commerce+reviews+set',
'datasets/Amazon+Access+Samples',
'datasets/Reuter_50_50',
'datasets/Farm+Ads',
'datasets/DBWorld+e-mails',
'datasets/KEGG+Metabolic+Relation+Network+%28Directed%29',
'datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29',
'datasets/Bank+Marketing',
'datasets/YouTube+Comedy+Slam+Preference+Data',
'datasets/Gas+Sensor+Array+Drift+Dataset',
'datasets/ILPD+%28Indian+Liver+Patient+Dataset%29',
'datasets/OPPORTUNITY+Activity+Recognition',
'datasets/Nomao',
'datasets/SMS+Spam+Collection',
'datasets/Skin+Segmentation',
'datasets/Planning+Relax',
'datasets/PAMAP2+Physical+Activity+Monitoring',
'datasets/Restaurant+%26+consumer+data',
'datasets/CNAE-9',
'datasets/Individual+household+electric+power+consumption',
'datasets/seeds',
'datasets/Northix',
'datasets/QtyT40I10D100K',
'datasets/Legal+Case+Reports',
'datasets/Human+Activity+Recognition+Using+Smartphones',
'datasets/One-hundred+plant+species+leaves+data+set',
'datasets/Energy+efficiency',
'datasets/Yacht+Hydrodynamics',
'datasets/Fertility',
'datasets/Daphnet+Freezing+of+Gait',
'datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29',
'datasets/ISTANBUL+STOCK+EXCHANGE',
'datasets/Buzz+in+social+media+',
'datasets/First-order+theorem+proving',
'datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29',
'datasets/Gas+sensor+arrays+in+open+sampling+settings',
'datasets/Climate+Model+Simulation+Crashes',
'datasets/MicroMass',
'datasets/QSAR+biodegradation',
'datasets/BLOGGER',
'datasets/Daily+and+Sports+Activities',
'datasets/User+Knowledge+Modeling',
'datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection',

'datasets/NYSK',
'datasets/Turkiye+Student+Evaluation',
'datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machines%29',
'datasets/EEG+Eye+State',
'datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure',
'datasets/seismic-bumps',
'datasets/banknote+authentication',
'datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Problem%3A+Pat',
'datasets/YouTube+Multiview+Video+Games+Dataset',
'datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations',
'datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors',
'datasets/SkillCraft1+Master+Table+Dataset',
'datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units',
'datasets/SML2010',
'datasets/Bike+Sharing+Dataset',
'datasets/Predict+keywords+activities+in+a+online+social+media',
'datasets/Thoracic+Surgery+Data',
'datasets/EMG+dataset+in+Lower+Limb',
'datasets/SUSY',
'datasets/HIGGS',
'datasets/Qualitative_Bankruptcy',
'datasets/LSVT+Voice+Rehabilitation',
'datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer',
'datasets/Wilt',
'datasets/User+Identification+From+Walking+Activity',
'datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer',
'datasets/Leaf',
'datasets/Dresses_Attribute_Sales',
'datasets/Tamilnadu+Electricity+Board+Hourly+Readings',
'datasets/Airfoil+Self-Noise',
'datasets/Wholesale+customers',
'datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis',
'datasets/Combined+Cycle+Power+Plant',
'datasets/Urban+Land+Cover',
'datasets/Diabetes+130-US+hospitals+for+years+1999-2008',
'datasets/Bach+Choral+Harmony',
'datasets/StoneFlakes',
'datasets/Tennis+Major+Tournament+Match+Statistics',
'datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings',
'datasets/Gesture+Phase+Segmentation',
'datasets/Perfume+Data',
'datasets/BlogFeedback',
'datasets/REALDISP+Activity+Recognition+Dataset',
'datasets/Newspaper+and+magazine+images+segmentation+dataset',
'datasets/AAAI+2014+Accepted+Papers',
'datasets/Gas+sensor+array+under+flow+modulation',
'datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures',
'datasets/UJIIndoorLoc',
'datasets/Sentence+Classification',
'datasets/Dow+Jones+Index',
'datasets/sEMG+for+Basic+Hand+movements',
'datasets/AAAI+2013+Accepted+Papers',
'datasets/Geographical+Original+of+Music',
'datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants',
'datasets/Grammatical+Facial+Expressions',
'datasets/NoisyOffice',
'datasets/MHEALTH+Dataset',
'datasets/Student+Performance',
'datasets/ElectricityLoadDiagrams20112014',
'datasets/Gas+sensor+array+under+dynamic+gas+mixtures',

'datasets/microblogPCU',
'datasets/Firm-Teacher_Clave-Direction_Classification',
'datasets/Dataset+for+Sensorless+Drive+Diagnosis',
'datasets/TV+News+Channel+Commercial+Detection+Dataset',
'datasets/Phishing+Websites',
'datasets/Greenhouse+Gas+Observing+Network',
'datasets/Diabetic+Retinopathy+Debrecen+Data+Set',
'datasets/HIV-1+protease+cleavage',
'datasets/Sentiment+Labelled+Sentences',
'datasets/Online+News+Popularity',
'datasets/Forest+type+mapping',
'datasets/wiki4HE',
'datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset',
'datasets/Chronic_Kidney_Disease',
'datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014',
'datasets/Folio',
'datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015',
'datasets/Cuff-Less+Blood+Pressure+Estimation',
'datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions',
'datasets/Mice+Protein+Expression',
'datasets/UJIIndoorLoc-Mag',
'datasets/Heterogeneity+Activity+Recognition',
'datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set',
'datasets/HEPMASS',
'datasets/Indoor+User+Movement+Prediction+from+RSS+data',
'datasets/Open+University+Learning+Analytics+dataset',
'datasets/default+of+credit+card+clients',
'datasets/Mesothelioma%E2%80%99s+disease+data+set+',
'datasets/Online+Retail',
'datasets/SIFT10M',
'datasets/GPS+Trajectories',
'datasets/Detect+Malacious+Executable%28AntiVirus%29',
'datasets/Occupancy+Detection+',
'datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Parkinson%E2%80%99s+D
isease',
'datasets/News+Aggregator',
'datasets/Air+Quality',
'datasets/Twin+gas+sensor+arrays',
'datasets/Gas+sensors+for+home+activity+monitoring',
'datasets/Facebook+Comment+Volume+Dataset',
'datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Living+
%28AAL%29',
'datasets/Polish+companies+bankruptcy+data',
'datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29',
'datasets/Dota2+Games+Results',
'datasets/Facebook+metrics',
'datasets/UbiqLog+%28smartphone+lifelogging%29',
'datasets/NIPS+Conference+Papers+1987-2015',
'datasets/HTRU2',
'datasets/Drug+consumption+%28quantified%29',
'datasets/Appliances+energy+prediction',
'datasets/Miskolc+IIS+Hybrid+IPS',
'datasets/KDC-4007+dataset+Collection',
'datasets/Geo-Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone',
'datasets/DrivFace',
'datasets/Website+Phishing',
'datasets/YouTube+Spam+Collection',
'datasets/Beijing+PM2.5+Data',
'datasets/Cargo+2000+Freight+Tracking+and+Tracing',
'datasets/Cervical+cancer+%28Risk+Factors%29',
'datasets/Quality+Assessment+of+Digital+Colposcopies',
'datasets/KASANDR',

'datasets/FMA%3A+A+Dataset+For+Music+Analysis',
'datasets/Air+quality',
'datasets/Epileptic+Seizure+Recognition',
'datasets/Devanagari+Handwritten+Character+Dataset',
'datasets/Stock+portfolio+performance',
'datasets/MoCap+Hand+Postures',
'datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech',
'datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting',
'datasets/PM2.5+Data+of+Five+Chinese+Cities',
'datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet',
'datasets/Sales_Transactions_Dataset_Weekly',
'datasets/Las+Vegas+Strip',
'datasets/Eco-hotel',
'datasets/MEU-Mobile+KSD',
'datasets/Crowdsourced+Mapping',
'datasets/gene+expression+cancer+RNA-Seq',
'datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer',
'datasets/chestnut+%E2%80%93+LARVIC',
'datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%28OBS%29+Ne
twork',
'datasets/Motion+Capture+Hand+Postures',
'datasets/Anuran+Calls+%28MFCCs%29',
'datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization',
'datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy',
'datasets/Daily+Demand+Forecasting+Orders',
'datasets/Paper+Reviews',
'datasets/extention+of+Z-Alizadeh+sani+dataset',
'datasets/Z-Alizadeh+Sani',
'datasets/Dynamic+Features+of+VirusShare+Executables',
'datasets/IDA2016Challenge',
'datasets/DSRC+Vehicle+Communications',
'datasets/Mturk+User-Perceived+Clusters+over+Images',
'datasets/Character+Font+Images',
'datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Instance+Labels',
'datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++',
'datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++',
'datasets/APS+Failure+at+Scania+Trucks',
'datasets/Wireless+Indoor+Localization',
'datasets/HCC+Survival',
'datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015',
'datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29',
'datasets/Autism+Screening+Adult',
'datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor',
'datasets/Immunotherapy+Dataset',
'datasets/Cryotherapy+Dataset+',
'datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes',
'datasets/Discrete+Tone+Image+Dataset',
'datasets/News+Popularity+in+Multiple+Social+Media+Platforms',
'datasets/Ultrasonic+flowmeter+diagnostics',
'datasets/ICMLA+2014+Accepted+Papers+Data+Set',
'datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation',
'datasets/Container+Crane+Controller+Data+Set',
'datasets/Residential+Building+Data+Set',
'datasets/Health+News+in+Twitter',
'datasets/chipseq',
'datasets/SGEMM+GPU+kernel+performance',
'datasets/Repeat+Consumption+Matrices',
'datasets/detection_of_IoT_botnet_attacks_N_BaIoT',
'datasets/Absenteeism+at+work',
'datasets/SCADI',
'datasets/Condition+monitoring+of+hydraulic+systems',
'datasets/Carbon+Nanotubes',

'datasets/Optical+Interconnection+Network+',
'datasets/Sports+articles+for+objectivity+analysis',
'datasets/Breast+Cancer+Coimbra',
'datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data',
'datasets/Dishonest+Internet+users+Dataset',
'datasets/Victorian+Era+Authorship+Attribution',
'datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set',
'datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing',
'datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals',
'datasets/Roman+Urdu+Data+Set',
'datasets/Avila',
'datasets/PANDOR',
'datasets/Drug+Review+Dataset+%28Druglib.com%29',
'datasets/Drug+Review+Dataset+%28Drugs.com%29',
'datasets/Physical+Unclonable+Functions',
'datasets/Superconductivty+Data',
'datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29',
'datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2',
'datasets/Student+Academics+Performance',
'datasets/Online+Shoppers+Purchasing+Intention+Dataset',
'datasets/PMU-UD',
'datasets/Parkinson%27s+Disease+Classification',
'datasets/Electrical+Grid+Stability+Simulated+Data+',
'datasets/Caesarian+Section+Classification+Dataset',
'datasets/BAUM-1',
'datasets/BAUM-2',
'datasets/Audit+Data',
'datasets/BuddyMove+Data+Set',
'datasets/Real+estate+valuation+data+set',
'datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connected+speech+Data+Set+',
'datasets/Somerville+Happiness+Survey',
'datasets/2.4+GHZ+Indoor+Channel+Measurements',
'datasets/EMG+data+for+gestures',
'datasets/Parking+Birmingham',
'datasets/Behavior+of+the+urban+traffic+of+the+city+of+Sao+Paulo+in+Brazil',
'datasets/Travel+Reviews',
'datasets/Tarvel+Review+Ratings',
'datasets/Rice+Leaf+Diseases']

In [31]:

```
len(data_name)
#En esta lista ya no se repite ningún link
```

Out[31]:

469

In [35]:

```
type(data_name)
```

Out[35]:

list

In [36]:

```
# Convertir a string
str_database = str(data_name)
str_database
```

Out[36]:

"['datasets/Abalone', 'datasets/Adult', 'datasets/Annealing', 'datasets/Anonymous+Microsoft+Web+Data', 'datasets/Arrhythmia', 'datasets/Artificial+Characters', 'datasets/Audiology+%28Original%29', 'datasets/Audiology+%28Standardized%29', 'datasets/Auto+MPG', 'datasets/Automobile', 'datasets/Badges', 'datasets/Balance+Scale', 'datasets/Balloons', 'datasets/Breast+Cancer', 'datasets/Breast+Cancer+Wisconsin+%28Original%29', 'datasets/Breast+Cancer+Wisconsin+%28Prognostic%29', 'datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29', 'datasets/Pittsburgh+Bridges', 'datasets/Car+Evaluation', 'datasets/Census+Income', 'datasets/Chess+%28King-Rook+vs.+King-Knight%29', 'datasets/Chess+%28King-Rook+vs.+King-Pawn%29', 'datasets/Chess+%28King-Rook+vs.+King%29', 'datasets/Chess+%28Domain+Theories%29', 'datasets/Bach+Chorales', 'datasets/Connect-4', 'datasets/Credit+Approval', 'datasets/Japanese+Credit+Screening', 'datasets/Computer+Hardware', 'datasets/Contraceptive+Method+Choice', 'datasets/Covertype', 'datasets/Cylinder+Bands', 'datasets/Dermatology', 'datasets/Diabetes', 'datasets/DGP2+-+The+Second+Data+Generation+Program', 'datasets/Document+Understanding', 'datasets/EBL+Domain+Theories', 'datasets/Echocardiogram', 'datasets/Ecoli', 'datasets/Flags', 'datasets/Function+Finding', 'datasets/Glass+Identification', 'datasets/Haberman%27s+Survival', 'datasets/Hayes-Roth', 'datasets/Heart+Disease', 'datasets/Hepatitis', 'datasets/Horse+Colic', 'datasets/ICU', 'datasets/Image+Segmentation', 'datasets/Internet+Advertisements', 'datasets/Ionosphere', 'datasets/Iris', 'datasets/ISOLET', 'datasets/Kinship', 'datasets/Labor+Relations', 'datasets/LED+Display+Domain', 'datasets/Lenses', 'datasets/Letter+Recognition', 'datasets/Liver+Disorders', 'datasets/Logic+Theorist', 'datasets/Lung+Cancer', 'datasets/Lymphography', 'datasets/Mechanical+Analysis', 'datasets/Meta-data', 'datasets/Mobile+Robots', 'datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29', 'datasets/Molecular+Biology+%28Protein+Secondary+Structure%29', 'datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29', 'datasets/MONK%27s+Problems', 'datasets/Moral+Reasoner', 'datasets/Multiple+Features', 'datasets/Mushroom', 'datasets/Musk+%28Version+1%29', 'datasets/Musk+%28Version+2%29', 'datasets/Nursery', 'datasets/Othello+Domain+Theory', 'datasets/Page+Blocks+Classification', 'datasets/Optical+Recognition+of+Handwritten+Digits', 'datasets/Pen-Based+Recognition+of+Handwritten+Digits', 'datasets/Post-Operative+Patient', 'datasets/Primary+Tumor', 'datasets/Prodigy', 'datasets/Qualitative+Structure+Activity+Relationships', 'datasets/Quadruped+Mammals', 'datasets/Servo', 'datasets/Shuttle+Landing+Control', 'datasets/Solar+Flare', 'datasets/Soybean+%28Large%29', 'datasets/Soybean+%28Small%29', 'datasets/Challenger+USA+Space+Shuttle+O-Ring', 'datasets/Low+Resolution+Spectrometer', 'datasets/Spambase', 'datasets/SPECT+Heart', 'datasets/SPECTF+Heart', 'datasets/Sponge', 'datasets/Statlog+Project', 'datasets/Student+Loan+Relational', 'datasets/Teaching+Assistant+Evaluation', 'datasets/Tic-Tac-Toe+Endgame', 'datasets/Thyroid+Disease', 'datasets/Trains', 'datasets/University', 'datasets/Congressional+Voting+Records', 'datasets/Water+Treatment+Plant', 'datasets/Waveform+Database+Generator+%28Version+1%29', 'datasets/Waveform+Database+Generator+%28Version+2%29', 'datasets/Wine', 'datasets/Yeast', 'datasets/Zoo', 'datasets/Undocumented', 'datasets/Twenty+Newsgroups', 'datasets/Australian+Sign+Language+signs', 'datasets/Australian+Sign+Language+signs+%28High+Quality%29', 'datasets/US+Census+Data+%281990%29', 'datasets/Census-Income+%28KDD%29', 'datasets/Coil+1999+Competition+Data', 'datasets/Corel+Image+Features', 'datasets/E.+Coli+Genes', 'datasets/EEG+Database', 'datasets/El+Nino', 'datasets/Entree+Chicago+Recommendation+Data', 'datasets/CMU+Face+Images', 'datasets/Insurance+Company+Benchmark+%28COIL+2000%29', 'datasets/Internet+Usage+Data', 'datasets/IPUMS+Census+Database', 'datasets/Japanese+Vowels', 'datasets/KDD+Cup+1998+Data', 'datasets/KDD+Cup+1999+Data', 'datasets/M.+Tuberculosis+Genes', 'datasets/Movie', 'datasets/MSNBC.com+Anonymous+Web+Data', 'datasets/NSF+Research+Award+Abstracts+1990-2003', 'datasets/Pioneer-1+Mobile+Robot+Data', 'datasets/Pseudo+Periodic+Synthetic+Time+Series', 'datasets/Reuters-21578+Text+Categorization+Collection', 'datasets/Robot+Execution+Failures', 'datasets/Synthetic+Control+Chart+Time+Series', 'datasets/Syskill+and+Webert+Web+Page+Ratings', 'datasets/UNIX+User+Data', 'datasets/Volcanoes+on+Venus+-+JARtool+experiment', 'datasets/Statlog+%28Australian+Credit+Approval%29', 'datasets/Statlog+%28German+Credit+Data%29', 'datasets/Statlog+%28Heart%29', 'datasets/Statlog+%28Landsat+Satellite%29', 'datasets/Statlog+%28Image+Segmentation%29', 'datasets/Statlog+%28Shuttle%29', 'datasets/Statlog+%28Vehicle+Silhouettes%29', 'datasets/Connectionist+Bench+%28Nettalk+Corpus%29', 'datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29', 'datasets/Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29', 'datasets/Economic+Sanctions', 'datasets/Protein+Data', 'datasets/Cloud', 'datasets/CalIt2+Building+People+Counts', 'datasets/Dodgers+Loop+Sensor', 'datasets/Poker+Hand', 'datasets/MAGIC+Gamma+Telescope', 'datasets/UJI+Pen+Characters', 'datasets/Mammographic+Mass', 'datasets/Forest+Fires', 'datasets/Reuters+Transcribed+Subset', 'datasets/Bag+of+Words', 'datasets/Concrete+Compressive+Strength', 'datasets/Hill-Valley', 'datasets/Arcene', 'datasets/Dexter', 'datasets/Dorothea', 'datasets/Gisette', 'datasets/Madelon', 'datasets/Ozone+Level+Detection', 'datasets/Abscisic+Acid+Signaling+Network', 'datasets/Parkinsons', 'datasets/Character+Trajectories', 'datasets/Blood+Transfusion+Service+Center', 'datasets/UJI+Pen+Characters+%28Version+2%29', 'datasets/Semeion+Handwritten+Digit', 'datasets/SECOM', 'datasets/Plants', 'datasets/Libras+Movement', 'datasets/Concrete+Slump+Test', 'datasets/Communities+and+Crime', 'datasets/Acute+Inflammations', 'datasets/Wine+Quality', 'datasets/URL+Reputation', 'datasets/p53+Mutants', 'datasets/Parkinsons+Telemonitoring', 'datasets/Demospongiae', 'datasets/Opinosis+Opinion+%26frasl%3B+Review', 'datasets/Breast+Tissue', 'datasets/Cardiotocography', 'datasets/Wall-Following+Robot+Navigation+Data', 'datasets/Spoken+Arabic+Digit', 'datasets/Localization+Data+for+Person+Activity', 'datasets/AutoUniv', 'datasets/Steel+Plates+Faults', 'datasets/MiniBooNE+particle+i

+Data+for+Person+Activity', 'datasets/AutoUniv', 'datasets/Steel+Plates+Faults', 'datasets/MiniBooNE+particle+identification', 'datasets/YearPredictionMSD', 'datasets/PEMS-SF', 'datasets/OpinRank+Review+Dataset', 'datasets/Relative+location+of+CT+slices+on+axial+axis', 'datasets/Online+Handwritten+Assamese+Characters+Dataset', 'datasets/PubChem+Bioassay+Data', 'datasets/Record+Linkage+Comparison+Patterns', 'datasets/Communities+and+Crime+Unnormalized', 'datasets/Vertebral+Column', 'datasets/EMG+Physical+Action+Data+Set', 'datasets/Vicon+Physical+Action+Data+Set', 'datasets/Amazon+Commerce+reviews+set', 'datasets/Amazon+Access+Samples', 'datasets/Reuter_50_50', 'datasets/Farm+Ads', 'datasets/DBWorld+e-mails', 'datasets/KEGG+Metabolic+Relation+Network+%28Directed%29', 'datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29', 'datasets/Bank+Marketing', 'datasets/YouTube+Comedy+Slam+Preference+Data', 'datasets/Gas+Sensor+Array+Drift+Dataset', 'datasets/ILPD+%28Indian+Liver+Patient+Dataset%29', 'datasets/OPPORTUNITY+Activity+Recognition', 'datasets/Nomao', 'datasets/SMS+Spam+Collection', 'datasets/Skin+Segmentation', 'datasets/Planning+Relax', 'datasets/PAMAP2+Physical+Activity+Monitoring', 'datasets/Restaurant+%26+consumer+data', 'datasets/CNAE-9', 'datasets/Individual+household+electric+power+consumption', 'datasets/seeds', 'datasets/Northix', 'datasets/QtyT40I10D100K', 'datasets/Legal+Case+Reports', 'datasets/Human+Activity+Recognition+Using+Smartphones', 'datasets/One-hundred+plant+species+leaves+data+set', 'datasets/Energy+efficiency', 'datasets/Yacht+Hydrodynamics', 'datasets/Fertility', 'datasets/Daphnet+Freezing+of+Gait', 'datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29', 'datasets/ISTANBUL+STOCK+EXCHANGE', 'datasets/Buzz+in+social+media+', 'datasets/First-order+theorem+proving', 'datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29', 'datasets/Gas+sensor+arrays+in+open+sampling+settings', 'datasets/Climate+Model+Simulation+Crashes', 'datasets/MicroMass', 'datasets/QSAR+biodegradation', 'datasets/BLOGGER', 'datasets/Daily+and+Sports+Activities', 'datasets/User+Knowledge+Modeling', 'datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection', 'datasets/NYSK', 'datasets/Turkiye+Student+Evaluation', 'datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machines%29', 'datasets/EEG+Eye+State', 'datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure', 'datasets/seismic-bumps', 'datasets/banknote+authentication', 'datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Problem%3A+Pat', 'datasets/YouTube+Multiview+Video+Games+Dataset', 'datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations', 'datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors', 'datasets/SkillCraft1+Master+Table+Dataset', 'datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units', 'datasets/SML2010', 'datasets/Bike+Sharing+Dataset', 'datasets/Predict+keywords+activities+in+a+online+social+media', 'datasets/Thoracic+Surgery+Data', 'datasets/EMG+dataset+in+Lower+Limb', 'datasets/SUSY', 'datasets/HIGGS', 'datasets/Qualitative_Bankruptcy', 'datasets/LSVT+Voice+Rehabilitation', 'datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer', 'datasets/Wilt', 'datasets/User+Identification+From+Walking+Activity', 'datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer', 'datasets/Leaf', 'datasets/Dresses_Attribute_Sales', 'datasets/Tamilnadu+Electricity+Board+Hourly+Readings', 'datasets/Airfoil+Self-Noise', 'datasets/Wholesale+customers', 'datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis', 'datasets/Combined+Cycle+Power+Plant', 'datasets/Urban+Land+Cover', 'datasets/Diabetes+130-US+hospitals+for+years+1999-2008', 'datasets/Bach+Choral+Harmony', 'datasets/StoneFlakes', 'datasets/Tennis+Major+Tournament+Match+Statistics', 'datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings', 'datasets/Gesture+Phase+Segmentation', 'datasets/Perfume+Data', 'datasets/BlogFeedback', 'datasets/REALDISP+Activity+Recognition+Dataset', 'datasets/Newspaper+and+magazine+images+segmentation+dataset', 'datasets/AAAI+2014+Accepted+Papers', 'datasets/Gas+sensor+array+under+flow+modulation', 'datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures', 'datasets/UJIIndoorLoc', 'datasets/Sentence+Classification', 'datasets/Dow+Jones+Index', 'datasets/sEMG+for+Basic+Hand+movements', 'datasets/AAAI+2013+Accepted+Papers', 'datasets/Geographical+Original+of+Music', 'datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants', 'datasets/Grammatical+Facial+Expressions', 'datasets/NoisyOffice', 'datasets/MHEALTH+Dataset', 'datasets/Student+Performance', 'datasets/ElectricityLoadDiagrams20112014', 'datasets/Gas+sensor+array+under+dynamic+gas+mixtures', 'datasets/microblogPCU', 'datasets/Firm-Teacher_Clave-Direction_Classification', 'datasets/Dataset+for+Sensorless+Drive+Diagnosis', 'datasets/TV+News+Channel+Commercial+Detection+Dataset', 'datasets/Phishing+Websites', 'datasets/Greenhouse+Gas+Observing+Network', 'datasets/Diabetic+Retinopathy+Debrecen+Data+Set', 'datasets/HIV-1+protease+cleavage', 'datasets/Sentiment+Labelled+Sentences', 'datasets/Online+News+Popularity', 'datasets/Forest+type+mapping', 'datasets/wiki4HE', 'datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset', 'datasets/Chronic_Kidney_Disease', 'datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014', 'datasets/Folio', 'datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015', 'datasets/Cuff-Less+Blood+Pressure+Estimation', 'datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions', 'datasets/Mice+Protein+Expression', 'datasets/UJIIndoorLoc-Mag', 'datasets/Heterogeneity+Activity+Recognition', 'datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set', 'datasets/HEPMASS', 'datasets/Indoor+User+Movement+Prediction+from+RSS+data', 'datasets/Open+University+Learning+Analytics+dataset', 'datasets/default+of+credit+card+clients', 'datasets/Mesothelioma%E2%80%99s+disease+data+set+', 'datasets/Online+Retail', 'datasets/SIFT10M', 'datasets/GPS+Trajectories', 'datasets/Detect+Malacious+Executable%28AntiVirus%29', 'datasets/Occupancy+Detection+', 'datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monito

ring+Parkinson%E2%80%99s+Disease', 'datasets/News+Aggregator', 'datasets/Air+Quality', 'datasets/Twin+gas+sensor+arrays', 'datasets/Gas+sensors+for+home+activity+monitoring', 'datasets/Facebook+Comment+Volume+Dataset', 'datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Living+%28AAL%29', 'datasets/Polish+companies+bankruptcy+data', 'datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29', 'datasets/Dota2+Games+Results', 'datasets/Facebook+metrics', 'datasets/UbiqLog+%28smartphone+lifelogging%29', 'datasets/NIPS+Conference+Papers+1987-2015', 'datasets/HTRU2', 'datasets/Drug+consumption+%28quantified%29', 'datasets/Appliances+energy+prediction', 'datasets/Miskolc+IIS+Hybrid+IPS', 'datasets/KDC-4007+dataset+Collection', 'datasets/Geo-Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone', 'datasets/DrivFace', 'datasets/Website+Phishing', 'datasets/YouTube+Spam+Collection', 'datasets/Beijing+PM2.5+Data', 'datasets/Cargo+2000+Freight+Tracking+and+Tracing', 'datasets/Cervical+cancer+%28Risk+Factors%29', 'datasets/Quality+Assessment+of+Digital+Colposcopies', 'datasets/KASANDR', 'datasets/FMA%3A+A+Dataset+For+Music+Analysis', 'datasets/Air+quality', 'datasets/Epileptic+Seizure+Recognition', 'datasets/Devanagari+Handwritten+Character+Dataset', 'datasets/Stock+portfolio+performance', 'datasets/MoCap+Hand+Postures', 'datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech', 'datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting', 'datasets/PM2.5+Data+of+Five+Chinese+Cities', 'datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet', 'datasets/Sales_Transactions_Dataset_Weekly', 'datasets/Las+Vegas+Strip', 'datasets/Eco-hotel', 'datasets/MEU-Mobile+KSD', 'datasets/Crowdsourced+Mapping', 'datasets/gene+expression+cancer+RNA-Seq', 'datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer', 'datasets/chestnut+%E2%80%93+LARVIC', 'datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%28OBS%29+Network', 'datasets/Motion+Capture+Hand+Postures', 'datasets/Anuran+Calls+%28MFCCs%29', 'datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization', 'datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy', 'datasets/Daily+Demand+Forecasting+Orders', 'datasets/Paper+Reviews', 'datasets/extention+of+Z-Alizadeh+sani+dataset', 'datasets/Z-Alizadeh+Sani', 'datasets/Dynamic+Features+of+VirusShare+Executables', 'datasets/IDA2016Challenge', 'datasets/DSRC+Vehicle+Communications', 'datasets/Mturk+User-Perceived+Clusters+over+Images', 'datasets/Character+Font+Images', 'datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Instance+Labels', 'datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++', 'datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++', 'datasets/APS+Failure+at+Scania+Trucks', 'datasets/Wireless+Indoor+Localization', 'datasets/HCC+Survival', 'datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015', 'datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29', 'datasets/Autism+Screening+Adult', 'datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor', 'datasets/Immunotherapy+Dataset', 'datasets/Cryotherapy+Dataset+', 'datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes', 'datasets/Discrete+Tone+Image+Dataset', 'datasets/News+Popularity+in+Multiple+Social+Media+Platforms', 'datasets/Ultrasonic+flowmeter+diagnostics', 'datasets/ICMLA+2014+Accepted+Papers+Data+Set', 'datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation', 'datasets/Container+Crane+Controller+Data+Set', 'datasets/Residential+Building+Data+Set', 'datasets/Health+News+in+Twitter', 'datasets/chipseq', 'datasets/SGEMM+GPU+kernel+performance', 'datasets/Repeat+Consumption+Matrices', 'datasets/detection_of_IoT_botnet_attacks_N_BaIoT', 'datasets/Absenteeism+at+work', 'datasets/SCADI', 'datasets/Condition+monitoring+of+hydraulic+systems', 'datasets/Carbon+Nanotubes', 'datasets/Optical+Interconnection+Network+', 'datasets/Sports+articles+for+objectivity+analysis', 'datasets/Breast+Cancer+Coimbra', 'datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data', 'datasets/Dishonest+Internet+users+Dataset', 'datasets/Victorian+Era+Authorship+Attribution', 'datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set', 'datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing', 'datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals', 'datasets/Roman+Urdu+Data+Set', 'datasets/Avila', 'datasets/PANDOR', 'datasets/Drug+Review+Dataset+%28Druglib.com%29', 'datasets/Drug+Review+Dataset+%28Drugs.com%29', 'datasets/Physical+Unclonable+Functions', 'datasets/Superconductivty+Data', 'datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29', 'datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2', 'datasets/Student+Academics+Performance', 'datasets/Online+Shoppers+Purchasing+Intention+Dataset', 'datasets/PMU-UD', 'datasets/Parkinson%27s+Disease+Classification', 'datasets/Electrical+Grid+Stability+Simulated+Data+', 'datasets/Caesarian+Section+Classification+Dataset', 'datasets/BAUM-1', 'datasets/BAUM-2', 'datasets/Audit+Data', 'datasets/BuddyMove+Data+Set', 'datasets/Real+estate+valuation+data+set', 'datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connected+speech+Data+Set+', 'datasets/Somerville+Happiness+Survey', 'datasets/2.4+GHZ+Indoor+Channel+Measurements', 'datasets/EMG+data+for+gestures', 'datasets/Parking+Birmingham', 'datasets/Behavior+of+the+urban+traffic+of+the+city+of+Sao+Paulo+in+Brazil', 'datasets/Travel+Reviews', 'datasets/Tarvel+Review+Ratings', 'datasets/Rice+Leaf+Diseases']"

In [37]:

```
import re
```

In [45]:

```
name_database = re.findall("\/(\S+)\',",str_database)
name_database
```

Out[45]:

```
['Abalone',
 'Adult',
 'Annealing',
 'Anonymous+Microsoft+Web+Data',
 'Arrhythmia',
 'Artificial+Characters',
 'Audiology+%28Original%29',
 'Audiology+%28Standardized%29',
 'Auto+MPG',
 'Automobile',
 'Badges',
 'Balance+Scale',
 'Balloons',
 'Breast+Cancer',
 'Breast+Cancer+Wisconsin+%28Original%29',
 'Breast+Cancer+Wisconsin+%28Prognostic%29',
 'Breast+Cancer+Wisconsin+%28Diagnostic%29',
 'Pittsburgh+Bridges',
 'Car+Evaluation',
 'Census+Income',
 'Chess+%28King-Rook+vs.+King-Knight%29',
 'Chess+%28King-Rook+vs.+King-Pawn%29',
 'Chess+%28King-Rook+vs.+King%29',
 'Chess+%28Domain+Theories%29',
 'Bach+Chorales',
 'Connect-4',
 'Credit+Approval',
 'Japanese+Credit+Screening',
 'Computer+Hardware',
 'Contraceptive+Method+Choice',
 'Covertype',
 'Cylinder+Bands',
 'Dermatology',
 'Diabetes',
 'DGP2+-+The+Second+Data+Generation+Program',
 'Document+Understanding',
 'EBL+Domain+Theories',
 'Echocardiogram',
 'Ecoli',
 'Flags',
 'Function+Finding',
 'Glass+Identification',
 'Haberman%27s+Survival',
 'Hayes-Roth',
 'Heart+Disease',
 'Hepatitis',
 'Horse+Colic',
 'ICU',
 'Image+Segmentation',
 'Internet+Advertisements',
 'Ionosphere',
 'Iris',
 'ISOLET',
 'Kinship',
 'Labor+Relations',
 'LED+Display+Domain',
```

'Lenses',
'Letter+Recognition',
'Liver+Disorders',
'Logic+Theorist',
'Lung+Cancer',
'Lymphography',
'Mechanical+Analysis',
'Meta-data',
'Mobile+Robots',
'Molecular+Biology+%28Promoter+Gene+Sequences%29',
'Molecular+Biology+%28Protein+Secondary+Structure%29',
'Molecular+Biology+%28Splice-junction+Gene+Sequences%29',
'MONK%27s+Problems',
'Moral+Reasoner',
'Multiple+Features',
'Mushroom',
'Musk+%28Version+1%29',
'Musk+%28Version+2%29',
'Nursery',
'Othello+Domain+Theory',
'Page+Blocks+Classification',
'Optical+Recognition+of+Handwritten+Digits',
'Pen-Based+Recognition+of+Handwritten+Digits',
'Post-Operative+Patient',
'Primary+Tumor',
'Prodigy',
'Qualitative+Structure+Activity+Relationships',
'Quadruped+Mammals',
'Servo',
'Shuttle+Landing+Control',
'Solar+Flare',
'Soybean+%28Large%29',
'Soybean+%28Small%29',
'Challenger+USA+Space+Shuttle+O-Ring',
'Low+Resolution+Spectrometer',
'Spambase',
'SPECT+Heart',
'SPECTF+Heart',
'Sponge',
'Statlog+Project',
'Student+Loan+Relational',
'Teaching+Assistant+Evaluation',
'Tic-Tac-Toe+Endgame',
'Thyroid+Disease',
'Trains',
'University',
'Congressional+Voting+Records',
'Water+Treatment+Plant',
'Waveform+Database+Generator+%28Version+1%29',
'Waveform+Database+Generator+%28Version+2%29',
'Wine',
'Yeast',
'Zoo',
'Undocumented',
'Twenty+Newsgroups',
'Australian+Sign+Language+signs',
'Australian+Sign+Language+signs+%28High+Quality%29',
'US+Census+Data+%281990%29',
'Census-Income+%28KDD%29',
'Coil+1999+Competition+Data',
'Corel+Image+Features',
'E.+Coli+Genes'

'EEG+Database',
'El+Nino',
'Entree+Chicago+Recommendation+Data',
'CMU+Face+Images',
'Insurance+Company+Benchmark+%28COIL+2000%29',
'Internet+Usage+Data',
'IPUMS+Census+Database',
'Japanese+Vowels',
'KDD+Cup+1998+Data',
'KDD+Cup+1999+Data',
'M.+Tuberculosis+Genes',
'Movie',
'MSNBC.com+Anonymous+Web+Data',
'NSF+Research+Award+Abstracts+1990-2003',
'Pioneer-1+Mobile+Robot+Data',
'Pseudo+Periodic+Synthetic+Time+Series',
'Reuters-21578+Text+Categorization+Collection',
'Robot+Execution+Failures',
'Synthetic+Control+Chart+Time+Series',
'Syskill+and+Webert+Web+Page+Ratings',
'UNIX+User+Data',
'Volcanoes+on+Venus+-+JARtool+experiment',
'Statlog+%28Australian+Credit+Approval%29',
'Statlog+%28German+Credit+Data%29',
'Statlog+%28Heart%29',
'Statlog+%28Landsat+Satellite%29',
'Statlog+%28Image+Segmentation%29',
'Statlog+%28Shuttle%29',
'Statlog+%28Vehicle+Silhouettes%29',
'Connectionist+Bench+%28Nettalk+Corpus%29',
'Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29',
'Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29',
'Economic+Sanctions',
'Protein+Data',
'Cloud',
'CalIt2+Building+People+Counts',
'Dodgers+Loop+Sensor',
'Poker+Hand',
'MAGIC+Gamma+Telescope',
'UJI+Pen+Characters',
'Mammographic+Mass',
'Forest+Fires',
'Reuters+Transcribed+Subset',
'Bag+of+Words',
'Concrete+Compressive+Strength',
'Hill-Valley',
'Arcene',
'Dexter',
'Dorothea',
'Gisette',
'Madelon',
'Ozone+Level+Detection',
'Abscisic+Acid+Signaling+Network',
'Parkinsons',
'Character+Trajectories',
'Blood+Transfusion+Service+Center',
'UJI+Pen+Characters+%28Version+2%29',
'Semeion+Handwritten+Digit',
'SECOM',
'Plants',
'Libras+Movement',
'Concrete+Slump+Test'

'Concrete+Slump+Test',
'Communities+and+Crime',
'Acute+Inflammations',
'Wine+Quality',
'URL+Reputation',
'p53+Mutants',
'Parkinsons+Telemonitoring',
'Demospongiae',
'Opinosis+Opinion+%26frasl%3B+Review',
'Breast+Tissue',
'Cardiotocography',
'Wall-Following+Robot+Navigation+Data',
'Spoken+Arabic+Digit',
'Localization+Data+for+Person+Activity',
'AutoUniv',
'Steel+Plates+Faults',
'MiniBooNE+particle+identification',
'YearPredictionMSD',
'PEMS-SF',
'OpinRank+Review+Dataset',
'Relative+location+of+CT+slices+on+axial+axis',
'Online+Handwritten+Assamese+Characters+Dataset',
'PubChem+Bioassay+Data',
'Record+Linkage+Comparison+Patterns',
'Communities+and+Crime+Unnormalized',
'Vertebral+Column',
'EMG+Physical+Action+Data+Set',
'Vicon+Physical+Action+Data+Set',
'Amazon+Commerce+reviews+set',
'Amazon+Access+Samples',
'Reuter_50_50',
'Farm+Ads',
'DBWorld+e-mails',
'KEGG+Metabolic+Relation+Network+%28Directed%29',
'KEGG+Metabolic+Reaction+Network+%28Undirected%29',
'Bank+Marketing',
'YouTube+Comedy+Slam+Preference+Data',
'Gas+Sensor+Array+Drift+Dataset',
'ILPD+%28Indian+Liver+Patient+Dataset%29',
'OPPORTUNITY+Activity+Recognition',
'Nomao',
'SMS+Spam+Collection',
'Skin+Segmentation',
'Planning+Relax',
'PAMAP2+Physical+Activity+Monitoring',
'Restaurant+%26+consumer+data',
'CNAE-9',
'Individual+household+electric+power+consumption',
'seeds',
'Northix',
'QtyT40I10D100K',
'Legal+Case+Reports',
'Human+Activity+Recognition+Using+Smartphones',
'One-hundred+plant+species+leaves+data+set',
'Energy+efficiency',
'Yacht+Hydrodynamics',
'Fertility',
'Daphnet+Freezing+of+Gait',
'3D+Road+Network+%28North+Jutland%2C+Denmark%29',
'ISTANBUL+STOCK+EXCHANGE',
'Buzz+in+social+media+',
'First-order+theorem+proving',
'Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29',

'Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29',
'Gas+sensor+arrays+in+open+sampling+settings',
'Climate+Model+Simulation+Crashes',
'MicroMass',
'QSAR+biodegradation',
'BLOGGER',
'Daily+and+Sports+Activities',
'User+Knowledge+Modeling',
'Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection',
'NYSK',
'Turkiye+Student+Evaluation',
'ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machines%29',
'EEG+Eye+State',
'Physicochemical+Properties+of+Protein+Tertiary+Structure',
'seismic-bumps',
'banknote+authentication',
'USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Problem%3A+Pat',
'YouTube+Multiview+Video+Games+Dataset',
'Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations',
'Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors',
'SkillCraft1+Master+Table+Dataset',
'Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units',
'SML2010',
'Bike+Sharing+Dataset',
'Predict+keywords+activities+in+a+online+social+media',
'Thoracic+Surgery+Data',
'EMG+dataset+in+Lower+Limb',
'SUSY',
'HIGGS',
'Qualitative_Bankruptcy',
'LSVT+Voice+Rehabilitation',
'Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer',
'Wilt',
'User+Identification+From+Walking+Activity',
'Activity+Recognition+from+Single+Chest-Mounted+Accelerometer',
'Leaf',
'Dresses_Attribute_Sales',
'Tamilnadu+Electricity+Board+Hourly+Readings',
'Airfoil+Self-Noise',
'Wholesale+customers',
'Twitter+Data+set+for+Arabic+Sentiment+Analysis',
'Combined+Cycle+Power+Plant',
'Urban+Land+Cover',
'Diabetes+130-US+hospitals+for+years+1999-2008',
'Bach+Choral+Harmony',
'StoneFlakes',
'Tennis+Major+Tournament+Match+Statistics',
'Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings',
'Gesture+Phase+Segmentation',
'Perfume+Data',
'BlogFeedback',
'REALDISP+Activity+Recognition+Dataset',
'Newspaper+and+magazine+images+segmentation+dataset',
'AAAI+2014+Accepted+Papers',
'Gas+sensor+array+under+flow+modulation',
'Gas+sensor+array+exposed+to+turbulent+gas+mixtures',
'UJIIndoorLoc',
'Sentence+Classification',
'Dow+Jones+Index',
'sEMG+for+Basic+Hand+movements',
'AAAI+2013+Accepted+Papers',
'Geographical+Original+of+Music'

'Geographical+Original+of+Music',
'Condition+Based+Maintenance+of+Naval+Propulsion+Plants',
'Grammatical+Facial+Expressions',
'NoisyOffice',
'MHEALTH+Dataset',
'Student+Performance',
'ElectricityLoadDiagrams20112014',
'Gas+sensor+array+under+dynamic+gas+mixtures',
'microblogPCU',
'Firm-Teacher_Clave-Direction_Classification',
'Dataset+for+Sensorless+Drive+Diagnosis',
'TV+News+Channel+Commercial+Detection+Dataset',
'Phishing+Websites',
'Greenhouse+Gas+Observing+Network',
'Diabetic+Retinopathy+Debrecen+Data+Set',
'HIV-1+protease+cleavage',
'Sentiment+Labelled+Sentences',
'Online+News+Popularity',
'Forest+type+mapping',
'wiki4HE',
'Online+Video+Characteristics+and+Transcoding+Time+Dataset',
'Chronic_Kidney_Disease',
'Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014',
'Folio',
'Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015',
'Cuff-Less+Blood+Pressure+Estimation',
'Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions',
'Mice+Protein+Expression',
'UJIIndoorLoc-Mag',
'Heterogeneity+Activity+Recognition',
'Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set',
'HEPMASS',
'Indoor+User+Movement+Prediction+from+RSS+data',
'Open+University+Learning+Analytics+dataset',
'default+of+credit+card+clients',
'Mesothelioma%E2%80%99s+disease+data+set+',
'Online+Retail',
'SIFT10M',
'GPS+Trajectories',
'Detect+Malacious+Executable%28AntiVirus%29',
'Occupancy+Detection+',
'Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Parkinson%E2%80%99s+Disease',
'News+Aggregator',
'Air+Quality',
'Twin+gas+sensor+arrays',
'Gas+sensors+for+home+activity+monitoring',
'Facebook+Comment+Volume+Dataset',
'Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Living+%28AAL%29',
'Polish+companies+bankruptcy+data',
'Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29',
'Dota2+Games+Results',
'Facebook+metrics',
'UbiqLog+%28smartphone+lifelogging%29',
'NIPS+Conference+Papers+1987-2015',
'HTRU2',
'Drug+consumption+%28quantified%29',
'Appliances+energy+prediction',
'Miskolc+IIS+Hybrid+IPS',
'KDC-4007+dataset+Collection',
'Geo-Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone',
'DrivFace',

'Website+Phishing',
'YouTube+Spam+Collection',
'Beijing+PM2.5+Data',
'Cargo+2000+Freight+Tracking+and+Tracing',
'Cervical+cancer+%28Risk+Factors%29',
'Quality+Assessment+of+Digital+Colposcopies',
'KASANDR',
'FMA%3A+A+Dataset+For+Music+Analysis',
'Air+quality',
'Epileptic+Seizure+Recognition',
'Devanagari+Handwritten+Character+Dataset',
'Stock+portfolio+performance',
'MoCap+Hand+Postures',
'Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech',
'Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting',
'PM2.5+Data+of+Five+Chinese+Cities',
'Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet',
'Sales_Transactions_Dataset_Weekly',
'Las+Vegas+Strip',
'Eco-hotel',
'MEU-Mobile+KSD',
'Crowdsourced+Mapping',
'gene+expression+cancer+RNA-Seq',
'Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer',
'chestnut+%E2%80%93+LARVIC',
'Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%28OBS%29+Network',
'Motion+Capture+Hand+Postures',
'Anuran+Calls+%28MFCCs%29',
'TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization',
'Gastrointestinal+Lesions+in+Regular+Colonoscopy',
'Daily+Demand+Forecasting+Orders',
'Paper+Reviews',
'extention+of+Z-Alizadeh+sani+dataset',
'Z-Alizadeh+Sani',
'Dynamic+Features+of+VirusShare+Executables',
'IDA2016Challenge',
'DSRC+Vehicle+Communications',
'Mturk+User-Perceived+Clusters+over+Images',
'Character+Font+Images',
'DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Instance+Labels',
'Autistic+Spectrum+Disorder+Screening+Data+for+Children++',
'Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++',
'APS+Failure+at+Scania+Trucks',
'Wireless+Indoor+Localization',
'HCC+Survival',
'CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015',
'University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29',
'Autism+Screening+Adult',
'Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor',
'Immunotherapy+Dataset',
'Cryotherapy+Dataset+',
'OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes',
'Discrete+Tone+Image+Dataset',
'News+Popularity+in+Multiple+Social+Media+Platforms',
'Ultrasonic+flowmeter+diagnostics',
'ICMLA+2014+Accepted+Papers+Data+Set',
'BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation',
'Container+Crane+Controller+Data+Set',
'Residential+Building+Data+Set',
'Health+News+in+Twitter',
'chipseq',
'SGEMM+GPU+kernel+performance',

'Repeat+Consumption+Matrices',
'detection_of_IoT_botnet_attacks_N_BaIoT',
'Absenteeism+at+work',
'SCADI',
'Condition+monitoring+of+hydraulic+systems',
'Carbon+Nanotubes',
'Optical+Interconnection+Network+',
'Sports+articles+for+objectivity+analysis',
'Breast+Cancer+Coimbra',
'GNFUV+Unmanned+Surface+Vehicles+Sensor+Data',
'Dishonest+Internet+users+Dataset',
'Victorian+Era+Authorship+Attribution',
'Simulated+Falls+and+Daily+Living+Activities+Data+Set',
'Multimodal+Damage+Identification+for+Humanitarian+Computing',
'EEG+Steady-State+Visual+Evoked+Potential+Signals',
'Roman+Urdu+Data+Set',
'Avila',
'PANDOR',
'Drug+Review+Dataset+%28Druglib.com%29',
'Drug+Review+Dataset+%28Drugs.com%29',
'Physical+Unclonable+Functions',
'Superconductivty+Data',
'WESAD+%28Wearable+Stress+and+Affect+Detection%29',
'GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2',
'Student+Academics+Performance',
'Online+Shoppers+Purchasing+Intention+Dataset',
'PMU-UD',
'Parkinson%27s+Disease+Classification',
'Electrical+Grid+Stability+Simulated+Data+',
'Caesarian+Section+Classification+Dataset',
'BAUM-1',
'BAUM-2',
'Audit+Data',
'BuddyMove+Data+Set',
'Real+estate+valuation+data+set',
'Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connected+speech+Data+Set+',
'Somerville+Happiness+Survey',
'2.4+GHZ+Indoor+Channel+Measurements',
'EMG+data+for+gestures',
'Parking+Birmingham',
'Behavior+of+the+urban+traffic+of+the+city+of+Sao+Paulo+in+Brazil',
'Travel+Reviews',
'Tarvel+Review+Ratings']

In [46]:

```
len(name_database)
```

Out[46]:

468

In [47]:

```
type(name_database)
```

Out[47]:

list

In [49]:

```
#Todos los links a las bases de datos tienen: https://archive.ics.uci.edu/ml/datasets/ + nombre de la base
```

```
#La lista de links será la unión de ambas partes
link_database = []
for database in name_database:
    link_database.append("https://archive.ics.uci.edu/ml/datasets/" + str(database))
```

In [50]:

```
link_database
```

Out[50]:

```
['https://archive.ics.uci.edu/ml/datasets/Abalone',
 'https://archive.ics.uci.edu/ml/datasets/Adult',
 'https://archive.ics.uci.edu/ml/datasets/Annealing',
 'https://archive.ics.uci.edu/ml/datasets/Anonymous+Microsoft+Web+Data',
 'https://archive.ics.uci.edu/ml/datasets/Arrhythmia',
 'https://archive.ics.uci.edu/ml/datasets/Artificial+Characters',
 'https://archive.ics.uci.edu/ml/datasets/Audiology+%28Original%29',
 'https://archive.ics.uci.edu/ml/datasets/Audiology+%28Standardized%29',
 'https://archive.ics.uci.edu/ml/datasets/Auto+MPG',
 'https://archive.ics.uci.edu/ml/datasets/Automobile',
 'https://archive.ics.uci.edu/ml/datasets/Badges',
 'https://archive.ics.uci.edu/ml/datasets/Balance+Scale',
 'https://archive.ics.uci.edu/ml/datasets/Balloons',
 'https://archive.ics.uci.edu/ml/datasets/Breast+Cancer',
 'https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29',
 'https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Prognostic%29',
 'https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29',
 'https://archive.ics.uci.edu/ml/datasets/Pittsburgh+Bridges',
 'https://archive.ics.uci.edu/ml/datasets/Car+Evaluation',
 'https://archive.ics.uci.edu/ml/datasets/Census+Income',
 'https://archive.ics.uci.edu/ml/datasets/Chess+%28King-Rook+vs.+King-Knight%29',
 'https://archive.ics.uci.edu/ml/datasets/Chess+%28King-Rook+vs.+King-Pawn%29',
 'https://archive.ics.uci.edu/ml/datasets/Chess+%28King-Rook+vs.+King%29',
 'https://archive.ics.uci.edu/ml/datasets/Chess+%28Domain+Theories%29',
 'https://archive.ics.uci.edu/ml/datasets/Bach+Chorales',
 'https://archive.ics.uci.edu/ml/datasets/Connect-4',
 'https://archive.ics.uci.edu/ml/datasets/Credit+Approval',
 'https://archive.ics.uci.edu/ml/datasets/Japanese+Credit+Screening',
 'https://archive.ics.uci.edu/ml/datasets/Computer+Hardware',
 'https://archive.ics.uci.edu/ml/datasets/Contraceptive+Method+Choice',
 'https://archive.ics.uci.edu/ml/datasets/Covertype',
 'https://archive.ics.uci.edu/ml/datasets/Cylinder+Bands',
 'https://archive.ics.uci.edu/ml/datasets/Dermatology',
 'https://archive.ics.uci.edu/ml/datasets/Diabetes',
 'https://archive.ics.uci.edu/ml/datasets/DGP2+-+The+Second+Data+Generation+Program',
 'https://archive.ics.uci.edu/ml/datasets/Document+Understanding',
 'https://archive.ics.uci.edu/ml/datasets/EBL+Domain+Theories',
 'https://archive.ics.uci.edu/ml/datasets/Echocardiogram',
 'https://archive.ics.uci.edu/ml/datasets/Ecoli',
 'https://archive.ics.uci.edu/ml/datasets/Flags',
 'https://archive.ics.uci.edu/ml/datasets/Function+Finding',
 'https://archive.ics.uci.edu/ml/datasets/Glass+Identification',
 'https://archive.ics.uci.edu/ml/datasets/Haberman%27s+Survival',
 'https://archive.ics.uci.edu/ml/datasets/Hayes-Roth',
 'https://archive.ics.uci.edu/ml/datasets/Heart+Disease',
 'https://archive.ics.uci.edu/ml/datasets/Hepatitis',
 'https://archive.ics.uci.edu/ml/datasets/Horse+Colic',
 'https://archive.ics.uci.edu/ml/datasets/ICU',
 'https://archive.ics.uci.edu/ml/datasets/Image+Segmentation',
 'https://archive.ics.uci.edu/ml/datasets/Internet+Advertisements',
 'https://archive.ics.uci.edu/ml/datasets/Ionosphere',
 'https://archive.ics.uci.edu/ml/datasets/Iris',
```

'https://archive.ics.uci.edu/ml/datasets/ISOLET',
'https://archive.ics.uci.edu/ml/datasets/Kinship',
'https://archive.ics.uci.edu/ml/datasets/Labor+Relations',
'https://archive.ics.uci.edu/ml/datasets/LED+Display+Domain',
'https://archive.ics.uci.edu/ml/datasets/Lenses',
'https://archive.ics.uci.edu/ml/datasets/Letter+Recognition',
'https://archive.ics.uci.edu/ml/datasets/Liver+Disorders',
'https://archive.ics.uci.edu/ml/datasets/Logic+Theorist',
'https://archive.ics.uci.edu/ml/datasets/Lung+Cancer',
'https://archive.ics.uci.edu/ml/datasets/Lymphography',
'https://archive.ics.uci.edu/ml/datasets/Mechanical+Analysis',
'https://archive.ics.uci.edu/ml/datasets/Meta-data',
'https://archive.ics.uci.edu/ml/datasets/Mobile+Robots',
'https://archive.ics.uci.edu/ml/datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29',
'https://archive.ics.uci.edu/ml/datasets/Molecular+Biology+%28Protein+Secondary+Structure%29',
'https://archive.ics.uci.edu/ml/datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29',
'https://archive.ics.uci.edu/ml/datasets/MONK%27s+Problems',
'https://archive.ics.uci.edu/ml/datasets/Moral+Reasoner',
'https://archive.ics.uci.edu/ml/datasets/Multiple+Features',
'https://archive.ics.uci.edu/ml/datasets/Mushroom',
'https://archive.ics.uci.edu/ml/datasets/Musk+%28Version+1%29',
'https://archive.ics.uci.edu/ml/datasets/Musk+%28Version+2%29',
'https://archive.ics.uci.edu/ml/datasets/Nursery',
'https://archive.ics.uci.edu/ml/datasets/Othello+Domain+Theory',
'https://archive.ics.uci.edu/ml/datasets/Page+Blocks+Classification',
'https://archive.ics.uci.edu/ml/datasets/Optical+Recognition+of+Handwritten+Digits',
'https://archive.ics.uci.edu/ml/datasets/Pen-Based+Recognition+of+Handwritten+Digits',
'https://archive.ics.uci.edu/ml/datasets/Post-Operative+Patient',
'https://archive.ics.uci.edu/ml/datasets/Primary+Tumor',
'https://archive.ics.uci.edu/ml/datasets/Prodigy',
'https://archive.ics.uci.edu/ml/datasets/Qualitative+Structure+Activity+Relationships',
'https://archive.ics.uci.edu/ml/datasets/Quadruped+Mammals',
'https://archive.ics.uci.edu/ml/datasets/Servo',
'https://archive.ics.uci.edu/ml/datasets/Shuttle+Landing+Control',
'https://archive.ics.uci.edu/ml/datasets/Solar+Flare',
'https://archive.ics.uci.edu/ml/datasets/Soybean+%28Large%29',
'https://archive.ics.uci.edu/ml/datasets/Soybean+%28Small%29',
'https://archive.ics.uci.edu/ml/datasets/Challenger+USA+Space+Shuttle+O-Ring',
'https://archive.ics.uci.edu/ml/datasets/Low+Resolution+Spectrometer',
'https://archive.ics.uci.edu/ml/datasets/Spambase',
'https://archive.ics.uci.edu/ml/datasets/SPECT+Heart',
'https://archive.ics.uci.edu/ml/datasets/SPECTF+Heart',
'https://archive.ics.uci.edu/ml/datasets/Sponge',
'https://archive.ics.uci.edu/ml/datasets/Statlog+Project',
'https://archive.ics.uci.edu/ml/datasets/Student+Loan+Relational',
'https://archive.ics.uci.edu/ml/datasets/Teaching+Assistant+Evaluation',
'https://archive.ics.uci.edu/ml/datasets/Tic-Tac-Toe+Endgame',
'https://archive.ics.uci.edu/ml/datasets/Thyroid+Disease',
'https://archive.ics.uci.edu/ml/datasets/Trains',
'https://archive.ics.uci.edu/ml/datasets/University',
'https://archive.ics.uci.edu/ml/datasets/Congressional+Voting+Records',
'https://archive.ics.uci.edu/ml/datasets/Water+Treatment+Plant',
'https://archive.ics.uci.edu/ml/datasets/Waveform+Database+Generator+%28Version+1%29',
'https://archive.ics.uci.edu/ml/datasets/Waveform+Database+Generator+%28Version+2%29',
'https://archive.ics.uci.edu/ml/datasets/Wine',
'https://archive.ics.uci.edu/ml/datasets/Yeast',
'https://archive.ics.uci.edu/ml/datasets/Zoo',
'https://archive.ics.uci.edu/ml/datasets/Undocumented',
'https://archive.ics.uci.edu/ml/datasets/Twenty+Newsgroups',
'https://archive.ics.uci.edu/ml/datasets/Australian+Sign+Language+signs',
'https://archive.ics.uci.edu/ml/datasets/Australian+Sign+Language+signs+%28High+Quality%29',
'https://archive.ics.uci.edu/ml/datasets/US+Census+Data+%281990%29',

'https://archive.ics.uci.edu/ml/datasets/Census-Income+%28KDD%29',
'https://archive.ics.uci.edu/ml/datasets/Coil+1999+Competition+Data',
'https://archive.ics.uci.edu/ml/datasets/Corel+Image+Features',
'https://archive.ics.uci.edu/ml/datasets/E.+Coli+Genes',
'https://archive.ics.uci.edu/ml/datasets/EEG+Database',
'https://archive.ics.uci.edu/ml/datasets/El+Nino',
'https://archive.ics.uci.edu/ml/datasets/Entree+Chicago+Recommendation+Data',
'https://archive.ics.uci.edu/ml/datasets/CMU+Face+Images',
'https://archive.ics.uci.edu/ml/datasets/Insurance+Company+Benchmark+%28COIL+2000%29',
'https://archive.ics.uci.edu/ml/datasets/Internet+Usage+Data',
'https://archive.ics.uci.edu/ml/datasets/IPUMS+Census+Database',
'https://archive.ics.uci.edu/ml/datasets/Japanese+Vowels',
'https://archive.ics.uci.edu/ml/datasets/KDD+Cup+1998+Data',
'https://archive.ics.uci.edu/ml/datasets/KDD+Cup+1999+Data',
'https://archive.ics.uci.edu/ml/datasets/M.+Tuberculosis+Genes',
'https://archive.ics.uci.edu/ml/datasets/Movie',
'https://archive.ics.uci.edu/ml/datasets/MSNBC.com+Anonymous+Web+Data',
'https://archive.ics.uci.edu/ml/datasets/NSF+Research+Award+Abstracts+1990-2003',
'https://archive.ics.uci.edu/ml/datasets/Pioneer-1+Mobile+Robot+Data',
'https://archive.ics.uci.edu/ml/datasets/Pseudo+Periodic+Synthetic+Time+Series',
'https://archive.ics.uci.edu/ml/datasets/Reuters-21578+Text+Categorization+Collection',
'https://archive.ics.uci.edu/ml/datasets/Robot+Execution+Failures',
'https://archive.ics.uci.edu/ml/datasets/Synthetic+Control+Chart+Time+Series',
'https://archive.ics.uci.edu/ml/datasets/Syskill+and+Webert+Web+Page+Ratings',
'https://archive.ics.uci.edu/ml/datasets/UNIX+User+Data',
'https://archive.ics.uci.edu/ml/datasets/Volcanoes+on+Venus+-+JARtool+experiment',
'https://archive.ics.uci.edu/ml/datasets/Statlog+%28Australian+Credit+Approval%29',
'https://archive.ics.uci.edu/ml/datasets/Statlog+%28German+Credit+Data%29',
'https://archive.ics.uci.edu/ml/datasets/Statlog+%28Heart%29',
'https://archive.ics.uci.edu/ml/datasets/Statlog+%28Landsat+Satellite%29',
'https://archive.ics.uci.edu/ml/datasets/Statlog+%28Image+Segmentation%29',
'https://archive.ics.uci.edu/ml/datasets/Statlog+%28Shuttle%29',
'https://archive.ics.uci.edu/ml/datasets/Statlog+%28Vehicle+Silhouettes%29',
'https://archive.ics.uci.edu/ml/datasets/Connectionist+Bench+%28Nettalk+Corpus%29',
'https://archive.ics.uci.edu/ml/datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29',
'https://archive.ics.uci.edu/ml/datasets/Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29',
'https://archive.ics.uci.edu/ml/datasets/Economic+Sanctions',
'https://archive.ics.uci.edu/ml/datasets/Protein+Data',
'https://archive.ics.uci.edu/ml/datasets/Cloud',
'https://archive.ics.uci.edu/ml/datasets/CalIt2+Building+People+Counts',
'https://archive.ics.uci.edu/ml/datasets/Dodgers+Loop+Sensor',
'https://archive.ics.uci.edu/ml/datasets/Poker+Hand',
'https://archive.ics.uci.edu/ml/datasets/MAGIC+Gamma+Telescope',
'https://archive.ics.uci.edu/ml/datasets/UJI+Pen+Characters',
'https://archive.ics.uci.edu/ml/datasets/Mammographic+Mass',
'https://archive.ics.uci.edu/ml/datasets/Forest+Fires',
'https://archive.ics.uci.edu/ml/datasets/Reuters+Transcribed+Subset',
'https://archive.ics.uci.edu/ml/datasets/Bag+of+Words',
'https://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength',
'https://archive.ics.uci.edu/ml/datasets/Hill-Valley',
'https://archive.ics.uci.edu/ml/datasets/Arcene',
'https://archive.ics.uci.edu/ml/datasets/Dexter',
'https://archive.ics.uci.edu/ml/datasets/Dorothea',
'https://archive.ics.uci.edu/ml/datasets/Gisette',
'https://archive.ics.uci.edu/ml/datasets/Madelon',
'https://archive.ics.uci.edu/ml/datasets/Ozone+Level+Detection',
'https://archive.ics.uci.edu/ml/datasets/Abscisic+Acid+Signaling+Network',
'https://archive.ics.uci.edu/ml/datasets/Parkinsons',
'https://archive.ics.uci.edu/ml/datasets/Character+Trajectories',
'https://archive.ics.uci.edu/ml/datasets/Blood+Transfusion+Service+Center',
'https://archive.ics.uci.edu/ml/datasets/UJI+Pen+Characters+%28Version+2%29',
'https://archive.ics.uci.edu/ml/datasets/Semeion+Handwritten+Digit',

'https://archive.ics.uci.edu/ml/datasets/Semeion+Handwritten+Digit',
'https://archive.ics.uci.edu/ml/datasets/SECOM',
'https://archive.ics.uci.edu/ml/datasets/Plants',
'https://archive.ics.uci.edu/ml/datasets/Libras+Movement',
'https://archive.ics.uci.edu/ml/datasets/Concrete+Slump+Test',
'https://archive.ics.uci.edu/ml/datasets/Communities+and+Crime',
'https://archive.ics.uci.edu/ml/datasets/Acute+Inflammations',
'https://archive.ics.uci.edu/ml/datasets/Wine+Quality',
'https://archive.ics.uci.edu/ml/datasets/URL+Reputation',
'https://archive.ics.uci.edu/ml/datasets/p53+Mutants',
'https://archive.ics.uci.edu/ml/datasets/Parkinsons+Telemonitoring',
'https://archive.ics.uci.edu/ml/datasets/Demospongiae',
'https://archive.ics.uci.edu/ml/datasets/Opinosis+Opinion+%26frasl%3B+Review',
'https://archive.ics.uci.edu/ml/datasets/Breast+Tissue',
'https://archive.ics.uci.edu/ml/datasets/Cardiotocography',
'https://archive.ics.uci.edu/ml/datasets/Wall-Following+Robot+Navigation+Data',
'https://archive.ics.uci.edu/ml/datasets/Spoken+Arabic+Digit',
'https://archive.ics.uci.edu/ml/datasets/Localization+Data+for+Person+Activity',
'https://archive.ics.uci.edu/ml/datasets/AutoUniv',
'https://archive.ics.uci.edu/ml/datasets/Steel+Plates+Faults',
'https://archive.ics.uci.edu/ml/datasets/MiniBooNE+particle+identification',
'https://archive.ics.uci.edu/ml/datasets/YearPredictionMSD',
'https://archive.ics.uci.edu/ml/datasets/PEMS-SF',
'https://archive.ics.uci.edu/ml/datasets/OpinRank+Review+Dataset',
'https://archive.ics.uci.edu/ml/datasets/Relative+location+of+CT+slices+on+axial+axis',
'https://archive.ics.uci.edu/ml/datasets/Online+Handwritten+Assamese+Characters+Dataset',
'https://archive.ics.uci.edu/ml/datasets/PubChem+Bioassay+Data',
'https://archive.ics.uci.edu/ml/datasets/Record+Linkage+Comparison+Patterns',
'https://archive.ics.uci.edu/ml/datasets/Communities+and+Crime+Unnormalized',
'https://archive.ics.uci.edu/ml/datasets/Vertebral+Column',
'https://archive.ics.uci.edu/ml/datasets/EMG+Physical+Action+Data+Set',
'https://archive.ics.uci.edu/ml/datasets/Vicon+Physical+Action+Data+Set',
'https://archive.ics.uci.edu/ml/datasets/Amazon+Commerce+reviews+set',
'https://archive.ics.uci.edu/ml/datasets/Amazon+Access+Samples',
'https://archive.ics.uci.edu/ml/datasets/Reuter_50_50',
'https://archive.ics.uci.edu/ml/datasets/Farm+Ads',
'https://archive.ics.uci.edu/ml/datasets/DBWorld+e-mails',
'https://archive.ics.uci.edu/ml/datasets/KEGG+Metabolic+Relation+Network+%28Directed%29',
'https://archive.ics.uci.edu/ml/datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29',
'https://archive.ics.uci.edu/ml/datasets/Bank+Marketing',
'https://archive.ics.uci.edu/ml/datasets/YouTube+Comedy+Slam+Preference+Data',
'https://archive.ics.uci.edu/ml/datasets/Gas+Sensor+Array+Drift+Dataset',
'https://archive.ics.uci.edu/ml/datasets/ILPD+%28Indian+Liver+Patient+Dataset%29',
'https://archive.ics.uci.edu/ml/datasets/OPPORTUNITY+Activity+Recognition',
'https://archive.ics.uci.edu/ml/datasets/Nomao',
'https://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection',
'https://archive.ics.uci.edu/ml/datasets/Skin+Segmentation',
'https://archive.ics.uci.edu/ml/datasets/Planning+Relax',
'https://archive.ics.uci.edu/ml/datasets/PAMAP2+Physical+Activity+Monitoring',
'https://archive.ics.uci.edu/ml/datasets/Restaurant+%26+consumer+data',
'https://archive.ics.uci.edu/ml/datasets/CNAE-9',
'https://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption',
'https://archive.ics.uci.edu/ml/datasets/seeds',
'https://archive.ics.uci.edu/ml/datasets/Northix',
'https://archive.ics.uci.edu/ml/datasets/QtyT40I10D100K',
'https://archive.ics.uci.edu/ml/datasets/Legal+Case+Reports',
'https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones',
'https://archive.ics.uci.edu/ml/datasets/One-hundred+plant+species+leaves+data+set',
'https://archive.ics.uci.edu/ml/datasets/Energy+efficiency',
'https://archive.ics.uci.edu/ml/datasets/Yacht+Hydrodynamics',
'https://archive.ics.uci.edu/ml/datasets/Fertility',
'https://archive.ics.uci.edu/ml/datasets/Daphnet+Freezing+of+Gait',

'https://archive.ics.uci.edu/ml/datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29',
'https://archive.ics.uci.edu/ml/datasets/ISTANBUL+STOCK+EXCHANGE',
'https://archive.ics.uci.edu/ml/datasets/Buzz+in+social+media+',
'https://archive.ics.uci.edu/ml/datasets/First-order+theorem+proving',
'https://archive.ics.uci.edu/ml/datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29',
'https://archive.ics.uci.edu/ml/datasets/Gas+sensor+arrays+in+open+sampling+settings',
'https://archive.ics.uci.edu/ml/datasets/Climate+Model+Simulation+Crashes',
'https://archive.ics.uci.edu/ml/datasets/MicroMass',
'https://archive.ics.uci.edu/ml/datasets/QSAR+biodegradation',
'https://archive.ics.uci.edu/ml/datasets/BLOGGER',
'https://archive.ics.uci.edu/ml/datasets/Daily+and+Sports+Activities',
'https://archive.ics.uci.edu/ml/datasets/User+Knowledge+Modeling',
'https://archive.ics.uci.edu/ml/datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection',
'https://archive.ics.uci.edu/ml/datasets/NYSK',
'https://archive.ics.uci.edu/ml/datasets/Turkiye+Student+Evaluation',
'https://archive.ics.uci.edu/ml/datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machines%29',
'https://archive.ics.uci.edu/ml/datasets/EEG+Eye+State',
'https://archive.ics.uci.edu/ml/datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure',
'https://archive.ics.uci.edu/ml/datasets/seismic-bumps',
'https://archive.ics.uci.edu/ml/datasets/banknote+authentication',
'https://archive.ics.uci.edu/ml/datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Problem%3A+Pat',
'https://archive.ics.uci.edu/ml/datasets/YouTube+Multiview+Video+Games+Dataset',
'https://archive.ics.uci.edu/ml/datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations',
'https://archive.ics.uci.edu/ml/datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors',
'https://archive.ics.uci.edu/ml/datasets/SkillCraft1+Master+Table+Dataset',
'https://archive.ics.uci.edu/ml/datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units',
'https://archive.ics.uci.edu/ml/datasets/SML2010',
'https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset',
'https://archive.ics.uci.edu/ml/datasets/Predict+keywords+activities+in+a+online+social+media',
'https://archive.ics.uci.edu/ml/datasets/Thoracic+Surgery+Data',
'https://archive.ics.uci.edu/ml/datasets/EMG+dataset+in+Lower+Limb',
'https://archive.ics.uci.edu/ml/datasets/SUSY',
'https://archive.ics.uci.edu/ml/datasets/HIGGS',
'https://archive.ics.uci.edu/ml/datasets/Qualitative_Bankruptcy',
'https://archive.ics.uci.edu/ml/datasets/LSVT+Voice+Rehabilitation',
'https://archive.ics.uci.edu/ml/datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer',
'https://archive.ics.uci.edu/ml/datasets/Wilt',
'https://archive.ics.uci.edu/ml/datasets/User+Identification+From+Walking+Activity',
'https://archive.ics.uci.edu/ml/datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer',
'https://archive.ics.uci.edu/ml/datasets/Leaf',
'https://archive.ics.uci.edu/ml/datasets/Dresses_Attribute_Sales',
'https://archive.ics.uci.edu/ml/datasets/Tamilnadu+Electricity+Board+Hourly+Readings',
'https://archive.ics.uci.edu/ml/datasets/Airfoil+Self-Noise',
'https://archive.ics.uci.edu/ml/datasets/Wholesale+customers',
'https://archive.ics.uci.edu/ml/datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis',
'https://archive.ics.uci.edu/ml/datasets/Combined+Cycle+Power+Plant',
'https://archive.ics.uci.edu/ml/datasets/Urban+Land+Cover',
'https://archive.ics.uci.edu/ml/datasets/Diabetes+130-US+hospitals+for+years+1999-2008',
'https://archive.ics.uci.edu/ml/datasets/Bach+Choral+Harmony',
'https://archive.ics.uci.edu/ml/datasets/StoneFlakes',
'https://archive.ics.uci.edu/ml/datasets/Tennis+Major+Tournament+Match+Statistics',
'https://archive.ics.uci.edu/ml/datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings',
'https://archive.ics.uci.edu/ml/datasets/Gesture+Phase+Segmentation',
'https://archive.ics.uci.edu/ml/datasets/Perfume+Data',
'https://archive.ics.uci.edu/ml/datasets/BlogFeedback',

'https://archive.ics.uci.edu/ml/datasets/REALDISP+Activity+Recognition+Dataset',
'https://archive.ics.uci.edu/ml/datasets/Newspaper+and+magazine+images+segmentation+dataset',
'https://archive.ics.uci.edu/ml/datasets/AAAI+2014+Accepted+Papers',
'https://archive.ics.uci.edu/ml/datasets/Gas+sensor+array+under+flow+modulation',
'https://archive.ics.uci.edu/ml/datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures',
'https://archive.ics.uci.edu/ml/datasets/UJIIndoorLoc',
'https://archive.ics.uci.edu/ml/datasets/Sentence+Classification',
'https://archive.ics.uci.edu/ml/datasets/Dow+Jones+Index',
'https://archive.ics.uci.edu/ml/datasets/sEMG+for+Basic+Hand+movements',
'https://archive.ics.uci.edu/ml/datasets/AAAI+2013+Accepted+Papers',
'https://archive.ics.uci.edu/ml/datasets/Geographical+Original+of+Music',
'https://archive.ics.uci.edu/ml/datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants',
'https://archive.ics.uci.edu/ml/datasets/Grammatical+Facial+Expressions',
'https://archive.ics.uci.edu/ml/datasets/NoisyOffice',
'https://archive.ics.uci.edu/ml/datasets/MHEALTH+Dataset',
'https://archive.ics.uci.edu/ml/datasets/Student+Performance',
'https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014',
'https://archive.ics.uci.edu/ml/datasets/Gas+sensor+array+under+dynamic+gas+mixtures',
'https://archive.ics.uci.edu/ml/datasets/microblogPCU',
'https://archive.ics.uci.edu/ml/datasets/Firm-Teacher_Clave-Direction_Classification',
'https://archive.ics.uci.edu/ml/datasets/Dataset+for+Sensorless+Drive+Diagnosis',
'https://archive.ics.uci.edu/ml/datasets/TV+News+Channel+Commercial+Detection+Dataset',
'https://archive.ics.uci.edu/ml/datasets/Phishing+Websites',
'https://archive.ics.uci.edu/ml/datasets/Greenhouse+Gas+Observing+Network',
'https://archive.ics.uci.edu/ml/datasets/Diabetic+Retinopathy+Debrecen+Data+Set',
'https://archive.ics.uci.edu/ml/datasets/HIV-1+protease+cleavage',
'https://archive.ics.uci.edu/ml/datasets/Sentiment+Labelled+Sentences',
'https://archive.ics.uci.edu/ml/datasets/Online+News+Popularity',
'https://archive.ics.uci.edu/ml/datasets/Forest+type+mapping',
'https://archive.ics.uci.edu/ml/datasets/wiki4HE',
'https://archive.ics.uci.edu/ml/datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset',
'https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease',
'https://archive.ics.uci.edu/ml/datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014',
'https://archive.ics.uci.edu/ml/datasets/Folio',
'https://archive.ics.uci.edu/ml/datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015',
'https://archive.ics.uci.edu/ml/datasets/Cuff-Less+Blood+Pressure+Estimation',
'https://archive.ics.uci.edu/ml/datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions',
'https://archive.ics.uci.edu/ml/datasets/Mice+Protein+Expression',
'https://archive.ics.uci.edu/ml/datasets/UJIIndoorLoc-Mag',
'https://archive.ics.uci.edu/ml/datasets/Heterogeneity+Activity+Recognition',
'https://archive.ics.uci.edu/ml/datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set',
'https://archive.ics.uci.edu/ml/datasets/HEPMASS',
'https://archive.ics.uci.edu/ml/datasets/Indoor+User+Movement+Prediction+from+RSS+data',
'https://archive.ics.uci.edu/ml/datasets/Open+University+Learning+Analytics+dataset',
'https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients',
'https://archive.ics.uci.edu/ml/datasets/Mesothelioma%E2%80%99s+disease+data+set+',
'https://archive.ics.uci.edu/ml/datasets/Online+Retail',
'https://archive.ics.uci.edu/ml/datasets/SIFT10M',
'https://archive.ics.uci.edu/ml/datasets/GPS+Trajectories',
'https://archive.ics.uci.edu/ml/datasets/Detect+Malacious+Executable%28AntiVirus%29',
'https://archive.ics.uci.edu/ml/datasets/Occupancy+Detection+',
'https://archive.ics.uci.edu/ml/datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Parkinson%E2%80%99s+Disease',
'https://archive.ics.uci.edu/ml/datasets/News+Aggregator',
'https://archive.ics.uci.edu/ml/datasets/Air+Quality',
'https://archive.ics.uci.edu/ml/datasets/Twin+gas+sensor+arrays',
'https://archive.ics.uci.edu/ml/datasets/Gas+sensors+for+home+activity+monitoring',
'https://archive.ics.uci.edu/ml/datasets/Facebook+Comment+Volume+Dataset',

'https://archive.ics.uci.edu/ml/datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Living+%28AAL%29',
'https://archive.ics.uci.edu/ml/datasets/Polish+companies+bankruptcy+data',
'https://archive.ics.uci.edu/ml/datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29',
'https://archive.ics.uci.edu/ml/datasets/Dota2+Games+Results',
'https://archive.ics.uci.edu/ml/datasets/Facebook+metrics',
'https://archive.ics.uci.edu/ml/datasets/UbiqLog+%28smartphone+lifelogging%29',
'https://archive.ics.uci.edu/ml/datasets/NIPS+Conference+Papers+1987-2015',
'https://archive.ics.uci.edu/ml/datasets/HTRU2',
'https://archive.ics.uci.edu/ml/datasets/Drug+consumption+%28quantified%29',
'https://archive.ics.uci.edu/ml/datasets/Appliances+energy+prediction',
'https://archive.ics.uci.edu/ml/datasets/Miskolc+IIS+Hybrid+IPS',
'https://archive.ics.uci.edu/ml/datasets/KDC-4007+dataset+Collection',
'https://archive.ics.uci.edu/ml/datasets/Geo-Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone',
'https://archive.ics.uci.edu/ml/datasets/DrivFace',
'https://archive.ics.uci.edu/ml/datasets/Website+Phishing',
'https://archive.ics.uci.edu/ml/datasets/YouTube+Spam+Collection',
'https://archive.ics.uci.edu/ml/datasets/Beijing+PM2.5+Data',
'https://archive.ics.uci.edu/ml/datasets/Cargo+2000+Freight+Tracking+and+Tracing',
'https://archive.ics.uci.edu/ml/datasets/Cervical+cancer+%28Risk+Factors%29',
'https://archive.ics.uci.edu/ml/datasets/Quality+Assessment+of+Digital+Colposcopies',
'https://archive.ics.uci.edu/ml/datasets/KASANDR',
'https://archive.ics.uci.edu/ml/datasets/FMA%3A+A+Dataset+For+Music+Analysis',
'https://archive.ics.uci.edu/ml/datasets/Air+quality',
'https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition',
'https://archive.ics.uci.edu/ml/datasets/Devanagari+Handwritten+Character+Dataset',
'https://archive.ics.uci.edu/ml/datasets/Stock+portfolio+performance',
'https://archive.ics.uci.edu/ml/datasets/MoCap+Hand+Postures',
'https://archive.ics.uci.edu/ml/datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech',
'https://archive.ics.uci.edu/ml/datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting',
'https://archive.ics.uci.edu/ml/datasets/PM2.5+Data+of+Five+Chinese+Cities',
'https://archive.ics.uci.edu/ml/datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet',
'https://archive.ics.uci.edu/ml/datasets/Sales_Transactions_Dataset_Weekly',
'https://archive.ics.uci.edu/ml/datasets/Las+Vegas+Strip',
'https://archive.ics.uci.edu/ml/datasets/Eco-hotel',
'https://archive.ics.uci.edu/ml/datasets/MEU-Mobile+KSD',
'https://archive.ics.uci.edu/ml/datasets/Crowdsourced+Mapping',
'https://archive.ics.uci.edu/ml/datasets/gene+expression+cancer+RNA-Seq',
'https://archive.ics.uci.edu/ml/datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer',
'https://archive.ics.uci.edu/ml/datasets/chestnut+%E2%80%93+LARVIC',
'https://archive.ics.uci.edu/ml/datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%28OBS%29+Network',
'https://archive.ics.uci.edu/ml/datasets/Motion+Capture+Hand+Postures',
'https://archive.ics.uci.edu/ml/datasets/Anuran+Calls+%28MFCCs%29',
'https://archive.ics.uci.edu/ml/datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization',
'https://archive.ics.uci.edu/ml/datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy',
'https://archive.ics.uci.edu/ml/datasets/Daily+Demand+Forecasting+Orders',
'https://archive.ics.uci.edu/ml/datasets/Paper+Reviews',
'https://archive.ics.uci.edu/ml/datasets/extention+of+Z-Alizadeh+sani+dataset',
'https://archive.ics.uci.edu/ml/datasets/Z-Alizadeh+Sani',
'https://archive.ics.uci.edu/ml/datasets/Dynamic+Features+of+VirusShare+Executables',
'https://archive.ics.uci.edu/ml/datasets/IDA2016Challenge',
'https://archive.ics.uci.edu/ml/datasets/DSRC+Vehicle+Communications',
'https://archive.ics.uci.edu/ml/datasets/Mturk+User-Perceived+Clusters+over+Images',
'https://archive.ics.uci.edu/ml/datasets/Character+Font+Images',
'https://archive.ics.uci.edu/ml/datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learnin

g+with+Instance+Labels',
 'https://archive.ics.uci.edu/ml/datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++',
 'https://archive.ics.uci.edu/ml/datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++',
 'https://archive.ics.uci.edu/ml/datasets/APS+Failure+at+Scania+Trucks',
 'https://archive.ics.uci.edu/ml/datasets/Wireless+Indoor+Localization',
 'https://archive.ics.uci.edu/ml/datasets/HCC+Survival',
 'https://archive.ics.uci.edu/ml/datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014
+and+2015',
 'https://archive.ics.uci.edu/ml/datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29',
 'https://archive.ics.uci.edu/ml/datasets/Autism+Screening+Adult',
 'https://archive.ics.uci.edu/ml/datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+w
earable+sensor',
 'https://archive.ics.uci.edu/ml/datasets/Immunotherapy+Dataset',
 'https://archive.ics.uci.edu/ml/datasets/Cryotherapy+Dataset+',
 'https://archive.ics.uci.edu/ml/datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes',
 'https://archive.ics.uci.edu/ml/datasets/Discrete+Tone+Image+Dataset',
 'https://archive.ics.uci.edu/ml/datasets/News+Popularity+in+Multiple+Social+Media+Platforms',
 'https://archive.ics.uci.edu/ml/datasets/Ultrasonic+flowmeter+diagnostics',
 'https://archive.ics.uci.edu/ml/datasets/ICMLA+2014+Accepted+Papers+Data+Set',
 'https://archive.ics.uci.edu/ml/datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation',
 'https://archive.ics.uci.edu/ml/datasets/Container+Crane+Controller+Data+Set',
 'https://archive.ics.uci.edu/ml/datasets/Residential+Building+Data+Set',
 'https://archive.ics.uci.edu/ml/datasets/Health+News+in+Twitter',
 'https://archive.ics.uci.edu/ml/datasets/chipseq',
 'https://archive.ics.uci.edu/ml/datasets/SGEMM+GPU+kernel+performance',
 'https://archive.ics.uci.edu/ml/datasets/Repeat+Consumption+Matrices',
 'https://archive.ics.uci.edu/ml/datasets/detection_of_IoT_botnet_attacks_N_BaIoT',
 'https://archive.ics.uci.edu/ml/datasets/Absenteeism+at+work',
 'https://archive.ics.uci.edu/ml/datasets/SCADI',
 'https://archive.ics.uci.edu/ml/datasets/Condition+monitoring+of+hydraulic+systems',
 'https://archive.ics.uci.edu/ml/datasets/Carbon+Nanotubes',
 'https://archive.ics.uci.edu/ml/datasets/Optical+Interconnection+Network+',
 'https://archive.ics.uci.edu/ml/datasets/Sports+articles+for+objectivity+analysis',
 'https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra',
 'https://archive.ics.uci.edu/ml/datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data',
 'https://archive.ics.uci.edu/ml/datasets/Dishonest+Internet+users+Dataset',
 'https://archive.ics.uci.edu/ml/datasets/Victorian+Era+Authorship+Attribution',
 'https://archive.ics.uci.edu/ml/datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set',
 'https://archive.ics.uci.edu/ml/datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing',
 'https://archive.ics.uci.edu/ml/datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals',
 'https://archive.ics.uci.edu/ml/datasets/Roman+Urdu+Data+Set',
 'https://archive.ics.uci.edu/ml/datasets/Avila',
 'https://archive.ics.uci.edu/ml/datasets/PANDOR',
 'https://archive.ics.uci.edu/ml/datasets/Drug+Review+Dataset+%28Druglib.com%29',
 'https://archive.ics.uci.edu/ml/datasets/Drug+Review+Dataset+%28Drugs.com%29',
 'https://archive.ics.uci.edu/ml/datasets/Physical+Unclonable+Functions',
 'https://archive.ics.uci.edu/ml/datasets/Superconductivty+Data',
 'https://archive.ics.uci.edu/ml/datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29',
 'https://archive.ics.uci.edu/ml/datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2',
 'https://archive.ics.uci.edu/ml/datasets/Student+Academics+Performance',
 'https://archive.ics.uci.edu/ml/datasets/Online+Shoppers+Purchasing+Intention+Dataset',
 'https://archive.ics.uci.edu/ml/datasets/PMU-UD',
 'https://archive.ics.uci.edu/ml/datasets/Parkinson%27s+Disease+Classification',
 'https://archive.ics.uci.edu/ml/datasets/Electrical+Grid+Stability+Simulated+Data+',
 'https://archive.ics.uci.edu/ml/datasets/Caesarian+Section+Classification+Dataset',
 'https://archive.ics.uci.edu/ml/datasets/BAUM-1',
 'https://archive.ics.uci.edu/ml/datasets/BAUM-2',
 'https://archive.ics.uci.edu/ml/datasets/Audit+Data',
 'https://archive.ics.uci.edu/ml/datasets/BuddyMove+Data+Set',
 'https://archive.ics.uci.edu/ml/datasets/Real+estate+valuation+data+set',
 'https://archive.ics.uci.edu/ml/datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+nat
ural+connected+speech+Data+Set+',

'https://archive.ics.uci.edu/ml/datasets/Somerville+Happiness+Survey',
'https://archive.ics.uci.edu/ml/datasets/2.4+GHZ+Indoor+Channel+Measurements',
'https://archive.ics.uci.edu/ml/datasets/EMG+data+for+gestures',
'https://archive.ics.uci.edu/ml/datasets/Parking+Birmingham',
'https://archive.ics.uci.edu/ml/datasets/Behavior+of+the+urban+traffic+of+the+city+of+Sao+Paulo+in+Brazil',
'https://archive.ics.uci.edu/ml/datasets/Travel+Reviews',
'https://archive.ics.uci.edu/ml/datasets/Tarvel+Review+Ratings']

In [51]:

```
#El tamañano de lso nombres y los links coincide con 468 unidades
len(link_database)
```

Out[51]:

468

Para el Data type tiene un código:
Univariate, Text
Es el datatype seguido de " " en cada caso Sin datatyope:

Datatype multi:
Multivariate
Datatype uni,txt:
Univariate, Text

In [55]:

```
data_ = soup.find_all("p")
data_
```

Out[55]:

["-//W3C//DTD HTML 4.01 Transitional//EN\">

,
Browse Through: ,
Default Task ,
<a href="datasets.php?format=&task=cla&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table">Classification</a> <font color="red">(350)</font>
<a href="datasets.php?format=&task=reg&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table">Regression</a> <font color="red">(96)</font>
<a href="datasets.php?format=&task=clu&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table">Clustering</a> <font color="red">(84)</font>
<a href="datasets.php?format=&task=other&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table">Other</a> <font color="red">(55)</font> ,
Attribute Type ,
<a href="datasets.php?format=&task=&att=cat&area=&numAtt=&numIns=&type=&sort=nameUp&view=table">Categorical</a> <font color="red">(38)</font>
<a href="datasets.php?format=&task=&att=num&area=&numAtt=&numIns=&type=&sort=nameUp&view=table">Numerical</a> <font color="red">(307)</font>
<a href="datasets.php?format=&task=&att=mix&area=&numAtt=&numIns=&type=&sort=nameUp&view=table">Mixed</a> <font color="red">(55)</font> ,
Data Type ,
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=mvar&sort=nameUp&view=table">Multivariate</a> <font color="red">(357)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=uvar&sort=nameUp&view=table">Univariate</a> <font color="red">(23)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=seq&sort=nameUp&view=table">Sequential</a> <font color="red">(47)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=ts&sort=nameUp&view=table">Time-Series</a> <font color="red">(91)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=text&sort=nameUp&view=table">Text</a> <font color="

red">(53)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=dt&sort=nameUp&view=table">Domain-Theory</a> <font color="red">(23)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=other&sort=nameUp&view=table">Other</a> <font color="red">(21)</font>
 ,
Area ,
<a href="datasets.php?format=&task=&att=&area=life&numAtt=&numIns=&type=&sort=nameUp&view=table">Life Sciences</a> <font color="red">(107)</font>
<a href="datasets.php?format=&task=&att=&area=phys&numAtt=&numIns=&type=&sort=nameUp&view=table">Physical Sciences</a> <font color="red">(49)</font>
<a href="datasets.php?format=&task=&att=&area=comp&numAtt=&numIns=&type=&sort=nameUp&view=table">CS / Engineering</a> <font color="red">(170)</font>
<a href="datasets.php?format=&task=&att=&area=soc&numAtt=&numIns=&type=&sort=nameUp&view=table">Social Sciences</a> <font color="red">(26)</font>
<a href="datasets.php?format=&task=&att=&area=bus&numAtt=&numIns=&type=&sort=nameUp&view=table">Business</a> <font color="red">(29)</font>
<a href="datasets.php?format=&task=&att=&area=game&numAtt=&numIns=&type=&sort=nameUp&view=table">Game</a> <font color="red">(10)</font>
<a href="datasets.php?format=&task=&att=&area=other&numAtt=&numIns=&type=&sort=nameUp&view=table">Other</a> <font color="red">(73)</font> ,
Attributes ,
<a href="datasets.php?format=&task=&att=&area=&numAtt=less10&numIns=&type=&sort=nameUp&view=table">Less than 10</a> <font color="red">(113)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=10to100&numIns=&type=&sort=nameUp&view=table">10 to 100</a> <font color="red">(210)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=greater100&numIns=&type=&sort=nameUp&view=table">Greater than 100</a> <font color="red">(84)</font> ,
Instances ,

<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=less100&type=&sort=nameUp&view=table">Less than 100</a> <font color="red">(27)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=100to1000&type=&sort=nameUp&view=table">100 to 1000</a> <font color="red">(162)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=greater1000&type=&sort=nameUp&view=table">Greater than 1000</a> <font color="red">(246)</font> ,
Format Type ,
<a href="datasets.php?format=mat&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table">Matrix</a> <font color="red">(324)</font>
<a href="datasets.php?format=nonmat&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table">Non-Matrix</a> <font color="red">(145)</font> ,
469 Data Sets ,
<font color="gray">Table View</font> <a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list">List View</a> ,
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameDown&view=table">Name</a> ,
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=typeUp&view=table">Data Types</a> ,
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=taskUp&view=table">Default Task</a> ,
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=attTypeUp&view=table">Attribute Types</a> ,
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=instUp&view=table"># Instances</a> ,
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=attUp&view=table"># Attributes</a> ,
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=dateUp&view=table">Year</a> ,
<a href="datasets/Abalone">Abalone</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
4177 ,
8 ,
1995 ,

1995 ,
<a href="datasets/Adult">Adult</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
48842 ,
14 ,
1996 ,
<a href="datasets/Annealing">Annealing</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
798 ,
38 ,
 ,
<a href="datasets/Anonymous+Microsoft+Web+Data">Anonymous Microsoft Web Data
</a> ,
 ,
Recommender-Systems ,
Categorical ,
37711 ,
294 ,
1998 ,
<a href="datasets/Arrhythmia">Arrhythmia</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
452 ,
279 ,
1998 ,
<a href="datasets/Artificial+Characters">Artificial Characters</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
6000 ,
7 ,
1992 ,
<a href="datasets/Audiology+%28Original%29">Audiology (Original)</a> ,
Multivariate ,
Classification ,
Categorical ,
226 ,
 ,
1987 ,
<a href="datasets/Audiology+%28Standardized%29">Audiology (Standardized)</a>
,
Multivariate ,
Classification ,
Categorical ,
226 ,
69 ,
1992 ,
<a href="datasets/Auto+MPG">Auto MPG</a> ,
Multivariate ,
Regression ,
Categorical, Real ,
398 ,
8 ,
1993 ,
<a href="datasets/Automobile">Automobile</a> ,
Multivariate ,
Regression ,
Categorical, Integer, Real ,

Categorical, Integer, Real ,
205 ,
26 ,
1987 ,
<a href="datasets/Badges">Badges</a> ,
Univariate, Text ,
Classification ,
 ,
294 ,
1 ,
1994 ,
<a href="datasets/Balance+Scale">Balance Scale</a> ,
Multivariate ,
Classification ,
Categorical ,
625 ,
4 ,
1994 ,
<a href="datasets/Balloons">Balloons</a> ,
Multivariate ,
Classification ,
Categorical ,
16 ,
4 ,
 ,
<a href="datasets/Breast+Cancer">Breast Cancer</a> ,
Multivariate ,
Classification ,
Categorical ,
286 ,
9 ,
1988 ,
<a href="datasets/Breast+Cancer+Wisconsin+%28Original%29">Breast Cancer Wiscon
sin (Original)</a> ,
Multivariate ,
Classification ,
Integer ,
699 ,
10 ,
1992 ,
<a href="datasets/Breast+Cancer+Wisconsin+%28Prognostic%29">Breast Cancer Wisc
onsin (Prognostic)</a> ,
Multivariate ,
Classification, Regression ,
Real ,
198 ,
34 ,
1995 ,
<a href="datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29">Breast Cancer Wisc
onsin (Diagnostic)</a> ,
Multivariate ,
Classification ,
Real ,
569 ,
32 ,
1995 ,
<a href="datasets/Pittsburgh+Bridges">Pittsburgh Bridges</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
108 ,
13 ,

1990 ,
<a href="datasets/Car+Evaluation">Car Evaluation</a> ,
Multivariate ,
Classification ,
Categorical ,
1728 ,
6 ,
1997 ,
<a href="datasets/Census+Income">Census Income</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
48842 ,
14 ,
1996 ,
<a href="datasets/Chess+%28King-Rook+vs.+King-Knight%29">Chess (King-Rook vs. King-Knight)</a> ,
Multivariate, Data-Generator ,
Classification ,
Categorical, Integer ,
 ,
22 ,
1988 ,
<a href="datasets/Chess+%28King-Rook+vs.+King-Pawn%29">Chess (King-Rook vs. King-Pawn)</a> ,
Multivariate ,
Classification ,
Categorical ,
3196 ,
36 ,
1989 ,
<a href="datasets/Chess+%28King-Rook+vs.+King%29">Chess (King-Rook vs. King)</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
28056 ,
6 ,
1994 ,
<a href="datasets/Chess+%28Domain+Theories%29">Chess (Domain Theories)</a> ,
Domain-Theory ,
 ,
 ,
 ,
 ,
 ,
<a href="datasets/Bach+Chorales">Bach Chorales</a> ,
Univariate, Time-Series ,
 ,
Categorical, Integer ,
100 ,
6 ,
 ,
<a href="datasets/Connect-4">Connect-4</a> ,
Multivariate, Spatial ,
Classification ,
Categorical ,
67557 ,
42 ,
1995 ,
<a href="datasets/Credit+Approval">Credit Approval</a> ,

Multivariate ,
Classification ,
Categorical, Integer, Real ,
690 ,
15 ,
 ,
<a href="datasets/Japanese+Credit+Screening">Japanese Credit Screening</a> ,
Multivariate, Domain-Theory ,
Classification ,
Categorical, Real, Integer ,
125 ,
 ,
1992 ,
<a href="datasets/Computer+Hardware">Computer Hardware</a> ,
Multivariate ,
Regression ,
Integer ,
209 ,
9 ,
1987 ,
<a href="datasets/Contraceptive+Method+Choice">Contraceptive Method Choice</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
1473 ,
9 ,
1997 ,
<a href="datasets/Covertype">Covertype</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
581012 ,
54 ,
1998 ,
<a href="datasets/Cylinder+Bands">Cylinder Bands</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
512 ,
39 ,
1995 ,
<a href="datasets/Dermatology">Dermatology</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
366 ,
33 ,
1998 ,
<a href="datasets/Diabetes">Diabetes</a> ,
Multivariate, Time-Series ,
 ,
Categorical, Integer ,
 ,
20 ,
 ,
<a href="datasets/DGP2+-+The+Second+Data+Generation+Program">DGP2 - The Second Data Generation Program</a> ,
Data-Generator ,
 ,
Real ,

 ,
 ,
 ,
<a href="datasets/Document+Understanding">Document Understanding</a> ,
 ,
 ,
 ,
 ,
 ,
1994 ,
<a href="datasets/EBL+Domain+Theories">EBL Domain Theories</a> ,
 ,
 ,
 ,
 ,
 ,
 ,
<a href="datasets/Echocardiogram">Echocardiogram</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
132 ,
12 ,
1989 ,
<a href="datasets/Ecoli">Ecoli</a> ,
Multivariate ,
Classification ,
Real ,
336 ,
8 ,
1996 ,
<a href="datasets/Flags">Flags</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
194 ,
30 ,
1990 ,
<a href="datasets/Function+Finding">Function Finding</a> ,
 ,
Function-Learning ,
Real ,
352 ,
 ,
1990 ,
<a href="datasets/Glass+Identification">Glass Identification</a> ,
Multivariate ,
Classification ,
Real ,
214 ,
10 ,
1987 ,
<a href="datasets/Haberman%27s+Survival">Haberman's Survival</a> ,
Multivariate ,
Classification ,
Integer ,
306 ,
3 ,
1999 ,
<a href="datasets/Hayes-Roth">Hayes-Roth</a> ,
Multivariate ,
Classification ,

Categorical ,
160 ,
5 ,
1989 ,
<a href="datasets/Heart+Disease">Heart Disease</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
303 ,
75 ,
1988 ,
<a href="datasets/Hepatitis">Hepatitis</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
155 ,
19 ,
1988 ,
<a href="datasets/Horse+Colic">Horse Colic</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
368 ,
27 ,
1989 ,
<a href="datasets/ICU">ICU</a> ,
Multivariate, Time-Series ,
 ,
Real ,
 ,
 ,
 ,
<a href="datasets/Image+Segmentation">Image Segmentation</a> ,
Multivariate ,
Classification ,
Real ,
2310 ,
19 ,
1990 ,
<a href="datasets/Internet+Advertisements">Internet Advertisements</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
3279 ,
1558 ,
1998 ,
<a href="datasets/Ionosphere">Ionosphere</a> ,
Multivariate ,
Classification ,
Integer, Real ,
351 ,
34 ,
1989 ,
<a href="datasets/Iris">Iris</a> ,
Multivariate ,
Classification ,
Real ,
150 ,
4 ,
1988 ,
<a href="datasets/ISOLET">ISOLET</a> ,
Multivariate ,

Classification ,
Real ,
7797 ,
617 ,
1994 ,
<a href="datasets/Kinship">Kinship</a> ,
Relational ,
Relational-Learning ,
Categorical ,
104 ,
12 ,
1990 ,
<a href="datasets/Labor+Relations">Labor Relations</a> ,
Multivariate ,
 ,
Categorical, Integer, Real ,
57 ,
16 ,
1988 ,
<a href="datasets/LED+Display+Domain">LED Display Domain</a> ,
Multivariate, Data-Generator ,
Classification ,
Categorical ,
 ,
7 ,
1988 ,
<a href="datasets/Lenses">Lenses</a> ,
Multivariate ,
Classification ,
Categorical ,
24 ,
4 ,
1990 ,
<a href="datasets/Letter+Recognition">Letter Recognition</a> ,
Multivariate ,
Classification ,
Integer ,
20000 ,
16 ,
1991 ,
<a href="datasets/Liver+Disorders">Liver Disorders</a> ,
Multivariate ,
 ,
Categorical, Integer, Real ,
345 ,
7 ,
1990 ,
<a href="datasets/Logic+Theorist">Logic Theorist</a> ,
Domain-Theory ,
 ,
 ,
 ,
 ,
 ,
<a href="datasets/Lung+Cancer">Lung Cancer</a> ,
Multivariate ,
Classification ,
Integer ,
32 ,
56 ,
1992 ,
<a href="datasets/Lymphography">Lymphography</a> ,

Multivariate ,
Classification ,
Categorical ,
148 ,
18 ,
1988 ,
<a href="datasets/Mechanical+Analysis">Mechanical Analysis</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
209 ,
8 ,
1990 ,
<a href="datasets/Meta-data">Meta-data</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
528 ,
22 ,
1996 ,
<a href="datasets/Mobile+Robots">Mobile Robots</a> ,
Domain-Theory ,
 ,
Categorical, Integer, Real ,
 ,
 ,
1995 ,
<a href="datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29">Molecular Biology (Promoter Gene Sequences)</a> ,
Sequential, Domain-Theory ,
Classification ,
Categorical ,
106 ,
58 ,
1990 ,
<a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular Biology (Protein Secondary Structure)</a> ,
Sequential ,
Classification ,
Categorical ,
128 ,
 ,
 ,
<a href="datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29">Molecular Biology (Splice-junction Gene Sequences)</a> ,
Sequential, Domain-Theory ,
Classification ,
Categorical ,
3190 ,
61 ,
1992 ,
<a href="datasets/MONK%27s+Problems">MONK's Problems</a> ,
Multivariate ,
Classification ,
Categorical ,
432 ,
7 ,
1992 ,
<a href="datasets/Moral+Reasoner">Moral Reasoner</a> ,
Domain-Theory ,
 ,

,
202 ,
 ,
1994 ,
<a href="datasets/Multiple+Features">Multiple Features</a> ,
Multivariate ,
Classification ,
Integer, Real ,
2000 ,
649 ,
 ,
<a href="datasets/Mushroom">Mushroom</a> ,
Multivariate ,
Classification ,
Categorical ,
8124 ,
22 ,
1987 ,
<a href="datasets/Musk+%28Version+1%29">Musk (Version 1)</a> ,
Multivariate ,
Classification ,
Integer ,
476 ,
168 ,
1994 ,
<a href="datasets/Musk+%28Version+2%29">Musk (Version 2)</a> ,
Multivariate ,
Classification ,
Integer ,
6598 ,
168 ,
1994 ,
<a href="datasets/Nursery">Nursery</a> ,
Multivariate ,
Classification ,
Categorical ,
12960 ,
8 ,
1997 ,
<a href="datasets/Othello+Domain+Theory">Othello Domain Theory</a> ,
Domain-Theory ,
 ,
 ,
 ,
 ,
1991 ,
<a href="datasets/Page+Blocks+Classification">Page Blocks Classification</a> ,
Multivariate ,
Classification ,
Integer, Real ,
5473 ,
10 ,
1995 ,
<a href="datasets/Optical+Recognition+of+Handwritten+Digits">Optical Recognition of Handwritten Digits</a> ,
Multivariate ,
Classification ,
Integer ,
5620 ,
64 ,
1998 ,
<a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits">Pen-Based Recognition of Handwritten Digits</a> ,

tion of Handwritten Digits</a> ,
Multivariate ,
Classification ,
Integer ,
10992 ,
16 ,
1998 ,
<a href="datasets/Post-Operative+Patient">Post-Operative Patient</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
90 ,
8 ,
1993 ,
<a href="datasets/Primary+Tumor">Primary Tumor</a> ,
Multivariate ,
Classification ,
Categorical ,
339 ,
17 ,
1988 ,
<a href="datasets/Prodigy">Prodigy</a> ,
Domain-Theory ,
 ,
 ,
 ,
 ,
 ,
<a href="datasets/Qualitative+Structure+Activity+Relationships">Qualitative Structure A
ctivity Relationships</a> ,
Domain-Theory ,
 ,
 ,
 ,
 ,
 ,
<a href="datasets/Quadruped+Mammals">Quadruped Mammals</a> ,
Multivariate, Data-Generator ,
Classification ,
Real ,
 ,
72 ,
1992 ,
<a href="datasets/Servo">Servo</a> ,
Multivariate ,
Regression ,
Categorical, Integer ,
167 ,
4 ,
1993 ,
<a href="datasets/Shuttle+Landing+Control">Shuttle Landing Control</a> ,
Multivariate ,
Classification ,
Categorical ,
15 ,
6 ,
1988 ,
<a href="datasets/Solar+Flare">Solar Flare</a> ,
Multivariate ,
Regression ,
Categorical ,
1389 ,

10 ,
1989 ,
<a href="datasets/Soybean+%28Large%29">Soybean (Large)</a> ,
Multivariate ,
Classification ,
Categorical ,
307 ,
35 ,
1988 ,
<a href="datasets/Soybean+%28Small%29">Soybean (Small)</a> ,
Multivariate ,
Classification ,
Categorical ,
47 ,
35 ,
1987 ,
<a href="datasets/Challenger+USA+Space+Shuttle+O-Ring">Challenger USA Space Sh
uttle O-Ring</a> ,
Multivariate ,
Regression ,
Integer ,
23 ,
4 ,
1993 ,
<a href="datasets/Low+Resolution+Spectrometer">Low Resolution Spectrometer</a></
b> ,
Multivariate ,
Classification ,
Integer, Real ,
531 ,
102 ,
1988 ,
<a href="datasets/Spambase">Spambase</a> ,
Multivariate ,
Classification ,
Integer, Real ,
4601 ,
57 ,
1999 ,
<a href="datasets/SPECT+Heart">SPECT Heart</a> ,
Multivariate ,
Classification ,
Categorical ,
267 ,
22 ,
2001 ,
<a href="datasets/SPECTF+Heart">SPECTF Heart</a> ,
Multivariate ,
Classification ,
Integer ,
267 ,
44 ,
2001 ,
<a href="datasets/Sponge">Sponge</a> ,
Multivariate ,
Clustering ,
Categorical, Integer ,
76 ,
45 ,
 ,
<a href="datasets/Statlog+Project">Statlog Project</a> ,
 ,

 ,
 ,
 ,
 ,
1992 ,
<a href="datasets/Student+Loan+Relational">Student Loan Relational</a> ,
Domain-Theory ,
 ,
 ,
1000 ,
 ,
1993 ,
<a href="datasets/Teaching+Assistant+Evaluation">Teaching Assistant Evaluation</a><
/b> ,
Multivariate ,
Classification ,
Categorical, Integer ,
151 ,
5 ,
1997 ,
<a href="datasets/Tic-Tac-Toe+Endgame">Tic-Tac-Toe Endgame</a> ,
Multivariate ,
Classification ,
Categorical ,
958 ,
9 ,
1991 ,
<a href="datasets/Thyroid+Disease">Thyroid Disease</a> ,
Multivariate, Domain-Theory ,
Classification ,
Categorical, Real ,
7200 ,
21 ,
1987 ,
<a href="datasets/Trains">Trains</a> ,
Multivariate ,
Classification ,
Categorical ,
10 ,
32 ,
1994 ,
<a href="datasets/University">University</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
285 ,
17 ,
1988 ,
<a href="datasets/Congressional+Voting+Records">Congressional Voting Records</a>
,
Multivariate ,
Classification ,
Categorical ,
435 ,
16 ,
1987 ,
<a href="datasets/Water+Treatment+Plant">Water Treatment Plant</a> ,
Multivariate ,
Clustering ,
Integer, Real ,
527 ,
38 ,

1993 ,
<a href="datasets/Waveform+Database+Generator+%28Version+1%29">Waveform Database Generator (Version 1)</a> ,
Multivariate, Data-Generator ,
Classification ,
Real ,
5000 ,
21 ,
1988 ,
<a href="datasets/Waveform+Database+Generator+%28Version+2%29">Waveform Database Generator (Version 2)</a> ,
Multivariate, Data-Generator ,
Classification ,
Real ,
5000 ,
40 ,
1988 ,
<a href="datasets/Wine">Wine</a> ,
Multivariate ,
Classification ,
Integer, Real ,
178 ,
13 ,
1991 ,
<a href="datasets/Yeast">Yeast</a> ,
Multivariate ,
Classification ,
Real ,
1484 ,
8 ,
1996 ,
<a href="datasets/Zoo">Zoo</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
101 ,
17 ,
1990 ,
<a href="datasets/Undocumented">Undocumented</a> ,
 ,
 ,
 ,
 ,
 ,
 ,
<a href="datasets/Twenty+Newsgroups">Twenty Newsgroups</a> ,
Text ,
 ,
 ,
20000 ,
 ,
1999 ,
<a href="datasets/Australian+Sign+Language+signs">Australian Sign Language signs</a> ,
Multivariate, Time-Series ,
Classification ,
Categorical, Real ,
6650 ,
15 ,
1999 ,
<a href="datasets/Australian+Sign+Language+signs+%28High+Quality%29">Australian Sign Language signs (High Quality)</a> ,

Multivariate, Time-Series ,
Classification ,
Real ,
2565 ,
22 ,
2002 ,
<a href="datasets/US+Census+Data+%281990%29">US Census Data (1990)</a><
/p>,
Multivariate ,
Clustering ,
Categorical ,
2458285 ,
68 ,
 ,
<a href="datasets/Census-Income+%28KDD%29">Census-Income (KDD)</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
299285 ,
40 ,
2000 ,
<a href="datasets/Coil+1999+Competition+Data">Coil 1999 Competition Data</a><
/p>,
Multivariate ,
 ,
Categorical, Real ,
340 ,
17 ,
1999 ,
<a href="datasets/Corel+Image+Features">Corel Image Features</a> ,
Multivariate ,
 ,
Real ,
68040 ,
89 ,
1999 ,
<a href="datasets/E.+Coli+Genes">E. Coli Genes</a> ,
Relational ,
 ,
 ,
 ,
 ,
2001 ,
<a href="datasets/EEG+Database">EEG Database</a> ,
Multivariate, Time-Series ,
 ,
Categorical, Integer, Real ,
122 ,
4 ,
1999 ,
<a href="datasets/El+Nino">El Nino</a> ,
Spatio-temporal ,
 ,
Integer, Real ,
178080 ,
12 ,
1999 ,
<a href="datasets/Entree+Chicago+Recommendation+Data">Entree Chicago Recomm
endation Data</a> ,
Transactional, Sequential ,
Recommender-Systems ,
Categorical ,

50672 ,
 ,
2000 ,
<a href="datasets/CMU+Face+Images">CMU Face Images</a> ,
Image ,
Classification ,
Integer ,
640 ,
 ,
1999 ,
<a href="datasets/Insurance+Company+Benchmark+%28COIL+2000%29">Insurance C
ompany Benchmark (COIL 2000)</a> ,
Multivariate ,
Regression, Description ,
Categorical, Integer ,
9000 ,
86 ,
2000 ,
<a href="datasets/Internet+Usage+Data">Internet Usage Data</a> ,
Multivariate ,
 ,
Categorical, Integer ,
10104 ,
72 ,
1999 ,
<a href="datasets/IPUMS+Census+Database">IPUMS Census Database</a> ,
Multivariate ,
 ,
Categorical, Integer ,
256932 ,
61 ,
1999 ,
<a href="datasets/Japanese+Vowels">Japanese Vowels</a> ,
Multivariate, Time-Series ,
Classification ,
Real ,
640 ,
12 ,
 ,
<a href="datasets/KDD+Cup+1998+Data">KDD Cup 1998 Data</a> ,
Multivariate ,
Regression ,
Categorical, Integer ,
191779 ,
481 ,
1998 ,
<a href="datasets/KDD+Cup+1999+Data">KDD Cup 1999 Data</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
4000000 ,
42 ,
1999 ,
<a href="datasets/M.+Tuberculosis+Genes">M. Tuberculosis Genes</a> ,
Relational ,
 ,
 ,
 ,
 ,
2001 ,
<a href="datasets/Movie">Movie</a> ,
Multivariate, Relational ,

 ,
 ,
10000 ,
 ,
1999 ,
<a href="datasets/MSNBC.com+Anonymous+Web+Data">MSNBC.com Anonymous Web Data</a> ,
Sequential ,
 ,
Categorical ,
989818 ,
 ,
 ,
<a href="datasets/NSF+Research+Award+Abstracts+1990-2003">NSF Research Award Abstracts 1990-2003</a> ,
Text ,
 ,
 ,
129000 ,
 ,
2003 ,
<a href="datasets/Pioneer-1+Mobile+Robot+Data">Pioneer-1 Mobile Robot Data</a> ,
Multivariate, Time-Series ,
 ,
Categorical, Real ,
 ,
 ,
1999 ,
<a href="datasets/Pseudo+Periodic+Synthetic+Time+Series">Pseudo Periodic Synthetic Time Series</a> ,
Univariate, Time-Series ,
 ,
 ,
100000 ,
 ,
1999 ,
<a href="datasets/Reuters-21578+Text+Categorization+Collection">Reuters-21578 Text Categorization Collection</a> ,
Text ,
Classification ,
Categorical ,
21578 ,
5 ,
1997 ,
<a href="datasets/Robot+Execution+Failures">Robot Execution Failures</a> ,
Multivariate, Time-Series ,
Classification ,
Integer ,
463 ,
90 ,
1999 ,
<a href="datasets/Synthetic+Control+Chart+Time+Series">Synthetic Control Chart Time Series</a> ,
Time-Series ,
Classification, Clustering ,
Real ,
600 ,
 ,
1999 ,
<a href="datasets/Syskill+and+Webert+Web+Page+Ratings">Syskill and Webert Web Page Ratings</a> ,

Multivariate, Text ,
Classification ,
Categorical ,
332 ,
5 ,
1998 ,
<a href="datasets/UNIX+User+Data">UNIX User Data</a> ,
Text, Sequential ,
 ,
 ,
 ,
 ,
 ,
<a href="datasets/Volcanoes+on+Venus+-+JARtool+experiment">Volcanoes on Venus - JARtool experiment</a> ,
Image ,
...]

In [74]:

```
data_2 = data_[25:] #los elementos que me interesan
data_2
```

Out[74]:

[<a href="datasets/Abalone">Abalone</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
4177 ,
8 ,
1995 ,
<a href="datasets/Adult">Adult</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
48842 ,
14 ,
1996 ,
<a href="datasets/Annealing">Annealing</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
798 ,
38 ,
 ,
<a href="datasets/Anonymous+Microsoft+Web+Data">Anonymous Microsoft Web Data</a> ,
 ,
Recommender-Systems ,
Categorical ,
37711 ,
294 ,
1998 ,
<a href="datasets/Arrhythmia">Arrhythmia</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
452 ,
279 ,
1998 ,
<a href="datasets/Artificial+Characters">Artificial Characters</a> ,

Multivariate ,
Classification ,
Categorical, Integer, Real ,
6000 ,
7 ,
1992 ,
<a href="datasets/Audiology+%28Original%29">Audiology (Original)</a> ,
Multivariate ,
Classification ,
Categorical ,
226 ,
 ,
1987 ,
<a href="datasets/Audiology+%28Standardized%29">Audiology (Standardized)</a>
,
Multivariate ,
Classification ,
Categorical ,
226 ,
69 ,
1992 ,
<a href="datasets/Auto+MPG">Auto MPG</a> ,
Multivariate ,
Regression ,
Categorical, Real ,
398 ,
8 ,
1993 ,
<a href="datasets/Automobile">Automobile</a> ,
Multivariate ,
Regression ,
Categorical, Integer, Real ,
205 ,
26 ,
1987 ,
<a href="datasets/Badges">Badges</a> ,
Univariate, Text ,
Classification ,
 ,
294 ,
1 ,
1994 ,
<a href="datasets/Balance+Scale">Balance Scale</a> ,
Multivariate ,
Classification ,
Categorical ,
625 ,
4 ,
1994 ,
<a href="datasets/Balloons">Balloons</a> ,
Multivariate ,
Classification ,
Categorical ,
16 ,
4 ,
 ,
<a href="datasets/Breast+Cancer">Breast Cancer</a> ,
Multivariate ,
Classification ,
Categorical ,
286 ,
9 ,

1988 ,
<a href="datasets/Breast+Cancer+Wisconsin+%28Original%29">Breast Cancer Wisconsin (Original)</a> ,
Multivariate ,
Classification ,
Integer ,
699 ,
10 ,
1992 ,
<a href="datasets/Breast+Cancer+Wisconsin+%28Prognostic%29">Breast Cancer Wisconsin (Prognostic)</a> ,
Multivariate ,
Classification, Regression ,
Real ,
198 ,
34 ,
1995 ,
<a href="datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29">Breast Cancer Wisconsin (Diagnostic)</a> ,
Multivariate ,
Classification ,
Real ,
569 ,
32 ,
1995 ,
<a href="datasets/Pittsburgh+Bridges">Pittsburgh Bridges</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
108 ,
13 ,
1990 ,
<a href="datasets/Car+Evaluation">Car Evaluation</a> ,
Multivariate ,
Classification ,
Categorical ,
1728 ,
6 ,
1997 ,
<a href="datasets/Census+Income">Census Income</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
48842 ,
14 ,
1996 ,
<a href="datasets/Chess+%28King-Rook+vs.+King-Knight%29">Chess (King-Rook vs. King-Knight)</a> ,
Multivariate, Data-Generator ,
Classification ,
Categorical, Integer ,
 ,
22 ,
1988 ,
<a href="datasets/Chess+%28King-Rook+vs.+King-Pawn%29">Chess (King-Rook vs. King-Pawn)</a> ,
Multivariate ,
Classification ,
Categorical ,
3196 ,
36 ,
1989 ,

<a href="datasets/Chess+%28King-Rook+vs.+King%29">Chess (King-Rook vs. King)</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
28056 ,
6 ,
1994 ,
<a href="datasets/Chess+%28Domain+Theories%29">Chess (Domain Theories)</a> ,
Domain-Theory ,
 ,
 ,
 ,
 ,
 ,
<a href="datasets/Bach+Chorales">Bach Chorales</a> ,
Univariate, Time-Series ,
 ,
Categorical, Integer ,
100 ,
6 ,
 ,
<a href="datasets/Connect-4">Connect-4</a> ,
Multivariate, Spatial ,
Classification ,
Categorical ,
67557 ,
42 ,
1995 ,
<a href="datasets/Credit+Approval">Credit Approval</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
690 ,
15 ,
 ,
<a href="datasets/Japanese+Credit+Screening">Japanese Credit Screening</a> ,
Multivariate, Domain-Theory ,
Classification ,
Categorical, Real, Integer ,
125 ,
 ,
1992 ,
<a href="datasets/Computer+Hardware">Computer Hardware</a> ,
Multivariate ,
Regression ,
Integer ,
209 ,
9 ,
1987 ,
<a href="datasets/Contraceptive+Method+Choice">Contraceptive Method Choice</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
1473 ,
9 ,
1997 ,
<a href="datasets/Covertype">Covertype</a> ,
Multivariate ,

Classification ,
Categorical, Integer ,
581012 ,
54 ,
1998 ,
<a href="datasets/Cylinder+Bands">Cylinder Bands</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
512 ,
39 ,
1995 ,
<a href="datasets/Dermatology">Dermatology</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
366 ,
33 ,
1998 ,
<a href="datasets/Diabetes">Diabetes</a> ,
Multivariate, Time-Series ,
 ,
Categorical, Integer ,
 ,
20 ,
 ,
<a href="datasets/DGP2+-+The+Second+Data+Generation+Program">DGP2 - The Sec
ond Data Generation Program</a> ,
Data-Generator ,
 ,
Real ,
 ,
 ,
 ,
<a href="datasets/Document+Understanding">Document Understanding</a> ,
 ,
 ,
 ,
 ,
 ,
1994 ,
<a href="datasets/EBL+Domain+Theories">EBL Domain Theories</a> ,
 ,
 ,
 ,
 ,
 ,
 ,
<a href="datasets/Echocardiogram">Echocardiogram</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
132 ,
12 ,
1989 ,
<a href="datasets/Ecoli">Ecoli</a> ,
Multivariate ,
Classification ,
Real ,
336 ,
8 ,
1996 ,

<a href="datasets/Flags">Flags</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
194 ,
30 ,
1990 ,
<a href="datasets/Function+Finding">Function Finding</a> ,
 ,
Function-Learning ,
Real ,
352 ,
 ,
1990 ,
<a href="datasets/Glass+Identification">Glass Identification</a> ,
Multivariate ,
Classification ,
Real ,
214 ,
10 ,
1987 ,
<a href="datasets/Haberman%27s+Survival">Haberman's Survival</a> ,
Multivariate ,
Classification ,
Integer ,
306 ,
3 ,
1999 ,
<a href="datasets/Hayes-Roth">Hayes-Roth</a> ,
Multivariate ,
Classification ,
Categorical ,
160 ,
5 ,
1989 ,
<a href="datasets/Heart+Disease">Heart Disease</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
303 ,
75 ,
1988 ,
<a href="datasets/Hepatitis">Hepatitis</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
155 ,
19 ,
1988 ,
<a href="datasets/Horse+Colic">Horse Colic</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
368 ,
27 ,
1989 ,
<a href="datasets/ICU">ICU</a> ,
Multivariate, Time-Series ,
 ,
Real ,
 ,
 ,

,
 ,
<a href="datasets/Image+Segmentation">Image Segmentation</a> ,
Multivariate ,
Classification ,
Real ,
2310 ,
19 ,
1990 ,
<a href="datasets/Internet+Advertisements">Internet Advertisements</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
3279 ,
1558 ,
1998 ,
<a href="datasets/Ionosphere">Ionosphere</a> ,
Multivariate ,
Classification ,
Integer, Real ,
351 ,
34 ,
1989 ,
<a href="datasets/Iris">Iris</a> ,
Multivariate ,
Classification ,
Real ,
150 ,
4 ,
1988 ,
<a href="datasets/ISOLET">ISOLET</a> ,
Multivariate ,
Classification ,
Real ,
7797 ,
617 ,
1994 ,
<a href="datasets/Kinship">Kinship</a> ,
Relational ,
Relational-Learning ,
Categorical ,
104 ,
12 ,
1990 ,
<a href="datasets/Labor+Relations">Labor Relations</a> ,
Multivariate ,
 ,
Categorical, Integer, Real ,
57 ,
16 ,
1988 ,
<a href="datasets/LED+Display+Domain">LED Display Domain</a> ,
Multivariate, Data-Generator ,
Classification ,
Categorical ,
 ,
7 ,
1988 ,
<a href="datasets/Lenses">Lenses</a> ,
Multivariate ,
Classification ,
Categorical ,
24 ,

24 ,
4 ,
1990 ,
<a href="datasets/Letter+Recognition">Letter Recognition</a> ,
Multivariate ,
Classification ,
Integer ,
20000 ,
16 ,
1991 ,
<a href="datasets/Liver+Disorders">Liver Disorders</a> ,
Multivariate ,
 ,
Categorical, Integer, Real ,
345 ,
7 ,
1990 ,
<a href="datasets/Logic+Theorist">Logic Theorist</a> ,
Domain-Theory ,
 ,
 ,
 ,
 ,
 ,
<a href="datasets/Lung+Cancer">Lung Cancer</a> ,
Multivariate ,
Classification ,
Integer ,
32 ,
56 ,
1992 ,
<a href="datasets/Lymphography">Lymphography</a> ,
Multivariate ,
Classification ,
Categorical ,
148 ,
18 ,
1988 ,
<a href="datasets/Mechanical+Analysis">Mechanical Analysis</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
209 ,
8 ,
1990 ,
<a href="datasets/Meta-data">Meta-data</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
528 ,
22 ,
1996 ,
<a href="datasets/Mobile+Robots">Mobile Robots</a> ,
Domain-Theory ,
 ,
Categorical, Integer, Real ,
 ,
 ,
1995 ,
<a href="datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29">Molecular Biology (Promoter Gene Sequences)</a> ,
Sequential, Domain-Theory ,

Classification ,
Categorical ,
106 ,
58 ,
1990 ,
<a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular Biology (Protein Secondary Structure)</a> ,
Sequential ,
Classification ,
Categorical ,
128 ,
 ,
 ,
<a href="datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29">Molecular Biology (Splice-junction Gene Sequences)</a> ,
Sequential, Domain-Theory ,
Classification ,
Categorical ,
3190 ,
61 ,
1992 ,
<a href="datasets/MONK%27s+Problems">MONK's Problems</a> ,
Multivariate ,
Classification ,
Categorical ,
432 ,
7 ,
1992 ,
<a href="datasets/Moral+Reasoner">Moral Reasoner</a> ,
Domain-Theory ,
 ,
 ,
202 ,
 ,
1994 ,
<a href="datasets/Multiple+Features">Multiple Features</a> ,
Multivariate ,
Classification ,
Integer, Real ,
2000 ,
649 ,
 ,
<a href="datasets/Mushroom">Mushroom</a> ,
Multivariate ,
Classification ,
Categorical ,
8124 ,
22 ,
1987 ,
<a href="datasets/Musk+%28Version+1%29">Musk (Version 1)</a> ,
Multivariate ,
Classification ,
Integer ,
476 ,
168 ,
1994 ,
<a href="datasets/Musk+%28Version+2%29">Musk (Version 2)</a> ,
Multivariate ,
Classification ,
Integer ,
6598 ,
168 ,

1994 ,
<a href="datasets/Nursery">Nursery</a> ,
Multivariate ,
Classification ,
Categorical ,
12960 ,
8 ,
1997 ,
<a href="datasets/Othello+Domain+Theory">Othello Domain Theory</a> ,
Domain-Theory ,
 ,
 ,
 ,
 ,
1991 ,
<a href="datasets/Page+Blocks+Classification">Page Blocks Classification</a> ,
Multivariate ,
Classification ,
Integer, Real ,
5473 ,
10 ,
1995 ,
<a href="datasets/Optical+Recognition+of+Handwritten+Digits">Optical Recognition of Handwritten Digits</a> ,
Multivariate ,
Classification ,
Integer ,
5620 ,
64 ,
1998 ,
<a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits">Pen-Based Recognition of Handwritten Digits</a> ,
Multivariate ,
Classification ,
Integer ,
10992 ,
16 ,
1998 ,
<a href="datasets/Post-Operative+Patient">Post-Operative Patient</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
90 ,
8 ,
1993 ,
<a href="datasets/Primary+Tumor">Primary Tumor</a> ,
Multivariate ,
Classification ,
Categorical ,
339 ,
17 ,
1988 ,
<a href="datasets/Prodigy">Prodigy</a> ,
Domain-Theory ,
 ,
 ,
 ,
 ,
 ,
<a href="datasets/Qualitative+Structure+Activity+Relationships">Qualitative Structure Activity Relationships</a> ,
Domain-Theory ,

 ,
 ,
 ,
 ,
 ,
<a href="datasets/Quadruped+Mammals">Quadruped Mammals</a> ,
Multivariate, Data-Generator ,
Classification ,
Real ,
 ,
72 ,
1992 ,
<a href="datasets/Servo">Servo</a> ,
Multivariate ,
Regression ,
Categorical, Integer ,
167 ,
4 ,
1993 ,
<a href="datasets/Shuttle+Landing+Control">Shuttle Landing Control</a> ,
Multivariate ,
Classification ,
Categorical ,
15 ,
6 ,
1988 ,
<a href="datasets/Solar+Flare">Solar Flare</a> ,
Multivariate ,
Regression ,
Categorical ,
1389 ,
10 ,
1989 ,
<a href="datasets/Soybean+%28Large%29">Soybean (Large)</a> ,
Multivariate ,
Classification ,
Categorical ,
307 ,
35 ,
1988 ,
<a href="datasets/Soybean+%28Small%29">Soybean (Small)</a> ,
Multivariate ,
Classification ,
Categorical ,
47 ,
35 ,
1987 ,
<a href="datasets/Challenger+USA+Space+Shuttle+O-Ring">Challenger USA Space Sh
uttle O-Ring</a> ,
Multivariate ,
Regression ,
Integer ,
23 ,
4 ,
1993 ,
<a href="datasets/Low+Resolution+Spectrometer">Low Resolution Spectrometer</a></
b> ,
Multivariate ,
Classification ,
Integer, Real ,
531 ,
102 ,

1988 ,
<a href="datasets/Spambase">Spambase</a> ,
Multivariate ,
Classification ,
Integer, Real ,
4601 ,
57 ,
1999 ,
<a href="datasets/SPECT+Heart">SPECT Heart</a> ,
Multivariate ,
Classification ,
Categorical ,
267 ,
22 ,
2001 ,
<a href="datasets/SPECTF+Heart">SPECTF Heart</a> ,
Multivariate ,
Classification ,
Integer ,
267 ,
44 ,
2001 ,
<a href="datasets/Sponge">Sponge</a> ,
Multivariate ,
Clustering ,
Categorical, Integer ,
76 ,
45 ,
 ,
<a href="datasets/Statlog+Project">Statlog Project</a> ,
 ,
 ,
 ,
 ,
 ,
1992 ,
<a href="datasets/Student+Loan+Relational">Student Loan Relational</a> ,
Domain-Theory ,
 ,
 ,
1000 ,
 ,
1993 ,
<a href="datasets/Teaching+Assistant+Evaluation">Teaching Assistant Evaluation</a><
/b> ,
Multivariate ,
Classification ,
Categorical, Integer ,
151 ,
5 ,
1997 ,
<a href="datasets/Tic-Tac-Toe+Endgame">Tic-Tac-Toe Endgame</a> ,
Multivariate ,
Classification ,
Categorical ,
958 ,
9 ,
1991 ,
<a href="datasets/Thyroid+Disease">Thyroid Disease</a> ,
Multivariate, Domain-Theory ,
Classification ,
Categorical, Real ,

7200 ,
21 ,
1987 ,
<a href="datasets/Trains">Trains</a> ,
Multivariate ,
Classification ,
Categorical ,
10 ,
32 ,
1994 ,
<a href="datasets/University">University</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
285 ,
17 ,
1988 ,
<a href="datasets/Congressional+Voting+Records">Congressional Voting Records</a>
,
Multivariate ,
Classification ,
Categorical ,
435 ,
16 ,
1987 ,
<a href="datasets/Water+Treatment+Plant">Water Treatment Plant</a> ,
Multivariate ,
Clustering ,
Integer, Real ,
527 ,
38 ,
1993 ,
<a href="datasets/Waveform+Database+Generator+%28Version+1%29">Waveform Dat
abase Generator (Version 1)</a> ,
Multivariate, Data-Generator ,
Classification ,
Real ,
5000 ,
21 ,
1988 ,
<a href="datasets/Waveform+Database+Generator+%28Version+2%29">Waveform Dat
abase Generator (Version 2)</a> ,
Multivariate, Data-Generator ,
Classification ,
Real ,
5000 ,
40 ,
1988 ,
<a href="datasets/Wine">Wine</a> ,
Multivariate ,
Classification ,
Integer, Real ,
178 ,
13 ,
1991 ,
<a href="datasets/Yeast">Yeast</a> ,
Multivariate ,
Classification ,
Real ,
1484 ,
8 ,
1996 ,

<a href="datasets/Zoo">Zoo</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
101 ,
17 ,
1990 ,
<a href="datasets/Undocumented">Undocumented</a> ,
 ,
 ,
 ,
 ,
 ,
 ,
<a href="datasets/Twenty+Newsgroups">Twenty Newsgroups</a> ,
Text ,
 ,
 ,
20000 ,
 ,
1999 ,
<a href="datasets/Australian+Sign+Language+signs">Australian Sign Language signs</a> ,
Multivariate, Time-Series ,
Classification ,
Categorical, Real ,
6650 ,
15 ,
1999 ,
<a href="datasets/Australian+Sign+Language+signs+%28High+Quality%29">Australian Sign Language signs (High Quality)</a> ,
Multivariate, Time-Series ,
Classification ,
Real ,
2565 ,
22 ,
2002 ,
<a href="datasets/US+Census+Data+%281990%29">US Census Data (1990)</a> ,
Multivariate ,
Clustering ,
Categorical ,
2458285 ,
68 ,
 ,
<a href="datasets/Census-Income+%28KDD%29">Census-Income (KDD)</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
299285 ,
40 ,
2000 ,
<a href="datasets/Coil+1999+Competition+Data">Coil 1999 Competition Data</a> ,
Multivariate ,
 ,
Categorical, Real ,
340 ,
17 ,
1999 ,
<a href="datasets/Corel+Image+Features">Corel Image Features</a> ,
Multivariate ,

Multivariate ,
 ,
Real ,
68040 ,
89 ,
1999 ,
<a href="datasets/E.+Coli+Genes">E. Coli Genes</a> ,
Relational ,
 ,
 ,
 ,
 ,
2001 ,
<a href="datasets/EEG+Database">EEG Database</a> ,
Multivariate, Time-Series ,
 ,
Categorical, Integer, Real ,
122 ,
4 ,
1999 ,
<a href="datasets/El+Nino">El Nino</a> ,
Spatio-temporal ,
 ,
Integer, Real ,
178080 ,
12 ,
1999 ,
<a href="datasets/Entree+Chicago+Recommendation+Data">Entree Chicago Recomm
endation Data</a> ,
Transactional, Sequential ,
Recommender-Systems ,
Categorical ,
50672 ,
 ,
2000 ,
<a href="datasets/CMU+Face+Images">CMU Face Images</a> ,
Image ,
Classification ,
Integer ,
640 ,
 ,
1999 ,
<a href="datasets/Insurance+Company+Benchmark+%28COIL+2000%29">Insurance C
ompany Benchmark (COIL 2000)</a> ,
Multivariate ,
Regression, Description ,
Categorical, Integer ,
9000 ,
86 ,
2000 ,
<a href="datasets/Internet+Usage+Data">Internet Usage Data</a> ,
Multivariate ,
 ,
Categorical, Integer ,
10104 ,
72 ,
1999 ,
<a href="datasets/IPUMS+Census+Database">IPUMS Census Database</a> ,
Multivariate ,
 ,
Categorical, Integer ,
256932 ,
61 ,

61 ,
1999 ,
<a href="datasets/Japanese+Vowels">Japanese Vowels</a> ,
Multivariate, Time-Series ,
Classification ,
Real ,
640 ,
12 ,
 ,
<a href="datasets/KDD+Cup+1998+Data">KDD Cup 1998 Data</a> ,
Multivariate ,
Regression ,
Categorical, Integer ,
191779 ,
481 ,
1998 ,
<a href="datasets/KDD+Cup+1999+Data">KDD Cup 1999 Data</a> ,
Multivariate ,
Classification ,
Categorical, Integer ,
4000000 ,
42 ,
1999 ,
<a href="datasets/M.+Tuberculosis+Genes">M. Tuberculosis Genes</a> ,
Relational ,
 ,
 ,
 ,
 ,
2001 ,
<a href="datasets/Movie">Movie</a> ,
Multivariate, Relational ,
 ,
 ,
10000 ,
 ,
1999 ,
<a href="datasets/MSNBC.com+Anonymous+Web+Data">MSNBC.com Anonymous Web Data</a> ,
Sequential ,
 ,
Categorical ,
989818 ,
 ,
 ,
<a href="datasets/NSF+Research+Award+Abstracts+1990-2003">NSF Research Award Abstracts 1990-2003</a> ,
Text ,
 ,
 ,
129000 ,
 ,
2003 ,
<a href="datasets/Pioneer-1+Mobile+Robot+Data">Pioneer-1 Mobile Robot Data</a> ,
Multivariate, Time-Series ,
 ,
Categorical, Real ,
 ,
 ,
1999 ,
<a href="datasets/Pseudo+Periodic+Synthetic+Time+Series">Pseudo Periodic Synthetic Time Series</a> ,

Time Series</a> ,
Univariate, Time-Series ,
 ,
 ,
100000 ,
 ,
1999 ,
<a href="datasets/Reuters-21578+Text+Categorization+Collection">Reuters-21578 Tex
t Categorization Collection</a> ,
Text ,
Classification ,
Categorical ,
21578 ,
5 ,
1997 ,
<a href="datasets/Robot+Execution+Failures">Robot Execution Failures</a> ,
Multivariate, Time-Series ,
Classification ,
Integer ,
463 ,
90 ,
1999 ,
<a href="datasets/Synthetic+Control+Chart+Time+Series">Synthetic Control Chart Time
Series</a> ,
Time-Series ,
Classification, Clustering ,
Real ,
600 ,
 ,
1999 ,
<a href="datasets/Syskill+and+Webert+Web+Page+Ratings">Syskill and Webert Web P
age Ratings</a> ,
Multivariate, Text ,
Classification ,
Categorical ,
332 ,
5 ,
1998 ,
<a href="datasets/UNIX+User+Data">UNIX User Data</a> ,
Text, Sequential ,
 ,
 ,
 ,
 ,
 ,
<a href="datasets/Volcanoes+on+Venus+-+JARtool+experiment">Volcanoes on Venus -
JARtool experiment</a> ,
Image ,
Classification ,
 ,
 ,
 ,
 ,
<a href="datasets/Statlog+%28Australian+Credit+Approval%29">Statlog (Australian Cre
dit Approval)</a> ,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
690 ,
14 ,
 ,
<a href="datasets/Statlog+%28German+Credit+Data%29">Statlog (German Credit Data
</a> ,

)</a> ,
 Multivariate ,
 Classification ,
 Categorical, Integer ,
 1000 ,
 20 ,
 1994 ,
 <a href="datasets/Statlog+%28Heart%29">Statlog (Heart)</a> ,
 Multivariate ,
 Classification ,
 Categorical, Real ,
 270 ,
 13 ,
 ...]

1 elemento es Name 2 elemento es Data Type 3 elemento es Default Task 4 elemento es Attriute Types 5 elemento es # Instances 6 elemento es # Attributes 7 elemento es Year Falta descripción de View List

In [87]:

```
#lista preliminar con Data Types
type_1= data_2[1::7]
type_1
```

Out[87]:

[Multivariate ,
 Multivariate ,
 Multivariate ,
  ,
 Multivariate ,
 Multivariate ,
 Multivariate ,
 Multivariate ,
 Multivariate ,
 Multivariate ,
 Univariate, Text ,
 Multivariate ,
 Multivariate ,
 Multivariate ,
 Multivariate ,
 Multivariate ,
 Multivariate ,
 Multivariate ,
 Multivariate ,
 Multivariate ,
 Multivariate, Data-Generator ,
 Multivariate ,
 Multivariate ,
 Domain-Theory ,
 Univariate, Time-Series ,
 Multivariate, Spatial ,
 Multivariate ,
 Multivariate, Domain-Theory ,
 Multivariate ,
 Multivariate ,
 Multivariate ,
 Multivariate ,
 Multivariate ,
 Multivariate, Time-Series ,
 Data-Generator ,
  ,
  ,

Multivariate ,
Multivariate ,
 ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate, Time-Series ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Relational ,
Multivariate ,
Multivariate, Data-Generator ,
Multivariate ,
Multivariate ,
Multivariate ,
Domain-Theory ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Domain-Theory ,
Sequential, Domain-Theory ,
Sequential ,
Sequential, Domain-Theory ,
Multivariate ,
Domain-Theory ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Domain-Theory ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Domain-Theory ,
Domain-Theory ,
Multivariate, Data-Generator ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
 ,
Domain-Theory ,
Multivariate ,
Multivariate ,

Multivariate, Domain-Theory ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate, Data-Generator ,
Multivariate, Data-Generator ,
Multivariate ,
Multivariate ,
Multivariate ,
 ,
Text ,
Multivariate, Time-Series ,
Multivariate, Time-Series ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Relational ,
Multivariate, Time-Series ,
Spatio-temporal ,
Transactional, Sequential ,
Image ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate, Time-Series ,
Multivariate ,
Multivariate ,
Relational ,
Multivariate, Relational ,
Sequential ,
Text ,
Multivariate, Time-Series ,
Univariate, Time-Series ,
Text ,
Multivariate, Time-Series ,
Time-Series ,
Multivariate, Text ,
Text, Sequential ,
Image ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
 ,
Domain-Theory ,
 ,
Multivariate ,
Multivariate, Time-Series ,
Multivariate, Time-Series ,
Multivariate ,
Multivariate ,
Multivariate, Sequential ,
Multivariate ,
Multivariate ,
Text ,

Text ,
Multivariate ,
Sequential ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate, Sequential, Time-Series ,
Multivariate ,
Multivariate ,
Time-Series ,
Multivariate ,
Multivariate, Sequential ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate, Sequential ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate, Time-Series ,
Multivariate ,
Multivariate ,
Multivariate ,
Text ,
Multivariate ,
Multivariate ,
Multivariate, Sequential ,
Multivariate, Time-Series ,
Univariate, Sequential, Time-Series ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate, Time-Series ,
Text ,
Domain-Theory ,
Multivariate, Sequential ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Time-Series ,
Time-Series ,
Multivariate, Text, Domain-Theory ,
Time-Series, Domain-Theory ,
Multivariate, Text, Domain-Theory ,
Text ,
Text ,
Multivariate, Univariate, Text ,
Multivariate, Univariate, Text ,
Multivariate ,
Text ,
Multivariate ,
Multivariate ,
Multivariate, Time-Series ,
Univariate ,
Multivariate, Text, Domain-Theory ,
Univariate ,
Univariate ,

Multivariate, Time-Series ,
Multivariate ,
Multivariate, Text ,
Multivariate, Time-Series ,
Multivariate ,
Multivariate, Univariate, Text ,
Sequential ,
Text ,
Multivariate, Time-Series ,
 ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate, Time-Series ,
Sequential, Text ,
Multivariate, Univariate, Time-Series ,
Time-Series, Multivariate ,
Multivariate ,
Sequential ,
Multivariate, Time-Series ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate, Time-Series ,
Multivariate ,
Multivariate ,
Multivariate, Sequential, Text ,
Multivariate ,
Multivariate ,
Multivariate, Sequential, Time-Series ,
Multivariate ,
Multivariate ,
Multivariate ,
Domain-Theory ,
Multivariate, Text ,
Multivariate, Time-Series ,
Multivariate, Sequential, Time-Series ,
Multivariate ,
Multivariate ,
Multivariate, Sequential, Time-Series, Text ,
Univariate ,
Multivariate, Sequential, Time-Series ,
Multivariate ,
Multivariate, Time-Series ,
 ,
 ,
Multivariate ,
Multivariate ,
Multivariate, Time-Series ,
Multivariate ,
Univariate, Sequential, Time-Series ,
Univariate, Sequential, Time-Series ,
Multivariate ,
Text ,
Multivariate ,
Multivariate ,
Multivariate ,
Text ,
Multivariate ,
Multivariate ,
Multivariate ,

Sequential ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate, Sequential, Time-Series ,
Univariate, Domain-Theory ,
Multivariate ,
Multivariate, Time-Series ,
 ,
Multivariate ,
Multivariate, Time-Series ,
Multivariate, Time-Series ,
Multivariate ,
Text ,
Time-Series ,
Time-Series ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate, Sequential ,
Multivariate ,
Multivariate, Time-Series ,
Multivariate ,
Time-Series ,
Multivariate, Time-Series ,
Multivariate, Univariate, Sequential, Text ,
Multivariate ,
Multivariate ,
Multivariate ,
 ,
Multivariate, Time-Series ,
Multivariate ,
Multivariate ,
Text ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Sequential, Time-Series ,
Multivariate ,
Multivariate, Sequential, Time-Series, Domain-Theory ,
Multivariate ,
Multivariate, Time-Series ,
Multivariate ,
Multivariate, Sequential, Time-Series ,
Multivariate, Time-Series ,
Multivariate, Sequential, Time-Series ,
Multivariate ,
Multivariate, Sequential, Time-Series ,
Multivariate, Sequential, Time-Series ,
Multivariate ,
Multivariate ,
Multivariate, Sequential, Time-Series ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate, Time-Series ,
Multivariate ,
Multivariate ,
Multivariate, Time-Series ,
Multivariate, Time-Series, Domain-Theory ,

Multivariate, Time-Series ,
Multivariate ,
Time-Series ,
Multivariate ,
Multivariate, Sequential, Time-Series ,
Multivariate ,
Multivariate ,
Multivariate ,
Text ,
Multivariate ,
Multivariate ,
Multivariate, Time-Series ,
Text ,
Multivariate, Text ,
Multivariate, Sequential, Time-Series ,
Multivariate ,
Multivariate ,
Text ,
Multivariate, Time-Series ,
Multivariate, Sequential ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate, Time-Series ,
Multivariate, Time-Series ,
Multivariate, Time-Series ,
 ,
Multivariate ,
Multivariate ,
Multivariate ,
Sequential, Time-Series ,
Multivariate, Time-Series ,
Multivariate ,
Multivariate, Time-Series ,
 ,
Text ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate, Sequential, Time-Series ,
 ,
Text ,
Multivariate ,
Multivariate ,
Text ,
Multivariate ,
Time-Series ,
Text ,
 ,
 ,
Multivariate, Time-Series ,
Multivariate ,
Sequential, Text ,
Multivariate, Text ,
Multivariate ,
Text ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,

Text ,
 ,
Sequential ,
Univariate ,
Univariate ,
Multivariate ,
Multivariate ,
Multivariate, Time-Series, Text ,
Multivariate ,
Multivariate ,
Multivariate, Sequential, Time-Series ,
Univariate, Domain-Theory ,
Multivariate ,
Text ,
Sequential ,
Multivariate ,
Multivariate ,
Multivariate, Sequential ,
Multivariate, Time-Series ,
Multivariate ,
Multivariate, Time-Series ,
Univariate ,
Multivariate ,
Multivariate, Text ,
Multivariate ,
Multivariate, Time-Series ,
Multivariate ,
Text ,
Time-Series ,
Multivariate, Text ,
Multivariate, Time-Series ,
Text ,
Multivariate ,
Multivariate ,
Multivariate, Text ,
Multivariate, Text ,
Multivariate ,
Multivariate ,
Multivariate, Time-Series ,
Multivariate, Sequential, Time-Series ,
Multivariate ,
Multivariate ,
Univariate ,
Multivariate ,
Multivariate ,
Univariate ,
Time-Series ,
Time-Series ,
Multivariate ,
Multivariate, Text ,
Multivariate ,
Multivariate ,
 ,
Multivariate ,
Time-Series ,
Multivariate, Univariate, Sequential, Time-Series ,
Multivariate, Time-Series ,
Multivariate, Text ,
Multivariate, Text ,
Multivariate ,
 In Collaboration With:]

In [88]:

```
#lista preliminar con Default Task
task_1= data_2[2::7]
task_1
```

Out[88]:

```
[<p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Recommender-Systems </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Regression </p>,
 <p class="normal">Regression </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification, Regression </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal"> </p>,
 <p class="normal"> </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Regression </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal"> </p>,
 <p class="normal"> </p>,
 <p class="normal"> </p>,
 <p class="normal"> </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Function-Learning </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal"> </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Classification </p>,
 <p class="normal">Relational-Learning </p>,
```

 ,
Classification ,
Classification ,
Classification ,
 ,
 ,
Classification ,
Classification ,
Classification ,
Classification ,
 ,
Classification ,
Classification ,
Classification ,
Classification ,
 ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
 ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
 ,
 ,
Classification ,
Regression ,
Classification ,
Regression ,
Classification ,
Classification ,
Regression ,
Classification ,
Classification ,
Classification ,
Classification ,
Clustering ,
 ,
 ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Clustering ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
 ,
 ,
Classification ,
Classification ,
Clustering ,
Classification ,
 ,

 ,
 ,
 ,
 ,
Recommender-Systems ,
Classification ,
Regression, Description ,
 ,
 ,
Classification ,
Regression ,
Classification ,
 ,
 ,
 ,
 ,
 ,
 ,
Classification ,
Classification ,
Classification, Clustering ,
Classification ,
 ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
 ,
Classification ,
Classification ,
 ,
 ,
 ,
 ,
 ,
Classification ,
Classification ,
Classification ,
Classification ,
Regression ,
Classification ,
Clustering ,
Regression ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Causal-Discovery ,
Classification ,
Classification, Clustering ,
Classification ,
Classification ,
Classification ,
Classification, Causal-Discovery ,
Clustering ,

Classification, Clustering ,
Regression ,
Regression ,
Classification ,
Classification, Regression ,
Classification ,
Classification ,
Regression ,
Classification ,
 ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Regression ,
Classification ,
 ,
Regression ,
Classification ,
Classification ,
Classification ,
Regression ,
Classification ,
Classification ,
Classification ,
Classification ,
Regression, Clustering, Causal-Discovery ,
Classification, Clustering ,
Classification ,
Classification ,
Classification, Regression, Clustering ,
Classification, Regression, Clustering ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification, Clustering ,
Classification ,
Classification ,
Classification ,
 ,
Classification ,
Regression, Clustering ,
Classification, Clustering ,
Classification ,
 ,
Classification ,
Classification, Clustering ,
Classification ,
Classification, Regression ,
Regression ,
Classification, Regression ,
Classification ,
Regression, Clustering ,
Classification, Regression ,
Regression, Classification ,

Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification, Clustering ,
Classification, Clustering ,
Classification ,
Clustering ,
Classification, Clustering ,
Classification ,
Classification ,
Regression ,
Classification ,
Classification ,
Classification ,
Classification, Clustering ,
Classification, Regression, Clustering, Causa ,
Classification, Clustering ,
Regression ,
Classification ,
Regression ,
Regression ,
 ,
Classification ,
 ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification, Clustering ,
Classification ,
Classification, Clustering ,
Classification, Clustering ,
Classification ,
Classification, Clustering ,
Classification, Regression, Clustering ,
Regression ,
Classification, Clustering ,
Classification ,
Regression ,
Classification ,
Classification, Clustering ,
Classification ,
Classification, Clustering, Causal-Discovery ,
Classification, Regression, Clustering ,
Classification, Regression ,
Classification, Clustering ,
Classification, Clustering ,
Regression ,
Classification ,
Classification ,
Clustering ,
Classification, Regression ,
Classification, Regression ,
Classification, Regression ,
Classification ,
Classification, Clustering ,
Classification ,
Clustering ,

Classification, Regression ,
Regression ,
Classification, Clustering ,
Classification, Regression ,
Classification ,
Classification, Regression ,
Regression, Clustering ,
Classification, Regression ,
Classification, Causal-Discovery ,
Classification ,
Classification ,
Classification, Clustering ,
Classification ,
Regression ,
Classification ,
Classification ,
Classification ,
Classification, Regression ,
Classification ,
Regression, Clustering, Causal-Discovery ,
Regression ,
Classification ,
Classification ,
Classification, Clustering ,
Clustering, Causal-Discovery ,
Classification, Regression ,
Classification ,
Classification, Clustering ,
Classification, Regression, Clustering ,
Classification, Clustering ,
Classification, Regression, Clustering ,
Classification ,
Classification ,
Classification, Regression, Clustering ,
Classification ,
Classification ,
Classification, Clustering ,
Causal-Discovery ,
Classification, Regression ,
Classification ,
Classification ,
Classification, Regression, Clustering ,
Classification, Clustering ,
Regression ,
Classification, Regression ,
Classification ,
Regression ,
Classification ,
Classification ,
Classification ,
Classification ,
Regression ,
Causal-Discovery ,
Clustering ,
Classification, Clustering ,
Classification ,
Regression ,
Classification, Clustering, Causal-Discovery ,
Classification, Regression ,
Classification, Regression, Clustering ,
Classification, Regression, Clustering ,
Classification ,

Classification ,
Regression ,
Classification, Regression ,
Classification ,
Classification ,
Causal-Discovery ,
Classification, Clustering ,
Regression ,
Classification, Clustering ,
Classification ,
Regression ,
Classification, Clustering ,
Classification, Regression ,
Classification ,
Regression ,
Classification, Regression, Clustering ,
Clustering ,
Classification, Regression ,
 ,
Classification ,
Classification ,
Classification, Clustering ,
Classification ,
Classification, Clustering ,
Classification ,
Classification, Clustering ,
Classification, Clustering ,
Classification, Clustering ,
Classification ,
Regression ,
Classification, Regression ,
Classification ,
Classification ,
Classification, Regression ,
Classification ,
Clustering ,
Clustering ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification, Regression ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Regression ,
Classification ,
Classification, Clustering ,
Classification, Clustering ,
Classification, Regression ,
Regression ,
Clustering ,
Classification ,
Regression ,
Clustering ,

Clustering ,
Classification, Clustering ,
Classification, Clustering ,
Classification, Clustering ,
Classification, Regression ,
Regression ,
Classification, Regression ,
Classification ,
Classification ,
Regression ,
Classification, Clustering ,
Classification ,
Classification ,
Classification ,
Classification, Regression ,
Classification ,
Classification ,
Recommendation ,
Classification, Regression, Clustering ,
Classification, Regression, Clustering ,
Classification ,
Regression ,
Classification, Regression ,
Regression ,
Classification ,
Classification, Clustering ,
Classification ,
Classification ,
Classification, Regression ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification, Clustering ,
Regression ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification, Regression, Clustering ,
Classification, Regression ,
Classification, Clustering ,
Classification, Clustering ,
Classification]

In [89]:

```
#lista preliminar con Attribute Types
attri_1= data_2[3::7]
attri_1
```

Out[89]:

[Categorical, Integer, Real ,
Categorical, Integer ,
Categorical, Integer, Real ,
Categorical ,
Categorical, Integer, Real ,
Categorical, Integer, Real ,
Categorical ,
Categorical ,
Categorical, Real ,
Categorical, Integer, Real ,

 ,
Categorical ,
Categorical ,
Categorical ,
Integer ,
Real ,
Real ,
Categorical, Integer ,
Categorical ,
Categorical, Integer ,
Categorical, Integer ,
Categorical ,
Categorical, Integer ,
 ,
Categorical, Integer ,
Categorical ,
Categorical, Integer, Real ,
Categorical, Real, Integer ,
Integer ,
Categorical, Integer ,
Categorical, Integer ,
Categorical, Integer, Real ,
Categorical, Integer ,
Categorical, Integer ,
Real ,
 ,
 ,
Categorical, Integer, Real ,
Real ,
Categorical, Integer ,
Real ,
Real ,
Integer ,
Categorical ,
Categorical, Integer, Real ,
Categorical, Integer, Real ,
Categorical, Integer, Real ,
Real ,
Real ,
Categorical, Integer, Real ,
Integer, Real ,
Real ,
Real ,
Categorical ,
Categorical, Integer, Real ,
Categorical ,
Categorical ,
Integer ,
Categorical, Integer, Real ,
 ,
Integer ,
Categorical ,
Categorical, Integer, Real ,
Categorical, Integer, Real ,
Categorical, Integer, Real ,
Categorical ,
Categorical ,
Categorical ,
Categorical ,
 ,
Integer, Real ,
Categorical ,

Integer ,
Integer ,
Categorical ,
 ,
Integer, Real ,
Integer ,
Integer ,
Categorical, Integer ,
Categorical ,
 ,
 ,
Real ,
Categorical, Integer ,
Categorical ,
Categorical ,
Categorical ,
Categorical ,
Integer ,
Integer, Real ,
Integer, Real ,
Categorical ,
Integer ,
Categorical, Integer ,
 ,
 ,
Categorical, Integer ,
Categorical ,
Categorical, Real ,
Categorical ,
Categorical, Integer ,
Categorical ,
Integer, Real ,
Real ,
Real ,
Integer, Real ,
Real ,
Categorical, Integer ,
 ,
 ,
Categorical, Real ,
Real ,
Categorical ,
Categorical, Integer ,
Categorical, Real ,
Real ,
 ,
Categorical, Integer, Real ,
Integer, Real ,
Categorical ,
Integer ,
Categorical, Integer ,
Categorical, Integer ,
Categorical, Integer ,
Real ,
Categorical, Integer ,
Categorical, Integer ,
 ,
 ,
Categorical ,
 ,
Categorical, Real ,
 ,

Categorical ,
Integer ,
Real ,
Categorical ,
 ,
 ,
Categorical, Integer, Real ,
Categorical, Integer ,
Categorical, Real ,
Integer ,
Real ,
Integer ,
Integer ,
Categorical ,
Real ,
Real ,
 ,
 ,
Real ,
Categorical, Integer ,
Categorical, Integer ,
Categorical, Integer ,
Real ,
Integer ,
Integer ,
Real ,
 ,
Integer ,
Real ,
Real ,
Real ,
Integer ,
Integer ,
Integer ,
Real ,
Real ,
Integer ,
Real ,
Real ,
Real ,
Integer ,
Integer ,
Real ,
Categorical ,
Real ,
Real ,
Real ,
Categorical, Integer ,
Real ,
Integer, Real ,
Real ,
Integer, Real ,
Integer ,
 ,
Real ,
Real ,
Real ,
Real ,
Real ,
Categorical, Integer, Real ,
Integer, Real ,
Real ,

Real ,
Real ,
 ,
Real ,
Integer ,
Integer, Real ,
Real ,
Real ,
Real ,
Real ,
Real ,
Real ,
 ,
Real ,
 ,
 ,
Integer, Real ,
Integer, Real ,
Real ,
 ,
Real ,
Integer, Real ,
Real ,
Real ,
Real ,
Real ,
Real ,
Real ,
 ,
Integer ,
Real ,
Real ,
Integer, Real ,
Integer ,
 ,
 ,
Real ,
Integer, Real ,
Real ,
Real ,
Real ,
Real ,
Real ,
Integer, Real ,
Real ,
Integer, Real ,
Real ,
Real ,
Real ,
Integer, Real ,
 ,
Real ,
Integer ,
Real ,
 ,
 ,
Real ,
Integer, Real ,
Real ,
Real ,
Real ,
Integer ,
Real ,
Integer, Real ,
Real ,
Integer, Real ,
Real ,
Real ,
Real ,
Integer, Real ,
Real ,
Real ,
Real ,
Real ,
Integer ,
Real ,
Real ,
Integer, Real ,
Real ,
Real ,
Real ,
Integer ,
Real ,
Real ,
Real ,
Integer, Real ,
Real ,
Real ,
Real ,
Integer, Real ,
Real ,
Real ,
Real ,
Integer, Real ,
Real ,
Real ,
Real ,
Integer, Real ,
Real ,
Integer ,
Real ,
 ,
 ,
Real ,
Integer, Real ,
Real ,
Real

Integer ,
Integer, Real ,
Real ,
 ,
Integer, Real ,
Real ,
Real ,
Integer, Real ,
Integer, Real ,
Integer, Real ,
Real ,
Real ,
Real ,
 ,
Real ,
 ,
 ,
Real ,
Real ,
Real ,
 ,
Real ,
Real ,
Integer ,
 ,
Real ,
 ,
Integer ,
 ,
Real ,
Integer, Real ,
Integer, Real ,
Real ,
Integer ,
Integer, Real ,
Real ,
 ,
 ,
Real ,
Real ,
Integer, Real ,
Integer ,
Integer, Real ,
Real ,
 ,
Real ,
Real ,
Real ,
Real ,
Real ,
Integer ,
Real ,
Real ,
Integer, Real ,
 ,
Real ,
Real ,
Integer ,
Real ,
Integer, Real ,
Categorical ,
 ,

Integer, Real ,
 ,
 ,
Integer, Real ,
Real ,
Real ,
 ,
Real ,
Real ,
Real ,
Real ,
Integer, Real ,
Real ,
Integer ,
Real ,
Real ,
Integer ,
Integer, Real ,
Real ,
Integer, Real ,
Integer ,
Real ,
Real ,
Real ,
Real ,
 ,
Real ,
Real ,
Real ,
Integer, Real ,
Real ,
Real ,
Real ,
 ,
Integer ,
 ,
Integer ,
Real ,
Real ,
Real ,
Integer ,
Integer ,
Integer, Real ,
Real ,
Integer ,
 ,
Integer, Real ,
Integer ,
Integer, Real ,
Real ,
Integer ,
Real ,
Real ,
Integer, Real ,
Integer ,
Real ,
Integer, Real ,
Integer, Real ,
Integer, Real ,
Integer, Real ,
Integer ,
Integer, Real ,

Integer ,
 ,
Integer, Real ,
 ,
Real ,
Real ,
 ,
Integer ,
Real ,
Real ,
Integer ,
Real ,
Integer ,
Integer ,
Integer, Real ,
Integer, Real ,
Integer ,
Integer ,
Real ,
Integer ,
Integer, Real ,
Integer ,
Integer ,
Integer ,
Integer, Real ,
Real ,
Integer, Real ,
Integer ,
 ,
Integer ,
Real ,
Integer, Real ,
Integer, Real ,
Real ,
 ,
Integer, Real ,
Real ,
 ,
Integer ,
Real ,
Real ,
Real ,
Integer ,
Integer ,
Real ,
Real ,
Integer, Real ,
 ,
Real ,
Real ,
Integer, Real ,
Integer ,
Integer ,
Real ,
 ,
 ,
Integer ,
Integer ,
Integer ,
 ,
Real ,
Categorical ,

Integer ,
Integer ,
Integer ,
Real ,
Real ,
Real ,
 ,
Integer, Real ,
 ,
Integer, Real ,
Real ,
Integer ,
 ,
 ,
Real ,
Real ,
Integer, Real ,
Real ,
Integer ,
Real ,
Real ,
Real ,
Integer, Real ,
Real ,
Real ,
Integer]

In [91]:

```
#lista preliminar con # Instances
inst_1= data_2[4::7]
inst_1
```

Out[91]:

[4177 ,
48842 ,
798 ,
37711 ,
452 ,
6000 ,
226 ,
226 ,
398 ,
205 ,
294 ,
625 ,
16 ,
286 ,
699 ,
198 ,
569 ,
108 ,
1728 ,
48842 ,
 ,
3196 ,
28056 ,
 ,
100 ,
67557 ,
690 ,
125

209 ,
1473 ,
581012 ,
512 ,
366 ,
 ,
 ,
 ,
 ,
132 ,
336 ,
194 ,
352 ,
214 ,
306 ,
160 ,
303 ,
155 ,
368 ,
 ,
2310 ,
3279 ,
351 ,
150 ,
7797 ,
104 ,
57 ,
 ,
24 ,
20000 ,
345 ,
 ,
32 ,
148 ,
209 ,
528 ,
 ,
106 ,
128 ,
3190 ,
432 ,
202 ,
2000 ,
8124 ,
476 ,
6598 ,
12960 ,
 ,
5473 ,
5620 ,
10992 ,
90 ,
339 ,
 ,
 ,
 ,
167 ,
15 ,
1389 ,
307 ,
47 ,
23 ,

531 ,
4601 ,
267 ,
267 ,
76 ,
 ,
1000 ,
151 ,
958 ,
7200 ,
10 ,
285 ,
435 ,
527 ,
5000 ,
5000 ,
178 ,
1484 ,
101 ,
 ,
20000 ,
6650 ,
2565 ,
2458285 ,
299285 ,
340 ,
68040 ,
 ,
122 ,
178080 ,
50672 ,
640 ,
9000 ,
10104 ,
256932 ,
640 ,
191779 ,
4000000 ,
 ,
10000 ,
989818 ,
129000 ,
 ,
100000 ,
21578 ,
463 ,
600 ,
332 ,
 ,
 ,
690 ,
1000 ,
270 ,
6435 ,
2310 ,
58000 ,
946 ,
20008 ,
208 ,
528 ,
 ,

 ,
1024 ,
10080 ,
50400 ,
1025010 ,
19020 ,
1364 ,
961 ,
517 ,
200 ,
8000000 ,
1030 ,
606 ,
900 ,
2600 ,
1950 ,
13500 ,
4400 ,
2536 ,
300 ,
197 ,
2858 ,
748 ,
11640 ,
1593 ,
1567 ,
22632 ,
360 ,
103 ,
1994 ,
120 ,
4898 ,
2396130 ,
16772 ,
5875 ,
503 ,
51 ,
106 ,
2126 ,
5456 ,
8800 ,
164860 ,
 ,
1941 ,
130065 ,
515345 ,
440 ,
 ,
53500 ,
8235 ,
 ,
5749132 ,
2215 ,
310 ,
10000 ,
3000 ,
1500 ,
30000 ,
2500 ,
4143 ,
64 ,
53414 ,

65554 ,
45211 ,
1138562 ,
13910 ,
583 ,
2551 ,
34465 ,
5574 ,
245057 ,
182 ,
3850505 ,
138 ,
1080 ,
2075259 ,
210 ,
115 ,
3960456 ,
 ,
10299 ,
1600 ,
768 ,
308 ,
100 ,
237 ,
434874 ,
536 ,
140000 ,
6118 ,
165632 ,
18000 ,
540 ,
931 ,
1055 ,
100 ,
9120 ,
403 ,
111740 ,
10421 ,
5820 ,
403 ,
14980 ,
45730 ,
2584 ,
1372 ,
306 ,
120000 ,
13910 ,
2747 ,
3395 ,
39242 ,
4137 ,
17389 ,
51 ,
470 ,
132 ,
5000000 ,
11000000 ,
250 ,
126 ,
 ,
4889 ,
 ,

 ,
340 ,
501 ,
45781 ,
1503 ,
440 ,
2000 ,
9568 ,
168 ,
100000 ,
5665 ,
79 ,
127 ,
1040 ,
9900 ,
560 ,
60021 ,
1419 ,
101 ,
399 ,
58 ,
180 ,
21048 ,
 ,
750 ,
3000 ,
150 ,
1059 ,
11934 ,
27965 ,
216 ,
120 ,
649 ,
370 ,
4178504 ,
221579 ,
10800 ,
58509 ,
129685 ,
2456 ,
2921 ,
1151 ,
6590 ,
3000 ,
39797 ,
326 ,
913 ,
168286 ,
400 ,
314080 ,
637 ,
1710671 ,
12000 ,
10929 ,
1080 ,
40000 ,
43930257 ,
230318 ,
10500000 ,
13197 ,
 ,
30000 ,

324 ,
541909 ,
11164866 ,
163 ,
373 ,
20560 ,
40 ,
422937 ,
9358 ,
640 ,
919438 ,
40949 ,
5744 ,
10503 ,
42240 ,
102944 ,
500 ,
9782222 ,
11463 ,
17898 ,
1885 ,
19735 ,
1540 ,
4007 ,
153540 ,
606 ,
1353 ,
1956 ,
43824 ,
3942 ,
858 ,
287 ,
17764280 ,
106574 ,
9358 ,
11500 ,
92000 ,
315 ,
78095 ,
130 ,
74 ,
52854 ,
77 ,
811 ,
504 ,
401 ,
2856 ,
10546 ,
801 ,
1540 ,
1451 ,
1075 ,
78095 ,
7195 ,
3600 ,
76 ,
60 ,
405 ,
303 ,
303 ,
107888 ,
76000 ,

10000 ,
180 ,
745000 ,
12234 ,
292 ,
104 ,
60000 ,
2000 ,
165 ,
217 ,
1175 ,
704 ,
75128 ,
90 ,
90 ,
50 ,
71 ,
93239 ,
540 ,
105 ,
6611 ,
15 ,
372 ,
58000 ,
4960 ,
241600 ,
130000 ,
7062606 ,
740 ,
70 ,
2205 ,
10721 ,
640 ,
1000 ,
116 ,
1672 ,
322 ,
93600 ,
3060 ,
5879 ,
9200 ,
20000 ,
20867 ,
 ,
4143 ,
215063 ,
6000000 ,
21263 ,
63000000 ,
10190 ,
300 ,
12330 ,
5180 ,
756 ,
10000 ,
80 ,
1184 ,
1047 ,
777 ,
249 ,
414 ,
 ,

143 ,
7840 ,
30000 ,
35717 ,
135 ,
980 ,
5456 ,
120]

In [92]:

```
#lista preliminar con # Attributes
numatt_1= data_2[5::7]
numatt_1
```

Out[92]:

[8 ,
14 ,
38 ,
294 ,
279 ,
7 ,
 ,
69 ,
8 ,
26 ,
1 ,
4 ,
4 ,
9 ,
10 ,
34 ,
32 ,
13 ,
6 ,
14 ,
22 ,
36 ,
6 ,
 ,
6 ,
42 ,
15 ,
 ,
9 ,
9 ,
54 ,
39 ,
33 ,
20 ,
 ,
 ,
 ,
12 ,
8 ,
30 ,
 ,
10 ,
3 ,
5 ,
75 ,

19 ,
27 ,
 ,
19 ,
1558 ,
34 ,
4 ,
617 ,
12 ,
16 ,
7 ,
4 ,
16 ,
7 ,
 ,
56 ,
18 ,
8 ,
22 ,
 ,
58 ,
 ,
61 ,
7 ,
 ,
649 ,
22 ,
168 ,
168 ,
8 ,
 ,
10 ,
64 ,
16 ,
8 ,
17 ,
 ,
 ,
72 ,
4 ,
6 ,
10 ,
35 ,
35 ,
4 ,
102 ,
57 ,
22 ,
44 ,
45 ,
 ,
 ,
5 ,
9 ,
21 ,
32 ,
17 ,
16 ,
38 ,
21 ,
40 ,
13 ,

8 ,
17 ,
 ,
 ,
15 ,
22 ,
68 ,
40 ,
17 ,
89 ,
 ,
4 ,
12 ,
 ,
 ,
86 ,
72 ,
61 ,
12 ,
481 ,
42 ,
 ,
 ,
 ,
 ,
 ,
 ,
5 ,
90 ,
 ,
5 ,
 ,
 ,
14 ,
20 ,
13 ,
36 ,
19 ,
9 ,
18 ,
4 ,
60 ,
10 ,
 ,
 ,
10 ,
4 ,
3 ,
11 ,
11 ,
 ,
6 ,
13 ,
 ,
100000 ,
9 ,
101 ,
10000 ,
20000 ,
100000 ,
5000 ,
500 ,

73 ,
43 ,
23 ,
3 ,
5 ,
 ,
256 ,
591 ,
70 ,
91 ,
10 ,
128 ,
6 ,
12 ,
3231961 ,
5409 ,
26 ,
 ,
 ,
10 ,
23 ,
24 ,
13 ,
8 ,
 ,
27 ,
50 ,
90 ,
138672 ,
 ,
386 ,
 ,
 ,
12 ,
147 ,
6 ,
8 ,
27 ,
10000 ,
20000 ,
10000 ,
54877 ,
4702 ,
24 ,
29 ,
17 ,
3 ,
128 ,
10 ,
242 ,
120 ,
 ,
4 ,
13 ,
52 ,
47 ,
857 ,
9 ,
7 ,
200 ,
4 ,
 ,

561 ,
64 ,
8 ,
7 ,
10 ,
9 ,
4 ,
8 ,
77 ,
51 ,
18 ,
1950000 ,
18 ,
1300 ,
41 ,
6 ,
5625 ,
5 ,
 ,
7 ,
33 ,
5 ,
15 ,
9 ,
19 ,
5 ,
5 ,
1000000 ,
129 ,
 ,
20 ,
152 ,
24 ,
16 ,
35 ,
17 ,
5 ,
18 ,
28 ,
7 ,
309 ,
3 ,
6 ,
 ,
 ,
16 ,
13 ,
5 ,
6 ,
8 ,
2 ,
4 ,
148 ,
55 ,
17 ,
8 ,
42 ,
26 ,
50 ,
2 ,
281 ,
120 ,

 ,
6 ,
120432 ,
150000 ,
529 ,
 ,
16 ,
2500 ,
5 ,
68 ,
16 ,
100 ,
216 ,
23 ,
33 ,
140256 ,
19 ,
20 ,
20 ,
49 ,
12 ,
30 ,
5232 ,
20 ,
1 ,
 ,
61 ,
27 ,
53 ,
11 ,
25 ,
0 ,
20 ,
9 ,
3 ,
561 ,
82 ,
13 ,
16 ,
13 ,
28 ,
4 ,
 ,
24 ,
34 ,
8 ,
128 ,
15 ,
513 ,
7 ,
7 ,
5 ,
15 ,
480000 ,
11 ,
54 ,
561 ,
64 ,
6 ,
116 ,
19 ,
 ,

5812 ,
9 ,
32 ,
29 ,
67 ,
 ,
25 ,
6400 ,
10 ,
5 ,
13 ,
98 ,
36 ,
69 ,
2158859 ,
518 ,
15 ,
179 ,
 ,
12 ,
38 ,
65 ,
102 ,
86 ,
7 ,
53 ,
20 ,
1 ,
71 ,
29 ,
20531 ,
65 ,
3 ,
22 ,
38 ,
22 ,
4814 ,
698 ,
13 ,
10 ,
59 ,
56 ,
482 ,
171 ,
5 ,
500 ,
411 ,
8519 ,
21 ,
21 ,
171 ,
7 ,
49 ,
12 ,
3 ,
21 ,
9 ,
8 ,
7 ,
2 ,
11 ,
11

173 ,
5 ,
15 ,
3 ,
105 ,
25000 ,
 ,
18 ,
21000 ,
115 ,
21 ,
206 ,
43680 ,
8 ,
10 ,
59 ,
10 ,
5 ,
5 ,
1000 ,
138 ,
 ,
16 ,
2 ,
10 ,
 ,
8 ,
6 ,
129 ,
81 ,
12 ,
6 ,
22 ,
18 ,
9 ,
754 ,
14 ,
5 ,
 ,
 ,
18 ,
7 ,
7 ,
 ,
7 ,
5 ,
6 ,
4 ,
18 ,
11 ,
25 ,
 ]

In [93]:

```
#lista preliminar con Year
year_1= data_2[6::7]
year_1
```

Out[93]:

[1995 ,

1996 ,
 ,
1998 ,
1998 ,
1992 ,
1987 ,
1992 ,
1993 ,
1987 ,
1994 ,
1994 ,
 ,
1988 ,
1992 ,
1995 ,
1995 ,
1990 ,
1997 ,
1996 ,
1988 ,
1989 ,
1994 ,
 ,
 ,
1995 ,
 ,
1992 ,
1987 ,
1997 ,
1998 ,
1995 ,
1998 ,
 ,
 ,
1994 ,
 ,
1989 ,
1996 ,
1990 ,
1990 ,
1987 ,
1999 ,
1989 ,
1988 ,
1988 ,
1989 ,
 ,
1990 ,
1998 ,
1989 ,
1988 ,
1994 ,
1990 ,
1988 ,
1988 ,
1990 ,
1991 ,
1990 ,
 ,
1992 ,
1988 ,
1990 ,

1996 ,
1995 ,
1990 ,
 ,
1992 ,
1992 ,
1994 ,
 ,
1987 ,
1994 ,
1994 ,
1997 ,
1991 ,
1995 ,
1998 ,
1998 ,
1993 ,
1988 ,
 ,
 ,
1992 ,
1993 ,
1988 ,
1989 ,
1988 ,
1987 ,
1993 ,
1988 ,
1999 ,
2001 ,
2001 ,
 ,
1992 ,
1993 ,
1997 ,
1991 ,
1987 ,
1994 ,
1988 ,
1987 ,
1993 ,
1988 ,
1988 ,
1991 ,
1996 ,
1990 ,
 ,
1999 ,
1999 ,
2002 ,
 ,
2000 ,
1999 ,
1999 ,
2001 ,
1999 ,
1999 ,
2000 ,
1999 ,
2000 ,
1999 ,
1999 ,

 ,
1998 ,
1999 ,
2001 ,
1999 ,
 ,
2003 ,
1999 ,
1999 ,
1997 ,
1999 ,
1999 ,
1998 ,
 ,
 ,
 ,
1994 ,
 ,
1993 ,
1990 ,
 ,
 ,
 ,
 ,
 ,
 ,
 ,
1989 ,
2006 ,
2006 ,
2007 ,
2007 ,
2007 ,
2007 ,
2008 ,
2008 ,
2008 ,
2007 ,
2008 ,
2008 ,
2008 ,
2008 ,
2008 ,
2008 ,
2008 ,
2008 ,
2008 ,
2008 ,
2008 ,
2008 ,
2008 ,
2009 ,
2008 ,
2008 ,
2008 ,
2008 ,
2009 ,
2009 ,
2009 ,
2009 ,
2009 ,
2009 ,
2010 ,
2009 ,
2010 ,

2010 ,
2010 ,
2010 ,
2010 ,
2010 ,
2010 ,
2010 ,
2010 ,
2010 ,
2011 ,
2011 ,
2011 ,
2011 ,
2011 ,
2011 ,
2011 ,
2011 ,
2011 ,
2011 ,
2011 ,
2011 ,
2011 ,
2011 ,
2011 ,
2011 ,
2012 ,
2012 ,
2012

2013 ,
2013 ,
2013 ,
2013 ,
2013 ,
2013 ,
2013 ,
2013 ,
2013 ,
2013 ,
2013 ,
2013 ,
2013 ,
2013 ,
2014 ,
2013 ,
2013 ,
2013 ,
2014 ,
2014 ,
2014 ,
2014 ,
2014 ,
2014 ,
2014 ,
2014 ,
2014 ,
2014 ,
2013 ,
2014 ,
2014 ,
2014 ,
2014 ,
2014 ,
2014 ,
2014 ,
2014 ,
2014 ,
2014

2015 ,
2015 ,
2015 ,
2015 ,
2015 ,
2014 ,
2015 ,
2015 ,
2015 ,
2015 ,
2015 ,
2015 ,
2015 ,
2015 ,
2015 ,
2015 ,
2015 ,
2015 ,
2015 ,
2015 ,
2015 ,
2015 ,
2015 ,
2016 ,
2016 ,
2015 ,
2016 ,
2016 ,
2015 ,
2016 ,
2016 ,
2016 ,
2016 ,
2016 ,
2016 ,
2016 ,
2016 ,
2016 ,
2016 ,
2016 ,
2016 ,
2016 ,
2016 ,
2016 ,
2016 ,
2016 ,
2017 ,
2016 ,
2017 ,
2016 ,
2017 ,
2017 ,
2016 ,
2016 ,
2017 ,
2017 ,
2016 ,
2017 ,
2017 ,
2017 ,
2017 ,
2016 ,
2017 ,

2017 ,
2016 ,
2016 ,
2016 ,
2017 ,
2017 ,
2017 ,
2017 ,
2017 ,
2017 ,
2017 ,
2016 ,
2016 ,
2016 ,
2016 ,
2017 ,
2017 ,
2017 ,
2017 ,
2017 ,
2016 ,
2017 ,
2017 ,
2017 ,
2017 ,
2017 ,
2017 ,
2017 ,
2016 ,
2016 ,
2016 ,
2017 ,
2017 ,
2017 ,
2017 ,
2017 ,
2017 ,
2017 ,
2017 ,
2016 ,
2018 ,
2018 ,
2016 ,
2018 ,
2018 ,
2018

2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2018 ,
2019 ,
2019 ,
2018 ,
2018 ,
2018 ,
2019]

In [94]:

```
#pasar las listas a string para sacar el valor deseado
type_2 = str(type_1)
task_2 = str(task_1)
attri_2 = str(attri_1)
inst_2 = str(inst_1)
numatt_2 = str(numatt_1)
year_2 = str(year_1)
```

In [133]:

```
#lista de Data Types
types_database = re.findall("\>(\w+)|\>(\s)\<",type_2)
types_database
```

Out[133]:

```
[('Multivariate', ''),
 ('Multivariate', ''),
 ('Multivariate', ''),
 ('', '\xa0'),
 ('Multivariate', ''),
 ('Multivariate', ''),
 ('Multivariate', ''),
 ('Multivariate', ''),
 ('Multivariate', ''),
```

('Multivariate', ''),
('Univariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Domain', ''),
('Univariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Data', ''),
('', '\xa0'),
('', '\xa0'),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('', '\xa0'),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Relational', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Domain', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Domain', ''),
('Sequential', ''),
('Sequential', ''),
('Sequential', ''),
('Multivariate', ''),
('Domain', ''),
('Multivariate', ''),

('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Domain', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Domain', ''),
('Domain', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('', '\xa0'),
('Domain', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('', '\xa0'),
('Text', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Relational', ''),
('Multivariate', ''),
('Spatio', ''),
('Transactional', ''),
('Image', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Relational', ''),
('Multivariate', ''),
('Sequential', ''),
('Text', ''),
('Multivariate', ''),

('Multivariate', ''),
('Univariate', ''),
('Text', ''),
('Multivariate', ''),
('Time', ''),
('Multivariate', ''),
('Text', ''),
('Image', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('', '\xa0'),
('Domain', ''),
('', '\xa0'),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Text', ''),
('Text', ''),
('Multivariate', ''),
('Sequential', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Time', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Text', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Univariate', ''),
('Multivariate', ''),
('Multivariate', ''),

('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Text', ''),
('Domain', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Time', ''),
('Time', ''),
('Multivariate', ''),
('Time', ''),
('Multivariate', ''),
('Text', ''),
('Text', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Text', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Univariate', ''),
('Multivariate', ''),
('Univariate', ''),
('Univariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Sequential', ''),
('Text', ''),
('Multivariate', ''),
('', '\xa0'),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Sequential', ''),
('Multivariate', ''),
('Time', ''),
('Multivariate', ''),
('Sequential', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate

('Multivariate', ''),
('Domain', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Univariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('', '\xa0'),
('', '\xa0'),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Univariate', ''),
('Univariate', ''),
('Multivariate', ''),
('Text', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Text', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Sequential', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Univariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('', '\xa0'),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Text', ''),
('Time', ''),
('Time', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Time', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('', '\xa0'),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),

('Text', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Sequential', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate

('Multivariate', ''),
('', '\xa0'),
('Text', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('', '\xa0'),
('Text', ''),
('Multivariate', ''),
('Multivariate', ''),
('Text', ''),
('Multivariate', ''),
('Time', ''),
('Text', ''),
('', '\xa0'),
('', '\xa0'),
('Multivariate', ''),
('Multivariate', ''),
('Sequential', ''),
('Multivariate', ''),
('Multivariate', ''),
('Text', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Text', ''),
('', '\xa0'),
('Sequential', ''),
('Univariate', ''),
('Univariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Univariate', ''),
('Multivariate', ''),
('Text', ''),
('Sequential', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Univariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Text', ''),
('Time', ''),
('Multivariate', ''),
('Multivariate', ''),
('Text', ''),
('Multivariate', ''),

('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Univariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Univariate', ''),
('Time', ''),
('Time', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('', '\xa0'),
('Multivariate', ''),
('Time', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', ''),
('Multivariate', '')]

In [134]:

```
len(types_database)
```

Out[134]:

469

In [142]:

```
types_database_1=[]
for element in types_database:
    i = element[0]
    types_database_1.append(i)

types_database_1
```

Out[142]:

['Multivariate',
 'Multivariate',
 'Multivariate',
 '',
 'Multivariate',
 'Multivariate',
 'Multivariate',
 'Multivariate',
 'Multivariate',
 'Multivariate',
 'Univariate',
 'Multivariate',
 'Multivariate',
 'Multivariate',
 'Multivariate',
 'Multivariate',

'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Domain',
'Univariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Data',
'',
'',
'Multivariate',
'Multivariate',
'Multivariate',
'',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Relational',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Domain',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Domain',
'Sequential',
'Sequential',
'Sequential',
'Multivariate',
'Domain',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Domain',
'Multivariate',
'Multivariate',

'Multivariate',
'Multivariate',
'Multivariate',
'Domain',
'Domain',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'',
'Domain',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'',
'Text',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Relational',
'Multivariate',
'Spatio',
'Transactional',
'Image',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Relational',
'Multivariate',
'Sequential',
'Text',
'Multivariate',
'Univariate',
'Text',
'Multivariate',
'Time',
'Multivariate',
'Text',
'Image',

'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'',
'Domain',
'',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Text',
'Text',
'Multivariate',
'Sequential',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Time',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Text',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Univariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Text',
'Domain',
'Multivariate',
'Multivariate',

'Multivariate',
'Multivariate',
'Multivariate',
'Time',
'Time',
'Multivariate',
'Time',
'Multivariate',
'Text',
'Text',
'Multivariate',
'Multivariate',
'Multivariate',
'Text',
'Multivariate',
'Multivariate',
'Multivariate',
'Univariate',
'Multivariate',
'Univariate',
'Univariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Sequential',
'Text',
'Multivariate',
'',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Sequential',
'Multivariate',
'Time',
'Multivariate',
'Sequential',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Domain',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',

'Univariate',
'Multivariate',
'Multivariate',
'Multivariate',
'',
'',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Univariate',
'Univariate',
'Multivariate',
'Text',
'Multivariate',
'Multivariate',
'Multivariate',
'Text',
'Multivariate',
'Multivariate',
'Multivariate',
'Sequential',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Univariate',
'Multivariate',
'Multivariate',
'',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Text',
'Time',
'Time',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Time',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'',
'Multivariate',
'Multivariate',
'Multivariate',
'Text',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Sequential',
'Multivariate',

'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Time',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Text',
'Multivariate',
'Multivariate',
'Multivariate',
'Text',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Text',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'',
'Multivariate',
'Multivariate',
'Multivariate',
'Sequential',
'Multivariate',
'Multivariate',
'Multivariate',
'',
'Text',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
''

'',
'Text',
'Multivariate',
'Multivariate',
'Text',
'Multivariate',
'Time',
'Text',
'',
'',
'Multivariate',
'Multivariate',
'Sequential',
'Multivariate',
'Multivariate',
'Text',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Text',
'',
'Sequential',
'Univariate',
'Univariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Univariate',
'Multivariate',
'Text',
'Sequential',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Univariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Text',
'Time',
'Multivariate',
'Multivariate',
'Text',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',

'Multivariate',
'Multivariate',
'Univariate',
'Multivariate',
'Multivariate',
'Univariate',
'Time',
'Time',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'',
'Multivariate',
'Time',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate',
'Multivariate']

In [143]:

```
task_database = re.findall("\>(\w+)|\>(\s)\<",task_2)
task_database
```

Out[143]:

```
[('Classification', ''),
 ('Classification', ''),
 ('Classification', ''),
 ('Recommender', ''),
 ('Classification', ''),
 ('Classification', ''),
 ('Classification', ''),
 ('Classification', ''),
 ('Regression', ''),
 ('Regression', ''),
 ('Classification', ''),
 ('Classification', ''),
 ('Classification', ''),
 ('Classification', ''),
 ('Classification', ''),
 ('Classification', ''),
 ('Classification', ''),
 ('Classification', ''),
 ('Classification', ''),
 ('Classification', ''),
 ('Classification', ''),
 ('Classification', ''),
 ('Classification', ''),
 ('Classification', ''),
 ('', '\xa0'),
 ('', '\xa0'),
 ('Classification', ''),
 ('Classification', ''),
 ('Classification', ''),
 ('Regression', ''),
 ('Classification', ''),
 ('Classification', ''),
 ('Classification', ''),
 ('Classification', ''),
 ('', '\xa0'),
 ('', '\xa0'),
```

('', '\xa0'),
('', '\xa0'),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Function', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('', '\xa0'),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Relational', ''),
('', '\xa0'),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('', '\xa0'),
('', '\xa0'),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('', '\xa0'),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('', '\xa0'),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('', '\xa0'),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('', '\xa0'),
('', '\xa0'),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Clustering', ''),
('', '\xa0'),
('', '\xa0'),

('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Clustering', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('', '\xa0'),
('', '\xa0'),
('Classification', ''),
('Classification', ''),
('Clustering', ''),
('Classification', ''),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('Recommender', ''),
('Classification', ''),
('Regression', ''),
('', '\xa0'),
('', '\xa0'),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('', '\xa0'),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('', '\xa0'),
('Classification', ''),
('Classification', ''),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),

('Classification', ''),
('Regression', ''),
('Classification', ''),
('Clustering', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Causal', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Clustering', ''),
('Classification', ''),
('Regression', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('', '\xa0'),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('', '\xa0'),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification

('Classification', ''),
('Classification', ''),
('Classification', ''),
('', '\xa0'),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('', '\xa0'),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Clustering', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Regression', ''),
('Regression', ''),
('', '\xa0'),
('Classification', ''),
('', '\xa0'),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),

('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Clustering', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Clustering', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Clustering', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Causal', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),

('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Causal', ''),
('Clustering', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Causal', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Clustering', ''),
('Classification', ''),
('', '\xa0'),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Clustering', ''),
('Clustering', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),

('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Clustering', ''),
('Classification', ''),
('Regression', ''),
('Clustering', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Recommendation', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Regression', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),
('Classification', ''),

('Classification', '')]

In [144]:

```
task_database_1=[]
for element in task_database:
    i = element[0]
    task_database_1.append(i)

task_database_1
```

Out[144]:

['Classification',
 'Classification',
 'Classification',
 'Recommender',
 'Classification',
 'Classification',
 'Classification',
 'Classification',
 'Regression',
 'Regression',
 'Classification',
 'Classification',
 'Classification',
 'Classification',
 'Classification',
 'Classification',
 'Classification',
 'Classification',
 'Classification',
 'Classification',
 'Classification',
 'Classification',
 'Classification',
 '',
 '',
 'Classification',
 'Classification',
 'Classification',
 'Regression',
 'Classification',
 'Classification',
 'Classification',
 'Classification',
 '',
 '',
 '',
 '',
 'Classification',
 'Classification',
 'Classification',
 'Function',
 'Classification',
 'Classification',
 'Classification',
 'Classification',
 'Classification',
 'Classification',
 'Classification',
 '',
 'Classification',

'Classification',
'Classification',
'Classification',
'Classification',
'Relational',
'',
'Classification',
'Classification',
'Classification',
'',
'',
'Classification',
'Classification',
'Classification',
'Classification',
'',
'Classification',
'Classification',
'Classification',
'Classification',
'',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'',
'',
'Classification',
'Regression',
'Classification',
'Regression',
'Classification',
'Classification',
'Regression',
'Classification',
'Classification',
'Classification',
'Classification',
'Clustering',
'',
'',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Clustering',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'',
'',
'Classification',

'Classification',
'Classification',
'Clustering',
'Classification',
'',
'',
'',
'',
'',
'',
'Recommender',
'Classification',
'Regression',
'',
'',
'Classification',
'Regression',
'Classification',
'',
'',
'',
'',
'',
'',
'Classification',
'Classification',
'Classification',
'Classification',
'',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'',
'Classification',
'Classification',
'',
'',
'',
'',
'',
'Classification',
'Classification',
'Classification',
'Classification',
'Regression',
'Classification',
'Clustering',
'Regression',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Causal',
'Classification',
'Classification',

'Classification',
'Classification',
'Classification',
'Classification',
'Clustering',
'Classification',
'Regression',
'Regression',
'Classification',
'Classification',
'Classification',
'Classification',
'Regression',
'Classification',
'',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Regression',
'Classification',
'',
'Regression',
'Classification',
'Classification',
'Classification',
'Regression',
'Classification',
'Classification',
'Classification',
'Classification',
'Regression',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'',
'Classification',
'Regression',
'Classification',
'Classification',
'',
'Classification',
'Classification',
'Classification',
'Classification',
'Regression',

'Classification',
'Classification',
'Regression',
'Classification',
'Regression',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Clustering',
'Classification',
'Classification',
'Classification',
'Regression',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Regression',
'Classification',
'Regression',
'Regression',
'',
'Classification',
'',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Regression',
'Classification',
'Classification',
'Regression',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Regression',
'Classification',
'Classification',
'Clustering',
'Classification',
'Classification',

'Classification',
'Classification',
'Classification',
'Classification',
'Clustering',
'Classification',
'Regression',
'Classification',
'Classification',
'Classification',
'Classification',
'Regression',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Regression',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Regression',
'Regression',
'Classification',
'Classification',
'Classification',
'Clustering',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Causal',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Regression',
'Classification',
'Classification',
'Regression',
'Classification',
'Classification',
'Classification',
'Classification',
'Regression',
'Causal',
'Clustering',
'Classification',
'Classification',
'Regression',

'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Regression',
'Classification',
'Classification',
'Classification',
'Causal',
'Classification',
'Regression',
'Classification',
'Classification',
'Regression',
'Classification',
'Classification',
'Classification',
'Regression',
'Classification',
'Clustering',
'Classification',
'',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Regression',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Clustering',
'Clustering',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Regression',
'Classification',
'Classification',
'Classification',
'Classification',

'Regression',
'Clustering',
'Classification',
'Regression',
'Clustering',
'Classification',
'Classification',
'Classification',
'Classification',
'Regression',
'Classification',
'Classification',
'Classification',
'Regression',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Recommendation',
'Classification',
'Classification',
'Classification',
'Regression',
'Classification',
'Regression',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Regression',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification']

In [150]:

```
attri_database = re.findall("\>(\w+)|\>(\s)\<",attri_2)
attri_database
```

Out[150]:

```
[('Categorical', ''),
 ('Categorical', ''),
 ('Categorical', ''),
 ('Categorical', ''),
 ('Categorical', ''),
 ('Categorical', ''),
```

('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('', '\xa0'),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Integer', ''),
('Real', ''),
('Real', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('', '\xa0'),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Integer', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Real', ''),
('', '\xa0'),
('', '\xa0'),
('Categorical', ''),
('Real', ''),
('Categorical', ''),
('Real', ''),
('Real', ''),
('Integer', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Real', ''),
('Real', ''),
('Categorical', ''),
('Integer', ''),
('Real', ''),
('Real', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Integer', ''),
('Categorical', ''),
('', '\xa0'),
('Integer', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),

('Categorical', ''),
('', '\xa0'),
('Integer', ''),
('Categorical', ''),
('Integer', ''),
('Integer', ''),
('Categorical', ''),
('', '\xa0'),
('Integer', ''),
('Integer', ''),
('Integer', ''),
('Categorical', ''),
('Categorical', ''),
('', '\xa0'),
('', '\xa0'),
('Real', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Integer', ''),
('Integer', ''),
('Integer', ''),
('Categorical', ''),
('Integer', ''),
('Categorical', ''),
('', '\xa0'),
('', '\xa0'),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Integer', ''),
('Real', ''),
('Real', ''),
('Integer', ''),
('Real', ''),
('Categorical', ''),
('', '\xa0'),
('', '\xa0'),
('Categorical', ''),
('Real', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Real', ''),
('', '\xa0'),
('Categorical', ''),
('Integer', ''),
('Categorical', ''),
('Integer', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Real', ''),
('Categorical', ''),
('Categorical', ''),
('', '\xa0'),
('', '\xa0'),

('Categorical', ''),
('', '\xa0'),
('Categorical', ''),
('', '\xa0'),
('Categorical', ''),
('Integer', ''),
('Real', ''),
('Categorical', ''),
('', '\xa0'),
('', '\xa0'),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Integer', ''),
('Real', ''),
('Integer', ''),
('Integer', ''),
('Categorical', ''),
('Real', ''),
('Real', ''),
('', '\xa0'),
('', '\xa0'),
('Real', ''),
('Categorical', ''),
('Categorical', ''),
('Categorical', ''),
('Real', ''),
('Integer', ''),
('Integer', ''),
('Real', ''),
('', '\xa0'),
('Integer', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('Integer', ''),
('Integer', ''),
('Integer', ''),
('Real', ''),
('Real', ''),
('Integer', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('Integer', ''),
('Integer', ''),
('Real', ''),
('Categorical', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('Categorical', ''),
('Real', ''),
('Integer', ''),
('Real', ''),
('Integer', ''),
('Integer', ''),
('', '\xa0'),
('Real', ''),
('Real', ''),
('Real', ''),
('Real', ''),

('Real', ''),
('Categorical', ''),
('Integer', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('', '\xa0'),
('Real', ''),
('Integer', ''),
('Integer', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('', '\xa0'),
('Real', ''),
('', '\xa0'),
('', '\xa0'),
('Integer', ''),
('Integer', ''),
('Real', ''),
('', '\xa0'),
('Real', ''),
('Integer', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('', '\xa0'),
('Integer', ''),
('Real', ''),
('Real', ''),
('Integer', ''),
('Integer', ''),
('', '\xa0'),
('', '\xa0'),
('Real', ''),
('Integer', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('Integer', ''),
('Real', ''),
('Integer', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('Integer', ''),
('', '\xa0'),
('Real', ''),
('Integer', ''),
('Real', ''),
('', '\xa0'),
('', '\xa0'),
('Real', ''),
('Integer', ''),

('Real', ''),
('Real', ''),
('Real', ''),
('Integer', ''),
('Integer', ''),
('Real', ''),
('', '\xa0'),
('Integer', ''),
('Real', ''),
('Real', ''),
('Integer', ''),
('Integer', ''),
('Integer', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('', '\xa0'),
('Real', ''),
('', '\xa0'),
('', '\xa0'),
('Real', ''),
('Real', ''),
('Real', ''),
('', '\xa0'),
('Real', ''),
('Real', ''),
('Integer', ''),
('', '\xa0'),
('Real', ''),
('', '\xa0'),
('Integer', ''),
('', '\xa0'),
('Real', ''),
('Integer', ''),
('Integer', ''),
('Real', ''),
('Integer', ''),
('Integer', ''),
('Real', ''),
('', '\xa0'),
('', '\xa0'),
('Real', ''),
('Real', ''),
('Integer', ''),
('Integer', ''),
('Integer', ''),
('Real', ''),
('', '\xa0'),
('Real', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('Integer', ''),
('Real', ''),
('Real', ''),
('Integer', ''),
('', '\xa0'),
('Real', ''),
('Real', ''),
('Integer', ''),
('Real', ''),

('Integer', ''),
('Categorical', ''),
('', '\xa0'),
('Integer', ''),
('', '\xa0'),
('', '\xa0'),
('Integer', ''),
('Real', ''),
('Real', ''),
('', '\xa0'),
('Real', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('Integer', ''),
('Real', ''),
('Integer', ''),
('Real', ''),
('Real', ''),
('Integer', ''),
('Integer', ''),
('Real', ''),
('Integer', ''),
('Integer', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('', '\xa0'),
('Real', ''),
('Real', ''),
('Real', ''),
('Integer', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('', '\xa0'),
('Integer', ''),
('', '\xa0'),
('Integer', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('Integer', ''),
('Integer', ''),
('Integer', ''),
('Real', ''),
('Integer', ''),
('', '\xa0'),
('Integer', ''),
('Integer', ''),
('Integer', ''),
('Real', ''),
('Integer', ''),
('Real', ''),
('Real', ''),
('Integer', ''),
('Integer', ''),
('Real', ''),
('Integer', ''),
('Integer', ''),
('Integer', ''),

('Integer', ''),
('Integer', ''),
('Integer', ''),
('Integer', ''),
('', '\xa0'),
('Integer', ''),
('', '\xa0'),
('Real', ''),
('Real', ''),
('', '\xa0'),
('Integer', ''),
('Real', ''),
('Real', ''),
('Integer', ''),
('Real', ''),
('Integer', ''),
('Integer', ''),
('Integer', ''),
('Integer', ''),
('Integer', ''),
('Integer', ''),
('Real', ''),
('Integer', ''),
('Integer', ''),
('Integer', ''),
('Integer', ''),
('Integer', ''),
('Integer', ''),
('Real', ''),
('Integer', ''),
('Integer', ''),
('', '\xa0'),
('Integer', ''),
('Real', ''),
('Integer', ''),
('Integer', ''),
('Real', ''),
('', '\xa0'),
('Integer', ''),
('Real', ''),
('', '\xa0'),
('Integer', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('Integer', ''),
('Integer', ''),
('Real', ''),
('Real', ''),
('Integer', ''),
('', '\xa0'),
('Real', ''),
('Real', ''),
('Integer', ''),
('Integer', ''),
('Integer', ''),
('Real', ''),
('', '\xa0'),
('', '\xa0'),
('Integer', ''),
('Integer', ''),

('Integer', ''),
('', '\xa0'),
('Real', ''),
('Categorical', ''),
('Integer', ''),
('Integer', ''),
('Integer', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('', '\xa0'),
('Integer', ''),
('', '\xa0'),
('Integer', ''),
('Real', ''),
('Integer', ''),
('', '\xa0'),
('', '\xa0'),
('Real', ''),
('Real', ''),
('Integer', ''),
('Real', ''),
('Integer', ''),
('Real', ''),
('Real', ''),
('Real', ''),
('Integer', ''),
('Real', ''),
('Real', ''),
('Integer', '')]

In [151]:

```
attri_database_1=[]
for element in attri_database:
    i = element[0]
    attri_database_1.append(i)
attri_database_1
```

Out[151]:

['Categorical',
 'Categorical',
 'Categorical',
 'Categorical',
 'Categorical',
 'Categorical',
 'Categorical',
 'Categorical',
 'Categorical',
 'Categorical',
 '',
 'Categorical',
 'Categorical',
 'Categorical',
 'Integer',
 'Real',
 'Real',
 'Categorical',
 'Categorical',
 'Categorical',
 'Categorical',
 'Categorical',
 'Categorical',

'Categorical',
'',
'Categorical',
'Categorical',
'Categorical',
'Categorical',
'Integer',
'Categorical',
'Categorical',
'Categorical',
'Categorical',
'Categorical',
'Real',
'',
'',
'Categorical',
'Real',
'Categorical',
'Real',
'Real',
'Integer',
'Categorical',
'Categorical',
'Categorical',
'Categorical',
'Real',
'Real',
'Categorical',
'Integer',
'Real',
'Real',
'Categorical',
'Categorical',
'Categorical',
'Categorical',
'Integer',
'Categorical',
'',
'Integer',
'Categorical',
'Categorical',
'Categorical',
'Categorical',
'Categorical',
'Categorical',
'Categorical',
'Categorical',
'',
'Integer',
'Categorical',
'Integer',
'Integer',
'Categorical',
'',
'Integer',
'Integer',
'Integer',
'Categorical',
'Categorical',
'',
'',
'Real',

'Categorical',
'Categorical',
'Categorical',
'Categorical',
'Categorical',
'Integer',
'Integer',
'Integer',
'Categorical',
'Integer',
'Categorical',
'',
'',
'Categorical',
'Categorical',
'Categorical',
'Categorical',
'Categorical',
'Categorical',
'Integer',
'Real',
'Real',
'Integer',
'Real',
'Categorical',
'',
'',
'Categorical',
'Real',
'Categorical',
'Categorical',
'Categorical',
'Real',
'',
'Categorical',
'Integer',
'Categorical',
'Integer',
'Categorical',
'Categorical',
'Categorical',
'Real',
'Categorical',
'Categorical',
'',
'',
'Categorical',
'',
'Categorical',
'',
'Categorical',
'Integer',
'Real',
'Categorical',
'',
'',
'Categorical',
'Categorical',
'Categorical',
'Integer',
'Real',
'Integer',

'Integer',
'Categorical',
'Real',
'Real',
'',
'',
'Real',
'Categorical',
'Categorical',
'Categorical',
'Real',
'Integer',
'Integer',
'Real',
'',
'Integer',
'Real',
'Real',
'Real',
'Integer',
'Integer',
'Integer',
'Real',
'Real',
'Integer',
'Real',
'Real',
'Real',
'Integer',
'Integer',
'Real',
'Categorical',
'Real',
'Real',
'Real',
'Categorical',
'Real',
'Integer',
'Real',
'Integer',
'Integer',
'',
'Real',
'Real',
'Real',
'Real',
'Real',
'Categorical',
'Integer',
'Real',
'Real',
'Real',
'',
'Real',
'Integer',
'Integer',
'Real',
'Real',
'Real',
'Real',
'Real',
'Real',

'Real',
'',
'Real',
'',
'',
'Integer',
'Integer',
'Real',
'',
'Real',
'Integer',
'Real',
'Real',
'Real',
'Real',
'Real',
'Real',
'',
'Integer',
'Real',
'Real',
'Integer',
'Integer',
'',
'',
'Real',
'Integer',
'Real',
'Real',
'Real',
'Real',
'Real',
'Integer',
'Real',
'Integer',
'Real',
'Real',
'Real',
'Integer',
'',
'Real',
'Integer',
'Real',
'',
'',
'Real',
'Integer',
'Real',
'Real',
'Real',
'Integer',
'Integer',
'Real',
'',
'Integer',
'Real',
'Real',
'Integer',
'Integer',
'Integer',
'Real',
'Real',
'Real',

'Real',
'',
'Real',
'',
'',
'Real',
'Real',
'Real',
'',
'Real',
'Real',
'Integer',
'',
'Real',
'',
'Integer',
'',
'Real',
'Integer',
'Integer',
'Real',
'Integer',
'Integer',
'Real',
'',
'',
'Real',
'Real',
'Integer',
'Integer',
'Integer',
'Real',
'',
'Real',
'Real',
'Real',
'Real',
'Real',
'Integer',
'Real',
'Real',
'Integer',
'',
'Real',
'Real',
'Integer',
'Real',
'Integer',
'Categorical',
'',
'Integer',
'',
'',
'Integer',
'Real',
'Real',
'',
'Real',
'Real',
'Real',
'Real',
'Integer',

'Real',
'Integer',
'Real',
'Real',
'Integer',
'Integer',
'Real',
'Integer',
'Integer',
'Real',
'Real',
'Real',
'Real',
'',
'Real',
'Real',
'Real',
'Integer',
'Real',
'Real',
'Real',
'',
'Integer',
'',
'Integer',
'Real',
'Real',
'Real',
'Integer',
'Integer',
'Integer',
'Real',
'Integer',
'',
'Integer',
'Integer',
'Integer',
'Real',
'Integer',
'Real',
'Real',
'Integer',
'Integer',
'Real',
'Integer',
'Integer',
'Integer',
'Integer',
'Integer',
'Integer',
'Integer',
'',
'Integer',
'',
'Real',
'Real',
'',
'Integer',
'Real',
'Real',
'Integer',
'Real',

'Integer',
'Integer',
'Integer',
'Integer',
'Integer',
'Integer',
'Real',
'Integer',
'Integer',
'Integer',
'Integer',
'Integer',
'Integer',
'Real',
'Integer',
'Integer',
'',
'Integer',
'Real',
'Integer',
'Integer',
'Real',
'',
'Integer',
'Real',
'',
'Integer',
'Real',
'Real',
'Real',
'Integer',
'Integer',
'Real',
'Real',
'Integer',
'',
'Real',
'Real',
'Integer',
'Integer',
'Integer',
'Real',
'',
'',
'Integer',
'Integer',
'Integer',
'',
'Real',
'Categorical',
'Integer',
'Integer',
'Integer',
'Real',
'Real',
'Real',
'',
'Integer',
'',
'Integer',
'Real',
'Integer',

''
'',
'Real',
'Real',
'Integer',
'Real',
'Integer',
'Real',
'Real',
'Real',
'Integer',
'Real',
'Real',
'Integer']

In [154]:

```
inst_database = re.findall("(\d+)|\>(\s)\<",inst_2)
inst_database
```

Out[154]:

```
[('4177', ''),
 ('48842', ''),
 ('798', ''),
 ('37711', ''),
 ('452', ''),
 ('6000', ''),
 ('226', ''),
 ('226', ''),
 ('398', ''),
 ('205', ''),
 ('294', ''),
 ('625', ''),
 ('16', ''),
 ('286', ''),
 ('699', ''),
 ('198', ''),
 ('569', ''),
 ('108', ''),
 ('1728', ''),
 ('48842', ''),
 ('', '\xa0'),
 ('3196', ''),
 ('28056', ''),
 ('', '\xa0'),
 ('100', ''),
 ('67557', ''),
 ('690', ''),
 ('125', ''),
 ('209', ''),
 ('1473', ''),
 ('581012', ''),
 ('512', ''),
 ('366', ''),
 ('', '\xa0'),
 ('', '\xa0'),
 ('', '\xa0'),
 ('', '\xa0'),
 ('132', ''),
 ('336', ''),
 ('194', ''),
 ('352', ''),
```

('214', ''),
('306', ''),
('160', ''),
('303', ''),
('155', ''),
('368', ''),
('', '\xa0'),
('2310', ''),
('3279', ''),
('351', ''),
('150', ''),
('7797', ''),
('104', ''),
('57', ''),
('', '\xa0'),
('24', ''),
('20000', ''),
('345', ''),
('', '\xa0'),
('32', ''),
('148', ''),
('209', ''),
('528', ''),
('', '\xa0'),
('106', ''),
('128', ''),
('3190', ''),
('432', ''),
('202', ''),
('2000', ''),
('8124', ''),
('476', ''),
('6598', ''),
('12960', ''),
('', '\xa0'),
('5473', ''),
('5620', ''),
('10992', ''),
('90', ''),
('339', ''),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('167', ''),
('15', ''),
('1389', ''),
('307', ''),
('47', ''),
('23', ''),
('531', ''),
('4601', ''),
('267', ''),
('267', ''),
('76', ''),
('', '\xa0'),
('1000', ''),
('151', ''),
('958', ''),
('7200', ''),
('10', ''),
('285', ''),
('435', ''),

('455', ''),
('527', ''),
('5000', ''),
('5000', ''),
('178', ''),
('1484', ''),
('101', ''),
('', '\xa0'),
('20000', ''),
('6650', ''),
('2565', ''),
('2458285', ''),
('299285', ''),
('340', ''),
('68040', ''),
('', '\xa0'),
('122', ''),
('178080', ''),
('50672', ''),
('640', ''),
('9000', ''),
('10104', ''),
('256932', ''),
('640', ''),
('191779', ''),
('4000000', ''),
('', '\xa0'),
('10000', ''),
('989818', ''),
('129000', ''),
('', '\xa0'),
('100000', ''),
('21578', ''),
('463', ''),
('600', ''),
('332', ''),
('', '\xa0'),
('', '\xa0'),
('690', ''),
('1000', ''),
('270', ''),
('6435', ''),
('2310', ''),
('58000', ''),
('946', ''),
('20008', ''),
('208', ''),
('528', ''),
('', '\xa0'),
('', '\xa0'),
('1024', ''),
('10080', ''),
('50400', ''),
('1025010', ''),
('19020', ''),
('1364', ''),
('961', ''),
('517', ''),
('200', ''),
('8000000', ''),
('1030', ''),
('606', ''),
('900', ''),

('900', ''),
('2600', ''),
('1950', ''),
('13500', ''),
('4400', ''),
('2536', ''),
('300', ''),
('197', ''),
('2858', ''),
('748', ''),
('11640', ''),
('1593', ''),
('1567', ''),
('22632', ''),
('360', ''),
('103', ''),
('1994', ''),
('120', ''),
('4898', ''),
('2396130', ''),
('16772', ''),
('5875', ''),
('503', ''),
('51', ''),
('106', ''),
('2126', ''),
('5456', ''),
('8800', ''),
('164860', ''),
('', '\xa0'),
('1941', ''),
('130065', ''),
('515345', ''),
('440', ''),
('', '\xa0'),
('53500', ''),
('8235', ''),
('', '\xa0'),
('5749132', ''),
('2215', ''),
('310', ''),
('10000', ''),
('3000', ''),
('1500', ''),
('30000', ''),
('2500', ''),
('4143', ''),
('64', ''),
('53414', ''),
('65554', ''),
('45211', ''),
('1138562', ''),
('13910', ''),
('583', ''),
('2551', ''),
('34465', ''),
('5574', ''),
('245057', ''),
('182', ''),
('3850505', ''),
('138', ''),
('1080', ''),

('2075259', ''),
('210', ''),
('115', ''),
('3960456', ''),
('', '\xa0'),
('10299', ''),
('1600', ''),
('768', ''),
('308', ''),
('100', ''),
('237', ''),
('434874', ''),
('536', ''),
('140000', ''),
('6118', ''),
('165632', ''),
('18000', ''),
('540', ''),
('931', ''),
('1055', ''),
('100', ''),
('9120', ''),
('403', ''),
('111740', ''),
('10421', ''),
('5820', ''),
('403', ''),
('14980', ''),
('45730', ''),
('2584', ''),
('1372', ''),
('306', ''),
('120000', ''),
('13910', ''),
('2747', ''),
('3395', ''),
('39242', ''),
('4137', ''),
('17389', ''),
('51', ''),
('470', ''),
('132', ''),
('5000000', ''),
('11000000', ''),
('250', ''),
('126', ''),
('', '\xa0'),
('4889', ''),
('', '\xa0'),
('', '\xa0'),
('340', ''),
('501', ''),
('45781', ''),
('1503', ''),
('440', ''),
('2000', ''),
('9568', ''),
('168', ''),
('100000', ''),
('5665', ''),
('79', ''),
('127', ''),

('1040', ''),
('9900', ''),
('560', ''),
('60021', ''),
('1419', ''),
('101', ''),
('399', ''),
('58', ''),
('180', ''),
('21048', ''),
('', '\xa0'),
('750', ''),
('3000', ''),
('150', ''),
('1059', ''),
('11934', ''),
('27965', ''),
('216', ''),
('120', ''),
('649', ''),
('370', ''),
('4178504', ''),
('221579', ''),
('10800', ''),
('58509', ''),
('129685', ''),
('2456', ''),
('2921', ''),
('1151', ''),
('6590', ''),
('3000', ''),
('39797', ''),
('326', ''),
('913', ''),
('168286', ''),
('400', ''),
('314080', ''),
('637', ''),
('1710671', ''),
('12000', ''),
('10929', ''),
('1080', ''),
('40000', ''),
('43930257', ''),
('230318', ''),
('10500000', ''),
('13197', ''),
('', '\xa0'),
('30000', ''),
('324', ''),
('541909', ''),
('11164866', ''),
('163', ''),
('373', ''),
('20560', ''),
('40', ''),
('422937', ''),
('9358', ''),
('640', ''),
('919438', ''),
('40949', ''),
('5744', ''),

('10503', ''),
('42240', ''),
('102944', ''),
('500', ''),
('9782222', ''),
('11463', ''),
('17898', ''),
('1885', ''),
('19735', ''),
('1540', ''),
('4007', ''),
('153540', ''),
('606', ''),
('1353', ''),
('1956', ''),
('43824', ''),
('3942', ''),
('858', ''),
('287', ''),
('17764280', ''),
('106574', ''),
('9358', ''),
('11500', ''),
('92000', ''),
('315', ''),
('78095', ''),
('130', ''),
('74', ''),
('52854', ''),
('77', ''),
('811', ''),
('504', ''),
('401', ''),
('2856', ''),
('10546', ''),
('801', ''),
('1540', ''),
('1451', ''),
('1075', ''),
('78095', ''),
('7195', ''),
('3600', ''),
('76', ''),
('60', ''),
('405', ''),
('303', ''),
('303', ''),
('107888', ''),
('76000', ''),
('10000', ''),
('180', ''),
('745000', ''),
('12234', ''),
('292', ''),
('104', ''),
('60000', ''),
('2000', ''),
('165', ''),
('217', ''),
('1175', ''),
('704', ''),
('75128', ''),

('90', ''),
('90', ''),
('50', ''),
('71', ''),
('93239', ''),
('540', ''),
('105', ''),
('6611', ''),
('15', ''),
('372', ''),
('58000', ''),
('4960', ''),
('241600', ''),
('130000', ''),
('7062606', ''),
('740', ''),
('70', ''),
('2205', ''),
('10721', ''),
('640', ''),
('1000', ''),
('116', ''),
('1672', ''),
('322', ''),
('93600', ''),
('3060', ''),
('5879', ''),
('9200', ''),
('20000', ''),
('20867', ''),
('', '\xa0'),
('4143', ''),
('215063', ''),
('6000000', ''),
('21263', ''),
('63000000', ''),
('10190', ''),
('300', ''),
('12330', ''),
('5180', ''),
('756', ''),
('10000', ''),
('80', ''),
('1184', ''),
('1047', ''),
('777', ''),
('249', ''),
('414', ''),
('', '\xa0'),
('143', ''),
('7840', ''),
('30000', ''),
('35717', ''),
('135', ''),
('980', ''),
('5456', ''),
('120', '')]

In [155]:

```
inst_database_1=[]
for element in inst_database:
```

```
    i = element[0]
    inst_database_1.append(i)
inst_database_1
```

Out[155]:

```
['4177',
 '48842',
 '798',
 '37711',
 '452',
 '6000',
 '226',
 '226',
 '398',
 '205',
 '294',
 '625',
 '16',
 '286',
 '699',
 '198',
 '569',
 '108',
 '1728',
 '48842',
 '',
 '3196',
 '28056',
 '',
 '100',
 '67557',
 '690',
 '125',
 '209',
 '1473',
 '581012',
 '512',
 '366',
 '',
 '',
 '',
 '',
 '132',
 '336',
 '194',
 '352',
 '214',
 '306',
 '160',
 '303',
 '155',
 '368',
 '',
 '2310',
 '3279',
 '351',
 '150',
 '7797',
 '104',
 '57',
 '',
```

'24',
'20000',
'345',
'',
'32',
'148',
'209',
'528',
'',
'106',
'128',
'3190',
'432',
'202',
'2000',
'8124',
'476',
'6598',
'12960',
'',
'5473',
'5620',
'10992',
'90',
'339',
'',
'',
'',
'167',
'15',
'1389',
'307',
'47',
'23',
'531',
'4601',
'267',
'267',
'76',
'',
'1000',
'151',
'958',
'7200',
'10',
'285',
'435',
'527',
'5000',
'5000',
'178',
'1484',
'101',
'',
'20000',
'6650',
'2565',
'2458285',
'299285',
'340',
'68040',
'',

'122',
'178080',
'50672',
'640',
'9000',
'10104',
'256932',
'640',
'191779',
'4000000',
'',
'10000',
'989818',
'129000',
'',
'100000',
'21578',
'463',
'600',
'332',
'',
'',
'690',
'1000',
'270',
'6435',
'2310',
'58000',
'946',
'20008',
'208',
'528',
'',
'',
'1024',
'10080',
'50400',
'1025010',
'19020',
'1364',
'961',
'517',
'200',
'8000000',
'1030',
'606',
'900',
'2600',
'1950',
'13500',
'4400',
'2536',
'300',
'197',
'2858',
'748',
'11640',
'1593',
'1567',
'22632',
'360',
'103',

'1994',
'120',
'4898',
'2396130',
'16772',
'5875',
'503',
'51',
'106',
'2126',
'5456',
'8800',
'164860',
'',
'1941',
'130065',
'515345',
'440',
'',
'53500',
'8235',
'',
'5749132',
'2215',
'310',
'10000',
'3000',
'1500',
'30000',
'2500',
'4143',
'64',
'53414',
'65554',
'45211',
'1138562',
'13910',
'583',
'2551',
'34465',
'5574',
'245057',
'182',
'3850505',
'138',
'1080',
'2075259',
'210',
'115',
'3960456',
'',
'10299',
'1600',
'768',
'308',
'100',
'237',
'434874',
'536',
'140000',
'6118',
'165632',

'18000',
'540',
'931',
'1055',
'100',
'9120',
'403',
'111740',
'10421',
'5820',
'403',
'14980',
'45730',
'2584',
'1372',
'306',
'120000',
'13910',
'2747',
'3395',
'39242',
'4137',
'17389',
'51',
'470',
'132',
'5000000',
'11000000',
'250',
'126',
'',
'4889',
'',
'',
'340',
'501',
'45781',
'1503',
'440',
'2000',
'9568',
'168',
'100000',
'5665',
'79',
'127',
'1040',
'9900',
'560',
'60021',
'1419',
'101',
'399',
'58',
'180',
'21048',
'',
'750',
'3000',
'150',
'1059',
'11934',

'27965',
'216',
'120',
'649',
'370',
'4178504',
'221579',
'10800',
'58509',
'129685',
'2456',
'2921',
'1151',
'6590',
'3000',
'39797',
'326',
'913',
'168286',
'400',
'314080',
'637',
'1710671',
'12000',
'10929',
'1080',
'40000',
'43930257',
'230318',
'10500000',
'13197',
'',
'30000',
'324',
'541909',
'11164866',
'163',
'373',
'20560',
'40',
'422937',
'9358',
'640',
'919438',
'40949',
'5744',
'10503',
'42240',
'102944',
'500',
'9782222',
'11463',
'17898',
'1885',
'19735',
'1540',
'4007',
'153540',
'606',
'1353',
'1956',
'43824',

'3942',
'858',
'287',
'17764280',
'106574',
'9358',
'11500',
'92000',
'315',
'78095',
'130',
'74',
'52854',
'77',
'811',
'504',
'401',
'2856',
'10546',
'801',
'1540',
'1451',
'1075',
'78095',
'7195',
'3600',
'76',
'60',
'405',
'303',
'303',
'107888',
'76000',
'10000',
'180',
'745000',
'12234',
'292',
'104',
'60000',
'2000',
'165',
'217',
'1175',
'704',
'75128',
'90',
'90',
'50',
'71',
'93239',
'540',
'105',
'6611',
'15',
'372',
'58000',
'4960',
'241600',
'130000',
'7062606',
'740',

'70',
'2205',
'10721',
'640',
'1000',
'116',
'1672',
'322',
'93600',
'3060',
'5879',
'9200',
'20000',
'20867',
'',
'4143',
'215063',
'6000000',
'21263',
'63000000',
'10190',
'300',
'12330',
'5180',
'756',
'10000',
'80',
'1184',
'1047',
'777',
'249',
'414',
'',
'143',
'7840',
'30000',
'35717',
'135',
'980',
'5456',
'120']

In [156]:

```
numatt_database = re.findall("(\d+)|\>(\s)\<",numatt_2)
numatt_database
```

Out[156]:

[('8', ''),
('14', ''),
('38', ''),
('294', ''),
('279', ''),
('7', ''),
('', '\xa0'),
('69', ''),
('8', ''),
('26', ''),
('1', ''),
('4', ''),
('4', ''),

('9', ''),
('10', ''),
('34', ''),
('32', ''),
('13', ''),
('6', ''),
('14', ''),
('22', ''),
('36', ''),
('6', ''),
('', '\xa0'),
('6', ''),
('42', ''),
('15', ''),
('', '\xa0'),
('9', ''),
('9', ''),
('54', ''),
('39', ''),
('33', ''),
('20', ''),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('12', ''),
('8', ''),
('30', ''),
('', '\xa0'),
('10', ''),
('3', ''),
('5', ''),
('75', ''),
('19', ''),
('27', ''),
('', '\xa0'),
('19', ''),
('1558', ''),
('34', ''),
('4', ''),
('617', ''),
('12', ''),
('16', ''),
('7', ''),
('4', ''),
('16', ''),
('7', ''),
('', '\xa0'),
('56', ''),
('18', ''),
('8', ''),
('22', ''),
('', '\xa0'),
('58', ''),
('', '\xa0'),
('61', ''),
('7', ''),
('', '\xa0'),
('649', ''),
('22', ''),
('168', ''),
('168', ''),
('8', ''),

('', '\xa0'),
('10', ''),
('64', ''),
('16', ''),
('8', ''),
('17', ''),
('', '\xa0'),
('', '\xa0'),
('72', ''),
('4', ''),
('6', ''),
('10', ''),
('35', ''),
('35', ''),
('4', ''),
('102', ''),
('57', ''),
('22', ''),
('44', ''),
('45', ''),
('', '\xa0'),
('', '\xa0'),
('5', ''),
('9', ''),
('21', ''),
('32', ''),
('17', ''),
('16', ''),
('38', ''),
('21', ''),
('40', ''),
('13', ''),
('8', ''),
('17', ''),
('', '\xa0'),
('', '\xa0'),
('15', ''),
('22', ''),
('68', ''),
('40', ''),
('17', ''),
('89', ''),
('', '\xa0'),
('4', ''),
('12', ''),
('', '\xa0'),
('', '\xa0'),
('86', ''),
('72', ''),
('61', ''),
('12', ''),
('481', ''),
('42', ''),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('5', ''),
('90', ''),
('', '\xa0'),

('5', ''),
('', '\xa0'),
('', '\xa0'),
('14', ''),
('20', ''),
('13', ''),
('36', ''),
('19', ''),
('9', ''),
('18', ''),
('4', ''),
('60', ''),
('10', ''),
('', '\xa0'),
('', '\xa0'),
('10', ''),
('4', ''),
('3', ''),
('11', ''),
('11', ''),
('', '\xa0'),
('6', ''),
('13', ''),
('', '\xa0'),
('100000', ''),
('9', ''),
('101', ''),
('10000', ''),
('20000', ''),
('100000', ''),
('5000', ''),
('500', ''),
('73', ''),
('43', ''),
('23', ''),
('3', ''),
('5', ''),
('', '\xa0'),
('256', ''),
('591', ''),
('70', ''),
('91', ''),
('10', ''),
('128', ''),
('6', ''),
('12', ''),
('3231961', ''),
('5409', ''),
('26', ''),
('', '\xa0'),
('', '\xa0'),
('10', ''),
('23', ''),
('24', ''),
('13', ''),
('8', ''),
('', '\xa0'),
('27', ''),
('50', ''),
('90', ''),
('138672', ''),
('', '\xa0'),

('386', ''),
('', '\xa0'),
('', '\xa0'),
('12', ''),
('147', ''),
('6', ''),
('8', ''),
('27', ''),
('10000', ''),
('20000', ''),
('10000', ''),
('54877', ''),
('4702', ''),
('24', ''),
('29', ''),
('17', ''),
('3', ''),
('128', ''),
('10', ''),
('242', ''),
('120', ''),
('', '\xa0'),
('4', ''),
('13', ''),
('52', ''),
('47', ''),
('857', ''),
('9', ''),
('7', ''),
('200', ''),
('4', ''),
('', '\xa0'),
('561', ''),
('64', ''),
('8', ''),
('7', ''),
('10', ''),
('9', ''),
('4', ''),
('8', ''),
('77', ''),
('51', ''),
('18', ''),
('1950000', ''),
('18', ''),
('1300', ''),
('41', ''),
('6', ''),
('5625', ''),
('5', ''),
('', '\xa0'),
('7', ''),
('33', ''),
('5', ''),
('15', ''),
('9', ''),
('19', ''),
('5', ''),
('5', ''),
('1000000', ''),
('129', ''),
('', '\xa0'),

('20', ''),
('152', ''),
('24', ''),
('16', ''),
('35', ''),
('17', ''),
('5', ''),
('18', ''),
('28', ''),
('7', ''),
('309', ''),
('3', ''),
('6', ''),
('', '\xa0'),
('', '\xa0'),
('16', ''),
('13', ''),
('5', ''),
('6', ''),
('8', ''),
('2', ''),
('4', ''),
('148', ''),
('55', ''),
('17', ''),
('8', ''),
('42', ''),
('26', ''),
('50', ''),
('2', ''),
('281', ''),
('120', ''),
('', '\xa0'),
('6', ''),
('120432', ''),
('150000', ''),
('529', ''),
('', '\xa0'),
('16', ''),
('2500', ''),
('5', ''),
('68', ''),
('16', ''),
('100', ''),
('216', ''),
('23', ''),
('33', ''),
('140256', ''),
('19', ''),
('20', ''),
('20', ''),
('49', ''),
('12', ''),
('30', ''),
('5232', ''),
('20', ''),
('1', ''),
('', '\xa0'),
('61', ''),
('27', ''),
('53', ''),
('11', ''),

('11', ''),
('25', ''),
('0', ''),
('20', ''),
('9', ''),
('3', ''),
('561', ''),
('82', ''),
('13', ''),
('16', ''),
('13', ''),
('28', ''),
('4', ''),
('', '\xa0'),
('24', ''),
('34', ''),
('8', ''),
('128', ''),
('15', ''),
('513', ''),
('7', ''),
('7', ''),
('5', ''),
('15', ''),
('480000', ''),
('11', ''),
('54', ''),
('561', ''),
('64', ''),
('6', ''),
('116', ''),
('19', ''),
('', '\xa0'),
('5812', ''),
('9', ''),
('32', ''),
('29', ''),
('67', ''),
('', '\xa0'),
('25', ''),
('6400', ''),
('10', ''),
('5', ''),
('13', ''),
('98', ''),
('36', ''),
('69', ''),
('2158859', ''),
('518', ''),
('15', ''),
('179', ''),
('', '\xa0'),
('12', ''),
('38', ''),
('65', ''),
('102', ''),
('86', ''),
('7', ''),
('53', ''),
('20', ''),
('1', ''),
('71', ''),
('29', ''),

('29', ''),
('20531', ''),
('65', ''),
('3', ''),
('22', ''),
('38', ''),
('22', ''),
('4814', ''),
('698', ''),
('13', ''),
('10', ''),
('59', ''),
('56', ''),
('482', ''),
('171', ''),
('5', ''),
('500', ''),
('411', ''),
('8519', ''),
('21', ''),
('21', ''),
('171', ''),
('7', ''),
('49', ''),
('12', ''),
('3', ''),
('21', ''),
('9', ''),
('8', ''),
('7', ''),
('2', ''),
('11', ''),
('11', ''),
('173', ''),
('5', ''),
('15', ''),
('3', ''),
('105', ''),
('25000', ''),
('', '\xa0'),
('18', ''),
('21000', ''),
('115', ''),
('21', ''),
('206', ''),
('43680', ''),
('8', ''),
('10', ''),
('59', ''),
('10', ''),
('5', ''),
('5', ''),
('1000', ''),
('138', ''),
('', '\xa0'),
('16', ''),
('2', ''),
('10', ''),
('', '\xa0'),
('8', ''),
('6', ''),
('129', ''),
('21', ''),

('81', ''),
('12', ''),
('6', ''),
('22', ''),
('18', ''),
('9', ''),
('754', ''),
('14', ''),
('5', ''),
('', '\xa0'),
('', '\xa0'),
('18', ''),
('7', ''),
('7', ''),
('', '\xa0'),
('7', ''),
('5', ''),
('6', ''),
('4', ''),
('18', ''),
('11', ''),
('25', ''),
('', '\xa0')]

In [157]:

```
numatt_database_1=[]
for element in numatt_database:
    i = element[0]
    numatt_database_1.append(i)
numatt_database_1
```

Out[157]:

['8',
'14',
'38',
'294',
'279',
'7',
'',
'69',
'8',
'26',
'1',
'4',
'4',
'9',
'10',
'34',
'32',
'13',
'6',
'14',
'22',
'36',
'6',
'',
'6',
'42',
'15',
'',
'9',

'9',
'54',
'39',
'33',
'20',
'',
'',
'',
'12',
'8',
'30',
'',
'10',
'3',
'5',
'75',
'19',
'27',
'',
'19',
'1558',
'34',
'4',
'617',
'12',
'16',
'7',
'4',
'16',
'7',
'',
'56',
'18',
'8',
'22',
'',
'58',
'',
'61',
'7',
'',
'649',
'22',
'168',
'168',
'8',
'',
'10',
'64',
'16',
'8',
'17',
'',
'',
'72',
'4',
'6',
'10',
'35',
'35',
'4',
'102'

'57',
'22',
'44',
'45',
'',
'',
'5',
'9',
'21',
'32',
'17',
'16',
'38',
'21',
'40',
'13',
'8',
'17',
'',
'',
'15',
'22',
'68',
'40',
'17',
'89',
'',
'4',
'12',
'',
'',
'86',
'72',
'61',
'12',
'481',
'42',
'',
'',
'',
'',
'',
'',
'',
'5',
'90',
'',
'5',
'',
'',
'',
'14',
'20',
'13',
'36',
'19',
'9',
'18',
'4',
'60',
'10',
'',
'',
'10',

'4',
'3',
'11',
'11',
'',
'6',
'13',
'',
'100000',
'9',
'101',
'10000',
'20000',
'100000',
'5000',
'500',
'73',
'43',
'23',
'3',
'5',
'',
'256',
'591',
'70',
'91',
'10',
'128',
'6',
'12',
'3231961',
'5409',
'26',
'',
'',
'10',
'23',
'24',
'13',
'8',
'',
'27',
'50',
'90',
'138672',
'',
'386',
'',
'',
'12',
'147',
'6',
'8',
'27',
'10000',
'20000',
'10000',
'54877',
'4702',
'24',
'29',
'17',

'17',
'3',
'128',
'10',
'242',
'120',
'',
'4',
'13',
'52',
'47',
'857',
'9',
'7',
'200',
'4',
'',
'561',
'64',
'8',
'7',
'10',
'9',
'4',
'8',
'77',
'51',
'18',
'1950000',
'18',
'1300',
'41',
'6',
'5625',
'5',
'',
'7',
'33',
'5',
'15',
'9',
'19',
'5',
'5',
'1000000',
'129',
'',
'20',
'152',
'24',
'16',
'35',
'17',
'5',
'18',
'28',
'7',
'309',
'3',
'6',
'',
'',
'10',

'16',
'13',
'5',
'6',
'8',
'2',
'4',
'148',
'55',
'17',
'8',
'42',
'26',
'50',
'2',
'281',
'120',
'',
'6',
'120432',
'150000',
'529',
'',
'16',
'2500',
'5',
'68',
'16',
'100',
'216',
'23',
'33',
'140256',
'19',
'20',
'20',
'49',
'12',
'30',
'5232',
'20',
'1',
'',
'61',
'27',
'53',
'11',
'25',
'0',
'20',
'9',
'3',
'561',
'82',
'13',
'16',
'13',
'28',
'4',
'',
'24',
'34',

'8',
'128',
'15',
'513',
'7',
'7',
'5',
'15',
'480000',
'11',
'54',
'561',
'64',
'6',
'116',
'19',
'',
'5812',
'9',
'32',
'29',
'67',
'',
'25',
'6400',
'10',
'5',
'13',
'98',
'36',
'69',
'2158859',
'518',
'15',
'179',
'',
'12',
'38',
'65',
'102',
'86',
'7',
'53',
'20',
'1',
'71',
'29',
'20531',
'65',
'3',
'22',
'38',
'22',
'4814',
'698',
'13',
'10',
'59',
'56',
'482',
'171',
'5',

'500',
'411',
'8519',
'21',
'21',
'171',
'7',
'49',
'12',
'3',
'21',
'9',
'8',
'7',
'2',
'11',
'11',
'173',
'5',
'15',
'3',
'105',
'25000',
'',
'18',
'21000',
'115',
'21',
'206',
'43680',
'8',
'10',
'59',
'10',
'5',
'5',
'1000',
'138',
'',
'16',
'2',
'10',
'',
'8',
'6',
'129',
'81',
'12',
'6',
'22',
'18',
'9',
'754',
'14',
'5',
'',
'',
'18',
'7',
'7',
'',
'7',

'5',
'6',
'4',
'18',
'11',
'25',
'']

In [158]:

```
year_database = re.findall("(\d+)|\>(\s)\<",year_2)
year_database
```

Out[158]:

```
[('1995', ''),
 ('1996', ''),
 ('', '\xa0'),
 ('1998', ''),
 ('1998', ''),
 ('1992', ''),
 ('1987', ''),
 ('1992', ''),
 ('1993', ''),
 ('1987', ''),
 ('1994', ''),
 ('1994', ''),
 ('', '\xa0'),
 ('1988', ''),
 ('1992', ''),
 ('1995', ''),
 ('1995', ''),
 ('1990', ''),
 ('1997', ''),
 ('1996', ''),
 ('1988', ''),
 ('1989', ''),
 ('1994', ''),
 ('', '\xa0'),
 ('', '\xa0'),
 ('1995', ''),
 ('', '\xa0'),
 ('1992', ''),
 ('1987', ''),
 ('1997', ''),
 ('1998', ''),
 ('1995', ''),
 ('1998', ''),
 ('', '\xa0'),
 ('', '\xa0'),
 ('1994', ''),
 ('', '\xa0'),
 ('1989', ''),
 ('1996', ''),
 ('1990', ''),
 ('1990', ''),
 ('1987', ''),
 ('1999', ''),
 ('1989', ''),
 ('1988', ''),
 ('1988', ''),
 ('1989', ''),
 ('', '\xa0'),
```

('', '\xa0'),
('1990', ''),
('1998', ''),
('1989', ''),
('1988', ''),
('1994', ''),
('1990', ''),
('1988', ''),
('1988', ''),
('1990', ''),
('1991', ''),
('1990', ''),
('', '\xa0'),
('1992', ''),
('1988', ''),
('1990', ''),
('1996', ''),
('1995', ''),
('1990', ''),
('', '\xa0'),
('1992', ''),
('1992', ''),
('1994', ''),
('', '\xa0'),
('1987', ''),
('1994', ''),
('1994', ''),
('1997', ''),
('1991', ''),
('1995', ''),
('1998', ''),
('1998', ''),
('1993', ''),
('1988', ''),
('', '\xa0'),
('', '\xa0'),
('1992', ''),
('1993', ''),
('1988', ''),
('1989', ''),
('1988', ''),
('1987', ''),
('1993', ''),
('1988', ''),
('1999', ''),
('2001', ''),
('2001', ''),
('', '\xa0'),
('1992', ''),
('1993', ''),
('1997', ''),
('1991', ''),
('1987', ''),
('1994', ''),
('1988', ''),
('1987', ''),
('1993', ''),
('1988', ''),
('1988', ''),
('1991', ''),
('1996', ''),
('1990', ''),
('', '\xa0'),

('', '\xa0'),
('1999', ''),
('1999', ''),
('2002', ''),
('', '\xa0'),
('2000', ''),
('1999', ''),
('1999', ''),
('2001', ''),
('1999', ''),
('1999', ''),
('2000', ''),
('1999', ''),
('2000', ''),
('1999', ''),
('1999', ''),
('', '\xa0'),
('1998', ''),
('1999', ''),
('2001', ''),
('1999', ''),
('', '\xa0'),
('2003', ''),
('1999', ''),
('1999', ''),
('1997', ''),
('1999', ''),
('1999', ''),
('1998', ''),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('1994', ''),
('', '\xa0'),
('1993', ''),
('1990', ''),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('', '\xa0'),
('1989', ''),
('2006', ''),
('2006', ''),
('2007', ''),
('2007', ''),
('2007', ''),
('2007', ''),
('2008', ''),
('2008', ''),
('2008', ''),
('2007', ''),
('2008', ''),
('2008', ''),
('2008', ''),
('2008', ''),
('2008', ''),
('2008', ''),
('2008', ''),
('2008', ''),
('2008', ''),

('2008', ''),
('2008', ''),
('2008', ''),
('2009', ''),
('2008', ''),
('2008', ''),
('2008', ''),
('2009', ''),
('2009', ''),
('2009', ''),
('2009', ''),
('2009', ''),
('2009', ''),
('2010', ''),
('2009', ''),
('2010', ''),
('2010', ''),
('2010', ''),
('2010', ''),
('2010', ''),
('2010', ''),
('2010', ''),
('2010', ''),
('2010', ''),
('2010', ''),
('2011', ''),
('2011', ''),
('2011', ''),
('2011', ''),
('2011', ''),
('2011', ''),
('2011', ''),
('2011', ''),
('2011', ''),
('2011', ''),
('2011', ''),
('2011', ''),
('2011', ''),
('2011', ''),
('2011', ''),
('2011', ''),
('2011', ''),
('2011', ''),
('2011', ''),
('2012', ''),
('2012', ''),
('2012

('2012', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2014', ''),
('2013', ''),
('2013', ''),
('2013', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2013', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2014

('2014', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2014', ''),
('2015', ''),
('2014', ''),
('2014', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2014', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2015', ''),
('2016', ''),
('2016', ''),
('2015', ''),
('2016', ''),
('2016', ''),
('2015', ''),
('2016', ''),
('2016', ''),
('2016

('2016', ''),
('2017', ''),
('2016', ''),
('2017', ''),
('2017', ''),
('2016', ''),
('2016', ''),
('2017', ''),
('2017', ''),
('2016', ''),
('2017', ''),
('2017', ''),
('2017', ''),
('2017', ''),
('2016', ''),
('2017', ''),
('2016', ''),
('2016', ''),
('2016', ''),
('2017', ''),
('2017', ''),
('2017', ''),
('2017', ''),
('2017', ''),
('2017', ''),
('2017', ''),
('2016', ''),
('2016', ''),
('2016', ''),
('2016', ''),
('2017', ''),
('2017', ''),
('2017', ''),
('2017', ''),
('2017', ''),
('2016', ''),
('2017', ''),
('2017', ''),
('2017', ''),
('2017', ''),
('2017', ''),
('2017', ''),
('2016', ''),
('2016', ''),
('2016', ''),
('2017', ''),
('2017', ''),
('2017', ''),
('2017', ''),
('2017', ''),
('2017', ''),
('2017', ''),
('2017', ''),
('2016', ''),
('2018', ''),
('2018', ''),
('2016', ''),
('2018', ''),
('2018', ''),
('2018', ''),
('2018', ''),

('2018', ''),
('2018', ''),
('2018', ''),

'1996',
'',
'1998',
'1998',
'1992',
'1987',
'1992',
'1993',
'1987',
'1994',
'1994',
'',
'1988',
'1992',
'1995',
'1995',
'1990',
'1997',
'1996',
'1988',
'1989',
'1994',
'',
'',
'1995',
'',
'1992',
'1987',
'1997',
'1998',
'1995',
'1998',
'',
'',
'1994',
'',
'1989',
'1996',
'1990',
'1990',
'1987',
'1999',
'1989',
'1988',
'1988',
'1989',
'',
'1990',
'1998',
'1989',
'1988',
'1994',
'1990',
'1988',
'1988',
'1990',
'1991',
'1990',
'',
'1992',
'1988',
'1990',

'1996',
'1995',
'1990',
'',
'1992',
'1992',
'1994',
'',
'1987',
'1994',
'1994',
'1997',
'1991',
'1995',
'1998',
'1998',
'1993',
'1988',
'',
'',
'1992',
'1993',
'1988',
'1989',
'1988',
'1987',
'1993',
'1988',
'1999',
'2001',
'2001',
'',
'1992',
'1993',
'1997',
'1991',
'1987',
'1994',
'1988',
'1987',
'1993',
'1988',
'1988',
'1991',
'1996',
'1990',
'',
'1999',
'1999',
'2002',
'',
'2000',
'1999',
'1999',
'2001',
'1999',
'1999',
'2000',
'1999',
'2000',
'1999',
'1999',

'',
'1998',
'1999',
'2001',
'1999',
'',
'2003',
'1999',
'1999',
'1997',
'1999',
'1999',
'1998',
'',
'',
'',
'1994',
'',
'1993',
'1990',
'',
'',
'',
'',
'',
'',
'',
'',
'1989',
'2006',
'2006',
'2007',
'2007',
'2007',
'2007',
'2008',
'2008',
'2008',
'2007',
'2008',
'2008',
'2008',
'2008',
'2008',
'2008',
'2008',
'2008',
'2008',
'2008',
'2008',
'2008',
'2009',
'2008',
'2008',
'2008',
'2009',
'2009',
'2009',
'2009',
'2009',
'2009',
'2009',
'2010',
'2009',
'2010',

'2010',
'2010',
'2010',
'2010',
'2010',
'2010',
'2010',
'2010',
'2010',
'2010',
'2011',
'2011',
'2011',
'2011',
'2011',
'2011',
'2011',
'2011',
'2011',
'2011',
'2011',
'2011',
'2011',
'2011',
'2011',
'2011',
'2011',
'2011',
'2011',
'2012',
'2012',
'2012',
'2012',
'2012',
'2012',
'2012',
'2012',
'2012',
'2012',
'2012',
'2012',
'2012',
'2012',
'2012',
'2012',
'2012',
'2012',
'2012',
'2012',
'2012',
'2012',
'2012',
'2013',
'2013',
'2013

'2013',
'2013',
'2013',
'2013',
'2013',
'2013',
'2013',
'2013',
'2013',
'2013',
'2013',
'2013',
'2013',
'2013',
'2013',
'2014',
'2013',
'2013',
'2013',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2013',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2014',
'2015',
'2014',
'2014',
'2015',
'2015',
'2015'

'2015',
'2015',
'2015',
'2015',
'2015',
'2014',
'2015',
'2015',
'2015',
'2015',
'2015',
'2015',
'2015',
'2015',
'2015',
'2015',
'2015',
'2015',
'2015',
'2015',
'2015',
'2015',
'2015',
'2016',
'2016',
'2015',
'2016',
'2016',
'2015',
'2016',
'2016',
'2016',
'2016',
'2016',
'2016',
'2016',
'2016',
'2016',
'2016',
'2016',
'2016',
'2016',
'2016',
'2016',
'2016',
'2016',
'2016',
'2016',
'2016',
'2017',
'2016',
'2017',
'2016',
'2017',
'2017',
'2016',
'2016',
'2017',
'2017',
'2016',
'2017',
'2017',
'2017',
'2017',
'2016',
'2017',

'2017',
'2016',
'2016',
'2016',
'2017',
'2017',
'2017',
'2017',
'2017',
'2017',
'2017',
'2017',
'2016',
'2016',
'2016',
'2016',
'2016',
'2017',
'2017',
'2017',
'2017',
'2017',
'2017',
'2016',
'2017',
'2017',
'2017',
'2017',
'2017',
'2017',
'2017',
'2016',
'2016',
'2016',
'2017',
'2017',
'2017',
'2017',
'2017',
'2017',
'2017',
'2017',
'2017',
'2017',
'2016',
'2018',
'2018',
'2016',
'2018',
'2018',
'2018

'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2018',
'2019',
'2019',
'2018',
'2018',
'2018',
'2019']

In [164]:

```
dataset = list(zip(name_database, types_database_1, task_database_1, types_database_1, inst_database_1,
                   numatt_database_1, year_database_1))
dataset
```

Out[164]:

```
[('Abalone',
  'Multivariate',
  'Classification',
  'Multivariate',
  '4177',
  '8',
  '1995'),
 ('Adult',
  'Multivariate',
  'Classification',
  'Multivariate',
  '48842',
  '14',
  '1996'),
 ('Annealing',
  'Multivariate',
  'Classification',
  'Multivariate',
  '798',
```

'38',
''),
('Anonymous+Microsoft+Web+Data',
'',
'Recommender',
'',
'37711',
'294',
'1998'),
('Arrhythmia',
'Multivariate',
'Classification',
'Multivariate',
'452',
'279',
'1998'),
('Artificial+Characters',
'Multivariate',
'Classification',
'Multivariate',
'6000',
'7',
'1992'),
('Audiology+%28Original%29',
'Multivariate',
'Classification',
'Multivariate',
'226',
'',
'1987'),
('Audiology+%28Standardized%29',
'Multivariate',
'Classification',
'Multivariate',
'226',
'69',
'1992'),
('Auto+MPG',
'Multivariate',
'Regression',
'Multivariate',
'398',
'8',
'1993'),
('Automobile',
'Multivariate',
'Regression',
'Multivariate',
'205',
'26',
'1987'),
('Badges', 'Univariate', 'Classification', 'Univariate', '294', '1', '1994'),
('Balance+Scale',
'Multivariate',
'Classification',
'Multivariate',
'625',
'4',
'1994'),
('Balloons', 'Multivariate', 'Classification', 'Multivariate', '16', '4', ''),
('Breast+Cancer',
'Multivariate',

'Classification',
'Multivariate',
'286',
'9',
'1988'),
('Breast+Cancer+Wisconsin+%28Original%29',
'Multivariate',
'Classification',
'Multivariate',
'699',
'10',
'1992'),
('Breast+Cancer+Wisconsin+%28Prognostic%29',
'Multivariate',
'Classification',
'Multivariate',
'198',
'34',
'1995'),
('Breast+Cancer+Wisconsin+%28Diagnostic%29',
'Multivariate',
'Classification',
'Multivariate',
'569',
'32',
'1995'),
('Pittsburgh+Bridges',
'Multivariate',
'Classification',
'Multivariate',
'108',
'13',
'1990'),
('Car+Evaluation',
'Multivariate',
'Classification',
'Multivariate',
'1728',
'6',
'1997'),
('Census+Income',
'Multivariate',
'Classification',
'Multivariate',
'48842',
'14',
'1996'),
('Chess+%28King-Rook+vs.+King-Knight%29',
'Multivariate',
'Classification',
'Multivariate',
'',
'22',
'1988'),
('Chess+%28King-Rook+vs.+King-Pawn%29',
'Multivariate',
'Classification',
'Multivariate',
'3196',
'36',
'1989'),
('Chess+%28King-Rook+vs.+King%29',

'Multivariate',
'Classification',
'Multivariate',
'28056',
'6',
'1994'),
('Chess+%28Domain+Theories%29', 'Domain', '', 'Domain', '', '', ''),
('Bach+Chorales', 'Univariate', '', 'Univariate', '100', '6', ''),
('Connect-4',
'Multivariate',
'Classification',
'Multivariate',
'67557',
'42',
'1995'),
('Credit+Approval',
'Multivariate',
'Classification',
'Multivariate',
'690',
'15',
''),
('Japanese+Credit+Screening',
'Multivariate',
'Classification',
'Multivariate',
'125',
'',
'1992'),
('Computer+Hardware',
'Multivariate',
'Regression',
'Multivariate',
'209',
'9',
'1987'),
('Contraceptive+Method+Choice',
'Multivariate',
'Classification',
'Multivariate',
'1473',
'9',
'1997'),
('Covertype',
'Multivariate',
'Classification',
'Multivariate',
'581012',
'54',
'1998'),
('Cylinder+Bands',
'Multivariate',
'Classification',
'Multivariate',
'512',
'39',
'1995'),
('Dermatology',
'Multivariate',
'Classification',
'Multivariate',
'366',

'33',
'1998'),
('Diabetes', 'Multivariate', '', 'Multivariate', '', '20', ''),
('DGP2+-+The+Second+Data+Generation+Program', 'Data', '', 'Data', '', '', ''),
('Document+Understanding', '', '', '', '', '', '1994'),
('EBL+Domain+Theories', '', '', '', '', '', ''),
('Echocardiogram',
'Multivariate',
'Classification',
'Multivariate',
'132',
'12',
'1989'),
('Ecoli',
'Multivariate',
'Classification',
'Multivariate',
'336',
'8',
'1996'),
('Flags',
'Multivariate',
'Classification',
'Multivariate',
'194',
'30',
'1990'),
('Function+Finding', '', 'Function', '', '352', '', '1990'),
('Glass+Identification',
'Multivariate',
'Classification',
'Multivariate',
'214',
'10',
'1987'),
('Haberman%27s+Survival',
'Multivariate',
'Classification',
'Multivariate',
'306',
'3',
'1999'),
('Hayes-Roth',
'Multivariate',
'Classification',
'Multivariate',
'160',
'5',
'1989'),
('Heart+Disease',
'Multivariate',
'Classification',
'Multivariate',
'303',
'75',
'1988'),
('Hepatitis',
'Multivariate',
'Classification',
'Multivariate',
'155',
'19',

'19',
'1988'),
('Horse+Colic',
'Multivariate',
'Classification',
'Multivariate',
'368',
'27',
'1989'),
('ICU', 'Multivariate', '', 'Multivariate', '', '', ''),
('Image+Segmentation',
'Multivariate',
'Classification',
'Multivariate',
'2310',
'19',
'1990'),
('Internet+Advertisements',
'Multivariate',
'Classification',
'Multivariate',
'3279',
'1558',
'1998'),
('Ionosphere',
'Multivariate',
'Classification',
'Multivariate',
'351',
'34',
'1989'),
('Iris',
'Multivariate',
'Classification',
'Multivariate',
'150',
'4',
'1988'),
('ISOLET',
'Multivariate',
'Classification',
'Multivariate',
'7797',
'617',
'1994'),
('Kinship', 'Relational', 'Relational', 'Relational', '104', '12', '1990'),
('Labor+Relations', 'Multivariate', '', 'Multivariate', '57', '16', '1988'),
('LED+Display+Domain',
'Multivariate',
'Classification',
'Multivariate',
'',
'7',
'1988'),
('Lenses',
'Multivariate',
'Classification',
'Multivariate',
'24',
'4',
'1990'),
('Letter+Recognition',

'Multivariate',
'Classification',
'Multivariate',
'20000',
'16',
'1991'),
('Liver+Disorders', 'Multivariate', '', 'Multivariate', '345', '7', '1990'),
('Logic+Theorist', 'Domain', '', 'Domain', '', '', ''),
('Lung+Cancer',
'Multivariate',
'Classification',
'Multivariate',
'32',
'56',
'1992'),
('Lymphography',
'Multivariate',
'Classification',
'Multivariate',
'148',
'18',
'1988'),
('Mechanical+Analysis',
'Multivariate',
'Classification',
'Multivariate',
'209',
'8',
'1990'),
('Meta-data',
'Multivariate',
'Classification',
'Multivariate',
'528',
'22',
'1996'),
('Mobile+Robots', 'Domain', '', 'Domain', '', '', '1995'),
('Molecular+Biology+%28Promoter+Gene+Sequences%29',
'Sequential',
'Classification',
'Sequential',
'106',
'58',
'1990'),
('Molecular+Biology+%28Protein+Secondary+Structure%29',
'Sequential',
'Classification',
'Sequential',
'128',
'',
''),
('Molecular+Biology+%28Splice-junction+Gene+Sequences%29',
'Sequential',
'Classification',
'Sequential',
'3190',
'61',
'1992'),
('MONK%27s+Problems',
'Multivariate',
'Classification',
'Multivariate',

'432',
'7',
'1992'),
('Moral+Reasoner', 'Domain', '', 'Domain', '202', '', '1994'),
('Multiple+Features',
'Multivariate',
'Classification',
'Multivariate',
'2000',
'649',
''),
('Mushroom',
'Multivariate',
'Classification',
'Multivariate',
'8124',
'22',
'1987'),
('Musk+%28Version+1%29',
'Multivariate',
'Classification',
'Multivariate',
'476',
'168',
'1994'),
('Musk+%28Version+2%29',
'Multivariate',
'Classification',
'Multivariate',
'6598',
'168',
'1994'),
('Nursery',
'Multivariate',
'Classification',
'Multivariate',
'12960',
'8',
'1997'),
('Othello+Domain+Theory', 'Domain', '', 'Domain', '', '', '1991'),
('Page+Blocks+Classification',
'Multivariate',
'Classification',
'Multivariate',
'5473',
'10',
'1995'),
('Optical+Recognition+of+Handwritten+Digits',
'Multivariate',
'Classification',
'Multivariate',
'5620',
'64',
'1998'),
('Pen-Based+Recognition+of+Handwritten+Digits',
'Multivariate',
'Classification',
'Multivariate',
'10992',
'16',
'1998'),
('Post-Operative+Patient',

'Multivariate',
'Classification',
'Multivariate',
'90',
'8',
'1993'),
('Primary+Tumor',
'Multivariate',
'Classification',
'Multivariate',
'339',
'17',
'1988'),
('Prodigy', 'Domain', '', 'Domain', '', '', ''),
('Qualitative+Structure+Activity+Relationships',
'Domain',
'',
'Domain',
'',
'',
''),
('Quadruped+Mammals',
'Multivariate',
'Classification',
'Multivariate',
'',
'72',
'1992'),
('Servo', 'Multivariate', 'Regression', 'Multivariate', '167', '4', '1993'),
('Shuttle+Landing+Control',
'Multivariate',
'Classification',
'Multivariate',
'15',
'6',
'1988'),
('Solar+Flare',
'Multivariate',
'Regression',
'Multivariate',
'1389',
'10',
'1989'),
('Soybean+%28Large%29',
'Multivariate',
'Classification',
'Multivariate',
'307',
'35',
'1988'),
('Soybean+%28Small%29',
'Multivariate',
'Classification',
'Multivariate',
'47',
'35',
'1987'),
('Challenger+USA+Space+Shuttle+O-Ring',
'Multivariate',
'Regression',
'Multivariate',
'23',

'4',
'1993'),
('Low+Resolution+Spectrometer',
'Multivariate',
'Classification',
'Multivariate',
'531',
'102',
'1988'),
('Spambase',
'Multivariate',
'Classification',
'Multivariate',
'4601',
'57',
'1999'),
('SPECT+Heart',
'Multivariate',
'Classification',
'Multivariate',
'267',
'22',
'2001'),
('SPECTF+Heart',
'Multivariate',
'Classification',
'Multivariate',
'267',
'44',
'2001'),
('Sponge', 'Multivariate', 'Clustering', 'Multivariate', '76', '45', ''),
('Statlog+Project', '', '', '', '', '', '1992'),
('Student+Loan+Relational', 'Domain', '', 'Domain', '1000', '', '1993'),
('Teaching+Assistant+Evaluation',
'Multivariate',
'Classification',
'Multivariate',
'151',
'5',
'1997'),
('Tic-Tac-Toe+Endgame',
'Multivariate',
'Classification',
'Multivariate',
'958',
'9',
'1991'),
('Thyroid+Disease',
'Multivariate',
'Classification',
'Multivariate',
'7200',
'21',
'1987'),
('Trains',
'Multivariate',
'Classification',
'Multivariate',
'10',
'32',
'1994'),
('University',

 'Multivariate',
 'Classification',
 'Multivariate',
 '285',
 '17',
 '1988'),
('Congressional+Voting+Records',
 'Multivariate',
 'Classification',
 'Multivariate',
 '435',
 '16',
 '1987'),
('Water+Treatment+Plant',
 'Multivariate',
 'Clustering',
 'Multivariate',
 '527',
 '38',
 '1993'),
('Waveform+Database+Generator+%28Version+1%29',
 'Multivariate',
 'Classification',
 'Multivariate',
 '5000',
 '21',
 '1988'),
('Waveform+Database+Generator+%28Version+2%29',
 'Multivariate',
 'Classification',
 'Multivariate',
 '5000',
 '40',
 '1988'),
('Wine',
 'Multivariate',
 'Classification',
 'Multivariate',
 '178',
 '13',
 '1991'),
('Yeast',
 'Multivariate',
 'Classification',
 'Multivariate',
 '1484',
 '8',
 '1996'),
('Zoo',
 'Multivariate',
 'Classification',
 'Multivariate',
 '101',
 '17',
 '1990'),
('Undocumented', '', '', '', '', '', ''),
('Twenty+Newsgroups', 'Text', '', 'Text', '20000', '', '1999'),
('Australian+Sign+Language+signs',
 'Multivariate',
 'Classification',
 'Multivariate',
 '6650',

'15',
'1999'),
('Australian+Sign+Language+signs+%28High+Quality%29',
'Multivariate',
'Classification',
'Multivariate',
'2565',
'22',
'2002'),
('US+Census+Data+%281990%29',
'Multivariate',
'Clustering',
'Multivariate',
'2458285',
'68',
''),
('Census-Income+%28KDD%29',
'Multivariate',
'Classification',
'Multivariate',
'299285',
'40',
'2000'),
('Coil+1999+Competition+Data',
'Multivariate',
'',
'Multivariate',
'340',
'17',
'1999'),
('Corel+Image+Features',
'Multivariate',
'',
'Multivariate',
'68040',
'89',
'1999'),
('E.+Coli+Genes', 'Relational', '', 'Relational', '', '', '2001'),
('EEG+Database', 'Multivariate', '', 'Multivariate', '122', '4', '1999'),
('El+Nino', 'Spatio', '', 'Spatio', '178080', '12', '1999'),
('Entree+Chicago+Recommendation+Data',
'Transactional',
'Recommender',
'Transactional',
'50672',
'',
'2000'),
('CMU+Face+Images', 'Image', 'Classification', 'Image', '640', '', '1999'),
('Insurance+Company+Benchmark+%28COIL+2000%29',
'Multivariate',
'Regression',
'Multivariate',
'9000',
'86',
'2000'),
('Internet+Usage+Data',
'Multivariate',
'',
'Multivariate',
'10104',
'72',
'1999'),

('IPUMS+Census+Database',
 'Multivariate',
 '',
 'Multivariate',
 '256932',
 '61',
 '1999'),
('Japanese+Vowels',
 'Multivariate',
 'Classification',
 'Multivariate',
 '640',
 '12',
 ''),
('KDD+Cup+1998+Data',
 'Multivariate',
 'Regression',
 'Multivariate',
 '191779',
 '481',
 '1998'),
('KDD+Cup+1999+Data',
 'Multivariate',
 'Classification',
 'Multivariate',
 '4000000',
 '42',
 '1999'),
('M.+Tuberculosis+Genes', 'Relational', '', 'Relational', '', '', '2001'),
('Movie', 'Multivariate', '', 'Multivariate', '10000', '', '1999'),
('MSNBC.com+Anonymous+Web+Data',
 'Sequential',
 '',
 'Sequential',
 '989818',
 '',
 ''),
('NSF+Research+Award+Abstracts+1990-2003',
 'Text',
 '',
 'Text',
 '129000',
 '',
 '2003'),
('Pioneer-1+Mobile+Robot+Data',
 'Multivariate',
 '',
 'Multivariate',
 '',
 '',
 '1999'),
('Pseudo+Periodic+Synthetic+Time+Series',
 'Univariate',
 '',
 'Univariate',
 '100000',
 '',
 '1999'),
('Reuters-21578+Text+Categorization+Collection',
 'Text',
 'Classification',
 'Text',

'Text',
'21578',
'5',
'1997'),
('Robot+Execution+Failures',
'Multivariate',
'Classification',
'Multivariate',
'463',
'90',
'1999'),
('Synthetic+Control+Chart+Time+Series',
'Time',
'Classification',
'Time',
'600',
'',
'1999'),
('Syskill+and+Webert+Web+Page+Ratings',
'Multivariate',
'Classification',
'Multivariate',
'332',
'5',
'1998'),
('UNIX+User+Data', 'Text', '', 'Text', '', '', ''),
('Volcanoes+on+Venus+-+JARtool+experiment',
'Image',
'Classification',
'Image',
'',
'',
''),
('Statlog+%28Australian+Credit+Approval%29',
'Multivariate',
'Classification',
'Multivariate',
'690',
'14',
''),
('Statlog+%28German+Credit+Data%29',
'Multivariate',
'Classification',
'Multivariate',
'1000',
'20',
'1994'),
('Statlog+%28Heart%29',
'Multivariate',
'Classification',
'Multivariate',
'270',
'13',
''),
('Statlog+%28Landsat+Satellite%29',
'Multivariate',
'Classification',
'Multivariate',
'6435',
'36',
'1993'),
('Statlog+%28Image+Segmentation%29',
'Multivariate',

'Multivariate',
'Classification',
'Multivariate',
'2310',
'19',
'1990'),
('Statlog+%28Shuttle%29',
'Multivariate',
'Classification',
'Multivariate',
'58000',
'9',
''),
('Statlog+%28Vehicle+Silhouettes%29',
'Multivariate',
'Classification',
'Multivariate',
'946',
'18',
''),
('Connectionist+Bench+%28Nettalk+Corpus%29',
'Multivariate',
'',
'Multivariate',
'20008',
'4',
''),
('Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29',
'Multivariate',
'Classification',
'Multivariate',
'208',
'60',
''),
('Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29',
'',
'Classification',
'',
'528',
'10',
''),
('Economic+Sanctions', 'Domain', '', 'Domain', '', '', ''),
('Protein+Data', '', '', '', '', '', ''),
('Cloud', 'Multivariate', '', 'Multivariate', '1024', '10', '1989'),
('CalIt2+Building+People+Counts',
'Multivariate',
'',
'Multivariate',
'10080',
'4',
'2006'),
('Dodgers+Loop+Sensor',
'Multivariate',
'',
'Multivariate',
'50400',
'3',
'2006'),
('Poker+Hand',
'Multivariate',
'Classification',
'Multivariate',

'1025010',
'11',
'2007'),
('MAGIC+Gamma+Telescope',
'Multivariate',
'Classification',
'Multivariate',
'19020',
'11',
'2007'),
('UJI+Pen+Characters',
'Multivariate',
'Classification',
'Multivariate',
'1364',
'',
'2007'),
('Mammographic+Mass',
'Multivariate',
'Classification',
'Multivariate',
'961',
'6',
'2007'),
('Forest+Fires',
'Multivariate',
'Regression',
'Multivariate',
'517',
'13',
'2008'),
('Reuters+Transcribed+Subset',
'Text',
'Classification',
'Text',
'200',
'',
'2008'),
('Bag+of+Words', 'Text', 'Clustering', 'Text', '8000000', '100000', '2008'),
('Concrete+Compressive+Strength',
'Multivariate',
'Regression',
'Multivariate',
'1030',
'9',
'2007'),
('Hill-Valley',
'Sequential',
'Classification',
'Sequential',
'606',
'101',
'2008'),
('Arcene',
'Multivariate',
'Classification',
'Multivariate',
'900',
'10000',
'2008'),
('Dexter',
'Multivariate',

'Classification',
'Multivariate',
'2600',
'20000',
'2008'),
('Dorothea',
'Multivariate',
'Classification',
'Multivariate',
'1950',
'100000',
'2008'),
('Gisette',
'Multivariate',
'Classification',
'Multivariate',
'13500',
'5000',
'2008'),
('Madelon',
'Multivariate',
'Classification',
'Multivariate',
'4400',
'500',
'2008'),
('Ozone+Level+Detection',
'Multivariate',
'Classification',
'Multivariate',
'2536',
'73',
'2008'),
('Abscisic+Acid+Signaling+Network',
'Multivariate',
'Causal',
'Multivariate',
'300',
'43',
'2008'),
('Parkinsons',
'Multivariate',
'Classification',
'Multivariate',
'197',
'23',
'2008'),
('Character+Trajectories',
'Time',
'Classification',
'Time',
'2858',
'3',
'2008'),
('Blood+Transfusion+Service+Center',
'Multivariate',
'Classification',
'Multivariate',
'748',
'5',
'2008'),
('UJI+Pen+Characters+%28Version+2%29',

'Multivariate',
'Classification',
'Multivariate',
'11640',
'',
'2009'),
('Semeion+Handwritten+Digit',
'Multivariate',
'Classification',
'Multivariate',
'1593',
'256',
'2008'),
('SECOM',
'Multivariate',
'Classification',
'Multivariate',
'1567',
'591',
'2008'),
('Plants',
'Multivariate',
'Clustering',
'Multivariate',
'22632',
'70',
'2008'),
('Libras+Movement',
'Multivariate',
'Classification',
'Multivariate',
'360',
'91',
'2009'),
('Concrete+Slump+Test',
'Multivariate',
'Regression',
'Multivariate',
'103',
'10',
'2009'),
('Communities+and+Crime',
'Multivariate',
'Regression',
'Multivariate',
'1994',
'128',
'2009'),
('Acute+Inflammations',
'Multivariate',
'Classification',
'Multivariate',
'120',
'6',
'2009'),
('Wine+Quality',
'Multivariate',
'Classification',
'Multivariate',
'4898',
'12',
'2009'),

('URL+Reputation',
 'Multivariate',
 'Classification',
 'Multivariate',
 '2396130',
 '3231961',
 '2009'),
('p53+Mutants',
 'Multivariate',
 'Classification',
 'Multivariate',
 '16772',
 '5409',
 '2010'),
('Parkinsons+Telemonitoring',
 'Multivariate',
 'Regression',
 'Multivariate',
 '5875',
 '26',
 '2009'),
('Demospongiae',
 'Multivariate',
 'Classification',
 'Multivariate',
 '503',
 '',
 '2010'),
('Opinosis+Opinion+%26frasl%3B+Review', 'Text', '', 'Text', '51', '', '2010'),
('Breast+Tissue',
 'Multivariate',
 'Classification',
 'Multivariate',
 '106',
 '10',
 '2010'),
('Cardiotocography',
 'Multivariate',
 'Classification',
 'Multivariate',
 '2126',
 '23',
 '2010'),
('Wall-Following+Robot+Navigation+Data',
 'Multivariate',
 'Classification',
 'Multivariate',
 '5456',
 '24',
 '2010'),
('Spoken+Arabic+Digit',
 'Multivariate',
 'Classification',
 'Multivariate',
 '8800',
 '13',
 '2010'),
('Localization+Data+for+Person+Activity',
 'Univariate',
 'Classification',
 'Univariate',
 '164860',

'8',
'2010'),
('AutoUniv',
'Multivariate',
'Classification',
'Multivariate',
'',
'',
'2010'),
('Steel+Plates+Faults',
'Multivariate',
'Classification',
'Multivariate',
'1941',
'27',
'2010'),
('MiniBooNE+particle+identification',
'Multivariate',
'Classification',
'Multivariate',
'130065',
'50',
'2010'),
('YearPredictionMSD',
'Multivariate',
'Regression',
'Multivariate',
'515345',
'90',
'2011'),
('PEMS-SF',
'Multivariate',
'Classification',
'Multivariate',
'440',
'138672',
'2011'),
('OpinRank+Review+Dataset', 'Text', '', 'Text', '', '', '2011'),
('Relative+location+of+CT+slices+on+axial+axis',
'Domain',
'Regression',
'Domain',
'53500',
'386',
'2011'),
('Online+Handwritten+Assamese+Characters+Dataset',
'Multivariate',
'Classification',
'Multivariate',
'8235',
'',
'2011'),
('PubChem+Bioassay+Data',
'Multivariate',
'Classification',
'Multivariate',
'',
'',
'2011'),
('Record+Linkage+Comparison+Patterns',
'Multivariate',
'Classification',

'Multivariate',
'5749132',
'12',
'2011'),
('Communities+and+Crime+Unnormalized',
'Multivariate',
'Regression',
'Multivariate',
'2215',
'147',
'2011'),
('Vertebral+Column',
'Multivariate',
'Classification',
'Multivariate',
'310',
'6',
'2011'),
('EMG+Physical+Action+Data+Set',
'Time',
'Classification',
'Time',
'10000',
'8',
'2011'),
('Vicon+Physical+Action+Data+Set',
'Time',
'Classification',
'Time',
'3000',
'27',
'2011'),
('Amazon+Commerce+reviews+set',
'Multivariate',
'Classification',
'Multivariate',
'1500',
'10000',
'2011'),
('Amazon+Access+Samples',
'Time',
'Regression',
'Time',
'30000',
'20000',
'2011'),
('Reuter_50_50',
'Multivariate',
'Classification',
'Multivariate',
'2500',
'10000',
'2011'),
('Farm+Ads', 'Text', 'Classification', 'Text', '4143', '54877', '2011'),
('DBWorld+e-mails', 'Text', 'Classification', 'Text', '64', '4702', '2011'),
('KEGG+Metabolic+Relation+Network+%28Directed%29',
'Multivariate',
'Classification',
'Multivariate',
'53414',
'24',
'2011')

('KEGG+Metabolic+Reaction+Network+%28Undirected%29',
 'Multivariate',
 'Classification',
 'Multivariate',
 '65554',
 '29',
 '2011'),
('Bank+Marketing',
 'Multivariate',
 'Classification',
 'Multivariate',
 '45211',
 '17',
 '2012'),
('YouTube+Comedy+Slam+Preference+Data',
 'Text',
 'Classification',
 'Text',
 '1138562',
 '3',
 '2012'),
('Gas+Sensor+Array+Drift+Dataset',
 'Multivariate',
 'Classification',
 'Multivariate',
 '13910',
 '128',
 '2012'),
('ILPD+%28Indian+Liver+Patient+Dataset%29',
 'Multivariate',
 'Classification',
 'Multivariate',
 '583',
 '10',
 '2012'),
('OPPORTUNITY+Activity+Recognition',
 'Multivariate',
 'Classification',
 'Multivariate',
 '2551',
 '242',
 '2012'),
('Nomao',
 'Univariate',
 'Classification',
 'Univariate',
 '34465',
 '120',
 '2012'),
('SMS+Spam+Collection',
 'Multivariate',
 'Classification',
 'Multivariate',
 '5574',
 '',
 '2012'),
('Skin+Segmentation',
 'Univariate',
 'Classification',
 'Univariate',
 '245057',
 '4',

'4',
'2012'),
('Planning+Relax',
'Univariate',
'Classification',
'Univariate',
'182',
'13',
'2012'),
('PAMAP2+Physical+Activity+Monitoring',
'Multivariate',
'Classification',
'Multivariate',
'3850505',
'52',
'2012'),
('Restaurant+%26+consumer+data',
'Multivariate',
'',
'Multivariate',
'138',
'47',
'2012'),
('CNAE-9',
'Multivariate',
'Classification',
'Multivariate',
'1080',
'857',
'2012'),
('Individual+household+electric+power+consumption',
'Multivariate',
'Regression',
'Multivariate',
'2075259',
'9',
'2012'),
('seeds',
'Multivariate',
'Classification',
'Multivariate',
'210',
'7',
'2012'),
('Northix',
'Multivariate',
'Classification',
'Multivariate',
'115',
'200',
'2012'),
('QtyT40I10D100K', 'Sequential', '', 'Sequential', '3960456', '4', '2012'),
('Legal+Case+Reports', 'Text', 'Classification', 'Text', '', '', '2012'),
('Human+Activity+Recognition+Using+Smartphones',
'Multivariate',
'Classification',
'Multivariate',
'10299',
'561',
'2012'),
('One-hundred+plant+species+leaves+data+set',
'',
'Classification',

'Classification',
'',
'1600',
'64',
'2012'),
('Energy+efficiency',
'Multivariate',
'Classification',
'Multivariate',
'768',
'8',
'2012'),
('Yacht+Hydrodynamics',
'Multivariate',
'Regression',
'Multivariate',
'308',
'7',
'2013'),
('Fertility',
'Multivariate',
'Classification',
'Multivariate',
'100',
'10',
'2013'),
('Daphnet+Freezing+of+Gait',
'Multivariate',
'Classification',
'Multivariate',
'237',
'9',
'2013'),
('3D+Road+Network+%28North+Jutland%2C+Denmark%29',
'Sequential',
'Regression',
'Sequential',
'434874',
'4',
'2013'),
('ISTANBUL+STOCK+EXCHANGE',
'Multivariate',
'Classification',
'Multivariate',
'536',
'8',
'2013'),
('Buzz+in+social+media+',
'Time',
'Regression',
'Time',
'140000',
'77',
'2013'),
('First-order+theorem+proving',
'Multivariate',
'Classification',
'Multivariate',
'6118',
'51',
'2013'),
('Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29',

'Sequential',
'Classification',
'Sequential',
'165632',
'18',
'2013'),
('Gas+sensor+arrays+in+open+sampling+settings',
'Multivariate',
'Classification',
'Multivariate',
'18000',
'1950000',
'2013'),
('Climate+Model+Simulation+Crashes',
'Multivariate',
'Classification',
'Multivariate',
'540',
'18',
'2013'),
('MicroMass',
'Multivariate',
'Classification',
'Multivariate',
'931',
'1300',
'2013'),
('QSAR+biodegradation',
'Multivariate',
'Classification',
'Multivariate',
'1055',
'41',
'2013'),
('BLOGGER',
'Multivariate',
'Classification',
'Multivariate',
'100',
'6',
'2013'),
('Daily+and+Sports+Activities',
'Multivariate',
'Classification',
'Multivariate',
'9120',
'5625',
'2013'),
('User+Knowledge+Modeling',
'Multivariate',
'Classification',
'Multivariate',
'403',
'5',
'2013'),
('Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection',
'Multivariate',
'Classification',
'Multivariate',
'111740',
'',
'2013'),

```
 ('NYSK', 'Multivariate', 'Clustering', 'Multivariate', '10421', '7', '2013'),
 ('Turkiye+Student+Evaluation',
  'Multivariate',
  'Classification',
  'Multivariate',
  '5820',
  '33',
  '2013'),
 ('ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machines%29',
  'Multivariate',
  'Classification',
  'Multivariate',
  '403',
  '5',
  '2013'),
 ('EEG+Eye+State',
  'Multivariate',
  'Classification',
  'Multivariate',
  '14980',
  '15',
  '2013'),
 ('Physicochemical+Properties+of+Protein+Tertiary+Structure',
  'Multivariate',
  'Regression',
  'Multivariate',
  '45730',
  '9',
  '2013'),
 ('seismic-bumps',
  'Multivariate',
  'Classification',
  'Multivariate',
  '2584',
  '19',
  '2013'),
 ('banknote+authentication',
  'Multivariate',
  'Classification',
  'Multivariate',
  '1372',
  '5',
  '2013'),
 ('USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Problem%
3A+Pat',
  'Domain',
  'Classification',
  'Domain',
  '306',
  '5',
  '2013'),
 ('YouTube+Multiview+Video+Games+Dataset',
  'Multivariate',
  'Classification',
  'Multivariate',
  '120000',
  '1000000',
  '2013'),
 ('Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations',
  'Multivariate',
  'Classification',
  'Multivariate',
```

'13910',
'129',
'2013'),
('Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors',
'Multivariate',
'Classification',
'Multivariate',
'2747',
'',
'2013'),
('SkillCraft1+Master+Table+Dataset',
'Multivariate',
'Regression',
'Multivariate',
'3395',
'20',
'2013'),
('Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units',
'Multivariate',
'Classification',
'Multivariate',
'39242',
'152',
'2013'),
('SML2010',
'Multivariate',
'Regression',
'Multivariate',
'4137',
'24',
'2014'),
('Bike+Sharing+Dataset',
'Univariate',
'Regression',
'Univariate',
'17389',
'16',
'2013'),
('Predict+keywords+activities+in+a+online+social+media',
'Multivariate',
'',
'Multivariate',
'51',
'35',
'2013'),
('Thoracic+Surgery+Data',
'Multivariate',
'Classification',
'Multivariate',
'470',
'17',
'2013'),
('EMG+dataset+in+Lower+Limb',
'Multivariate',
'',
'Multivariate',
'132',
'5',
'2014'),
('SUSY', '', 'Classification', '', '5000000', '18', '2014'),
('HIGGS', '', 'Classification', '', '11000000', '28', '2014'),
('Qualitative_Bankruptcy',

'Multivariate',
'Classification',
'Multivariate',
'250',
'7',
'2014'),
('LSVT+Voice+Rehabilitation',
'Multivariate',
'Classification',
'Multivariate',
'126',
'309',
'2014'),
('Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer',
'Multivariate',
'Classification',
'Multivariate',
'',
'3',
'2014'),
('Wilt',
'Multivariate',
'Classification',
'Multivariate',
'4889',
'6',
'2014'),
('User+Identification+From+Walking+Activity',
'Univariate',
'Classification',
'Univariate',
'',
'',
'2014'),
('Activity+Recognition+from+Single+Chest-Mounted+Accelerometer',
'Univariate',
'Classification',
'Univariate',
'',
'',
'2014'),
('Leaf',
'Multivariate',
'Classification',
'Multivariate',
'340',
'16',
'2014'),
('Dresses_Attribute_Sales',
'Text',
'Classification',
'Text',
'501',
'13',
'2014'),
('Tamilnadu+Electricity+Board+Hourly+Readings',
'Multivariate',
'Classification',
'Multivariate',
'45781',
'5',
'2013'),

('Airfoil+Self-Noise',
 'Multivariate',
 'Regression',
 'Multivariate',
 '1503',
 '6',
 '2014'),
('Wholesale+customers',
 'Multivariate',
 'Classification',
 'Multivariate',
 '440',
 '8',
 '2014'),
('Twitter+Data+set+for+Arabic+Sentiment+Analysis',
 'Text',
 'Classification',
 'Text',
 '2000',
 '2',
 '2014'),
('Combined+Cycle+Power+Plant',
 'Multivariate',
 'Regression',
 'Multivariate',
 '9568',
 '4',
 '2014'),
('Urban+Land+Cover',
 'Multivariate',
 'Classification',
 'Multivariate',
 '168',
 '148',
 '2014'),
('Diabetes+130-US+hospitals+for+years+1999-2008',
 'Multivariate',
 'Classification',
 'Multivariate',
 '100000',
 '55',
 '2014'),
('Bach+Choral+Harmony',
 'Sequential',
 'Classification',
 'Sequential',
 '5665',
 '17',
 '2014'),
('StoneFlakes',
 'Multivariate',
 'Classification',
 'Multivariate',
 '79',
 '8',
 '2014'),
('Tennis+Major+Tournament+Match+Statistics',
 'Multivariate',
 'Classification',
 'Multivariate',
 '127',
 '42'

'2014'),
('Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings',
'Multivariate',
'Classification',
'Multivariate',
'1040',
'26',
'2014'),
('Gesture+Phase+Segmentation',
'Multivariate',
'Classification',
'Multivariate',
'9900',
'50',
'2014'),
('Perfume+Data',
'Univariate',
'Classification',
'Univariate',
'560',
'2',
'2014'),
('BlogFeedback',
'Multivariate',
'Regression',
'Multivariate',
'60021',
'281',
'2014'),
('REALDISP+Activity+Recognition+Dataset',
'Multivariate',
'Classification',
'Multivariate',
'1419',
'120',
'2014'),
('Newspaper+and+magazine+images+segmentation+dataset',
'',
'Classification',
'',
'101',
'',
'2014'),
('AAAI+2014+Accepted+Papers',
'Multivariate',
'Clustering',
'Multivariate',
'399',
'6',
'2014'),
('Gas+sensor+array+under+flow+modulation',
'Multivariate',
'Classification',
'Multivariate',
'58',
'120432',
'2014'),
('Gas+sensor+array+exposed+to+turbulent+gas+mixtures',
'Multivariate',
'Classification',
'Multivariate',
'180',

'150000',
'2014'),
('UJIIndoorLoc',
'Multivariate',
'Classification',
'Multivariate',
'21048',
'529',
'2014'),
('Sentence+Classification', 'Text', 'Classification', 'Text', '', '', '2014'),
('Dow+Jones+Index', 'Time', 'Classification', 'Time', '750', '16', '2014'),
('sEMG+for+Basic+Hand+movements',
'Time',
'Classification',
'Time',
'3000',
'2500',
'2014'),
('AAAI+2013+Accepted+Papers',
'Multivariate',
'Clustering',
'Multivariate',
'150',
'5',
'2014'),
('Geographical+Original+of+Music',
'Multivariate',
'Classification',
'Multivariate',
'1059',
'68',
'2014'),
('Condition+Based+Maintenance+of+Naval+Propulsion+Plants',
'Multivariate',
'Regression',
'Multivariate',
'11934',
'16',
'2014'),
('Grammatical+Facial+Expressions',
'Multivariate',
'Classification',
'Multivariate',
'27965',
'100',
'2014'),
('NoisyOffice',
'Multivariate',
'Classification',
'Multivariate',
'216',
'216',
'2015'),
('MHEALTH+Dataset',
'Multivariate',
'Classification',
'Multivariate',
'120',
'23',
'2014'),
('Student+Performance',
'Multivariate',

'Multivariate',
'Classification',
'Multivariate',
'649',
'33',
'2014'),
('ElectricityLoadDiagrams20112014',
'Time',
'Regression',
'Time',
'370',
'140256',
'2015'),
('Gas+sensor+array+under+dynamic+gas+mixtures',
'Multivariate',
'Classification',
'Multivariate',
'4178504',
'19',
'2015'),
('microblogPCU',
'Multivariate',
'Classification',
'Multivariate',
'221579',
'20',
'2015'),
('Firm-Teacher_Clave-Direction_Classification',
'Multivariate',
'Classification',
'Multivariate',
'10800',
'20',
'2015'),
('Dataset+for+Sensorless+Drive+Diagnosis',
'Multivariate',
'Classification',
'Multivariate',
'58509',
'49',
'2015'),
('TV+News+Channel+Commercial+Detection+Dataset',
'Multivariate',
'Classification',
'Multivariate',
'129685',
'12',
'2015'),
('Phishing+Websites', '', 'Classification', '', '2456', '30', '2015'),
('Greenhouse+Gas+Observing+Network',
'Multivariate',
'Regression',
'Multivariate',
'2921',
'5232',
'2015'),
('Diabetic+Retinopathy+Debrecen+Data+Set',
'Multivariate',
'Classification',
'Multivariate',
'1151',
'20',

'2014'),
('HIV-1+protease+cleavage',
'Multivariate',
'Classification',
'Multivariate',
'6590',
'1',
'2015'),
('Sentiment+Labelled+Sentences',
'Text',
'Classification',
'Text',
'3000',
'',
'2015'),
('Online+News+Popularity',
'Multivariate',
'Classification',
'Multivariate',
'39797',
'61',
'2015'),
('Forest+type+mapping',
'Multivariate',
'Classification',
'Multivariate',
'326',
'27',
'2015'),
('wiki4HE',
'Multivariate',
'Regression',
'Multivariate',
'913',
'53',
'2015'),
('Online+Video+Characteristics+and+Transcoding+Time+Dataset',
'Multivariate',
'Regression',
'Multivariate',
'168286',
'11',
'2015'),
('Chronic_Kidney_Disease',
'Multivariate',
'Classification',
'Multivariate',
'400',
'25',
'2015'),
('Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014',
'Sequential',
'Classification',
'Sequential',
'314080',
'0',
'2015'),
('Folio',
'Multivariate',
'Classification',
'Multivariate',
'637',

'20',
'2015'),
('Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015',
'Multivariate',
'Clustering',
'Multivariate',
'1710671',
'9',
'2015'),
('Cuff-Less+Blood+Pressure+Estimation',
'Multivariate',
'Classification',
'Multivariate',
'12000',
'3',
'2015'),
('Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions',
'Multivariate',
'Classification',
'Multivariate',
'10929',
'561',
'2015'),
('Mice+Protein+Expression',
'Multivariate',
'Classification',
'Multivariate',
'1080',
'82',
'2015'),
('UJIIndoorLoc-Mag',
'Multivariate',
'Classification',
'Multivariate',
'40000',
'13',
'2015'),
('Heterogeneity+Activity+Recognition',
'Multivariate',
'Classification',
'Multivariate',
'43930257',
'16',
'2015'),
('Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set',
'Multivariate',
'Classification',
'Multivariate',
'230318',
'13',
'2015'),
('HEPMASS',
'Multivariate',
'Classification',
'Multivariate',
'10500000',
'28',
'2016'),
('Indoor+User+Movement+Prediction+from+RSS+data',
'Multivariate',
'Classification',
'Multivariate',

'13197',
'4',
'2016'),
('Open+University+Learning+Analytics+dataset',
'Multivariate',
'Classification',
'Multivariate',
'',
'',
'2015'),
('default+of+credit+card+clients',
'Multivariate',
'Classification',
'Multivariate',
'30000',
'24',
'2016'),
('Mesothelioma%E2%80%99s+disease+data+set+',
'Multivariate',
'Classification',
'Multivariate',
'324',
'34',
'2016'),
('Online+Retail',
'Multivariate',
'Classification',
'Multivariate',
'541909',
'8',
'2015'),
('SIFT10M',
'Multivariate',
'Causal',
'Multivariate',
'11164866',
'128',
'2016'),
('GPS+Trajectories',
'Multivariate',
'Classification',
'Multivariate',
'163',
'15',
'2016'),
('Detect+Malacious+Executable%28AntiVirus%29',
'Multivariate',
'Classification',
'Multivariate',
'373',
'513',
'2016'),
('Occupancy+Detection+',
'Multivariate',
'Classification',
'Multivariate',
'20560',
'7',
'2016'),
('Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Parkinson%E2%80%99s+Disease',
'Multivariate',
'Classification',

 'Multivariate',
 '40',
 '7',
 '2016'),
('News+Aggregator',
 'Multivariate',
 'Classification',
 'Multivariate',
 '422937',
 '5',
 '2016'),
('Air+Quality',
 'Multivariate',
 'Regression',
 'Multivariate',
 '9358',
 '15',
 '2016'),
('Twin+gas+sensor+arrays',
 'Multivariate',
 'Classification',
 'Multivariate',
 '640',
 '480000',
 '2016'),
('Gas+sensors+for+home+activity+monitoring',
 'Multivariate',
 'Classification',
 'Multivariate',
 '919438',
 '11',
 '2016'),
('Facebook+Comment+Volume+Dataset',
 'Multivariate',
 'Regression',
 'Multivariate',
 '40949',
 '54',
 '2016'),
('Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Living+%28AAL
%29',
 'Time',
 'Classification',
 'Time',
 '5744',
 '561',
 '2016'),
('Polish+companies+bankruptcy+data',
 'Multivariate',
 'Classification',
 'Multivariate',
 '10503',
 '64',
 '2016'),
('Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29',
 'Multivariate',
 'Classification',
 'Multivariate',
 '42240',
 '6',
 '2016'),
('Dota2+Games+Results',

'Multivariate',
'Classification',
'Multivariate',
'102944',
'116',
'2016'),
('Facebook+metrics',
'Multivariate',
'Regression',
'Multivariate',
'500',
'19',
'2016'),
('UbiqLog+%28smartphone+lifelogging%29',
'Multivariate',
'Causal',
'Multivariate',
'9782222',
'',
'2016'),
('NIPS+Conference+Papers+1987-2015',
'Text',
'Clustering',
'Text',
'11463',
'5812',
'2016'),
('HTRU2',
'Multivariate',
'Classification',
'Multivariate',
'17898',
'9',
'2017'),
('Drug+consumption+%28quantified%29',
'Multivariate',
'Classification',
'Multivariate',
'1885',
'32',
'2016'),
('Appliances+energy+prediction',
'Multivariate',
'Regression',
'Multivariate',
'19735',
'29',
'2017'),
('Miskolc+IIS+Hybrid+IPS',
'Text',
'Classification',
'Text',
'1540',
'67',
'2016'),
('KDC-4007+dataset+Collection',
'Multivariate',
'Classification',
'Multivariate',
'4007',
'',
'2017'),

('Geo-Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone',
 'Multivariate',
 'Classification',
 'Multivariate',
 '153540',
 '25',
 '2017'),
('DrivFace',
 'Multivariate',
 'Classification',
 'Multivariate',
 '606',
 '6400',
 '2016'),
('Website+Phishing',
 'Multivariate',
 'Classification',
 'Multivariate',
 '1353',
 '10',
 '2016'),
('YouTube+Spam+Collection',
 'Text',
 'Classification',
 'Text',
 '1956',
 '5',
 '2017'),
('Beijing+PM2.5+Data',
 'Multivariate',
 'Regression',
 'Multivariate',
 '43824',
 '13',
 '2017'),
('Cargo+2000+Freight+Tracking+and+Tracing',
 'Multivariate',
 'Classification',
 'Multivariate',
 '3942',
 '98',
 '2016'),
('Cervical+cancer+%28Risk+Factors%29',
 'Multivariate',
 'Classification',
 'Multivariate',
 '858',
 '36',
 '2017'),
('Quality+Assessment+of+Digital+Colposcopies',
 'Multivariate',
 'Classification',
 'Multivariate',
 '287',
 '69',
 '2017'),
('KASANDR',
 'Multivariate',
 'Causal',
 'Multivariate',
 '17764280',
 '2158859',

'2017'),
('FMA%3A+A+Dataset+For+Music+Analysis',
'Multivariate',
'Classification',
'Multivariate',
'106574',
'518',
'2017'),
('Air+quality',
'Multivariate',
'Regression',
'Multivariate',
'9358',
'15',
'2016'),
('Epileptic+Seizure+Recognition',
'Multivariate',
'Classification',
'Multivariate',
'11500',
'179',
'2017'),
('Devanagari+Handwritten+Character+Dataset',
'',
'Classification',
'',
'92000',
'',
'2016'),
('Stock+portfolio+performance',
'Multivariate',
'Regression',
'Multivariate',
'315',
'12',
'2016'),
('MoCap+Hand+Postures',
'Multivariate',
'Classification',
'Multivariate',
'78095',
'38',
'2016'),
('Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech',
'Multivariate',
'Classification',
'Multivariate',
'130',
'65',
'2017'),
('Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting',
'Sequential',
'Classification',
'Sequential',
'74',
'102',
'2017'),
('PM2.5+Data+of+Five+Chinese+Cities',
'Multivariate',
'Regression',
'Multivariate',
'52854',

'52854',
'86',
'2017'),
('Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet',
'Multivariate',
'Classification',
'Multivariate',
'77',
'7',
'2017'),
('Sales_Transactions_Dataset_Weekly',
'Multivariate',
'Clustering',
'Multivariate',
'811',
'53',
'2017'),
('Las+Vegas+Strip', '', 'Classification', '', '504', '20', '2017'),
('Eco-hotel', 'Text', '', 'Text', '401', '1', '2017'),
('MEU-Mobile+KSD',
'Multivariate',
'Classification',
'Multivariate',
'2856',
'71',
'2016'),
('Crowdsourced+Mapping',
'Multivariate',
'Classification',
'Multivariate',
'10546',
'29',
'2016'),
('gene+expression+cancer+RNA-Seq',
'Multivariate',
'Classification',
'Multivariate',
'801',
'20531',
'2016'),
('Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer',
'Multivariate',
'Classification',
'Multivariate',
'1540',
'65',
'2016'),
('chestnut+%E2%80%93+LARVIC', '', 'Classification', '', '1451', '3', '2017'),
('Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%28OBS%29+Network',
'Text',
'Classification',
'Text',
'1075',
'22',
'2017'),
('Motion+Capture+Hand+Postures',
'Multivariate',
'Classification',
'Multivariate',
'78095',
'38',
'2017'),
('Anuran+Calls+%28MFCCs%29',

('Anuran+Calls+%28MFCCs%29',
 'Multivariate',
 'Classification',
 'Multivariate',
 '7195',
 '22',
 '2017'),
('TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization',
 'Text',
 'Classification',
 'Text',
 '3600',
 '4814',
 '2017'),
('Gastrointestinal+Lesions+in+Regular+Colonoscopy',
 'Multivariate',
 'Classification',
 'Multivariate',
 '76',
 '698',
 '2016'),
('Daily+Demand+Forecasting+Orders',
 'Time',
 'Regression',
 'Time',
 '60',
 '13',
 '2017'),
('Paper+Reviews', 'Text', 'Classification', 'Text', '405', '10', '2017'),
('extention+of+Z-Alizadeh+sani+dataset',
 '',
 'Classification',
 '',
 '303',
 '59',
 '2017'),
('Z-Alizadeh+Sani', '', 'Classification', '', '303', '56', '2017'),
('Dynamic+Features+of+VirusShare+Executables',
 'Multivariate',
 'Classification',
 'Multivariate',
 '107888',
 '482',
 '2017'),
('IDA2016Challenge',
 'Multivariate',
 'Classification',
 'Multivariate',
 '76000',
 '171',
 '2017'),
('DSRC+Vehicle+Communications',
 'Sequential',
 'Clustering',
 'Sequential',
 '10000',
 '5',
 '2017'),
('Mturk+User-Perceived+Clusters+over+Images',
 'Multivariate',
 'Clustering',
 'Multivariate',

'180',
'500',
'2016'),
('Character+Font+Images',
'Multivariate',
'Classification',
'Multivariate',
'745000',
'411',
'2016'),
('DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Instance+Labels',
'Text',
'Classification',
'Text',
'12234',
'8519',
'2016'),
('Autistic+Spectrum+Disorder+Screening+Data+for+Children++',
'Multivariate',
'Classification',
'Multivariate',
'292',
'21',
'2017'),
('Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++',
'Multivariate',
'Classification',
'Multivariate',
'104',
'21',
'2017'),
('APS+Failure+at+Scania+Trucks',
'Multivariate',
'Classification',
'Multivariate',
'60000',
'171',
'2017'),
('Wireless+Indoor+Localization',
'Multivariate',
'Classification',
'Multivariate',
'2000',
'7',
'2017'),
('HCC+Survival',
'Multivariate',
'Classification',
'Multivariate',
'165',
'49',
'2017'),
('CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015',
'Multivariate',
'Classification',
'Multivariate',
'217',
'12',
'2017'),
('University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29',
'Text',
'Classification',

'Text',
'1175',
'3',
'2017'),
('Autism+Screening+Adult', '', 'Classification', '', '704', '21', '2017'),
('Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor',
'Sequential',
'Classification',
'Sequential',
'75128',
'9',
'2016'),
('Immunotherapy+Dataset',
'Univariate',
'Classification',
'Univariate',
'90',
'8',
'2018'),
('Cryotherapy+Dataset+',
'Univariate',
'Classification',
'Univariate',
'90',
'7',
'2018'),
('OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes',
'Multivariate',
'Classification',
'Multivariate',
'50',
'2',
'2016'),
('Discrete+Tone+Image+Dataset',
'Multivariate',
'Classification',
'Multivariate',
'71',
'11',
'2018'),
('News+Popularity+in+Multiple+Social+Media+Platforms',
'Multivariate',
'Regression',
'Multivariate',
'93239',
'11',
'2018'),
('Ultrasonic+flowmeter+diagnostics',
'Multivariate',
'Classification',
'Multivariate',
'540',
'173',
'2018'),
('ICMLA+2014+Accepted+Papers+Data+Set',
'Multivariate',
'Classification',
'Multivariate',
'105',
'5',
'2018'),
('BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation',

'Multivariate',
'Classification',
'Multivariate',
'6611',
'15',
'2018'),
('Container+Crane+Controller+Data+Set',
'Univariate',
'Classification',
'Univariate',
'15',
'3',
'2018'),
('Residential+Building+Data+Set',
'Multivariate',
'Regression',
'Multivariate',
'372',
'105',
'2018'),
('Health+News+in+Twitter',
'Text',
'Clustering',
'Text',
'58000',
'25000',
'2018'),
('chipseq', 'Sequential', 'Classification', 'Sequential', '4960', '', '2018'),
('SGEMM+GPU+kernel+performance',
'Multivariate',
'Regression',
'Multivariate',
'241600',
'18',
'2018'),
('Repeat+Consumption+Matrices',
'Multivariate',
'Clustering',
'Multivariate',
'130000',
'21000',
'2018'),
('detection_of_IoT_botnet_attacks_N_BaIoT',
'Multivariate',
'Classification',
'Multivariate',
'7062606',
'115',
'2018'),
('Absenteeism+at+work',
'Multivariate',
'Classification',
'Multivariate',
'740',
'21',
'2018'),
('SCADI',
'Multivariate',
'Classification',
'Multivariate',
'70',
'206',

'2018'),
('Condition+monitoring+of+hydraulic+systems',
'Multivariate',
'Classification',
'Multivariate',
'2205',
'43680',
'2018'),
('Carbon+Nanotubes',
'Univariate',
'Regression',
'Univariate',
'10721',
'8',
'2018'),
('Optical+Interconnection+Network+',
'Multivariate',
'Classification',
'Multivariate',
'640',
'10',
'2018'),
('Sports+articles+for+objectivity+analysis',
'Multivariate',
'Classification',
'Multivariate',
'1000',
'59',
'2018'),
('Breast+Cancer+Coimbra',
'Multivariate',
'Classification',
'Multivariate',
'116',
'10',
'2018'),
('GNFUV+Unmanned+Surface+Vehicles+Sensor+Data',
'Multivariate',
'Regression',
'Multivariate',
'1672',
'5',
'2018'),
('Dishonest+Internet+users+Dataset',
'Multivariate',
'Classification',
'Multivariate',
'322',
'5',
'2018'),
('Victorian+Era+Authorship+Attribution',
'Text',
'Classification',
'Text',
'93600',
'1000',
'2018'),
('Simulated+Falls+and+Daily+Living+Activities+Data+Set',
'Time',
'Classification',
'Time',
'3060',

'138',
'2018'),
('Multimodal+Damage+Identification+for+Humanitarian+Computing',
'Multivariate',
'Classification',
'Multivariate',
'5879',
'',
'2018'),
('EEG+Steady-State+Visual+Evoked+Potential+Signals',
'Multivariate',
'Classification',
'Multivariate',
'9200',
'16',
'2018'),
('Roman+Urdu+Data+Set',
'Text',
'Classification',
'Text',
'20000',
'2',
'2018'),
('Avila',
'Multivariate',
'Classification',
'Multivariate',
'20867',
'10',
'2018'),
('PANDOR', 'Multivariate', 'Recommendation', 'Multivariate', '', '', '2018'),
('Drug+Review+Dataset+%28Druglib.com%29',
'Multivariate',
'Classification',
'Multivariate',
'4143',
'8',
'2018'),
('Drug+Review+Dataset+%28Drugs.com%29',
'Multivariate',
'Classification',
'Multivariate',
'215063',
'6',
'2018'),
('Physical+Unclonable+Functions',
'Multivariate',
'Classification',
'Multivariate',
'6000000',
'129',
'2018'),
('Superconductivty+Data',
'Multivariate',
'Regression',
'Multivariate',
'21263',
'81',
'2018'),
('WESAD+%28Wearable+Stress+and+Affect+Detection%29',
'Multivariate',
'Classification'

'Multivariate',
'63000000',
'12',
'2018'),
('GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2',
'Multivariate',
'Regression',
'Multivariate',
'10190',
'6',
'2018'),
('Student+Academics+Performance',
'Multivariate',
'Classification',
'Multivariate',
'300',
'22',
'2018'),
('Online+Shoppers+Purchasing+Intention+Dataset',
'Multivariate',
'Classification',
'Multivariate',
'12330',
'18',
'2018'),
('PMU-UD', 'Univariate', 'Classification', 'Univariate', '5180', '9', '2018'),
('Parkinson%27s+Disease+Classification',
'Multivariate',
'Classification',
'Multivariate',
'756',
'754',
'2018'),
('Electrical+Grid+Stability+Simulated+Data+',
'Multivariate',
'Classification',
'Multivariate',
'10000',
'14',
'2018'),
('Caesarian+Section+Classification+Dataset',
'Univariate',
'Classification',
'Univariate',
'80',
'5',
'2018'),
('BAUM-1', 'Time', 'Classification', 'Time', '1184', '', '2018'),
('BAUM-2', 'Time', 'Classification', 'Time', '1047', '', '2018'),
('Audit+Data',
'Multivariate',
'Classification',
'Multivariate',
'777',
'18',
'2018'),
('BuddyMove+Data+Set',
'Multivariate',
'Classification',
'Multivariate',
'249',
'7'

'2018'),
('Real+estate+valuation+data+set',
'Multivariate',
'Regression',
'Multivariate',
'414',
'7',
'2018'),
('Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connected+speech+Data+Set+',
'Multivariate',
'Classification',
'Multivariate',
'',
'',
'2018'),
('Somerville+Happiness+Survey', '', 'Classification', '', '143', '7', '2018'),
('2.4+GHZ+Indoor+Channel+Measurements',
'Multivariate',
'Classification',
'Multivariate',
'7840',
'5',
'2018'),
('EMG+data+for+gestures',
'Time',
'Classification',
'Time',
'30000',
'6',
'2019'),
('Parking+Birmingham',
'Multivariate',
'Classification',
'Multivariate',
'35717',
'4',
'2019'),
('Behavior+of+the+urban+traffic+of+the+city+of+Sao+Paulo+in+Brazil',
'Multivariate',
'Classification',
'Multivariate',
'135',
'18',
'2018'),
('Travel+Reviews',
'Multivariate',
'Classification',
'Multivariate',
'980',
'11',
'2018'),
('Tarvel+Review+Ratings',
'Multivariate',
'Classification',
'Multivariate',
'5456',
'25',
'2018')]

In [165]:

```
df =pd.DataFrame(dataset, columns = [ 'Name' , 'Data Type' , 'Default Task' , 'Attribute Type' , '# Instances' ,
"# Attributes", 'Year'])
df
```

Out[165]:

	Name	Data Type	Default Task	Attribute Type
0	Abalone	Multivariate	Classification	Multivariate
1	Adult	Multivariate	Classification	Multivariate
2	Annealing	Multivariate	Classification	Multivariate
3	Anonymous+Microsoft+Web+Data		Recommender	
4	Arrhythmia	Multivariate	Classification	Multivariate
5	Artificial+Characters	Multivariate	Classification	Multivariate
6	Audiology+%28Original%29	Multivariate	Classification	Multivariate
7	Audiology+%28Standardized%29	Multivariate	Classification	Multivariate
8	Auto+MPG	Multivariate	Regression	Multivariate
9	Automobile	Multivariate	Regression	Multivariate
10	Badges	Univariate	Classification	Univariate
11	Balance+Scale	Multivariate	Classification	Multivariate
12	Balloons	Multivariate	Classification	Multivariate
13	Breast+Cancer	Multivariate	Classification	Multivariate
14	Breast+Cancer+Wisconsin+%28Original%29	Multivariate	Classification	Multivariate
15	Breast+Cancer+Wisconsin+%28Prognostic%29	Multivariate	Classification	Multivariate
16	Breast+Cancer+Wisconsin+%28Diagnostic%29	Multivariate	Classification	Multivariate
17	Pittsburgh+Bridges	Multivariate	Classification	Multivariate
18	Car+Evaluation	Multivariate	Classification	Multivariate
19	Census+Income	Multivariate	Classification	Multivariate
20	Chess+%28King-Rook+vs.+King-Knight%29	Multivariate	Classification	Multivariate
21	Chess+%28King-Rook+vs.+King-Pawn%29	Multivariate	Classification	Multivariate
22	Chess+%28King-Rook+vs.+King%29	Multivariate	Classification	Multivariate
23	Chess+%28Domain+Theories%29	Domain		Domain
24	Bach+Chorales	Univariate		Univariate
25	Connect-4	Multivariate	Classification	Multivariate
26	Credit+Approval	Multivariate	Classification	Multivariate
27	Japanese+Credit+Screening	Multivariate	Classification	Multivariate
28	Computer+Hardware	Multivariate	Regression	Multivariate
29	Contraceptive+Method+Choice	Multivariate	Classification	Multivariate
...	...	...	...	..
438	Multimodal+Damage+Identification+for+Humanitar...	Multivariate	Classification	Multivariate
439	EEG+Steady-State+Visual+Evoked+Potential+Signals	Multivariate	Classification	Multivariate
440	Roman+Urdu+Data+Set	Text	Classification	Text
441	Avila	Multivariate	Classification	Multivariate

	Name	Data Type	Default Task	Attribute Type
442	PANDOR	Multivariate	Recommendation	Multivariate
443	Drug+Review+Dataset+%28Druglib.com%29	Multivariate	Classification	Multivariate
444	Drug+Review+Dataset+%28Drugs.com%29	Multivariate	Classification	Multivariate
445	Physical+Unclonable+Functions	Multivariate	Classification	Multivariate
446	Superconductivty+Data	Multivariate	Regression	Multivariate
447	WESAD+%28Wearable+Stress+and+Affect+Detection%29	Multivariate	Classification	Multivariate
448	GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2	Multivariate	Regression	Multivariate
449	Student+Academics+Performance	Multivariate	Classification	Multivariate
450	Online+Shoppers+Purchasing+Intention+Dataset	Multivariate	Classification	Multivariate
451	PMU-UD	Univariate	Classification	Univariate
452	Parkinson%27s+Disease+Classification	Multivariate	Classification	Multivariate
453	Electrical+Grid+Stability+Simulated+Data+	Multivariate	Classification	Multivariate
454	Caesarian+Section+Classification+Dataset	Univariate	Classification	Univariate
455	BAUM-1	Time	Classification	Time
456	BAUM-2	Time	Classification	Time
457	Audit+Data	Multivariate	Classification	Multivariate
458	BuddyMove+Data+Set	Multivariate	Classification	Multivariate
459	Real+estate+valuation+data+set	Multivariate	Regression	Multivariate
460	Early+biomarkers+of+Parkinson%E2%80%99s+diseas...	Multivariate	Classification	Multivariate
461	Somerville+Happiness+Survey		Classification	
462	2.4+GHZ+Indoor+Channel+Measurements	Multivariate	Classification	Multivariate
463	EMG+data+for+gestures	Time	Classification	Time
464	Parking+Birmingham	Multivariate	Classification	Multivariate
465	Behavior+of+the+urban+traffic+of+the+city+of+S...	Multivariate	Classification	Multivariate
466	Travel+Reviews	Multivariate	Classification	Multivariate
467	Tarvel+Review+Ratings	Multivariate	Classification	Multivariate

468 rows × 7 columns

In [166]:

```
html_2 = requests.get("https://archive.ics.uci.edu/ml/datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list").text
html_2
```

Out[166]:

'<!DOCTYPE HTML PUBLIC \\"-//W3C//DTD HTML 4.01 Transitional//EN\\">\n<html>\n<head>\n<title>UCI Machine Learning Repository: Data Sets</title>\n\n<!-- Stylesheet link -->\n<link rel="stylesheet" type="text/css" href="assets/ml.css" />\n\n<script language="JavaScript" type="text/javascript">\n<!--\nfunction checkform (form)\n{\n // see http://www.thesitewizard.com/archive/validation.shtml\n // for an explanation of this script and how to use it on your\n // own website\n\n // ** START **\n if (form.q.value == "")\n {\n   alert("Please enter search terms.");\n   form.q.focus();\n   return false ;\n }\n\n if (getCheckedValue(form.sitesearch) == "ics.uci.edu" && form.q.value.indexOf("site:archive.ics.uci.edu/ml") == -1)\n {\n   form.q.value = form.q.value + " site:archive.ics.uci.edu/ml";\n }\n\n // ** END **\n return true ;\n}\n\n// return the value of the radio button that is checked\n// return an empty string if none are checked, or\n// there are no radio buttons\nfunction getCheckedValue(radioObj) {\n\tif(!radioObj)\n\t\treturn "";\n\tvar radioLength = radioObj.length;\n\tif(radioLength == undefined)\n\t\tif(radioObj.checked)\n\t\t\treturn radioObj.value;\n\t\telse\n\t\t\treturn "";\n\tfor(var i = 0; i < radioLength; i++) {\n\t\tif(radioObj[i].checked) {\n\t\t\treturn radioObj[i].value;\n\t\t}\n\t}\n\treturn "";\n}\n\n\n\n\n\n//-->\n</script>\n\n</head>\n\

n<body>\n\n\n<!-- SITE HEADER (INCLUDES LOGO AND SEARCH BOX) -->\n\n<!-- SITE HEADER (INCLUDES LOGO AND SEARCH BOX) -->\n\n<table width=100% bgcolor="#003366">\n<tr>\n\t<td>\n\t\t<a href="index.html" alt="Home"><img src="assets/logo.gif" border=0></img></a>
          <a href="http://cml.ics.uci.edu"><font color="FFDD33">Center for Machine Learning and Intelligent Systems</font></a>\n\t</td>\n\t<td width=100% valign=top align="right">\n\t\t\n\t\t<a href="about.html">About</a> \n\t\t<a href="citation_policy.html">Citation Policy</a> \n\t\t<a href="donation_policy.html">Donate a Data Set</a> \n\t\t<a href="contact.html">Contact</a>\n\t\t\n\n\t\t
\n\t\t
\n\t\t<!-- Search Google -->\n\n\t\t<FORM method=GET action=http://www.google.com/custom onsubmit="return checkform(this);">\n\t\t<INPUT TYPE=text name=q size=30 maxlength=255 value="">\n\t\t<INPUT type=submit name=sa VALUE="Search">\n\t\t<INPUT type=hidden name=cof VALUE="AH:center;LH:130;L:http://archive.ics.uci.edu/assets/logo.gif;LW:384;AWFID:869c0b2eaa8d518e;">\n\t\t<input type=hidden name=domains value="ics.uci.edu">\n\t\t
\n\t\t<input type=radio name=sitesearch value="ics.uci.edu" checked> <font size="1">Repository</font>\n\t\t<input type=radio name=sitesearch value=""> <font size="1">Web</font>\n\t\t   \n\t\t<A HREF=http://www.google.com/search><IMG SRC=http://www.google.com/logos/Logo_25blk.gif border=0 ALT=Google align=middle height=27></A>\n\t\t
\n\t\t</FORM>\n\t\t<!-- Search Google -->\n\n\n\t\t<a href="datasets.php"><font size="3" color="#FFDD33">View ALL Data Sets</font></a>\n\t\t
\n\t</td>\n</tr>\n</table>\n\n
\n\n<table cellpadding=3>\n\t<tr>\n\t\t<td valign="top">\n\n\t\t\t<table><tr><td> Browse Through: </td></tr></table>\n\n\n\n<table border=2 cellpadding=2 width=150>\n\t\t<tr> <td bgcolor="#003366"> Default Task\t \n\t\t</td>\n\t</tr>\n\t<tr>\n\t\t<td valign=top> <a href=\'datasets.php?format=&task=cla&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list\'>Classification</a> <font color=red>(350)</font>
<a href=\'datasets.php?format=&task=reg&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list\'>Regression</a> <font color=red>(96)</font>
<a href=\'datasets.php?format=&task=clu&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list\'>Clustering</a> <font color=red>(84)</font>
<a href=\'datasets.php?format=&task=other&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list\'>Other</a> <font color=red>(55)</font>\t \n\t\t</td>\n\t</tr>\n\t<tr> <td bgcolor="#003366"> Attribute Type \t \n\t\t</td>\n\t</tr>\n\t<tr>\n\t\t<td valign=top> <a href=\'datasets.php?format=&task=&att=cat&area=&numAtt=&numIns=&type=&sort=nameUp&view=list\'>Categorical</a> <font color=red>(38)</font>
<a href=\'datasets.php?format=&task=&att=num&area=&numAtt=&numIns=&type=&sort=nameUp&view=list\'>Numerical</a> <font color=red>(307)</font>
<a href=\'datasets.php?format=&task=&att=mix&area=&numAtt=&numIns=&type=&sort=nameUp&view=list\'>Mixed</a> <font color=red>(55)</font>\t \n\t\t</td>\n\t</tr>\n\t<tr> <td bgcolor="#003366"> Data Type \t \n\t\t</td>\n\t</tr>\n\t<tr>\n\t\t<td valign=top> <a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=mvar&sort=nameUp&view=list\'>Multivariate</a> <font color=red>(357)</font>
<a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=uvar&sort=nameUp&view=list\'>Univariate</a> <font color=red>(23)</font>
<a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=seq&sort=nameUp&view=list\'>Sequential</a> <font color=red>(47)</font>
<a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=ts&sort=nameUp&view=list\'>Time-Series</a> <font color=red>(91)</font>
<a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=text&sort=nameUp&view=list\'>Text</a> <font color=red>(53)</font>
<a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=dt&sort=nameUp&view=list\'>Domain-Theory</a> <font color=red>(23)</font>
<a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=other&sort=nameUp&view=list\'>Other</a> <font color=red>(21)</font>
\t \n\t\t</td>\n\t</tr>\n\t<tr><td bgcolor="#003366"> Area\t \n\t\t</td>\n\t</tr>\n\t<tr>\n\t\t<td valign=top> <a href=\'datasets.php?format=&task=&att=&area=life&numAtt=&numIns=&type=&sort=nameUp&view=list\'>Life Sciences</a> <font color=red>(107)</font>
<a href=\'datasets.php?format=&task=&att=&area=phys&numAtt=&numIns=&type=&sort=nameUp&view=list\'>Physical Sciences</a> <font color=red>(49)</font>
<a href=\'datasets.php?format=&task=&att=&area=comp&numAtt=&numIns=&type=&sort=nameUp&view=list\'>CS / Engineering</a> <font color=red>(170)</font>
<a href=\'datasets.php?format=&task=&att=&area=soc&numAtt=&numIns=&type=&sort=nameUp&view=list\'>Social Sciences</a> <font color=red>(26)</font>
<a href=\'datasets.php?format=&task=&att=&area=bus&numAtt=&numIns=&type=&sort=nameUp&view=list\'>Business</a> <font color=red>(29)</font>
<a href=\'datasets.php?format=&task=&att=&area=game&numAtt=&numIns=&type=&sort=nameUp&view=list\'>Game</a> <font color=red>(10)</font>
<a href=\'datasets.php?format=&task=&att=&area=other&numAtt=&numIns=&type=&sort=nameUp&view=list\'>Other</a> <font color=red>(73)</font>\t \n\t\t</td>\n\t</tr>\n\t<tr><td bgcolor="#003366"> # Attributes\t \n\t\t</td>\n\t</tr>\n\t<tr>\n\t\t<td valign=top> <a href=\'datasets.php?format=&task=&att=&area=&numAtt=less10&numIns=&type=&sort=nameUp&view=list\'>Less than 10</a> <font color=red>(113)</font>
<a href=\'datasets.php?format=&task=&att=&area=&numAtt=10to100&numIns=&type=&sort=nameUp&view=list\'>10 to 100</a> <font color=red>(210)</font>
<a href=\'datasets.php?format=&task=&att=&area=&numAtt=greater100&numIns=&type=&sort=nameUp&view=list\'>Greater than 100</a> <font color=red>(84)</font>\t \n\t\t</td>\n\t</tr>\n\t<tr><td bgcolor="#003366"> # Instances </td>\n\t\t</tr>\n\n\t

\t<tr>\n\t\t\t<td valign=top> \n\t\t<a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=less100&type=&sort=nameUp&view=list\'>Less than 100</a> <font color=red>(27)</font>
<a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=100to1000&type=&sort=nameUp&view=list\'>100 to 1000</a> <font color=red>(162)</font>
<a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=greater1000&type=&sort=nameUp&view=list\'>Greater than 1000</a> <font color=red>(246)</font>\t \n\t\t</td>\n\t</tr>\n\t<tr><td bgcolor="#003366"> Format Type\t \n\t\t</td>\n\t</tr>\n\t<tr>\n\t\t<td valign=top> <a href=\'datasets.php?format=mat&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list\'>Matrix</a> <font color=red>(324)</font>
<a href=\'datasets.php?format=nonmat&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list\'>Non-Matrix</a> <font color=red>(145)</font>\t \n\t\t</td>\n\t</tr></table>\n\t</td>\n\n\t<td valign=top>\n\n<table width=100%>\n\t\t<tr>\n\t\t\t<td> 469 Data Sets </td>\n\t\t\t<td align="right"> <a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table\'>Table View</a> <font color=gray>List View</font> </td>\n\t\t</tr>\n\t</table><table cellpadding=3><tr><td><hr> 1. <a href="datasets/2.4+GHZ+Indoor+Channel+Measurements">2.4 GHZ Indoor Channel Measurements</a>: Measurement of the S21,consists of 10 sweeps, each sweep contains 601 frequency points with spacing of 0.167MHz to cover a 100MHz band centered at 2.4GHz. 2. <a href="datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29">3D Road Network (North Jutland, Denmark)</a>: 3D road network with highly accurate elevation information (+-20cm) from Denmark used in eco-routing and fuel/Co2-estimation routing algorithms. 3. <a href="datasets/AAAI+2013+Accepted+Papers">AAAI 2013 Accepted Papers</a>: This data set compromises the metadata for the 2013 AAAI conference\'s accepted papers (main track only), including paper titles, abstracts, and keywords of varying granularity. 4. <a href="datasets/AAAI+2014+Accepted+Papers">AAAI 2014 Accepted Papers</a>: This data set compromises the metadata for the 2014 AAAI conference\'s accepted papers, including paper titles, authors, abstracts, and keywords of varying granularity. 5. <a href="datasets/Abalone">Abalone</a>: Predict the age of abalone from physical measurements 6. <a href="datasets/Abscisic+Acid+Signaling+Network">Abscisic Acid Signaling Network</a>: The objective is to determine the set of boolean rules that describe the interactions of the nodes within this plant signaling network. The dataset includes 300 separate boolean pseudodynamic simulations using an asynchronous update scheme. 7. <a href="datasets/Absenteeism+at+work">Absenteeism at work</a>: The database was created with records of absenteeism at work from July 2007 to July 2010 at a courier company in Brazil. 8. <a href="datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors">Activities of Daily Living (ADLs) Recognition Using Binary Sensors</a>: This dataset comprises information regarding the ADLs performed by two users on a daily basis in their \r\nown homes. 9. <a href="datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer">Activity Recognition from Single Chest-Mounted Accelerometer</a>: The dataset collects data from a wearable accelerometer mounted on the chest. The dataset is intended for Activity Recognition research purposes. 10. <a href="datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29">Activity Recognition system based on Multisensor data fusion (AReM)</a>: This dataset contains temporal data from a Wireless Sensor Network worn by an actor performing the activities: bending, cycling, lying down, sitting, standing, walking. 11. <a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor">Activity recognition with healthy older people using a batteryless wearable sensor</a>: Sequential motion data from 14 healthy older people aged 66 to 86 years old using a batteryless, wearable sensor on top of their clothing for the recognition of activities in clinical environments. 12. <a href="datasets/Acute+Inflammations">Acute Inflammations</a>: The data was created by a medical expert as a data set to test the expert system, \r\nwhich will perform the presumptive diagnosis of two diseases of the urinary system.\r\n 13. <a href="datasets/Adult">Adult</a>: Predict whether income exceeds $50K/yr based on census data. Also known as "Census Income" dataset. 14. <a href="datasets/Air+Quality">Air Quality</a>: Contains the responses of a gas multisensor device deployed on the field in an Italian city. Hourly responses averages are recorded along with gas concentrations references from a certified analyzer. 15. <a href="datasets/Air+quality">Air quality</a>: Contains the responses of a gas multisensor device deployed on the field in an Italian city. 16. <a href="datasets/Airfoil+Self-Noise">Airfoil Self-Noise</a>: NASA data set, obtained from a series of aerodynamic and acoustic tests of two and three-dimensional airfoil blade sections conducted in an anechoic wind tunnel. 17. <a href="datasets/Amazon+Access+Samples">Amazon Access Samples</a>: Amazon\'s InfoSec is getting smarter about the way Access data is leveraged. This is an anonymized sample of access provisioned within the company. 18. <a href="datasets/Amazon+Commerce+reviews+set">Amazon Commerce reviews set</a>: The dataset is used for authorship identification in online Writeprint which is a new research field of pattern recognition. 19. <a href="datasets/Annealing">Annealing</a>: Steel annealing data 20. <a href="datasets/Anonymous+Microsoft+Web+Data">Anonymous Microsoft Web Data</a>: Log of anonymous users of www.microsoft.com; predict areas of the web site a user visited based on data on other areas the user visit

ed. 21. <a href="datasets/Anuran+Calls+%28MFCCs%29">Anuran Calls (MFCCs)</a>: Acoustic features extracted from syllables of anuran (frogs) calls, including the family, the genus, and the species labels (multilabel). 22. <a href="datasets/Appliances+energy+prediction">Appliances energy prediction</a>: Experimental data used to create regression models of appliances energy use in a low energy building. 23. <a href="datasets/APS+Failure+at+Scania+Trucks">APS Failure at Scania Trucks</a>: The datasets\' positive class consists of component failures for a specific component of the APS system. The negative class consists of trucks with failures for components not related to the APS. 24. <a href="datasets/Arcene">Arcene</a>: ARCENE\'s task is to distinguish cancer versus normal patterns from mass-spectrometric data. This is a two-class classification problem with continuous input variables. This dataset is one of 5 datasets of the NIPS 2003 feature selection challenge. 25. <a href="datasets/Arrhythmia">Arrhythmia</a>: Distinguish between the presence and absence of cardiac arrhythmia and classify it in one of the 16 groups. 26. <a href="datasets/Artificial+Characters">Artificial Characters</a>: Dataset artificially generated by using first order theory which describes structure of ten capital letters of English alphabet 27. <a href="datasets/Audiology+%28Original%29">Audiology (Original)</a>: Nominal audiology dataset from Baylor 28. <a href="datasets/Audiology+%28Standardized%29">Audiology (Standardized)</a>: Standardized version of the original audiology database 29. <a href="datasets/Audit+Data">Audit Data</a>: Exhaustive one year non-confidential data in the year 2015 to 2016 of firms is collected from the Auditor Office of India to build a predictor for classifying suspicious firms. 30. <a href="datasets/Australian+Sign+Language+signs">Australian Sign Language signs</a>: This data consists of sample of Auslan (Australian Sign Language) signs. Examples of 95 signs were collected from five signers with a total of 6650 sign samples. 31. <a href="datasets/Australian+Sign+Language+signs+%28High+Quality%29">Australian Sign Language signs (High Quality)</a>: This data consists of sample of Auslan (Australian Sign Language) signs. 27 examples of each of 95 Auslan signs were captured from a native signer using high-quality position trackers 32. <a href="datasets/Autism+Screening+Adult">Autism Screening Adult</a>: Autistic Spectrum Disorder Screening Data for Adult. This dataset is related to classification and predictive tasks. 33. <a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++">Autistic Spectrum Disorder Screening Data for Adolescent </a>: Autistic Spectrum Disorder Screening Data for Adolescent. This dataset is related to classification and predictive tasks. 34. <a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++">Autistic Spectrum Disorder Screening Data for Children </a>: Children screening data for autism suitable for classification and predictive tasks 35. <a href="datasets/Auto+MPG">Auto MPG</a>: Revised from CMU StatLib library, data concerns city-cycle fuel consumption 36. <a href="datasets/Automobile">Automobile</a>: From 1985 Ward\'s Automotive Yearbook 37. <a href="datasets/AutoUniv">AutoUniv</a>: AutoUniv is an advanced data generator for classifications tasks. The aim is to reflect the nuances and heterogeneity of real data. Data can be generated in .csv, ARFF or C4.5 formats. 38. <a href="datasets/Avila">Avila</a>: The Avila data set has been extracted from 800 images of the \'Avila Bible\', an XII century giant Latin copy of the Bible. The prediction task consists in associating each pattern to a copyist. 39. <a href="datasets/Bach+Choral+Harmony">Bach Choral Harmony</a>: The data set is composed of 60 chorales (5665 events) by J.S. Bach (1675-1750).\r\nEach event of each chorale is labelled using 1 among 101 chord labels and described\r\nthrough 14 features. 40. <a href="datasets/Bach+Chorales">Bach Chorales</a>: Time-series data based on chorales; challenge is to learn generative grammar; data in Lisp 41. <a href="datasets/Badges">Badges</a>: Badges labeled with a "+" or "-" as a function of a person\'s name 42. <a href="datasets/Bag+of+Words">Bag of Words</a>: This data set contains five text collections in the form of bags-of-words. 43. <a href="datasets/Balance+Scale">Balance Scale</a>: Balance scale weight & distance database 44. <a href="datasets/Balloons">Balloons</a>: Data previously used in cognitive psychology experiment; 4 data sets represent different conditions of an experiment 45. <a href="datasets/Bank+Marketing">Bank Marketing</a>: The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe a term deposit (variable y). 46. <a href="datasets/banknote+authentication">banknote authentication</a>: Data were extracted from images that were taken for the evaluation of an authentication procedure for banknotes. 47. <a href="datasets/BAUM-1">BAUM-1</a>: BAUM-1 dataset contains 1184 multimodal facial video clips collected from 31 subjects. The 1184 video clips contain spontaneous facial expressions and speech of 13 emotional and mental states. 48. <a href="datasets/BAUM-2">BAUM-2</a>: A multilingual audio-visual affective face database consisting of 1047 video clips of 286 subjects. 49. <a href="datasets/Behavior+of+the+urban+traffic+of+the+city+of+Sao+Paulo+in+Brazil">Behavior of the urban traffic of the city of Sao Paulo in Brazil</a>: The database was created with records of behavior of the urban traffic of the city of Sao Paulo in Brazil. 50. <a href="datasets/Beijing+PM2.5+Data">Beijing PM2.5 Data</a>: This hourly data set contains the PM2.5 data of US Embassy in Beijing. Meanwhile, met

eorological data from Beijing Capital International Airport are also included. 51. <a h ref="datasets/Bike+Sharing+Dataset">Bike Sharing Dataset</a>: This dataset contains the hourly and daily count of rental bikes between years 2011 and 2012 in Capital bikeshare system with the corresponding weathe r and seasonal information. 52. <a href="datasets/BLE+RSSI+Dataset+for+Indoor+lo calization+and+Navigation">BLE RSSI Dataset for Indoor localization and Navigation</a>: This dataset con tains RSSI readings gathered from an array of Bluetooth Low Energy (BLE) iBeacons in a real-world and operat ional indoor environment for localization and navigation purposes. 53. <a href="data sets/BlogFeedback">BlogFeedback</a>: Instances in this dataset contain features extracted from blog pos ts. The task associated with the data is to predict how many comments the post will receive. 54. <a href="datasets/BLOGGER">BLOGGER</a>: In this paper, we look for to recognize the caus es of users tend\r\nto cyber space in Kohkiloye and Boyer Ahmad Province in\r\nIran 55. <a href="datasets/Blood+Transfusion+Service+Center">Blood Transfusion Service Center</a>: Data ta ken from the Blood Transfusion Service Center in Hsin-Chu City in Taiwan -- this is a classification problem. 56. <a href="datasets/Breast+Cancer">Breast Cancer</a>: Breast Cancer Data (R estricted Access) 57. <a href="datasets/Breast+Cancer+Coimbra">Breast Cancer Co imbra</a>: Clinical features were observed or measured for 64 patients with breast cancer and 52 healthy controls. 58. <a href="datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29">Bre ast Cancer Wisconsin (Diagnostic)</a>: Diagnostic Wisconsin Breast Cancer Database 59. <a href="datasets/Breast+Cancer+Wisconsin+%28Original%29">Breast Cancer Wisconsin (Original) </a>: Original Wisconsin Breast Cancer Database 60. <a href="datasets/Breast+ Cancer+Wisconsin+%28Prognostic%29">Breast Cancer Wisconsin (Prognostic)</a>: Prognostic Wisconsin Breast Cancer Database 61. <a href="datasets/Breast+Tissue">Breast Tissue</a></ b>: Dataset with electrical impedance measurements of freshly excised tissue samples from the breast. 62. <a href="datasets/BuddyMove+Data+Set">BuddyMove Data Set</a>: User interest information extracted from user reviews published in holidayiq.com about various types of point of interests in S outh India 63. <a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attac k+on+Optical+Burst+Switching+%28OBS%29+Network">Burst Header Packet (BHP) flooding attack on Optical Burst Switching (OBS) Network</a>: One of the primary challenges in identifying the risks of the Burst Hea der Packet (BHP) flood attacks in Optical Burst Switching networks (OBS) is the scarcity of reliable historical dat a. 64. <a href="datasets/Buzz+in+social+media+">Buzz in social media </a>: T his data-set contains examples of buzz events from two different social networks: Twitter, and Tom\'s Hardware , a forum network focusing on new technology with more conservative dynamics. 65. <a href="datasets/Caesarian+Section+Classification+Dataset">Caesarian Section Classification Dataset</a></ b>: This dataset contains information about caesarian section results of 80 pregnant women with the most imp ortant characteristics of delivery problems in the medical field. 66. <a href="datasets/ CalIt2+Building+People+Counts">CalIt2 Building People Counts</a>: This data comes from the main door of the CalIt2 building at UCI. 67. <a href="datasets/Car+Evaluation">Car Evaluation< /a>: Derived from simple hierarchical decision model, this database may be useful for testing constructive i nduction and structure discovery methods. 68. <a href="datasets/Carbon+Nanotube s">Carbon Nanotubes</a>: This dataset contains 10721 initial and calculated atomic coordinates of carbon nanotubes. 69. <a href="datasets/Cardiotocography">Cardiotocography</a>: Th e dataset consists of measurements of fetal heart rate (FHR) and uterine contraction (UC) features on cardioto cograms classified by expert obstetricians. 70. <a href="datasets/Cargo+2000+Freig ht+Tracking+and+Tracing">Cargo 2000 Freight Tracking and Tracing</a>: Sanitized and anonymized Carg o 2000 (C2K) airfreight tracking and tracing events, covering five months of business execution (3,942 process instances, 7,932 transport legs, 56,082 activities). 71. <a href="datasets/Census+In come">Census Income</a>: Predict whether income exceeds $50K/yr based on census data. Also known as "Adult" dataset. 72. <a href="datasets/Census-Income+%28KDD%29">Census-In come (KDD)</a>: This data set contains weighted census data extracted from the 1994 and 1995 current p opulation surveys conducted by the U.S. Census Bureau. 73. <a href="datasets/Cerv ical+cancer+%28Risk+Factors%29">Cervical cancer (Risk Factors)</a>: This dataset focuses on the predic tion of indicators/diagnosis of cervical cancer. The features cover demographic information, habits, and historic medical records. 74. <a href="datasets/Challenger+USA+Space+Shuttle+O-Ring">C hallenger USA Space Shuttle O-Ring</a>: Task: predict the number of O-rings that experience thermal dist ress on a flight at 31 degrees F given data on the previous 23 shuttle flights 75. <a h ref="datasets/Character+Font+Images">Character Font Images</a>: Character images from scanned and computer generated fonts. 76. <a href="datasets/Character+Trajectories">Character Trajectories</a>: Multiple, labelled samples of pen tip trajectories recorded whilst writing individual charact ers. All samples are from the same writer, for the purposes of primitive extraction. Only characters with a single pen-down segment were considered. 77. <a href="datasets/Chess+%28Domain+Th eories%29">Chess (Domain Theories)</a>: 6 different domain theories for generating legal moves of ches s 78. <a href="datasets/Chess+%28King-Rook+vs.+King%29">Chess (King-Rook vs.

King)</a>: Chess Endgame Database for White King and Rook against Black King (KRK). 79. <a href="datasets/Chess+%28King-Rook+vs.+King-Knight%29">Chess (King-Rook vs. King-Knight)</a>: Knight Pin Chess End-Game Database Creator 80. <a href="datasets/Chess+%28King-Rook+vs.+King-Pawn%29">Chess (King-Rook vs. King-Pawn)</a>: King+Rook versus King+Pawn on a7 (usually abbreviated KRKPA7). 81. <a href="datasets/chestnut+%E2%80%93+LARVIC">chestnut – LARVIC</a>: The research project presents this database, shows the images of chestnuts that will be processed to determine the presence or absence of defects 82. <a href="datasets/chipseq">chipseq</a>: ChIP-seq experiments characterize protein modifications or binding at\r\nspecific genomic locations in specific samples. The machine learning\r\nproblem in these data is structured binary classification. 83. <a href="datasets/Chronic_Kidney_Disease">Chronic_Kidney_Disease</a>: This dataset can be used to predict the chronic kidney disease and it can be collected from the hospital nearly 2 months of period. 84. <a href="datasets/Climate+Model+Simulation+Crashes">Climate Model Simulation Crashes</a>: Given Latin hypercube samples of 18 climate model input parameter values, predict climate model simulation crashes and determine the parameter value combinations that cause the failures. 85. <a href="datasets/Cloud">Cloud</a>: Little Documentation 86. <a href="datasets/CMU+Face+Images">CMU Face Images</a>: This data consists of 640 black and white face images of people taken with varying pose (straight, left, right, up), expression (neutral, happy, sad, angry), eyes (wearing sunglasses or not), and size 87. <a href="datasets/CNAE-9">CNAE-9</a>: This is a data set containing 1080 documents of free text business descriptions of Brazilian companies categorized into a\r\nsubset of 9 categories 88. <a href="datasets/Coil+1999+Competition+Data">Coil 1999 Competition Data</a>: This data set is from the 1999 Computational Intelligence and Learning (COIL) competition. The data contains measurements of river chemical concentrations and algae densities. 89. <a href="datasets/Combined+Cycle+Power+Plant">Combined Cycle Power Plant</a>: The dataset contains 9568 data points collected from a Combined Cycle Power Plant over 6 years (2006-2011), when the plant was set to work with full load. 90. <a href="datasets/Communities+and+Crime">Communities and Crime</a>: Communities within the United States. The data combines socio-economic data from the 1990 US Census, law enforcement data from the 1990 US LEMAS survey, and crime data from the 1995 FBI UCR. 91. <a href="datasets/Communities+and+Crime+Unnormalized">Communities and Crime Unnormalized</a>: Communities in the US. Data combines socio-economic data from the \'90 Census, law enforcement data from the 1990 Law Enforcement Management and Admin Stats survey, and crime data from the 1995 FBI UCR 92. <a href="datasets/Computer+Hardware">Computer Hardware</a>: Relative CPU Performance Data, described in terms of its cycle time, memory size, etc. 93. <a href="datasets/Concrete+Compressive+Strength">Concrete Compressive Strength</a>: Concrete is the most important material in civil engineering. The concrete compressive strength is a highly nonlinear function of age and ingredients. 94. <a href="datasets/Concrete+Slump+Test">Concrete Slump Test</a>: Concrete is a highly complex material. The slump flow of concrete is not only determined by the water content, but that is also influenced by other concrete ingredients. 95. <a href="datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants">Condition Based Maintenance of Naval Propulsion Plants</a>: Data have been generated from a sophisticated simulator of a Gas Turbines (GT), mounted on a Frigate characterized by a COmbined Diesel eLectric And Gas (CODLAG) propulsion plant type. 96. <a href="datasets/Condition+monitoring+of+hydraulic+systems">Condition monitoring of hydraulic systems</a>: The data set addresses the condition assessment of a hydraulic test rig based on multi sensor data. Four fault types are superimposed with several severity grades impeding selective quantification. 97. <a href="datasets/Congressional+Voting+Records">Congressional Voting Records</a>: 1984 United Stated Congressional Voting Records; Classify as Republican or Democrat 98. <a href="datasets/Connect-4">Connect-4</a>: Contains connect-4 positions 99. <a href="datasets/Connectionist+Bench+%28Nettalk+Corpus%29">Connectionist Bench (Nettalk Corpus)</a>: The file "nettalk.data" contains a list of 20,008 English words, along with a phonetic transcription for each word. The task is to train a network to produce the proper phonemes 100. <a href="datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29">Connectionist Bench (Sonar, Mines vs. Rocks)</a>: The task is to train a network to discriminate between sonar signals bounced off a metal cylinder and those bounced off a roughly cylindrical rock. 101. <a href="datasets/Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29">Connectionist Bench (Vowel Recognition - Deterding Data)</a>: Speaker independent recognition of the eleven steady state vowels of British English using a specified training set of lpc derived log area ratios. 102. <a href="datasets/Container+Crane+Controller+Data+Set">Container Crane Controller Data Set</a>: A container crane has the function of transporting containers from one point to another point. 103. <a href="datasets/Contraceptive+Method+Choice">Contraceptive Method Choice</a>: Dataset is a subset of the 1987 National Indonesia Contraceptive Prevalence Survey. 104. <a href="datasets/Corel+Image+Features">Corel Image Features</a>: This dataset contains image features extracted from a Corel image collection. Four sets of features are available based on the

ntains image features extracted from a Corel image collection. Four sets of features are available based on the color histogram, color histogram layout, color moments, and co-occurence 105. <a href="datasets/Covertype">Covertype</a>: Forest CoverType dataset 106. <a href="datasets/Credit+Approval">Credit Approval</a>: This data concerns credit card applications; good mix of attributes 107. <a href="datasets/Crowdsourced+Mapping">Crowdsourced Mapping</a>: Crowdsourced data from OpenStreetMap is used to automate the classification of satellite images into different land cover classes (impervious, farm, forest, grass, orchard, water). 108. <a href="datasets/Cryotherapy+Dataset+">Cryotherapy Dataset </a>: This dataset contains information about wart treatment results of 90 patients using cryotherapy. 109. <a href="datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015">CSM (Conventional and Social Media Movies) Dataset 2014 and 2015</a>: 12 features categorized as conventional and social media features. Both conventional features, collected from movies databases on Web as well as social media features(YouTube,Twitter). 110. <a href="datasets/Cuff-Less+Blood+Pressure+Estimation">Cuff-Less Blood Pressure Estimation</a>: This Data set provides preprocessed and cleaned vital signals which can be used in designing algorithms for cuff-less estimation of the blood pressure. 111. <a href="datasets/Cylinder+Bands">Cylinder Bands</a>: Used in decision tree induction for mitigating process delays known as "cylinder bands" in rotogravure printing 112. <a href="datasets/Daily+and+Sports+Activities">Daily and Sports Activities</a>: The dataset comprises motion sensor data of 19 daily and sports activities each performed by 8 subjects in their own style for 5 minutes. Five Xsens MTx units are used on the torso, arms, and legs.\r\n 113. <a href="datasets/Daily+Demand+Forecasting+Orders">Daily Demand Forecasting Orders</a>: The dataset was collected during 60 days, this is a real database of a brazilian logistics company. 114. <a href="datasets/Daphnet+Freezing+of+Gait">Daphnet Freezing of Gait</a>: This dataset contains the annotated readings of 3 acceleration sensors at the hip and leg of Parkinson\'s disease patients that experience freezing of gait (FoG) during walking tasks.\r\n 115. <a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting">Data for Software Engineering Teamwork Assessment in Education Setting</a>: Data include over 100 Team Activity Measures and outcomes (ML classes) obtained from activities of 74 student teams during the creation of final class project in SW Eng. classes at SFSU, Fulda, FAU 116. <a href="datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer">Dataset for ADL Recognition with Wrist-worn Accelerometer</a>: Recordings of 16 volunteers performing 14 Activities of Daily Living (ADL) while carrying a single wrist-worn tri-axial accelerometer. 117. <a href="datasets/Dataset+for+Sensorless+Drive+Diagnosis">Dataset for Sensorless Drive Diagnosis</a>: Features are extracted from motor current. The motor has intact and defective components. This results in 11 different classes with different conditions. 118. <a href="datasets/DBWorld+e-mails">DBWorld e-mails</a>: It contains 64 e-mails which I have manually collected from DBWorld mailing list. They are classified in: \'announces of conferences\' and \'everything else\'. 119. <a href="datasets/default+of+credit+card+clients">default of credit card clients</a>: This research aimed at the case of customers’ default payments in Taiwan and compares the predictive accuracy of probability of default among six data mining methods. 120. <a href="datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Instance+Labels">DeliciousMIL: A Data Set for Multi-Label Multi-Instance Learning with Instance Labels</a>: This dataset includes 1) 12234 documents (8251 training, 3983 test) extracted from DeliciousT140 dataset, 2) class labels for all documents, 3) labels for a subset of sentences of the test documents. 121. <a href="datasets/Demospongiae">Demospongiae</a>: Marine sponges of the Demospongiae class classification domain. 122. <a href="datasets/Dermatology">Dermatology</a>: Aim for this dataset is to determine the type of Eryhemato-Squamous Disease. 123. <a href="datasets/Detect+Malacious+Executable%28AntiVirus%29">Detect Malacious Executable(AntiVirus)</a>: I extract features from malacious and non-malacious and create and training dataset to teach svm classifier.Dataset made of unknown executable to detect if it is virus or normal safe executable. 124. <a href="datasets/detection_of_IoT_botnet_attacks_N_BaIoT">detection_of_IoT_botnet_attacks_N_BaIoT</a>: This dataset addresses the lack of public botnet datasets, especially for the IoT. It suggests *real* traffic data, gathered from 9 commercial IoT devices authentically infected by Mirai and BASHLITE. 125. <a href="datasets/Devanagari+Handwritten+Character+Dataset">Devanagari Handwritten Character Dataset</a>: This is an image database of Handwritten Devanagari characters. There are 46 classes of characters with 2000 examples each. The dataset is split into training set(85%) and testing set(15%). 126. <a href="datasets/Dexter">Dexter</a>: DEXTER is a text classification problem in a bag-of-word representation. This is a two-class classification problem with sparse continuous input variables. This dataset is one of five datasets of the NIPS 2003 feature selection challenge.\r\n 127. <a href="datasets/DGP2+-+The+Second+Data+Generation+Program">DGP2 - The Second Data Generation Program</a>: Generates application domains based on specific parameters, number of features, and proportion of positive to negative examples 128. <a href="datasets/Diabetes">Diabetes</a>: This diabetes dataset is from AIM \'94 129. <a href="datasets/Diabetes+130-US+hospitals+for+years+1999-2008">Diabetes 130-US hospitals for years 1999-2008</a>: This data has been prepared to analy

s+1999-2008 >Diabetes 130-US hospitals for years 1999-2008</a>: This data has been prepared to analyze factors related to readmission as well as other \r\n\r\noutcomes pertaining to patients with diabetes. 130. <a href="datasets/Diabetic+Retinopathy+Debrecen+Data+Set">Diabetic Retinopathy Debrecen Data Set</a>: This dataset contains features extracted from the Messidor image set to predict whether an image contains signs of diabetic retinopathy or not. 131. <a href="datasets/Discrete+Tone+Image+Dataset">Discrete Tone Image Dataset</a>: Discrete Tone Images(DTI)are available which needs to be analyzed in detail. Here, we created this dataset for those who do research in DTI. \r\n 132. <a href="datasets/Dishonest+Internet+users+Dataset">Dishonest Internet users Dataset</a>: The dataset was used to test an architecture based on a trust model capable to cope with the evaluation of the trustworthiness of users interacting in pervasive environments. 133. <a href="datasets/Document+Understanding">Document Understanding</a>: Five concepts, expressed as predicates, to be learned 134. <a href="datasets/Dodgers+Loop+Sensor">Dodgers Loop Sensor</a>: Loop sensor data was collected for the Glendale on ramp for the 101 North freeway in Los Angeles 135. <a href="datasets/Dorothea">Dorothea</a>: DOROTHEA is a drug discovery dataset. Chemical compounds represented by structural molecular features must be classified as active (binding to thrombin) or inactive. This is one of 5 datasets of the NIPS 2003 feature selection challenge. 136. <a href="datasets/Dota2+Games+Results">Dota2 Games Results</a>: Dota 2 is a popular computer game with two teams of 5 players. At the start of the game each player chooses a unique hero with different strengths and weaknesses. 137. <a href="datasets/Dow+Jones+Index">Dow Jones Index</a>: This dataset contains weekly data for the Dow Jones Industrial Index. It has been used in computational investing research. 138. <a href="datasets/Dresses_Attribute_Sales">Dresses_Attribute_Sales</a>: This dataset contain Attributes of dresses and their recommendations according to their sales.Sales are monitor on the basis of alternate days. 139. <a href="datasets/DrivFace">DrivFace</a>: The DrivFace contains images sequences of subjects while driving in real scenarios. It is composed of 606 samples of 640×480, acquired over different days from 4 drivers with several facial features. 140. <a href="datasets/Drug+consumption+%28quantified%29">Drug consumption (quantified)</a>: Classify type of drug consumer by personality data 141. <a href="datasets/Drug+Review+Dataset+%28Druglib.com%29">Drug Review Dataset (Druglib.com)</a>: The dataset provides patient reviews on specific drugs along with related conditions. Reviews and ratings are grouped into reports on the three aspects benefits, side effects and overall comment. 142. <a href="datasets/Drug+Review+Dataset+%28Drugs.com%29">Drug Review Dataset (Drugs.com)</a>: The dataset provides patient reviews on specific drugs along with related conditions and a 10 star patient rating reflecting overall patient satisfaction. 143. <a href="datasets/DSRC+Vehicle+Communications">DSRC Vehicle Communications</a>: This set Provides data regarding wireless communications between vehicles and road side units. two separate data sets are provided (normal scenario) and in the presence of attacker (jammer). 144. <a href="datasets/Dynamic+Features+of+VirusShare+Executables">Dynamic Features of VirusShare Executables</a>: This dataset contains the dynamic features of 107,888 executables, collected by VirusShare from Nov/2010 to Jul/2014. 145. <a href="datasets/E.+Coli+Genes">E. Coli Genes</a>: Data giving characteristics of each ORF (potential gene) in the E. coli genome. Sequence, homology (similarity to other genes) and structural information, and function (if known) are provided. 146. <a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connected+speech+Data+Set+">Early biomarkers of Parkinson’s disease based on natural connected speech Data Set </a>: . 147. <a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech">Early biomarkers of Parkinson&#65533;s disease based on natural connected speech</a>: Predict a pattern of neurodegeneration in the dataset of speech features obtained from patients with early untreated Parkinson’s disease and patients at high risk developing Parkinson’s disease. 148. <a href="datasets/EBL+Domain+Theories">EBL Domain Theories</a>: Assorted small-scale domain theories 149. <a href="datasets/Echocardiogram">Echocardiogram</a>: Data for classifying if patients will survive for at least one year after a heart attack 150. <a href="datasets/Eco-hotel">Eco-hotel</a>: This dataset includes Online Textual Reviews from both online (e.g., TripAdvisor) and offline (e.g., Guests\' book) sources from the Areias do Seixo Eco-Resort. 151. <a href="datasets/Ecoli">Ecoli</a>: This data contains protein localization sites 152. <a href="datasets/Economic+Sanctions">Economic Sanctions</a>: Domain Theory on Economic Sanctions; Undocumented 153. <a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set">Educational Process Mining (EPM): A Learning Analytics Data Set</a>: Educational Process Mining data set is built from the recordings of 115 subjects\' activities through a logging application while learning with an educational simulator. 154. <a href="datasets/EEG+Database">EEG Database</a>: This data arises from a large study to examine EEG correlates of genetic predisposition to alcoholism. It contains measurements from 64 electrodes placed on the scalp sampled at 256 Hz 155. <a href="datasets/EEG+Eye+State">EEG Eye State</a>: The data set consists of 14 EEG values and a value indicating the eye state. 156. <a href="datasets/EEG+Steady+State+Visual+Evoked+Potential+Signals">EEG Steady State Visual Evoked Potential Sig

"datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals">EEG Steady-State Visual Evoked Potential Signals</a>: This database consists on 30 subjects performing Brain Computer Interface for Steady State Visual Evoked Potentials (BCI-SSVEP). 157. <a href="datasets/El+Nino">El Nino</a>: The data set contains oceanographic and surface meteorological readings taken from a series of buoys positioned throughout the equatorial Pacific. 158. <a href="datasets/Electrical+Grid+Stability+Simulated+Data+">Electrical Grid Stability Simulated Data </a>: The local stability analysis of the 4-node star system (electricity producer is in the center) implementing Decentral Smart Grid Control concept. 159. <a href="datasets/ElectricityLoadDiagrams20112014">ElectricityLoadDiagrams20112014</a>: This data set contains electricity consumption of 370 points/clients.\r\n 160. <a href="datasets/EMG+data+for+gestures">EMG data for gestures</a>: These are files of raw EMG data recorded by MYO Thalmic bracelet 161. <a href="datasets/EMG+dataset+in+Lower+Limb">EMG dataset in Lower Limb</a>: 3 different exercises: sitting, standing and walking in the muscles: biceps femoris, vastus medialis, rectus femoris and semitendinosus addition to goniometry in the exercises. 162. <a href="datasets/EMG+Physical+Action+Data+Set">EMG Physical Action Data Set</a>: The Physical Action Data Set includes 10 normal and 10 aggressive physical actions that measure the human activity. The data have been collected by 4 subjects using the Delsys EMG wireless apparatus. 163. <a href="datasets/Energy+efficiency">Energy efficiency</a>: This study looked into assessing the heating load and cooling load requirements of buildings (that is, energy efficiency) as a function of building parameters. 164. <a href="datasets/Entree+Chicago+Recommendation+Data">Entree Chicago Recommendation Data</a>: This data contains a record of user interactions with the Entree Chicago restaurant recommendation system. 165. <a href="datasets/Epileptic+Seizure+Recognition">Epileptic Seizure Recognition</a>: This dataset is a pre-processed and re-structured/reshaped version of a very commonly used dataset featuring epileptic seizure detection. 166. <a href="datasets/extention+of+Z-Alizadeh+sani+dataset">extention of Z-Alizadeh sani dataset</a>: It was collected for CAD diagnosis. 167. <a href="datasets/Facebook+Comment+Volume+Dataset">Facebook Comment Volume Dataset</a>: Instances in this dataset contain features extracted from facebook posts. The task associated with the data is to predict how many comments the post will receive. 168. <a href="datasets/Facebook+metrics">Facebook metrics</a>: Facebook performance metrics of a renowned cosmetic\'s brand Facebook page. 169. <a href="datasets/Farm+Ads">Farm Ads</a>: This data was collected from text ads found on twelve websites that deal with various farm animal related topics. The binary labels are based on whether or not the content owner approves of the ad. 170. <a href="datasets/Fertility">Fertility</a>: 100 volunteers provide a semen sample analyzed according to the WHO 2010 criteria. Sperm concentration are related to socio-demographic data, environmental factors, health status, and life habits 171. <a href="datasets/Firm-Teacher_Clave-Direction_Classification">Firm-Teacher_Clave-Direction_Classification</a>: The data are binary attack-point vectors and their clave-direction class(es) according to the partido-alto-based paradigm. 172. <a href="datasets/First-order+theorem+proving">First-order theorem proving</a>: Given a theorem, predict which of five heuristics will give the fastest proof when used by a first-order prover. A sixth prediction declines to attempt a proof, should the theorem be too difficult. 173. <a href="datasets/Flags">Flags</a>: From Collins Gem Guide to Flags, 1986 174. <a href="datasets/FMA%3A+A+Dataset+For+Music+Analysis">FMA: A Dataset For Music Analysis</a>: FMA features 106,574 tracks and includes song title, album, artist, genres; play counts, favorites, comments; description, biography, tags; together with audio (343 days, 917 GiB) and features. 175. <a href="datasets/Folio">Folio</a>: 20 photos of leaves for each of 32 different species. 176. <a href="datasets/Forest+Fires">Forest Fires</a>: This is a difficult regression task, where the aim is to predict the burned area of forest fires, in the northeast region of Portugal, by using meteorological and other data (see details at: http://www.dsi.uminho.pt/~pcortez/forestfires). 177. <a href="datasets/Forest+type+mapping">Forest type mapping</a>: Multi-temporal remote sensing data of a forested area in Japan. The goal is to map different forest types using spectral data. 178. <a href="datasets/Function+Finding">Function Finding</a>: Cases collected mostly from investigations in physical science; intention is to evaluate function-finding algorithms 179. <a href="datasets/Gas+Sensor+Array+Drift+Dataset">Gas Sensor Array Drift Dataset</a>: This archive contains 13910 measurements from 16 chemical sensors utilized in simulations for drift compensation in a discrimination task of 6 gases at various levels of concentrations. 180. <a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations">Gas Sensor Array Drift Dataset at Different Concentrations</a>: This archive contains 13910 measurements from 16 chemical sensors exposed to 6 different gases at various concentration levels. 181. <a href="datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures">Gas sensor array exposed to turbulent gas mixtures</a>: A chemical detection platform composed of 8 chemoresistive gas sensors was exposed to turbulent gas mixtures generated naturally in a wind tunnel. The acquired time series of the sensors are provided. 182. <a href="datasets/Gas+sensor+array+under+dynamic+gas+mixtures">Gas sensor array under dynamic gas mixtures</a>: The data set contains the recordings of 16 chemical sens

ors exposed to two dynamic gas mixtures at varying concentrations. For each mixture, signals were acquired c ontinuously during 12 hours. 183. <a href="datasets/Gas+sensor+array+under+flow +modulation">Gas sensor array under flow modulation</a>: The data set contains 58 time series acquired from 16 chemical sensors under gas flow modulation conditions. The sensors were exposed to different gaseou s binary mixtures of acetone and ethanol. 184. <a href="datasets/Gas+sensor+array s+in+open+sampling+settings">Gas sensor arrays in open sampling settings</a>: The dataset contains 18 000 time-series recordings from a chemical detection platform at six different locations in a wind tunnel facility i n response to ten high-priority chemical gaseous substances 185. <a href="datasets /Gas+sensors+for+home+activity+monitoring">Gas sensors for home activity monitoring</a>: 100 recordin gs of a sensor array under different conditions in a home setting: background, wine and banana presentations. The array includes 8 MOX gas sensors, and humidity and temperature sensors.\r\n 186. <a href="datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy">Gastrointestinal Lesions in Regular Colonoscopy</a>: This dataset contains features extracted from colonoscopy videos used to detect gastroi ntestinal lesions. It contains 76 lesions: 15 serrated adenomas, 21 hyperplastic lesions and 40 adenoma. 187. <a href="datasets/gene+expression+cancer+RNA-Seq">gene expression cancer RN A-Seq</a>: This collection of data is part of the RNA-Seq (HiSeq) PANCAN data set, it is a random extracti on of gene expressions of patients having different types of tumor: BRCA, KIRC, COAD, LUAD and PRAD. 188. <a href="datasets/Geo-Magnetic+field+and+WLAN+dataset+for+indoor+localisation+ from+wristband+and+smartphone">Geo-Magnetic field and WLAN dataset for indoor localisation from wristban d and smartphone</a>: A multisource and multivariate dataset for indoor localisation methods based on W LAN and Geo-Magnetic field fingerprinting 189. <a href="datasets/Geographical+Ori ginal+of+Music">Geographical Original of Music</a>: Instances in this dataset contain audio features extra cted from 1059 wave files. The task associated with the data is to predict the geographical origin of music.\r\n 190. <a href="datasets/Gesture+Phase+Segmentation">Gesture Phase Segmentation </a>: The dataset is composed by features extracted from 7 videos with people gesticulating, aiming at stu dying Gesture Phase Segmentation. It contains 50 attributes divided into two files for each video. 191. <a href="datasets/Gisette">Gisette</a>: GISETTE is a handwritten digit recognition probl em. The problem is to separate the highly confusible digits \'4\' and \'9\'. This dataset is one of five datasets of t he NIPS 2003 feature selection challenge.\r\n 192. <a href="datasets/Glass+Identific ation">Glass Identification</a>: From USA Forensic Science Service; 6 types of glass; defined in terms of t heir oxide content (i.e. Na, Fe, K, etc) 193. <a href="datasets/GNFUV+Unmanned+S urface+Vehicles+Sensor+Data">GNFUV Unmanned Surface Vehicles Sensor Data</a>: The data-set conta ins four (4) sets of mobile sensor readings data (humidity, temperature) corresponding to a swarm of four (4) U nmanned Surface Vehicles (USVs) in a test-bed in Athens (Greece). 194. <a href=" datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2">GNFUV Unmanned Surface Vehicles S ensor Data Set 2</a>: The data-set contains eight (2x4) data-sets of mobile sensor readings data (humidit y, temperature) corresponding to a swarm of four Unmanned Surface Vehicles (USVs) in a test-bed, Athens, Gr eece. 195. <a href="datasets/GPS+Trajectories">GPS Trajectories</a>: The dat aset has been feed by Android app called Go!Track. It is available at Goolge Play Store(https://play.google.com /store/apps/details?id=com.go.router). 196. <a href="datasets/Grammatical+Facial+ Expressions">Grammatical Facial Expressions</a>: This dataset supports the development of models that make possible to interpret Grammatical Facial Expressions from Brazilian Sign Language (Libras). 197. <a href="datasets/Greenhouse+Gas+Observing+Network">Greenhouse Gas Observing Net work</a>: Design an observing network to monitor emissions of a greenhouse gas (GHG) in California give n time series of synthetic observations and tracers from weather model simulations.\r\n 198. <a href="datasets/Haberman%27s+Survival">Haberman\'s Survival</a>: Dataset contains cases fr om study conducted on the survival of patients who had undergone surgery for breast cancer 199. <a href="datasets/Hayes-Roth">Hayes-Roth</a>: Topic: human subjects study 200. <a href="datasets/HCC+Survival">HCC Survival</a>: Hepatocellular Carcinoma dataset (HCC dataset) was collected at a University Hospital in Portugal. It contains real clinical data of 165 patients dia gnosed with HCC. 201. <a href="datasets/Health+News+in+Twitter">Health News in Twitter</a>: The data was collected in 2015 using Twitter API. This dataset contains health news from mor e than 15 major health news agencies such as BBC, CNN, and NYT. 202. <a href=" datasets/Heart+Disease">Heart Disease</a>: 4 databases: Cleveland, Hungary, Switzerland, and the VA L ong Beach 203. <a href="datasets/Hepatitis">Hepatitis</a>: From G.Gong: CMU ; Mostly Boolean or numeric-valued attribute types; Includes cost data (donated by Peter Turney) 204. <a href="datasets/HEPMASS">HEPMASS</a>: The search for exotic particles requires s orting through a large number of collisions to find the events of interest. This data set challenges one to detect a new particle of unknown mass. 205. <a href="datasets/Heterogeneity+Activity+Rec ognition">Heterogeneity Activity Recognition</a>: The Heterogeneity Human Activity Recognition (HHAR) d ataset from Smartphones and Smartwatches is a dataset devised to benchmark human activity recognition algo rithms (classification, automatic data segmentation, sensor fusion, feature extraction, etc.) in real-world context

s; specifically, the dataset is gathered with a variety of different device models and use-scenarios, in order to re flect sensing heterogeneities to be expected in real deployments. 206. <a href="data sets/HIGGS">HIGGS</a>: This is a classification problem to distinguish between a signal process which pr oduces Higgs bosons and a background process which does not. 207. <a href="dat asets/Hill-Valley">Hill-Valley</a>: Each record represents 100 points on a two-dimensional graph. When pl otted in order (from 1 through 100) as the Y co-ordinate, the points will create either a Hill (a &#65533;bump&#65533; in the te rrain) or a Valley (a &#65533;dip&#65533; in the terrain). 208. <a href="datasets/HIV-1+protease+c leavage">HIV-1 protease cleavage</a>: The data contains lists of octamers (8 amino acids) and a flag (-1 or 1) depending on whether HIV-1 protease will cleave in the central position (between amino acids 4 and 5).</ p> 209. <a href="datasets/Horse+Colic">Horse Colic</a>: Well documented attribute s; 368 instances with 28 attributes (continuous, discrete, and nominal); 30% missing values 210. <a href="datasets/HTRU2">HTRU2</a>: Pulsar candidates collected during the HTRU survey. Pulsars are a type of star, of considerable scientific interest. Candidates must be classified in to pulsar and non-pulsar classes to aid discovery. 211. <a href="datasets/Human+Activity+Recognition +Using+Smartphones">Human Activity Recognition Using Smartphones</a>: Human Activity Recognition d atabase built from the recordings of 30 subjects performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors. 212. <a href="datasets/Hybrid +Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer">Hybrid Indoor Positioning Dataset from WiFi RSSI, Bluetooth and magnetometer</a>: The dataset was created for the comparison a nd evaluation of hybrid indoor positioning methods. The dataset presented contains data from W-LAN and Blue tooth interfaces, and Magnetometer. 213. <a href="datasets/ICMLA+2014+Accepte d+Papers+Data+Set">ICMLA 2014 Accepted Papers Data Set</a>: This data set compromises the metada ta for the 2014 ICMLA conference\'s accepted papers, including ID, paper titles, author\'s keywords, abstracts a nd sessions in which they were exposed. 214. <a href="datasets/ICU">ICU</a>: Data set prepared for the use of participants for the 1994 AAAI Spring Symposium on Artificial Intelligence in M edicine. 215. <a href="datasets/IDA2016Challenge">IDA2016Challenge</a>: Th e dataset consists of data collected from heavy Scania trucks in everyday usage. 216. < b><a href="datasets/ILPD+%28Indian+Liver+Patient+Dataset%29">ILPD (Indian Liver Patient Dataset)</a> : This data set contains 10 variables that are age, gender, total Bilirubin, direct Bilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT and Alkphos. 217. <a href="datasets/Image+Segmentation" >Image Segmentation</a>: Image data described by high-level numeric-valued attributes, 7 classes 218. <a href="datasets/Immunotherapy+Dataset">Immunotherapy Dataset</a>: This d ataset contains information about wart treatment results of 90 patients using immunotherapy. 219. <a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Park inson%E2%80%99s+Disease">Improved Spiral Test Using Digitized Graphics Tablet for Monitoring Parkinson’s Disease</a>: Handwriting database consists of 25 PWP(People with Parkinson) and 15 healthy individuals. Three types of recordings (Static Spiral Test, Dynamic Spiral Test and Stability Test) are taken. 220. <a href="datasets/Individual+household+electric+power+consumption">Individual household el ectric power consumption</a>: Measurements of electric power consumption in one household with a one-minute sampling rate over a period of almost 4 years. Different electrical quantities and some sub-metering val ues are available. 221. <a href="datasets/Indoor+User+Movement+Prediction+from+ RSS+data">Indoor User Movement Prediction from RSS data</a>: This dataset contains temporal data fro m a Wireless Sensor Network deployed in real-world office environments. The task is intended as real-life benc hmark in the area of Ambient Assisted Living. 222. <a href="datasets/Insurance+Co mpany+Benchmark+%28COIL+2000%29">Insurance Company Benchmark (COIL 2000)</a>: This data se t used in the CoIL 2000 Challenge contains information on customers of an insurance company. The data consi sts of 86 variables and includes product usage data and socio-demographic data 223. < b><a href="datasets/Internet+Advertisements">Internet Advertisements</a>: This dataset represents a set of possible advertisements on Internet pages. 224. <a href="datasets/Internet+Usag e+Data">Internet Usage Data</a>: This data contains general demographic information on internet users i n 1997. 225. <a href="datasets/Ionosphere">Ionosphere</a>: Classification of r adar returns from the ionosphere 226. <a href="datasets/IPUMS+Census+Database ">IPUMS Census Database</a>: This data set contains unweighted PUMS census data from the Los Angel es and Long Beach areas for the years 1970, 1980, and 1990. 227. <a href="datase ts/Iris">Iris</a>: Famous database; from Fisher, 1936 228. <a href="datasets/IS OLET">ISOLET</a>: Goal: Predict which letter-name was spoken--a simple classification task. <p clas s="normal">229. <a href="datasets/ISTANBUL+STOCK+EXCHANGE">ISTANBUL STOCK EXCHANGE</a> : Data sets includes returns of Istanbul Stock Exchange with seven other international index; SP, DAX, FTS E, NIKKEI, BOVESPA, MSCE_EU, MSCI_EM from Jun 5, 2009 to Feb 22, 2011. 230. <a href="datasets/Japanese+Credit+Screening">Japanese Credit Screening</a>: Includes domain theory (generated by talking to Japanese domain experts); data in Lisp 231. <a href="datas ets/Japanese+Vowels">Japanese Vowels</a>: This dataset records 640 time series of 12 LPC cepstrum c

oefficients taken from nine male speakers. 232. <a href="datasets/KASANDR">KASANDR</a>: KASANDR is a novel, publicly available collection for recommendation systems that records the behavior of customers of the European leader in e-Commerce advertising, Kelkoo. 233. <a href="datasets/KDC-4007+dataset+Collection">KDC-4007 dataset Collection</a>: KDC-4007 dataset Collection is the Kurdish Documents Classification text used in categories regarding Kurdish Sorani news and articles. 234. <a href="datasets/KDD+Cup+1998+Data">KDD Cup 1998 Data</a>: This is the data set used for The Second International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-98 235. <a href="datasets/KDD+Cup+1999+Data">KDD Cup 1999 Data</a>: This is the data set used for The Third International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-99 236. <a href="datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29">KEGG Metabolic Reaction Network (Undirected)</a>: KEGG Metabolic pathways modeled as un-directed reaction network. Variety of graphical features presented. 237. <a href="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29">KEGG Metabolic Relation Network (Directed)</a>: KEGG Metabolic pathways modeled as directed relation network. Variety of graphical features presented. 238. <a href="datasets/Kinship">Kinship</a>: Relational dataset 239. <a href="datasets/Labor+Relations">Labor Relations</a>: From Collective Bargaining Review 240. <a href="datasets/Las+Vegas+Strip">Las Vegas Strip</a>: This dataset includes quantitative and categorical features from online reviews from 21 hotels located in Las Vegas Strip, extracted from TripAdvisor (http://www.tripadvisor.com). 241. <a href="datasets/Leaf">Leaf</a>: This dataset consists in a collection of shape and texture features extracted from digital images of leaf specimens originating from a total of 40 different plant species. 242. <a href="datasets/LED+Display+Domain">LED Display Domain</a>: From Classification and Regression Trees book; We provide here 2 C programs for generating sample databases 243. <a href="datasets/Legal+Case+Reports">Legal Case Reports</a>: A textual corpus of 4000 legal cases for automatic summarization and citation analysis. For each document we collect catchphrases, citations sentences, citation catchphrases and citation classes. 244. <a href="datasets/Lenses">Lenses</a>: Database for fitting contact lenses 245. <a href="datasets/Letter+Recognition">Letter Recognition</a>: Database of character image features; try to identify the letter 246. <a href="datasets/Libras+Movement">Libras Movement</a>: The data set contains 15 classes of 24 instances each. Each class references to a hand movement type in LIBRAS (Portuguese\r\nname \'L&#65533;ngua BRAsileira de Sinais\', oficial brazilian signal language). 247. <a href="datasets/Liver+Disorders">Liver Disorders</a>: BUPA Medical Research Ltd. database donated by Richard S. Forsyth 248. <a href="datasets/Localization+Data+for+Person+Activity">Localization Data for Person Activity</a>: Data contains recordings of five people performing different activities. Each person wore four sensors (tags) while performing the same scenario five times. 249. <a href="datasets/Logic+Theorist">Logic Theorist</a>: All code for Logic Theorist 250. <a href="datasets/Low+Resolution+Spectrometer">Low Resolution Spectrometer</a>: From IRAS data -- NASA Ames Research Center 251. <a href="datasets/LSVT+Voice+Rehabilitation">LSVT Voice Rehabilitation</a>: 126 samples from 14 participants, 309 features. Aim: assess whether voice rehabilitation treatment lead to phonations considered \'acceptable\' or \'unacceptable\' (binary class classification problem). 252. <a href="datasets/Lung+Cancer">Lung Cancer</a>: Lung cancer data; no attribute definitions 253. <a href="datasets/Lymphography">Lymphography</a>: This lymphography domain was obtained from the University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia. (Restricted access) 254. <a href="datasets/M.+Tuberculosis+Genes">M. Tuberculosis Genes</a>: Data giving characteristics of each ORF (potential gene) in the M. tuberculosis bacterium. Sequence, homology (similarity to other genes) and structural information, and function (if known) are provided 255. <a href="datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014">Machine Learning based ZZAlpha Ltd. Stock Recommendations 2012-2014</a>: The data here are the ZZAlpha® machine learning recommendations made for various US traded stock portfolios the morning of each day during the 3 year period Jan 1, 2012 - Dec 31, 2014. 256. <a href="datasets/Madelon">Madelon</a>: MADELON is an artificial dataset, which was part of the NIPS 2003 feature selection challenge. This is a two-class classification problem with continuous input variables. The difficulty is that the problem is multivariate and highly non-linear. 257. <a href="datasets/MAGIC+Gamma+Telescope">MAGIC Gamma Telescope</a>: Data are MC generated to simulate registration of high energy gamma particles in an atmospheric Cherenkov telescope 258. <a href="datasets/Mammographic+Mass">Mammographic Mass</a>: Discrimination of benign and malignant mammographic masses based on BI-RADS attributes and the patient\'s age. 259. <a href="datasets/Mechanical+Analysis">Mechanical Analysis</a>: Fault diagnosis problem of electromechanical devices; also PUMPS DATA SET is newer version with domain theory and results 260. <a href="datasets/Mesothelioma%E2%80%99s+disease+data+set+">Mesothelioma’s disease data set </a>: Mesothelioma’s disease data set were prepared at Dicle University Faculty of Medicine in Turkey.\r\nThree hundr

ed and twenty-four Mesothelioma patient data. In the dataset, all samples have 34 features. 261. <a href="datasets/Meta-data">Meta-data</a>: Meta-Data was used in order to give advice about which classification method is appropriate for a particular dataset (taken from results of Statlog project). 262. <a href="datasets/MEU-Mobile+KSD">MEU-Mobile KSD</a>: This dataset contains keystroke dynamics data collected on a touch mobile device (Nexus 7). The dataset contains 2856 records, 51 records per subject for 56 subjects. 263. <a href="datasets/MHEALTH+Dataset">MHEALTH Dataset</a>: The MHEALTH (Mobile Health) dataset is devised to benchmark techniques dealing with human behavior analysis based on multimodal body sensing. 264. <a href="datasets/Mice+Protein+Expression">Mice Protein Expression</a>: Expression levels of 77 proteins measured in the cerebral cortex of 8 classes of control and Down syndrome mice exposed to context fear conditioning, a task used to assess associative learning. 265. <a href="datasets/microblogPCU">microblogPCU</a>: MicroblogPCU data is crawled from sina weibo microblog[http://weibo.com/]. This data can be used to study machine learning methods as well as do some social network research. 266. <a href="datasets/MicroMass">MicroMass</a>: A dataset to explore machine learning approaches for the identification of microorganisms from mass-spectrometry data. 267. <a href="datasets/MiniBooNE+particle+identification">MiniBooNE particle identification</a>: This dataset is taken from the MiniBooNE experiment and is used to distinguish electron neutrinos (signal) from muon neutrinos (background). 268. <a href="datasets/Miskolc+IIS+Hybrid+IPS">Miskolc IIS Hybrid IPS</a>: The dataset was created for the comparison and evaluation of hybrid indoor positioning methods. The dataset presented contains data from W-LAN and Bluetooth interfaces, and Magnetometer. 269. <a href="datasets/Mobile+Robots">Mobile Robots</a>: Learning concepts from sensor data of a mobile robot; set of data sets 270. <a href="datasets/MoCap+Hand+Postures">MoCap Hand Postures</a>: 5 types of hand postures from 12 users were recorded using unlabeled markers attached to fingers of a glove in a motion capture environment. Due to resolution and occlusion, missing values are common. 271. <a href="datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29">Molecular Biology (Promoter Gene Sequences)</a>: E. Coli promoter gene sequences (DNA) with partial domain theory 272. <a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular Biology (Protein Secondary Structure)</a>: From CMU connectionist bench repository; Classifies secondary structure of certain globular proteins 273. <a href="datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29">Molecular Biology (Splice-junction Gene Sequences)</a>: Primate splice-junction gene sequences (DNA) with associated imperfect domain theory 274. <a href="datasets/MONK%27s+Problems">MONK\'s Problems</a>: A set of three artificial domains over the same attribute space; Used to test a wide range of induction algorithms 275. <a href="datasets/Moral+Reasoner">Moral Reasoner</a>: Horn-clause model that qualitatively simulates moral reasoning; Theory includes negated literals 276. <a href="datasets/Motion+Capture+Hand+Postures">Motion Capture Hand Postures</a>: 5 types of hand postures from 12 users were recorded using unlabeled markers on fingers of a glove in a motion capture environment. Due to resolution and occlusion, missing values are common. 277. <a href="datasets/Movie">Movie</a>: This data set contains a list of over 10000 films including many older, odd, and cult films. There is information on actors, casts, directors, producers, studios, etc. 278. <a href="datasets/MSNBC.com+Anonymous+Web+Data">MSNBC.com Anonymous Web Data</a>: This data describes the page visits of users who visited msnbc.com on September 28, 1999. Visits are recorded at the level of URL category (see description) and are recorded in time order. 279. <a href="datasets/Mturk+User-Perceived+Clusters+over+Images">Mturk User-Perceived Clusters over Images</a>: This dataset was collected by Shan-Hung Wu and DataLab members at NTHU, Taiwan. There\'re 325 user-perceived clusters from 100 users and their corresponding descriptions. 280. <a href="datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing">Multimodal Damage Identification for Humanitarian Computing</a>: 5879 captioned images (image and text) from social media related to damage during natural disasters/wars, and belong to 6 classes: Fires, Floods, Natural landscape, Infrastructural, Human, Non-damage. 281. <a href="datasets/Multiple+Features">Multiple Features</a>: This dataset consists of features of handwritten numerals (`0\'--`9\') extracted from a collection of Dutch utility maps 282. <a href="datasets/Mushroom">Mushroom</a>: From Audobon Society Field Guide; mushrooms described in terms of physical characteristics; classification: poisonous or edible 283. <a href="datasets/Musk+%28Version+1%29">Musk (Version 1)</a>: The goal is to learn to predict whether new molecules will be musks or non-musks 284. <a href="datasets/Musk+%28Version+2%29">Musk (Version 2)</a>: The goal is to learn to predict whether new molecules will be musks or non-musks 285. <a href="datasets/News+Aggregator">News Aggregator</a>: References to news pages collected from an web aggregator in the period from 10-March-2014 to 10-August-2014. The resources are grouped into clusters that represent pages discussing the same story. 286. <a href="datasets/News+Popularity+in+Multiple+Social+Media+Platforms">News Popularity in Multiple Social Media Platforms</a>: Large data set of news items and their respective social feedback on multiple platforms: Facebook, Google+ and LinkedIn. <p cl

ass="normal">287. <a href="datasets/Newspaper+and+magazine+images+segmentation+dataset">Newspaper and magazine images segmentation dataset</a>: Dataset is well suited for segmentation tasks. It contains 101 scanned pages from different newspapers and magazines in Russian with ground truth pixel-based masks. 288. <a href="datasets/NIPS+Conference+Papers+1987-2015">NIPS Conference Papers 1987-2015</a>: This data set contains the distribution of words in the full text of the NIPS conference papers published from 1987 to 2015. 289. <a href="datasets/NoisyOffice">NoisyOffice</a>: Corpus intended to do cleaning (or binarization) and enhancement of noisy grayscale printed text images using supervised learning methods. Noisy images and their corresponding ground truth provided. 290. <a href="datasets/Nomao">Nomao</a>: Nomao collects data about places (name, phone, localization...) from many sources.\r\nDeduplication consists in detecting what data refer to the same place.\r\nInstances in the dataset compare 2 spots. 291. <a href="datasets/Northix">Northix</a>: Northix is designed to be a schema matching benchmark problem for data integration of two entity relationship databases. 292. <a href="datasets/NSF+Research+Award+Abstracts+1990-2003">NSF Research Award Abstracts 1990-2003</a>: This data set consists of (a) 129,000 abstracts describing NSF awards for basic research, (b) bag-of-word data files extracted from the abstracts, (c) a list of words used for indexing the bag-of-word 293. <a href="datasets/Nursery">Nursery</a>: Nursery Database was derived from a hierarchical decision model originally developed to rank applications for nursery schools. 294. <a href="datasets/NYSK">NYSK</a>: NYSK (New York v. Strauss-Kahn) is a collection of English news articles about the case relating to allegations of sexual assault against the former IMF director Dominique Strauss-Kahn (May 2011). 295. <a href="datasets/Occupancy+Detection+">Occupancy Detection </a>: Experimental data used for binary classification (room occupancy) from Temperature,Humidity,Light and CO2. Ground-truth occupancy was obtained from time stamped pictures that were taken every minute. 296. <a href="datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes">OCT data & Color Fundus Images of Left & Right Eyes</a>: This dataset contains OCT data (in mat format) and color fundus data (in jpg format) of left & right eyes of 50 healthy persons. 297. <a href="datasets/One-hundred+plant+species+leaves+data+set">One-hundred plant species leaves data set</a>: Sixteen samples of leaf each of one-hundred plant species. For each sample, a shape descriptor, fine scale margin and texture histogram are given. 298. <a href="datasets/Online+Handwritten+Assamese+Characters+Dataset">Online Handwritten Assamese Characters Dataset</a>: This is a dataset of 8235 online handwritten assamese characters. The “online” process involves capturing of data as text is written on a digitizing tablet with an electronic pen. 299. <a href="datasets/Online+News+Popularity">Online News Popularity</a>: This dataset summarizes a heterogeneous set of features about articles published by Mashable in a period of two years. The goal is to predict the number of shares in social networks (popularity). 300. <a href="datasets/Online+Retail">Online Retail</a>: This is a transnational data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail. 301. <a href="datasets/Online+Shoppers+Purchasing+Intention+Dataset">Online Shoppers Purchasing Intention Dataset</a>: Of the 12,330 sessions in the dataset,\r\n84.5% (10,422) were negative class samples that did not\r\nend with shopping, and the rest (1908) were positive class\r\nsamples ending with shopping. 302. <a href="datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset">Online Video Characteristics and Transcoding Time Dataset</a>: The dataset contains a million randomly sampled video instances listing 10 fundamental video characteristics along with the YouTube video ID. 303. <a href="datasets/Open+University+Learning+Analytics+dataset">Open University Learning Analytics dataset</a>: Open University Learning Analytics Dataset contains data about courses, students and their interactions with Virtual Learning Environment for seven selected courses and more than 30000 students. 304. <a href="datasets/Opinosis+Opinion+%26frasl%3B+Review">Opinosis Opinion ⁄ Review</a>: This dataset contains sentences extracted from user reviews on a given topic. Example topics are “performance of Toyota Camry” and “sound quality of ipod nano”. 305. <a href="datasets/OpinRank+Review+Dataset">OpinRank Review Dataset</a>: This data set contains user reviews of cars and and hotels collected from Tripadvisor (~259,000 \r\nreviews) and Edmunds (~42,230 reviews). 306. <a href="datasets/OPPORTUNITY+Activity+Recognition">OPPORTUNITY Activity Recognition</a>: The OPPORTUNITY Dataset for Human Activity Recognition from Wearable, Object, and Ambient Sensors is a dataset devised to benchmark human activity recognition algorithms (classification, automatic data segmentation, sensor fusion, feature extraction, etc). 307. <a href="datasets/Optical+Interconnection+Network+">Optical Interconnection Network </a>: This dataset contains 640 performance measurements from a simulation of 2-Dimensional Multiprocessor Optical Interconnection Network. 308. <a href="datasets/Optical+Recognition+of+Handwritten+Digits">Optical Recognition of Handwritten Digits</a>: Two versions of this database available; see folder 309. <a href="datasets/Othello+Domain+Theory">Othello Domain Theory</a>: Used in research to generate features for an inductive learning system 310. <a href="datasets/Ozone+Level+Detection">Ozone Level Detection</a>: Two ground ozone level data sets are included in this collection. One is the eight hour peak set (eighthr.data), the ot

her is the one hour peak set (onehr.data). Those data were collected from 1998 to 2004 at the Houston, Galveston and Brazoria area. 311. <a href="datasets/p53+Mutants">p53 Mutants</a>: The goal is to model mutant p53 transcriptional activity (active vs inactive) based on data extracted from biophysical simulations.\r\n 312. <a href="datasets/Page+Blocks+Classification">Page Blocks Classification</a>: The problem consists of classifying all the blocks of the page layout of a document that has been detected by a segmentation process. 313. <a href="datasets/PAMAP2+Physical+Activity+Monitoring">PAMAP2 Physical Activity Monitoring</a>: The PAMAP2 Physical Activity Monitoring dataset contains data of 18 different physical activities, performed by 9 subjects wearing 3 inertial measurement units and a heart rate monitor. 314. <a href="datasets/PANDOR">PANDOR</a>: PANDOR is a novel and publicly available dataset for online recommendation provided by Purch (http://www.purch.com/). 315. <a href="datasets/Paper+Reviews">Paper Reviews</a>: This sentiment analysis data set contains scientific paper reviews from an international conference on computing and informatics. The task is to predict the orientation or the evaluation of a review. 316. <a href="datasets/Parking+Birmingham">Parking Birmingham</a>: Data collected from car parks in Birmingham that are operated by NCP from \r\nBirmingham City Council. UK Open Government Licence (OGL).\r\nhttps://data.birmingham.gov.uk/dataset/birmingham-parking 317. <a href="datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet">Parkinson Disease Spiral Drawings Using Digitized Graphics Tablet</a>: Handwriting database consists of 62 PWP(People with Parkinson) and 15 healthy individuals. Three types of recordings (Static Spiral Test, Dynamic Spiral Test and Stability Test) are taken. 318. <a href="datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings">Parkinson Speech Dataset with Multiple Types of Sound Recordings</a>: The training data belongs to 20 Parkinson\'s Disease (PD) patients and 20 healthy subjects. From all subjects, multiple types of sound recordings (26) are taken. 319. <a href="datasets/Parkinson%27s+Disease+Classification">Parkinson\'s Disease Classification</a>: The data used in this study were gathered from 188 patients with PD (107 men and 81 women) with ages ranging from 33 to 87 (65.1±10.9). 320. <a href="datasets/Parkinsons">Parkinsons</a>: Oxford Parkinson\'s Disease Detection Dataset 321. <a href="datasets/Parkinsons+Telemonitoring">Parkinsons Telemonitoring</a>: Oxford Parkinson\'s Disease Telemonitoring Dataset 322. <a href="datasets/PEMS-SF">PEMS-SF</a>: 15 months worth of daily data (440 daily records) that describes the occupancy rate, between 0 and 1, of different car lanes of the San Francisco bay area freeways across time. 323. <a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits">Pen-Based Recognition of Handwritten Digits</a>: Digit database of 250 samples from 44 writers 324. <a href="datasets/Perfume+Data">Perfume Data</a>: This data consists of odors of 20 different perfumes. Data was obtained by using a handheld odor meter (OMX-GR sensor) per second for 28 seconds period. 325. <a href="datasets/Phishing+Websites">Phishing Websites</a>: This dataset collected mainly from: PhishTank archive, MillerSmiles archive, Google’s searching operators. 326. <a href="datasets/Physical+Unclonable+Functions">Physical Unclonable Functions</a>: The dataset is generated from Physical Unclonable Functions (PUFs) simulation, specifically XOR Arbiter PUFs. PUFs are used for authentication purposes. For more info, refer to our paper below. 327. <a href="datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure">Physicochemical Properties of Protein Tertiary Structure</a>: This is a data set of Physicochemical Properties of Protein Tertiary Structure. The data set is taken from CASP 5-9. There are 45730 decoys and size varying from 0 to 21 armstrong. 328. <a href="datasets/Pioneer-1+Mobile+Robot+Data">Pioneer-1 Mobile Robot Data</a>: This dataset contains time series sensor readings of the Pioneer-1 mobile robot. The data is broken into "experiences" in which the robot takes action for some period of time and experiences a control 329. <a href="datasets/Pittsburgh+Bridges">Pittsburgh Bridges</a>: Bridges database that has original and numeric-discretized datasets 330. <a href="datasets/Planning+Relax">Planning Relax</a>: The dataset concerns with the classification of two mental stages from recorded EEG signals: Planning (during imagination of motor act) and Relax state. 331. <a href="datasets/Plants">Plants</a>: Data has been extracted from the USDA plants database. It contains all plants (species and genera) in the database and the states of USA and Canada where they occur. 332. <a href="datasets/PM2.5+Data+of+Five+Chinese+Cities">PM2.5 Data of Five Chinese Cities</a>: This hourly data set contains the PM2.5 data in Beijing, Shanghai, Guangzhou, Chengdu and Shenyang. Meanwhile, meteorological data for each city are also included. 333. <a href="datasets/PMU-UD">PMU-UD</a>: The handwritten dataset was collected from 170 participants with a total of 5,180 numeral patterns. The dataset is named Prince Mohammad Bin Fahd University - Urdu/Arabic Database (PMU-UD). 334. <a href="datasets/Poker+Hand">Poker Hand</a>: Purpose is to predict poker hands 335. <a href="datasets/Polish+companies+bankruptcy+data">Polish companies bankruptcy data</a>: The dataset is about bankruptcy prediction of Polish companies.The bankrupt companies were analyzed in the period 2000-2012, while the still operating companies were evaluated from 2007 to 2013. 336. <a href="datasets/Post-Operative+Patient">Post-Operative Patient</a>: Dataset of patient features 337. <a href="datasets/Predict+keyword

Patient</a>: Dataset of patient features 337. <a href="datasets/Predict+keywords+activities+in+a+online+social+media">Predict keywords activities in a online social media</a>: The data from Twitter was collected during 360 consecutive days. It was done by querying 1497 English keywords sampled from Wikipedia. This dataset is proposed in a Learning to rank setting. 338. <a href="datasets/Primary+Tumor">Primary Tumor</a>: From Ljubljana Oncology Institute 339. <a href="datasets/Prodigy">Prodigy</a>: Assorted domains like blocksworld, eightpuzzle, and schedworld. 340. <a href="datasets/Protein+Data">Protein Data</a>: Undocumented 341. <a href="datasets/Pseudo+Periodic+Synthetic+Time+Series">Pseudo Periodic Synthetic Time Series</a>: This data set is designed for testing indexing schemes in time series databases. The data appears highly periodic, but never exactly repeats itself. 342. <a href="datasets/PubChem+Bioassay+Data">PubChem Bioassay Data</a>: These highly imbalanced bioassay datasets are from the differing types of screening that can be performed using HTS technology. 21 datasets were created from 12 bioassays. 343. <a href="datasets/QSAR+biodegradation">QSAR biodegradation</a>: Data set containing values for 41 attributes (molecular descriptors) used to classify 1055 chemicals into 2 classes (ready and not ready biodegradable). 344. <a href="datasets/QtyT40I10D100K">QtyT40I10D100K</a>: Since there is no numerical sequential data stream available in standard data sets, this data set is generated from the original T40I10D100K data set 345. <a href="datasets/Quadruped+Mammals">Quadruped Mammals</a>: The file animals.c is a data generator of structured instances representing quadruped animals 346. <a href="datasets/Qualitative+Structure+Activity+Relationships">Qualitative Structure Activity Relationships</a>: Two sets of datasets are given: pyrimidines and triazines 347. <a href="datasets/Qualitative_Bankruptcy">Qualitative_Bankruptcy</a>: Predict the Bankruptcy from Qualitative parameters from experts. 348. <a href="datasets/Quality+Assessment+of+Digital+Colposcopies">Quality Assessment of Digital Colposcopies</a>: This dataset explores the subjective quality assessment of digital colposcopies. 349. <a href="datasets/Real+estate+valuation+data+set">Real estate valuation data set</a>: The “real estate valuation” is a regression problem. The market historical data set of real estate valuation are collected from Sindian Dist., New Taipei City, Taiwan. 350. <a href="datasets/REALDISP+Activity+Recognition+Dataset">REALDISP Activity Recognition Dataset</a>: The REALDISP dataset is devised to evaluate techniques dealing with the effects of sensor displacement in wearable activity recognition as well as to benchmark general activity recognition algorithms 351. <a href="datasets/Record+Linkage+Comparison+Patterns">Record Linkage Comparison Patterns</a>: Element-wise comparison of records with personal data from a record linkage setting. The task is to decide from a comparison pattern whether the underlying records belong to one person. 352. <a href="datasets/Relative+location+of+CT+slices+on+axial+axis">Relative location of CT slices on axial axis</a>: The dataset consists of 384 features extracted from CT images. The class variable is numeric and denotes the relative location of the CT slice on the axial axis of the human body. 353. <a href="datasets/Repeat+Consumption+Matrices">Repeat Consumption Matrices</a>: The dataset contains 7 datasets of User - Item matrices, where each entry represents how many times a user consumed an item. Item is used as an umbrella term for various categories. 354. <a href="datasets/Residential+Building+Data+Set">Residential Building Data Set</a>: Data set includes construction cost, sale prices, project variables, and economic variables corresponding to real estate single-family residential apartments in Tehran, Iran. 355. <a href="datasets/Restaurant+%26+consumer+data">Restaurant & consumer data</a>: The dataset was obtained from a recommender system prototype. The task was to generate a top-n list of restaurants according to the consumer preferences. 356. <a href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection">Reuters RCV1 RCV2 Multilingual, Multiview Text Categorization Test collection</a>: This test collection contains feature characteristics of documents originally written in five different languages and their translations, over a common set of 6 categories. 357. <a href="datasets/Reuters+Transcribed+Subset">Reuters Transcribed Subset</a>: This dataset is created by reading out 200 files from the 10 largest Reuters \r\nclasses and using an Automatic Speech Recognition system to create \r\ncorresponding transcriptions. 358. <a href="datasets/Reuters-21578+Text+Categorization+Collection">Reuters-21578 Text Categorization Collection</a>: This is a collection of documents that appeared on Reuters newswire in 1987. The documents were assembled and indexed with categories. 359. <a href="datasets/Reuter_50_50">Reuter_50_50</a>: The dataset is used for authorship identification in online Writeprint which is a new research field of pattern recognition. 360. <a href="datasets/Rice+Leaf+Diseases">Rice Leaf Diseases</a>: There are three classes/diseases: Bacterial leaf blight, Brown spot, and Leaf smut, each having 40 images. The format of all images is jpg. 361. <a href="datasets/Robot+Execution+Failures">Robot Execution Failures</a>: This dataset contains force and torque measurements on a robot after failure detection. Each failure is characterized by 15 force/torque samples collected at regular time intervals 362. <a href="datasets/Roman+Urdu+Data+Set">Roman Urdu Data Set</a>: Roman Urdu (the scripting style for Urdu language) is one of the limited resource languages.A data corpus comprising of more than 20000 records was collected. 363. <a href="datasets/Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_W

s= normal >363. <a href= datasets/Sales_Transactions_Dataset_Weekly >Sales_Transactions_Dataset_Weekly</a>: Contains weekly purchased quantities of 800 over products over 52 weeks. Normalised values are provided too. 364. <a href="datasets/SCADI">SCADI</a>: First self-care activities dataset based on ICF-CY. 365. <a href="datasets/SECOM">SECOM</a>: Data from a semi-conductor manufacturing process 366. <a href="datasets/seeds">seeds</a>: Measurements of geometrical properties of kernels belonging to three different varieties of wheat. A soft X-ray technique and GRAINS package were used to construct all seven, real-valued attributes. 367. <a href="datasets/seismic-bumps">seismic-bumps</a>: The data describe the problem of high energy (higher than 10^4 J) seismic bumps forecasting in a coal \r\nmine. Data come from two of longwalls located in a Polish coal mine. 368. <a href="datasets/Semeion+Handwritten+Digit">Semeion Handwritten Digit</a>: 1593 handwritten digits from around 80 persons were scanned, stretched in a rectangular box 16x16 in a gray scale of 256 values. 369. <a href="datasets/sEMG+for+Basic+Hand+movements">sEMG for Basic Hand movements</a>: The “sEMG for Basic Hand movements” includes 2 databases of surface electromyographic signals of 6 hand movements using Delsys\' EMG System. Healthy subjects conducted six daily life grasps. 370. <a href="datasets/Sentence+Classification">Sentence Classification</a>: Contains sentences from the abstract and introduction of 30 articles annotated with a modified Argumentative Zones annotation scheme. These articles come from biology, machine learning and psychology. 371. <a href="datasets/Sentiment+Labelled+Sentences">Sentiment Labelled Sentences</a>: The dataset contains sentences labelled with positive or negative sentiment. 372. <a href="datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machines%29">ser Knowledge Modeling Data (Students\' Knowledge Levels on DC Electrical Machines)</a>: The dataset is about the users\' learning activities and knowledge levels on subjects of DC Electrical Machines. The dataset had been obtained from online web-courses and reported in my Ph.D. Thesis. 373. <a href="datasets/Servo">Servo</a>: Data was from a simulation of a servo system 374. <a href="datasets/SGEMM+GPU+kernel+performance">SGEMM GPU kernel performance</a>: Running times for multiplying two 2048 x 2048 matrices using a GPU OpenCL SGEMM kernel with varying parameters (using the library \'CLTune\'). 375. <a href="datasets/Shuttle+Landing+Control">Shuttle Landing Control</a>: Tiny database; all nominal values 376. <a href="datasets/SIFT10M">SIFT10M</a>: In SIFT10M, each data point is a SIFT feature which is extracted from Caltech-256 by the open source VLFeat library. The corresponding patches of the SIFT features are provided. 377. <a href="datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set">Simulated Falls and Daily Living Activities Data Set</a>: 20 falls and 16 daily living activities were performed by 17 volunteers with 5 repetitions while wearing 6 sensors (3.060 instances) that attached to their head, chest, waist, wrist, thigh and ankle. 378. <a href="datasets/SkillCraft1+Master+Table+Dataset">SkillCraft1 Master Table Dataset</a>: This data was used in Thompson et al. (2013). A list of possible game actions is discussed in Thompson, Blair, Chen, & Henrey (2013). 379. <a href="datasets/Skin+Segmentation">Skin Segmentation</a>: The Skin Segmentation dataset is constructed over B, G, R color space. Skin and Nonskin dataset is generated using skin textures from face images of diversity of age, gender, and race people. 380. <a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Living+%28AAL%29">Smartphone Dataset for Human Activity Recognition (HAR) in Ambient Assisted Living (AAL)</a>: This data is an addition to an existing dataset on UCI. We collected more data to improve the accuracy of our human activity recognition algorithms applied in the domain of Ambient Assisted Living. 381. <a href="datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions">Smartphone-Based Recognition of Human Activities and Postural Transitions</a>: Activity recognition data set built from the recordings of 30 subjects performing basic activities and postural transitions while carrying a waist-mounted smartphone with embedded inertial sensors.\r\n 382. <a href="datasets/SML2010">SML2010</a>: This dataset is collected from a monitor system mounted in a domotic house. It corresponds to approximately 40 days of monitoring data. 383. <a href="datasets/SMS+Spam+Collection">SMS Spam Collection</a>: The SMS Spam Collection is a public set of SMS labeled messages that have been collected for mobile phone spam research. 384. <a href="datasets/Solar+Flare">Solar Flare</a>: Each class attribute counts the number of solar flares of a certain class that occur in a 24 hour period 385. <a href="datasets/Somerville+Happiness+Survey">Somerville Happiness Survey</a>: A data extract of a non-federal dataset posted here https://catalog.data.gov/dataset/somerville-happiness-survey-responses-2011-2013-2015 386. <a href="datasets/Soybean+%28Large%29">Soybean (Large)</a>: Michalski\'s famous soybean disease database 387. <a href="datasets/Soybean+%28Small%29">Soybean (Small)</a>: Michalski\'s famous soybean disease database 388. <a href="datasets/Spambase">Spambase</a>: Classifying Email as Spam or Non-Spam 389. <a href="datasets/SPECT+Heart">SPECT Heart</a>: Data on cardiac Single Proton Emission Computed Tomography (SPECT) images. Each patient classified into two categories: normal and abnormal. 390. <a href="datasets/SPECTF+Heart">SPECTF Heart</a>: Data on card

lac Single Proton Emission Computed Tomography (SPECT) images. Each patient classified into two categories : normal and abnormal. 391. <a href="datasets/Spoken+Arabic+Digit">Spoken Arabi c Digit</a>: This dataset contains timeseries of mel-frequency cepstrum coefficients (MFCCs) correspondin g to spoken Arabic digits. Includes data from 44 male and 44 female native Arabic speakers. 392. <a href="datasets/Sponge">Sponge</a>: Data on sponges; Attributes in spanish <p clas s="normal">393. <a href="datasets/Sports+articles+for+objectivity+analysis">Sports articles for objectivity a nalysis</a>: 1000 sports articles were labeled using Amazon Mechanical Turk as objective or subjective. T he raw texts, extracted features, and the URLs from which the articles were retrieved are provided. 394. <a href="datasets/Statlog+%28Australian+Credit+Approval%29">Statlog (Australian Credit Approval)</a>: This file concerns credit card applications. This database exists elsewhere in the repository (Credit Screening Database) in a slightly different form 395. <a href="datasets/Statlo g+%28German+Credit+Data%29">Statlog (German Credit Data)</a>: This dataset classifies people descri bed by a set of attributes as good or bad credit risks. Comes in two formats (one all numeric). Also comes with a cost matrix 396. <a href="datasets/Statlog+%28Heart%29">Statlog (Heart)</a>: This dataset is a heart disease database similar to a database already present in the repository (Heart Disea se databases) but in a slightly different form 397. <a href="datasets/Statlog+%28Ima ge+Segmentation%29">Statlog (Image Segmentation)</a>: This dataset is an image segmentation databa se similar to a database already present in the repository (Image segmentation database) but in a slightly differ ent form. 398. <a href="datasets/Statlog+%28Landsat+Satellite%29">Statlog (Lands at Satellite)</a>: Multi-spectral values of pixels in 3x3 neighbourhoods in a satellite image, and the classific ation associated with the central pixel in each neighbourhood 399. <a href="datasets /Statlog+%28Shuttle%29">Statlog (Shuttle)</a>: The shuttle dataset contains 9 attributes all of which are n umerical. Approximately 80% of the data belongs to class 1 400. <a href="datasets/S tatlog+%28Vehicle+Silhouettes%29">Statlog (Vehicle Silhouettes)</a>: 3D objects within a 2D image by ap plication of an ensemble of shape feature extractors to the 2D silhouettes of the objects. 401. <a href="datasets/Statlog+Project">Statlog Project</a>: Various Databases: Vehicle silhouttes, L andsat Sattelite, Shuttle, Australian Credit Approval, Heart Disease, Image Segmentation, German Credit < p class="normal">402. <a href="datasets/Steel+Plates+Faults">Steel Plates Faults</a>: A dataset of st eel plates' faults, classified into 7 different types. \r\nThe goal was to train machine learning for automatic patter n recognition.\r\n 403. <a href="datasets/Stock+portfolio+performance">Stock portfol io performance</a>: The data set of performances of weighted scoring stock portfolios are obtained with m ixture design from the US stock market historical database. 404. <a href="datasets/S toneFlakes">StoneFlakes</a>: Stone flakes are waste products of the stone tool production in\r\nthe prehis toric era. The variables are means of geometric and\r\nstylistic features of the flakes contained in different inve ntories. 405. <a href="datasets/Student+Academics+Performance">Student Academ ics Performance</a>: The dataset tried to find the end semester percentage prediction based on different s ocial, economic and academic attributes. 406. <a href="datasets/Student+Loan+Rel ational">Student Loan Relational</a>: Student Loan Relational Domain 407. <a href="datasets/Student+Performance">Student Performance</a>: Predict student performance in secondar y education (high school). 408. <a href="datasets/Superconductivty+Data">Superco nductivty Data</a>: Two file s contain data on 21263 superconductors and their relevant features. <p cl ass="normal">409. <a href="datasets/SUSY">SUSY</a>: This is a classification problem to distinguish between a signal process which produces supersymmetric particles and a background process which does not. 410. <a href="datasets/Synthetic+Control+Chart+Time+Series">Synthetic Control C hart Time Series</a>: This data consists of synthetically generated control charts. 4 11. <a href="datasets/Syskill+and+Webert+Web+Page+Ratings">Syskill and Webert Web Page Ratings</a> : This database contains HTML source of web pages plus the ratings of a single user on these web pages. Web pages are on four seperate subjects (Bands- recording artists; Goats; Sheep; and BioMedical) <p clas s="normal">412. <a href="datasets/Tamilnadu+Electricity+Board+Hourly+Readings">Tamilnadu Electricity B oard Hourly Readings</a>: This data can be effectively produced the result to fewer parameter of the Load profile can be reduced in the Database 413. <a href="datasets/Tarvel+Review+Rati ngs">Tarvel Review Ratings</a>: Google reviews on attractions from 24 categories across Europe are con sidered. Google user rating ranges from 1 to 5 and average user rating per category is calculated. 414. <a href="datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+20 15">Taxi Service Trajectory - Prediction Challenge, ECML PKDD 2015</a>: An accurate dataset describing trajectories performed by all the 442 taxis running in the city of Porto, in Portugal.\r\n 41 5. <a href="datasets/Teaching+Assistant+Evaluation">Teaching Assistant Evaluation</a>: The data con sist of evaluations of teaching performance; scores are "low", "medium", or "high" 416. < b><a href="datasets/Tennis+Major+Tournament+Match+Statistics">Tennis Major Tournament Match Statistics< /a>: This is a collection of 8 files containing the match statistics for both women and men at the four major t ennis tournaments of the year 2013. Each file has 42 columns and a minimum of 76 rows. 417. <a href="datasets/Thoracic+Surgery+Data">Thoracic Surgery Data</a>: The data is dedicated t

o classification problem related to the post-operative life expectancy in the lung cancer patients: class 1 - death within one year after surgery, class 2 - survival. 418. <a href="datasets/Thyroid+Disease">Thyroid Disease</a>: 10 separate databases from Garavan Institute 419. <a href="datasets/Tic-Tac-Toe+Endgame">Tic-Tac-Toe Endgame</a>: Binary classification task on possible configurations of tic-tac-toe game 420. <a href="datasets/Trains">Trains</a>: 2 data formats (structured, one-instance-per-line) 421. <a href="datasets/Travel+Reviews">Travel Reviews</a>: Reviews on destinations in 10 categories mentioned across East Asia. Each traveler rating is mapped as Excellent(4), Very Good(3), Average(2), Poor(1), and Terrible(0) and average rating is used. 422. <a href="datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization">TTC-3600: Benchmark dataset for Turkish text categorization</a>: The TTC-3600 data set is a collection of Turkish news and articles including categorized 3,600 documents from 6 well-known portals in Turkey. It has 4 different forms in ARFF Weka format. 423. <a href="datasets/Turkiye+Student+Evaluation">Turkiye Student Evaluation</a>: This data set contains a total 5820 evaluation scores provided by students from Gazi University in Ankara (Turkey). There is a total of 28 course specific questions and additional 5 attributes. 424. <a href="datasets/TV+News+Channel+Commercial+Detection+Dataset">TV News Channel Commercial Detection Dataset</a>: TV Commercials data set consists of standard audio-visual features of video shots extracted from 150 hours of TV news broadcast of 3 Indian and 2 international news channels (30 Hours each). 425. <a href="datasets/Twenty+Newsgroups">Twenty Newsgroups</a>: This data set consists of 20000 messages taken from 20 newsgroups. 426. <a href="datasets/Twin+gas+sensor+arrays">Twin gas sensor arrays</a>: 5 replicates of an 8-MOX gas sensor array were exposed to different gas conditions (4 volatiles at 10 concentration levels each). 427. <a href="datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis">Twitter Data set for Arabic Sentiment Analysis</a>: This problem of Sentiment Analysis (SA) has been studied well on the English language but not Arabic one. Two main approaches have been devised: corpus-based and lexicon-based. 428. <a href="datasets/UbiqLog+%28smartphone+lifelogging%29">UbiqLog (smartphone lifelogging)</a>: UbiqLog is the smartphone lifelogging tool that runs on the smartphone of 35 users for about 2 months. \r\n 429. <a href="datasets/UJI+Pen+Characters">UJI Pen Characters</a>: Data consists of written characters in a UNIPEN-like format 430. <a href="datasets/UJI+Pen+Characters+%28Version+2%29">UJI Pen Characters (Version 2)</a>: A pen-based database with more than 11k isolated handwritten characters 431. <a href="datasets/UJIIndoorLoc">UJIIndoorLoc</a>: The UJIIndoorLoc is a Multi-Building Multi-Floor indoor localization database to test Indoor Positioning System that rely on WLAN/WiFi fingerprint. 432. <a href="datasets/UJIIndoorLoc-Mag">UJIIndoorLoc-Mag</a>: The UJIIndoorLoc-Mag is an indoor localization database to test Indoor Positioning System that rely on Earth\'s magnetic field variations. 433. <a href="datasets/Ultrasonic+flowmeter+diagnostics">Ultrasonic flowmeter diagnostics</a>: Fault diagnosis of four liquid ultrasonic flowmeters 434. <a href="datasets/Undocumented">Undocumented</a>: Various datasets without documentation (feel free to explore!) 435. <a href="datasets/University">University</a>: Data in original (LISP-readable) form 436. <a href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29">University of Tehran Question Dataset 2016 (UTQD.2016)</a>: Persian questions gathered from a jeopardy game broadcasted on Iranian national television. 437. <a href="datasets/UNIX+User+Data">UNIX User Data</a>: This file contains 9 sets of sanitized user data drawn from the command histories of 8 UNIX computer users at Purdue over the course of up to 2 years. 438. <a href="datasets/Urban+Land+Cover">Urban Land Cover</a>: Classification of urban land cover using high resolution aerial imagery. Intended to assist sustainable urban planning efforts. 439. <a href="datasets/URL+Reputation">URL Reputation</a>: Anonymized 120-day subset of the ICML-09 URL data containing 2.4 million examples and 3.2 million features. 440. <a href="datasets/US+Census+Data+%281990%29">US Census Data (1990)</a>: The USCensus1990raw data set contains a one percent sample of the Public Use Microdata Samples (PUMS) person records drawn from the full 1990 census sample. 441. <a href="datasets/User+Identification+From+Walking+Activity">User Identification From Walking Activity</a>: The dataset collects data from an Android smartphone positioned in the chest pocket from 22 participants walking in the wild over a predefined path. \r\n 442. <a href="datasets/User+Knowledge+Modeling">User Knowledge Modeling</a>: It is the real dataset about the students\' knowledge status about the subject of Electrical DC Machines. The dataset had been obtained from Ph.D. Thesis. 443. <a href="datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Problem%3A+Pat">USPTO Algorithm Challenge, run by NASA-Harvard Tournament Lab and TopCoder Problem: Pat</a>: Data used for USPTO Algorithm Competition. Contains drawing pages from US patents with manually labeled figure and part labels. 444. <a href="datasets/Vertebral+Column">Vertebral Column</a>: Data set containing values for six biomechanical features used to classify orthopaedic patients into 3 classes (normal, disk hernia or spondilolysthesis) or 2 classes (normal or abnormal). 445. <a href="datasets/Vicon+Physical+Action+Data+Set">Vicon Physical Action Data Set</a></

b>: The Physical Action Data Set includes 10 normal and 10 aggressive physical actions that measure the human activity. The data have been collected by 10 subjects using the Vicon 3D tracker. 446. <a href="datasets/Victorian+Era+Authorship+Attribution">Victorian Era Authorship Attribution</a>: To create the largest authorship attribution dataset, we extracted works of 50 well-known authors. To have a non-exhaustive learning, in training there are 45 authors whereas, in the testing, it\'s 50 447. <a href="datasets/Volcanoes+on+Venus+-+JARtool+experiment">Volcanoes on Venus - JARtool experiment</a>: The JARtool project was a pioneering effort to develop an automatic system for cataloging small volcanoes in the large set of Venus images returned by the Magellan spacecraft. 448. <a href="datasets/Wall-Following+Robot+Navigation+Data">Wall-Following Robot Navigation Data</a>: The data were collected as the SCITOS G5 robot navigates through the room following the wall in a clockwise direction, for 4 rounds, using 24 ultrasound sensors arranged circularly around its \'waist\'. 449. <a href="datasets/Water+Treatment+Plant">Water Treatment Plant</a>: Multiple classes predict plant state 450. <a href="datasets/Waveform+Database+Generator+%28Version+1%29">Waveform Database Generator (Version 1)</a>: CART book\'s waveform domains 451. <a href="datasets/Waveform+Database+Generator+%28Version+2%29">Waveform Database Generator (Version 2)</a>: CART book\'s waveform domains 452. <a href="datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29">Wearable Computing: Classification of Body Postures and Movements (PUC-Rio)</a>: A dataset with 5 classes (sitting-down, standing-up, standing, walking, and sitting) collected on 8 hours of activities of 4 healthy subjects. We also established a baseline performance index. 453. <a href="datasets/Website+Phishing">Website Phishing</a>: \r\n\r\n 454. <a href="datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units">Weight Lifting Exercises monitored with Inertial Measurement Units</a>: Six young health subjects were asked to perform 5 variations of the biceps curl weight lifting exercise. One of the variations is the one predicted by the health professional. 455. <a href="datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29">WESAD (Wearable Stress and Affect Detection)</a>: WESAD (Wearable Stress and Affect Detection) contains data of 15 subjects during a stress-affect lab study, while wearing physiological and motion sensors. 456. <a href="datasets/Wholesale+customers">Wholesale customers</a>: The data set refers to clients of a wholesale distributor. It includes the annual spending in monetary units (m.u.) on diverse product categories 457. <a href="datasets/wiki4HE">wiki4HE</a>: Survey of faculty members from two Spanish universities on teaching uses of Wikipedia 458. <a href="datasets/Wilt">Wilt</a>: High-resolution Remote Sensing data set (Quickbird). Small number of training samples of diseased trees, large number for other land cover. Testing data set from stratified random sample of image. 459. <a href="datasets/Wine">Wine</a>: Using chemical analysis determine the origin of wines 460. <a href="datasets/Wine+Quality">Wine Quality</a>: Two datasets are included, related to red and white vinho verde wine samples, from the north of Portugal. The goal is to model wine quality based on physicochemical tests (see [Cortez et al., 2009], http://www3.dsi.uminho.pt/pcortez/wine/). 461. <a href="datasets/Wireless+Indoor+Localization">Wireless Indoor Localization</a>: Collected in indoor space by observing signal strengths of seven WiFi signals visible on a smartphone. The decision variable is one of the four rooms. 462. <a href="datasets/Yacht+Hydrodynamics">Yacht Hydrodynamics</a>: Delft data set, used to predict the hydodynamic performance of sailing yachts from dimensions and velocity. 463. <a href="datasets/YearPredictionMSD">YearPredictionMSD</a>: Prediction of the release year of a song from audio features. Songs are mostly western, commercial tracks ranging from 1922 to 2011, with a peak in the year 2000s. 464. <a href="datasets/Yeast">Yeast</a>: Predicting the Cellular Localization Sites of Proteins 465. <a href="datasets/YouTube+Comedy+Slam+Preference+Data">YouTube Comedy Slam Preference Data</a>: This dataset provides user vote data on which video from a pair of videos is funnier collected on YouTube Comedy Slam. The task is to automatically predict this preference based on video metadata. 466. <a href="datasets/YouTube+Multiview+Video+Games+Dataset">YouTube Multiview Video Games Dataset</a>: This dataset contains about 120k instances, each described by 13 feature types, with class information, specially useful for exploring multiview topics (cotraining, ensembles, clustering,..). 467. <a href="datasets/YouTube+Spam+Collection">YouTube Spam Collection</a>: It is a public set of comments collected for spam research. It has five datasets composed by 1,956 real messages extracted from five videos that were among the 10 most viewed on the collection period. 468. <a href="datasets/Z-Alizadeh+Sani">Z-Alizadeh Sani</a>: It was collected for CAD diagnosis. 469. <a href="datasets/Zoo">Zoo</a>: Artificial, 7 classes of animals </tr></td></table><hr>\n\t</td>\n</tr>\n</table>\n\n\t</td>\n</tr>\n</table>\n\n\n\n<table cellpadding=5 align=center><tr valign=center>\n\t\t<td> Supported By: </td>\n <td><img src="assets/nsfe.gif" height=60 /> </td>\n <td> In Collaboration With: </td>\n <td><img src="assets/rexaSmall.jpg" /></td>\n</tr></table>\n\n \n\n<a href="about.html">About</a> || \n<a href="citation_policy.html">Citation Policy</a> || \n<a href="donation_policy.html">Donation Policy</a> || \n<a href="contact.html">Contact</a> ||&n

bsp;\n<a href="http://cml.ics.uci.edu">CML</a>\n\n \n\n</body>\n</html>\n'

In [173]:

```
soup_2 = BeautifulSoup(html_2)
soup_2
```

Out[173]:

```
<!DOCTYPE HTML>
<html><body><p>"-//W3C//DTD HTML 4.01 Transitional//EN\"&gt;

</p>
<title>UCI Machine Learning Repository: Data Sets</title>
<!-- Stylesheet link -->
<link href="assets/ml.css" rel="stylesheet" type="text/css"/>
<script language="JavaScript" type="text/javascript">
<!--
function checkform ( form )
{
  // see http://www.thesitewizard.com/archive/validation.shtml
  // for an explanation of this script and how to use it on your
  // own website

  // ** START **
  if (form.q.value == "")
  {
    alert( "Please enter search terms." );
    form.q.focus();
    return false ;
  }

  if (getCheckedValue(form.sitesearch) == "ics.uci.edu" && form.q.value.indexOf("site:archive.ics.uci.edu/ml") ==
-1)
  {
    form.q.value = form.q.value + " site:archive.ics.uci.edu/ml";
  }

  // ** END **
  return true ;
}

// return the value of the radio button that is checked
// return an empty string if none are checked, or
// there are no radio buttons
function getCheckedValue(radioObj) {
 if(!radioObj)
  return "";
 var radioLength = radioObj.length;
 if(radioLength == undefined)
  if(radioObj.checked)
   return radioObj.value;
  else
   return "";
 for(var i = 0; i < radioLength; i++) {
  if(radioObj[i].checked) {
   return radioObj[i].value;
  }
 }
 return "";
}
```

```
//-->
</script>
<!-- SITE HEADER (INCLUDES LOGO AND SEARCH BOX) -->
<!-- SITE HEADER (INCLUDES LOGO AND SEARCH BOX) -->
<table bgcolor="#003366" width="100%">
<tr>
<td>
<span class="normal"><a alt="Home" href="index.html"><img border="0" src="assets/logo.gif"/></a><br/>
<a href="http://cml.ics.uci.edu"><font color="FFDD33">Center for Machine Learning and Intelligent Systems</fo
nt></a></span>
</td>
<td align="right" valign="top" width="100%">
<span class="whitetext">
<a href="about.html">About</a>
  <a href="citation_policy.html">Citation Policy</a>
  <a href="donation_policy.html">Donate a Data Set</a>
  <a href="contact.html">Contact</a>
</span>
<br/>
<br/>
<!-- Search Google -->
<form action="http://www.google.com/custom" method="GET" onsubmit="return checkform(this);">
<input maxlength="255" name="q" size="30" type="text" value=""/>
<input name="sa" type="submit" value="Search"/>
<input name="cof" type="hidden" value="AH:center;LH:130;L:http://archive.ics.uci.edu/assets/logo.gif;LW:384;A
WFID:869c0b2eaa8d518e;"/>
<input name="domains" type="hidden" value="ics.uci.edu"/>
<br/>
<input checked="" name="sitesearch" type="radio" value="ics.uci.edu"/> <span class="whitetext"><font size="1"
>Repository</font></span>
<input name="sitesearch" type="radio" value=""/> <span class="whitetext"><font size="1">Web</font></span>

  <a href="http://www.google.com/search"><img align="middle" alt="Google" border="0" height="27" src="http://
www.google.com/logos/Logo_25blk.gif"/></a>
<br/>
</form>
<!-- Search Google -->
<span class="whitetext"><a href="datasets.php"><font color="#FFDD33" size="3"><b>View ALL Data Sets</b><
/font></a></span>
<br/>
</td>
</tr>
</table>
<br/>
<table cellpadding="3">
<tr>
<td valign="top">
<table><tr><td><p align="center" class="big">Browse Through:</p></td></tr></table>
<table border="2" cellpadding="2" width="150">
<tr> <td bgcolor="#003366"><p class="whitetext"><b>Default Task</b> </p>
</td>
</tr>
<tr>
<td valign="top"><p class="normal"><a href="datasets.php?format=&amp;task=cla&amp;att=&amp;area=&amp;
numAtt=&amp;numIns=&amp;type=&amp;sort=nameUp&amp;view=list">Classification</a> <font color="red">(3
50)</font><br/><a href="datasets.php?format=&amp;task=reg&amp;att=&amp;area=&amp;numAtt=&amp;numI
ns=&amp;type=&amp;sort=nameUp&amp;view=list">Regression</a> <font color="red">(96)</font><br/><a href
="datasets.php?format=&amp;task=clu&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=&amp;s
ort=nameUp&amp;view=list">Clustering</a> <font color="red">(84)</font><br/><a href="datasets.php?format=&
```

ort=nameUp&view=list">Clustering</a> <font color="red">(84)</font>
<a href="datasets.php?format=&task=other&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list">Other</a> <font color="red">(55)</font>
</td>
</tr>
<tr> <td bgcolor="#003366"> Attribute Type
</td>
</tr>
<tr>
<td valign="top"> <a href="datasets.php?format=&task=&att=cat&area=&numAtt=&numIns=&type=&sort=nameUp&view=list">Categorical</a> <font color="red">(38)</font>
<a href="datasets.php?format=&task=&att=num&area=&numAtt=&numIns=&type=&sort=nameUp&view=list">Numerical</a> <font color="red">(307)</font>
<a href="datasets.php?format=&task=&att=mix&area=&numAtt=&numIns=&type=&sort=nameUp&view=list">Mixed</a> <font color="red">(55)</font>
</td>
</tr>
<tr> <td bgcolor="#003366"> Data Type
</td>
</tr>
<tr>
<td valign="top"> <a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=mvar&sort=nameUp&view=list">Multivariate</a> <font color="red">(357)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=uvar&sort=nameUp&view=list">Univariate</a> <font color="red">(23)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=seq&sort=nameUp&view=list">Sequential</a> <font color="red">(47)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=ts&sort=nameUp&view=list">Time-Series</a> <font color="red">(91)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=text&sort=nameUp&view=list">Text</a> <font color="red">(53)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=dt&sort=nameUp&view=list">Domain-Theory</a> <font color="red">(23)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=other&sort=nameUp&view=list">Other</a> <font color="red">(21)</font>
</td>
</tr>
<tr><td bgcolor="#003366"> Area
</td>
</tr>
<tr>
<td valign="top"> <a href="datasets.php?format=&task=&att=&area=life&numAtt=&numIns=&type=&sort=nameUp&view=list">Life Sciences</a> <font color="red">(107)</font>
<a href="datasets.php?format=&task=&att=&area=phys&numAtt=&numIns=&type=&sort=nameUp&view=list">Physical Sciences</a> <font color="red">(49)</font>
<a href="datasets.php?format=&task=&att=&area=comp&numAtt=&numIns=&type=&sort=nameUp&view=list">CS / Engineering</a> <font color="red">(170)</font>
<a href="datasets.php?format=&task=&att=&area=soc&numAtt=&numIns=&type=&sort=nameUp&view=list">Social Sciences</a> <font color="red">(26)</font>
<a href="datasets.php?format=&task=&att=&area=bus&numAtt=&numIns=&type=&sort=nameUp&view=list">Business</a> <font color="red">(29)</font>
<a href="datasets.php?format=&task=&att=&area=game&numAtt=&numIns=&type=&sort=nameUp&view=list">Game</a> <font color="red">(10)</font>
<a href="datasets.php?format=&task=&att=&area=other&numAtt=&numIns=&type=&sort=nameUp&view=list">Other</a> <font color="red">(73)</font>

</td>
</tr>
<tr><td bgcolor="#003366"> # Attributes
</td>
</tr>
<tr>
<td valign="top"> <a href="datasets.php?format=&task=&att=&area=&numAtt=less10&numIns=&type=&sort=nameUp&view=list">Less than 10</a> <font color="red"

>(113)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=10to100&numIns=&type=&sort=nameUp&view=list">10 to 100</a> <font color="red">(210)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=greater100&numIns=&type=&sort=nameUp&view=list">Greater than 100</a> <font color="red">(84)</font>
</td>
</tr>
<tr><td bgcolor="#003366"> # Instances </td>
</tr>
<tr>
<td valign="top">
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=less100&type=&sort=nameUp&view=list">Less than 100</a> <font color="red">(27)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=100to1000&type=&sort=nameUp&view=list">100 to 1000</a> <font color="red">(162)</font>
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=greater1000&type=&sort=nameUp&view=list">Greater than 1000</a> <font color="red">(246)</font>
</td>
</tr>
<tr><td bgcolor="#003366"> Format Type
</td>
</tr>
<tr>
<td valign="top"> <a href="datasets.php?format=mat&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list">Matrix</a> <font color="red">(324)</font>
<a href="datasets.php?format=nonmat&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list">Non-Matrix</a> <font color="red">(145)</font>
</td>
</tr></table>
</td>
<td valign="top">
<table width="100%">
<tr>
<td> 469 Data Sets </td>
<td align="right"> <a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table">Table View</a> <font color="gray">List View</font> </td>
</tr>
</table><table cellpadding="3"><tr><td><hr/> 1. <a href="datasets/2.4+GHZ+Indoor+Channel+Measurements">2.4 GHZ Indoor Channel Measurements</a>: Measurement of the S21,consists of 10 sweeps, each sweep contains 601 frequency points with spacing of 0.167MHz to cover a 100MHz band centered at 2.4GHz. 2. <a href="datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29">3D Road Network (North Jutland, Denmark)</a>: 3D road network with highly accurate elevation information (+-20cm) from Denmark used in eco-routing and fuel/Co2-estimation routing algorithms. 3. <a href="datasets/AAAI+2013+Accepted+Papers">AAAI 2013 Accepted Papers</a>: This data set compromises the metadata for the 2013 AAAI conference's accepted papers (main track only), including paper titles, abstracts, and keywords of varying granularity. 4. <a href="datasets/AAAI+2014+Accepted+Papers">AAAI 2014 Accepted Papers</a>: This data set compromises the metadata for the 2014 AAAI conference's accepted papers, including paper titles, authors, abstracts, and keywords of varying granularity. 5. <a href="datasets/Abalone">Abalone</a>: Predict the age of abalone from physical measurements 6. <a href="datasets/Abscisic+Acid+Signaling+Network">Abscisic Acid Signaling Network</a>: The objective is to determine the set of boolean rules that describe the interactions of the nodes within this plant signaling network. The dataset includes 300 separate boolean pseudodynamic simulations using an asynchronous update scheme. 7. <a href="datasets/Absenteeism+at+work">Absenteeism at work</a>: The database was created with records of absenteeism at work from July 2007 to July 2010 at a courier company in Brazil. 8. <a href="datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors">Activities of Daily Living (ADLs) Recognition Using Binary Sensors</a>: This dataset comprises information regarding the ADLs performed by two users on a daily basis in their
own homes. 9. <a href="datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer">Activity Recognition from Single Chest-Mounted Accelerometer</a>: The dataset collects data from a wearable accelerometer mounted on the chest. The dataset is intended for Activity Recognition research purposes. 10. <a href="datasets/Activity+Recognition+system+based+on+Mu

ltisensor+data+fusion+%28AReM%29">Activity Recognition system based on Multisensor data fusion (AReM)</a>: This dataset contains temporal data from a Wireless Sensor Network worn by an actor performing the activities: bending, cycling, lying down, sitting, standing, walking. 11. <a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor">Activity recognition with healthy older people using a batteryless wearable sensor</a>: Sequential motion data from 14 healthy older people aged 66 to 86 years old using a batteryless, wearable sensor on top of their clothing for the recognition of activities in clinical environments. 12. <a href="datasets/Acute+Inflammations">Acute Inflammations</a>: The data was created by a medical expert as a data set to test the expert system,
which will perform the presumptive diagnosis of two diseases of the urinary system.
13. <a href="datasets/Adult">Adult</a>: Predict whether income exceeds $50K/yr based on census data. Also known as "Census Income" dataset. 14. <a href="datasets/Air+Quality">Air Quality</a>: Contains the responses of a gas multisensor device deployed on the field in an Italian city. Hourly responses averages are recorded along with gas concentrations references from a certified analyzer. 15. <a href="datasets/Air+quality">Air quality</a>: Contains the responses of a gas multisensor device deployed on the field in an Italian city. 16. <a href="datasets/Airfoil+Self-Noise">Airfoil Self-Noise</a>: NASA data set, obtained from a series of aerodynamic and acoustic tests of two and three-dimensional airfoil blade sections conducted in an anechoic wind tunnel. 17. <a href="datasets/Amazon+Access+Samples">Amazon Access Samples</a>: Amazon's InfoSec is getting smarter about the way Access data is leveraged. This is an anonymized sample of access provisioned within the company. 18. <a href="datasets/Amazon+Commerce+reviews+set">Amazon Commerce reviews set</a>: The dataset is used for authorship identification in online Writeprint which is a new research field of pattern recognition. 19. <a href="datasets/Annealing">Annealing</a>: Steel annealing data 20. <a href="datasets/Anonymous+Microsoft+Web+Data">Anonymous Microsoft Web Data</a>: Log of anonymous users of www.microsoft.com; predict areas of the web site a user visited based on data on other areas the user visited. 21. <a href="datasets/Anuran+Calls+%28MFCCs%29">Anuran Calls (MFCCs)</a>: Acoustic features extracted from syllables of anuran (frogs) calls, including the family, the genus, and the species labels (multilabel). 22. <a href="datasets/Appliances+energy+prediction">Appliances energy prediction</a>: Experimental data used to create regression models of appliances energy use in a low energy building. 23. <a href="datasets/APS+Failure+at+Scania+Trucks">APS Failure at Scania Trucks</a>: The datasets' positive class consists of component failures for a specific component of the APS system. The negative class consists of trucks with failures for components not related to the APS. 24. <a href="datasets/Arcene">Arcene</a>: ARCENE's task is to distinguish cancer versus normal patterns from mass-spectrometric data. This is a two-class classification problem with continuous input variables. This dataset is one of 5 datasets of the NIPS 2003 feature selection challenge. 25. <a href="datasets/Arrhythmia">Arrhythmia</a>: Distinguish between the presence and absence of cardiac arrhythmia and classify it in one of the 16 groups. 26. <a href="datasets/Artificial+Characters">Artificial Characters</a>: Dataset artificially generated by using first order theory which describes structure of ten capital letters of English alphabet 27. <a href="datasets/Audiology+%28Original%29">Audiology (Original)</a>: Nominal audiology dataset from Baylor 28. <a href="datasets/Audiology+%28Standardized%29">Audiology (Standardized)</a>: Standardized version of the original audiology database 29. <a href="datasets/Audit+Data">Audit Data</a>: Exhaustive one year non-confidential data in the year 2015 to 2016 of firms is collected from the Auditor Office of India to build a predictor for classifying suspicious firms. 30. <a href="datasets/Australian+Sign+Language+signs">Australian Sign Language signs</a>: This data consists of sample of Auslan (Australian Sign Language) signs. Examples of 95 signs were collected from five signers with a total of 6650 sign samples. 31. <a href="datasets/Australian+Sign+Language+signs+%28High+Quality%29">Australian Sign Language signs (High Quality)</a>: This data consists of sample of Auslan (Australian Sign Language) signs. 27 examples of each of 95 Auslan signs were captured from a native signer using high-quality position trackers 32. <a href="datasets/Autism+Screening+Adult">Autism Screening Adult</a>: Autistic Spectrum Disorder Screening Data for Adult. This dataset is related to classification and predictive tasks. 33. <a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++">Autistic Spectrum Disorder Screening Data for Adolescent </a>: Autistic Spectrum Disorder Screening Data for Adolescent. This dataset is related to classification and predictive tasks. 34. <a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++">Autistic Spectrum Disorder Screening Data for Children </a>: Childrn screening data for autism suitable for classification and predictive tasks 35. <a href="datasets/Auto+MPG">Auto MPG</a>: Revised from CMU StatLib library, data concerns city-cycle fuel consumption 36. <a href="datasets/Automobile">Automobile</a>: From 1985 Ward's Automotive Yearbook 37. <a href="datasets/AutoUniv">AutoUniv</a>: AutoUniv is an advanced data generator for classifications tasks. The aim is to reflect the nuances and heterogeneit

y of real data. Data can be generated in .csv, ARFF or C4.5 formats. 38. <a href="datasets/Avila">Avila</a>: The Avila data set has been extracted from 800 images of the 'Avila Bible', an XII century giant Latin copy of the Bible. The prediction task consists in associating each pattern to a copyist. 39. <a href="datasets/Bach+Choral+Harmony">Bach Choral Harmony</a>: The data set is composed of 60 chorales (5665 events) by J.S. Bach (1675-1750).
Each event of each chorale is labelled using 1 among 101 chord labels and described
through 14 features. 40. <a href="datasets/Bach+Chorales">Bach Chorales</a>: Time-series data based on chorales; challenge is to learn generative grammar; data in Lisp 41. <a href="datasets/Badges">Badges</a>: Badges labeled with a "+" or "-" as a function of a person's name 42. <a href="datasets/Bag+of+Words">Bag of Words</a>: This data set contains five text collections in the form of bags-of-words. 43. <a href="datasets/Balance+Scale">Balance Scale</a>: Balance scale weight & distance database 44. <a href="datasets/Balloons">Balloons</a>: Data previously used in cognitive psychology experiment; 4 data sets represent different conditions of an experiment 45. <a href="datasets/Bank+Marketing">Bank Marketing</a>: The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe a term deposit (variable y). 46. <a href="datasets/banknote+authentication">banknote authentication</a>: Data were extracted from images that were taken for the evaluation of an authentication procedure for banknotes. 47. <a href="datasets/BAUM-1">BAUM-1</a>: BAUM-1 dataset contains 1184 multimodal facial video clips collected from 31 subjects. The 1184 video clips contain spontaneous facial expressions and speech of 13 emotional and mental states. 48. <a href="datasets/BAUM-2">BAUM-2</a>: A multilingual audio-visual affective face database consisting of 1047 video clips of 286 subjects. 49. <a href="datasets/Behavior+of+the+urban+traffic+of+the+city+of+Sao+Paulo+in+Brazil">Behavior of the urban traffic of the city of Sao Paulo in Brazil</a>: The database was created with records of behavior of the urban traffic of the city of Sao Paulo in Brazil. 50. <a href="datasets/Beijing+PM2.5+Data">Beijing PM2.5 Data</a>: This hourly data set contains the PM2.5 data of US Embassy in Beijing. Meanwhile, meteorological data from Beijing Capital International Airport are also included. 51. <a href="datasets/Bike+Sharing+Dataset">Bike Sharing Dataset</a>: This dataset contains the hourly and daily count of rental bikes between years 2011 and 2012 in Capital bikeshare system with the corresponding weather and seasonal information. 52. <a href="datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation">BLE RSSI Dataset for Indoor localization and Navigation</a>: This dataset contains RSSI readings gathered from an array of Bluetooth Low Energy (BLE) iBeacons in a real-world and operational indoor environment for localization and navigation purposes. 53. <a href="datasets/BlogFeedback">BlogFeedback</a>: Instances in this dataset contain features extracted from blog posts. The task associated with the data is to predict how many comments the post will receive. 54. <a href="datasets/BLOGGER">BLOGGER</a>: In this paper, we look for to recognize the causes of users tend
to cyber space in Kohkiloye and Boyer Ahmad Province in
Iran 55. <a href="datasets/Blood+Transfusion+Service+Center">Blood Transfusion Service Center</a>: Data taken from the Blood Transfusion Service Center in Hsin-Chu City in Taiwan -- this is a classification problem. 56. <a href="datasets/Breast+Cancer">Breast Cancer</a>: Breast Cancer Data (Restricted Access) 57. <a href="datasets/Breast+Cancer+Coimbra">Breast Cancer Coimbra</a>: Clinical features were observed or measured for 64 patients with breast cancer and 52 healthy controls. 58. <a href="datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29">Breast Cancer Wisconsin (Diagnostic)</a>: Diagnostic Wisconsin Breast Cancer Database 59. <a href="datasets/Breast+Cancer+Wisconsin+%28Original%29">Breast Cancer Wisconsin (Original)</a>: Original Wisconsin Breast Cancer Database 60. <a href="datasets/Breast+Cancer+Wisconsin+%28Prognostic%29">Breast Cancer Wisconsin (Prognostic)</a>: Prognostic Wisconsin Breast Cancer Database 61. <a href="datasets/Breast+Tissue">Breast Tissue</a>: Dataset with electrical impedance measurements of freshly excised tissue samples from the breast. 62. <a href="datasets/BuddyMove+Data+Set">BuddyMove Data Set</a>: User interest information extracted from user reviews published in holidayiq.com about various types of point of interests in South India 63. <a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%28OBS%29+Network">Burst Header Packet (BHP) flooding attack on Optical Burst Switching (OBS) Network</a>: One of the primary challenges in identifying the risks of the Burst Header Packet (BHP) flood attacks in Optical Burst Switching networks (OBS) is the scarcity of reliable historical data. 64. <a href="datasets/Buzz+in+social+media+">Buzz in social media </a>: This data-set contains examples of buzz events from two different social networks: Twitter, and Tom's Hardware, a forum network focusing on new technology with more conservative dynamics. 65. <a href="datasets/Caesarian+Section+Classification+Dataset">Caesarian Section Classification Dataset</a>: This dataset contains information about caesarian section results of 80 pregnant women with the most important characteristics of delivery problems in the medical field. <p class="normal

">66. <a href="datasets/CalIt2+Building+People+Counts">CalIt2 Building People Counts</a>: This data comes from the main door of the CalIt2 building at UCI. 67. <a href="datasets/Car+Evaluation">Car Evaluation</a>: Derived from simple hierarchical decision model, this database may be useful for testing constructive induction and structure discovery methods. 68. <a href="datasets/Carbon+Nanotubes">Carbon Nanotubes</a>: This dataset contains 10721 initial and calculated atomic coordinates of carbon nanotubes. 69. <a href="datasets/Cardiotocography">Cardiotocography</a>: The dataset consists of measurements of fetal heart rate (FHR) and uterine contraction (UC) features on cardiotocograms classified by expert obstetricians. 70. <a href="datasets/Cargo+2000+Freight+Tracking+and+Tracing">Cargo 2000 Freight Tracking and Tracing</a>: Sanitized and anonymized Cargo 2000 (C2K) airfreight tracking and tracing events, covering five months of business execution (3,942 process instances, 7,932 transport legs, 56,082 activities). 71. <a href="datasets/Census+Income">Census Income</a>: Predict whether income exceeds $50K/yr based on census data. Also known as "Adult" dataset. 72. <a href="datasets/Census-Income+%28KDD%29">Census-Income (KDD)</a>: This data set contains weighted census data extracted from the 1994 and 1995 current population surveys conducted by the U.S. Census Bureau. 73. <a href="datasets/Cervical+cancer+%28Risk+Factors%29">Cervical cancer (Risk Factors)</a>: This dataset focuses on the prediction of indicators/diagnosis of cervical cancer. The features cover demographic information, habits, and historic medical records. 74. <a href="datasets/Challenger+USA+Space+Shuttle+O-Ring">Challenger USA Space Shuttle O-Ring</a>: Task: predict the number of O-rings that experience thermal distress on a flight at 31 degrees F given data on the previous 23 shuttle flights 75. <a href="datasets/Character+Font+Images">Character Font Images</a>: Character images from scanned and computer generated fonts. 76. <a href="datasets/Character+Trajectories">Character Trajectories</a>: Multiple, labelled samples of pen tip trajectories recorded whilst writing individual characters. All samples are from the same writer, for the purposes of primitive extraction. Only characters with a single pen-down segment were considered. 77. <a href="datasets/Chess+%28Domain+Theories%29">Chess (Domain Theories)</a>: 6 different domain theories for generating legal moves of chess 78. <a href="datasets/Chess+%28King-Rook+vs.+King%29">Chess (King-Rook vs. King)</a>: Chess Endgame Database for White King and Rook against Black King (KRK). 79. <a href="datasets/Chess+%28King-Rook+vs.+King-Knight%29">Chess (King-Rook vs. King-Knight)</a>: Knight Pin Chess End-Game Database Creator 80. <a href="datasets/Chess+%28King-Rook+vs.+King-Pawn%29">Chess (King-Rook vs. King-Pawn)</a>: King+Rook versus King+Pawn on a7 (usually abbreviated KRKPA7). 81. <a href="datasets/chestnut+%E2%80%93+LARVIC">chestnut – LARVIC</a>: The research project presents this database, shows the images of chestnuts that will be processed to determine the presence or absence of defects 82. <a href="datasets/chipseq">chipseq</a>: ChIP-seq experiments characterize protein modifications or binding at
specific genomic locations in specific samples. The machine learning
problem in these data is structured binary classification. 83. <a href="datasets/Chronic_Kidney_Disease">Chronic_Kidney_Disease</a>: This dataset can be used to predict the chronic kidney disease and it can be collected from the hospital nearly 2 months of period. 84. <a href="datasets/Climate+Model+Simulation+Crashes">Climate Model Simulation Crashes</a>: Given Latin hypercube samples of 18 climate model input parameter values, predict climate model simulation crashes and determine the parameter value combinations that cause the failures. 85. <a href="datasets/Cloud">Cloud</a>: Little Documentation 86. <a href="datasets/CMU+Face+Images">CMU Face Images</a>: This data consists of 640 black and white face images of people taken with varying pose (straight, left, right, up), expression (neutral, happy, sad, angry), eyes (wearing sunglasses or not), and size 87. <a href="datasets/CNAE-9">CNAE-9</a>: This is a data set containing 1080 documents of free text business descriptions of Brazilian companies categorized into a
subset of 9 categories 88. <a href="datasets/Coil+1999+Competition+Data">Coil 1999 Competition Data</a>: This data set is from the 1999 Computational Intelligence and Learning (COIL) competition. The data contains measurements of river chemical concentrations and algae densities. 89. <a href="datasets/Combined+Cycle+Power+Plant">Combined Cycle Power Plant</a>: The dataset contains 9568 data points collected from a Combined Cycle Power Plant over 6 years (2006-2011), when the plant was set to work with full load. 90. <a href="datasets/Communities+and+Crime">Communities and Crime</a>: Communities within the United States. The data combines socio-economic data from the 1990 US Census, law enforcement data from the 1990 US LEMAS survey, and crime data from the 1995 FBI UCR. 91. <a href="datasets/Communities+and+Crime+Unnormalized">Communities and Crime Unnormalized</a>: Communities in the US. Data combines socio-economic data from the '90 Census, law enforcement data from the 1990 Law Enforcement Management and Admin Stats survey, and crime data from the 1995 FBI UCR 92. <a href="datasets/Computer+Hardware">Computer Hardware</a>: Relative CPU Performance Data, described in terms of its cycle time, memory size, etc. 93. <a href="datasets/Concrete+Compressive+Strength">Concr

ete Compressive Strength</a>: Concrete is the most important material in civil engineering. The concrete c ompressive strength is a highly nonlinear function of age and ingredients. 94. <a hr ef="datasets/Concrete+Slump+Test">Concrete Slump Test</a>: Concrete is a highly complex material. Th e slump flow of concrete is not only determined by the water content, but that is also influenced by other concre te ingredients. 95. <a href="datasets/Condition+Based+Maintenance+of+Naval+Prop ulsion+Plants">Condition Based Maintenance of Naval Propulsion Plants</a>: Data have been generated fr om a sophisticated simulator of a Gas Turbines (GT), mounted on a Frigate characterized by a COmbined Dies el eLectric And Gas (CODLAG) propulsion plant type. 96. <a href="datasets/Conditio n+monitoring+of+hydraulic+systems">Condition monitoring of hydraulic systems</a>: The data set address es the condition assessment of a hydraulic test rig based on multi sensor data. Four fault types are superimpos ed with several severity grades impeding selective quantification. 97. <a href="datas ets/Congressional+Voting+Records">Congressional Voting Records</a>: 1984 United Stated Congression al Voting Records; Classify as Republican or Democrat 98. <a href="datasets/Conne ct-4">Connect-4</a>: Contains connect-4 positions 99. <a href="datasets/Conn ectionist+Bench+%28Nettalk+Corpus%29">Connectionist Bench (Nettalk Corpus)</a>: The file "nettalk.dat a" contains a list of 20,008 English words, along with a phonetic transcription for each word. The task is to train a network to produce the proper phonemes 100. <a href="datasets/Connectionist+B ench+%28Sonar%2C+Mines+vs.+Rocks%29">Connectionist Bench (Sonar, Mines vs. Rocks)</a>: The tas k is to train a network to discriminate between sonar signals bounced off a metal cylinder and those bounced of f a roughly cylindrical rock. 101. <a href="datasets/Connectionist+Bench+%28Vowel +Recognition+-+Deterding+Data%29">Connectionist Bench (Vowel Recognition - Deterding Data)</a>: Spe aker independent recognition of the eleven steady state vowels of British English using a specified training set o f lpc derived log area ratios. 102. <a href="datasets/Container+Crane+Controller+Da ta+Set">Container Crane Controller Data Set</a>: A container crane has the function of transporting contai ners from one point to another point. 103. <a href="datasets/Contraceptive+Method +Choice">Contraceptive Method Choice</a>: Dataset is a subset of the 1987 National Indonesia Contrace ptive Prevalence Survey. 104. <a href="datasets/Corel+Image+Features">Corel Ima ge Features</a>: This dataset contains image features extracted from a Corel image collection. Four sets of features are available based on the color histogram, color histogram layout, color moments, and co-occuren ce 105. <a href="datasets/Covertype">Covertype</a>: Forest CoverType datas et 106. <a href="datasets/Credit+Approval">Credit Approval</a>: This data conc erns credit card applications; good mix of attributes 107. <a href="datasets/Crowdso urced+Mapping">Crowdsourced Mapping</a>: Crowdsourced data from OpenStreetMap is used to autom ate the classification of satellite images into different land cover classes (impervious, farm, forest, grass, orchar d, water). 108. <a href="datasets/Cryotherapy+Dataset+">Cryotherapy Dataset </a> : This dataset contains information about wart treatment results of 90 patients using cryotherapy. <p cla ss="normal">109. <a href="datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+201 4+and+2015">CSM (Conventional and Social Media Movies) Dataset 2014 and 2015</a>: 12 features cate gorized as conventional and social media features. Both conventional features, collected from movies database s on Web as well as social media features(YouTube,Twitter). 110. <a href="datasets /Cuff-Less+Blood+Pressure+Estimation">Cuff-Less Blood Pressure Estimation</a>: This Data set provides preprocessed and cleaned vital signals which can be used in designing algorithms for cuff-less estimation of the blood pressure. 111. <a href="datasets/Cylinder+Bands">Cylinder Bands</a>: U sed in decision tree induction for mitigating process delays known as "cylinder bands" in rotogravure printing 112. <a href="datasets/Daily+and+Sports+Activities">Daily and Sports Activities</a>: The dataset comprises motion sensor data of 19 daily and sports activities each performed by 8 subjects in t heir own style for 5 minutes. Five Xsens MTx units are used on the torso, arms, and legs.
113. <a href="datasets/Daily+Demand+Forecasting+Orders">Daily Demand Foreca sting Orders</a>: The dataset was collected during 60 days, this is a real database of a brazilian logistics c ompany. 114. <a href="datasets/Daphnet+Freezing+of+Gait">Daphnet Freezing of Gait</a>: This dataset contains the annotated readings of 3 acceleration sensors at the hip and leg of Parki nson's disease patients that experience freezing of gait (FoG) during walking tasks.
115. <a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in +Education+Setting">Data for Software Engineering Teamwork Assessment in Education Setting</a>: Data include over 100 Team Activity Measures and outcomes (ML classes) obtained from activities of 74 student tea ms during the creation of final class project in SW Eng. classes at SFSU, Fulda, FAU 11 6. <a href="datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer">Dataset for ADL Rec ognition with Wrist-worn Accelerometer</a>: Recordings of 16 volunteers performing 14 Activities of Daily L iving (ADL) while carrying a single wrist-worn tri-axial accelerometer. 117. <a href="d atasets/Dataset+for+Sensorless+Drive+Diagnosis">Dataset for Sensorless Drive Diagnosis</a>: Features are extracted from motor current. The motor has intact and defective components. This results in 11 different cl asses with different conditions. 118. <a href="datasets/DBWorld+e-mails">DBWorld

e-mails</a>: It contains 64 e-mails which I have manually collected from DBWorld mailing list. They are cla ssified in: 'announces of conferences' and 'everything else'. 119. <a href="datasets/d efault+of+credit+card+clients">default of credit card clients</a>: This research aimed at the case of custom ers' default payments in Taiwan and compares the predictive accuracy of probability of default among six data mining methods. 120. <a href="datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-La bel+Multi-Instance+Learning+with+Instance+Labels">DeliciousMIL: A Data Set for Multi-Label Multi-Instance Le arning with Instance Labels</a>: This dataset includes 1) 12234 documents (8251 training, 3983 test) extr acted from DeliciousT140 dataset, 2) class labels for all documents, 3) labels for a subset of sentences of the t est documents. 121. <a href="datasets/Demospongiae">Demospongiae</a>: M arine sponges of the Demospongiae class classification domain. 122. <a href="datas ets/Dermatology">Dermatology</a>: Aim for this dataset is to determine the type of Eryhemato-Squamous Disease. 123. <a href="datasets/Detect+Malacious+Executable%28AntiVirus%29">D etect Malacious Executable(AntiVirus)</a>: I extract features from malacious and non-malacious and creat e and training dataset to teach svm classifier.Dataset made of unknown executable to detect if it is virus or nor mal safe executable. 124. <a href="datasets/detection_of_IoT_botnet_attacks_N_Ba IoT">detection_of_IoT_botnet_attacks_N_BaIoT</a>: This dataset addresses the lack of public botnet data sets, especially for the IoT. It suggests *real* traffic data, gathered from 9 commercial IoT devices authentically infected by Mirai and BASHLITE. 125. <a href="datasets/Devanagari+Handwritten+C haracter+Dataset">Devanagari Handwritten Character Dataset</a>: This is an image database of Handwrit ten Devanagari characters. There are 46 classes of characters with 2000 examples each. The dataset is split in to training set(85%) and testing set(15%). 126. <a href="datasets/Dexter">Dexter</ a>: DEXTER is a text classification problem in a bag-of-word representation. This is a two-class classificatio n problem with sparse continuous input variables. This dataset is one of five datasets of the NIPS 2003 feature selection challenge.
127. <a href="datasets/DGP2+-+The+Second+Data+Generation+Program">DGP2 - The Second Data Generation Program</a>: Generates application domains based on specific parameters, number of features, and proportion of positive to negative examples 128. <a href="d atasets/Diabetes">Diabetes</a>: This diabetes dataset is from AIM '94 129. <a href="datasets/Diabetes+130-US+hospitals+for+years+1999-2008">Diabetes 130-US hospitals for years 1999-2008</a>: This data has been prepared to analyze factors related to readmission as well as other

outcomes pertaining to patients with diabetes. 130. <a href="datasets/Diabetic+Retin opathy+Debrecen+Data+Set">Diabetic Retinopathy Debrecen Data Set</a>: This dataset contains features extracted from the Messidor image set to predict whether an image contains signs of diabetic retinopathy or not . 131. <a href="datasets/Discrete+Tone+Image+Dataset">Discrete Tone Image Dat aset</a>: Discrete Tone Images(DTI)are available which needs to be analyzed in detail. Here, we created t his dataset for those who do research in DTI.
132. <a href="datasets/Dishonest+Internet+users+Dataset">Dishonest Internet user s Dataset</a>: The dataset was used to test an architecture based on a trust model capable to cope with t he evaluation of the trustworthiness of users interacting in pervasive environments. 133. <a href="datasets/Document+Understanding">Document Understanding</a>: Five concepts, expresse d as predicates, to be learned 134. <a href="datasets/Dodgers+Loop+Sensor">Dodg ers Loop Sensor</a>: Loop sensor data was collected for the Glendale on ramp for the 101 North freeway i n Los Angeles 135. <a href="datasets/Dorothea">Dorothea</a>: DOROTHEA is a drug discovery dataset. Chemical compounds represented by structural molecular features must be classified as active (binding to thrombin) or inactive. This is one of 5 datasets of the NIPS 2003 feature selection challeng e. 136. <a href="datasets/Dota2+Games+Results">Dota2 Games Results</a>: Dota 2 is a popular computer game with two teams of 5 players. At the start of the game each player chooses a unique hero with different strengths and weaknesses. 137. <a href="datasets/Dow+J ones+Index">Dow Jones Index</a>: This dataset contains weekly data for the Dow Jones Industrial Index. It has been used in computational investing research. 138. <a href="datasets/Dresse s_Attribute_Sales">Dresses_Attribute_Sales</a>: This dataset contain Attributes of dresses and their reco mmendations according to their sales.Sales are monitor on the basis of alternate days. 139. <a href="datasets/DrivFace">DrivFace</a>: The DrivFace contains images sequences of subjects while driving in real scenarios. It is composed of 606 samples of 640×480, acquired over different days from 4 d rivers with several facial features. 140. <a href="datasets/Drug+consumption+%28qu antified%29">Drug consumption (quantified)</a>: Classify type of drug consumer by personality data 141. <a href="datasets/Drug+Review+Dataset+%28Druglib.com%29">Drug Review Datas et (Druglib.com)</a>: The dataset provides patient reviews on specific drugs along with related conditions. Reviews and ratings are grouped into reports on the three aspects benefits, side effects and overall comment.< /p> 142. <a href="datasets/Drug+Review+Dataset+%28Drugs.com%29">Drug Review Da taset (Drugs.com)</a>: The dataset provides patient reviews on specific drugs along with related condition

s and a 10 star patient rating reflecting overall patient satisfaction. 143. <a href="datasets/DSRC+Vehicle+Communications">DSRC Vehicle Communications</a>: This set Provides data regarding wireless communications between vehicles and road side units. two separate data sets are provided (normal scenario) and in the presence of attacker (jammer). 144. <a href="datasets/Dynamic+Features+of+VirusShare+Executables">Dynamic Features of VirusShare Executables</a>: This dataset contains the dynamic features of 107,888 executables, collected by VirusShare from Nov/2010 to Jul/2014. 145. <a href="datasets/E.+Coli+Genes">E. Coli Genes</a>: Data giving characteristics of each ORF (potential gene) in the E. coli genome. Sequence, homology (similarity to other genes) and structural information, and function (if known) are provided. 146. <a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connected+speech+Data+Set+">Early biomarkers of Parkinson’s disease based on natural connected speech Data Set </a>: . 147. <a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech">Early biomarkers of Parkinson’s disease based on natural connected speech</a>: Predict a pattern of neurodegeneration in the dataset of speech features obtained from patients with early untreated Parkinson’s disease and patients at high risk developing Parkinson’s disease. 148. <a href="datasets/EBL+Domain+Theories">EBL Domain Theories</a>: Assorted small-scale domain theories 149. <a href="datasets/Echocardiogram">Echocardiogram</a>: Data for classifying if patients will survive for at least one year after a heart attack 150. <a href="datasets/Eco-hotel">Eco-hotel</a>: This dataset includes Online Textual Reviews from both online (e.g., TripAdvisor) and offline (e.g., Guests' book) sources from the Areias do Seixo Eco-Resort. 151. <a href="datasets/Ecoli">Ecoli</a>: This data contains protein localization sites 152. <a href="datasets/Economic+Sanctions">Economic Sanctions</a>: Domain Theory on Economic Sanctions; Undocumented 153. <a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set">Educational Process Mining (EPM): A Learning Analytics Data Set</a>: Educational Process Mining data set is built from the recordings of 115 subjects' activities through a logging application while learning with an educational simulator. 154. <a href="datasets/EEG+Database">EEG Database</a>: This data arises from a large study to examine EEG correlates of genetic predisposition to alcoholism. It contains measurements from 64 electrodes placed on the scalp sampled at 256 Hz 155. <a href="datasets/EEG+Eye+State">EEG Eye State</a>: The data set consists of 14 EEG values and a value indicating the eye state. 156. <a href="datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals">EEG Steady-State Visual Evoked Potential Signals</a>: This database consists on 30 subjects performing Brain Computer Interface for Steady State Visual Evoked Potentials (BCI-SSVEP). 157. <a href="datasets/El+Nino">El Nino</a>: The data set contains oceanographic and surface meteorological readings taken from a series of buoys positioned throughout the equatorial Pacific. 158. <a href="datasets/Electrical+Grid+Stability+Simulated+Data+">Electrical Grid Stability Simulated Data </a>: The local stability analysis of the 4-node star system (electricity producer is in the center) implementing Decentral Smart Grid Control concept. 159. <a href="datasets/ElectricityLoadDiagrams20112014">ElectricityLoadDiagrams20112014</a>: This data set contains electricity consumption of 370 points/clients.
160. <a href="datasets/EMG+data+for+gestures">EMG data for gestures</a>: These are files of raw EMG data recorded by MYO Thalmic bracelet 161. <a href="datasets/EMG+dataset+in+Lower+Limb">EMG dataset in Lower Limb</a>: 3 different exercises: sitting, standing and walking in the muscles: biceps femoris, vastus medialis, rectus femoris and semitendinosus addition to goniometry in the exercises. 162. <a href="datasets/EMG+Physical+Action+Data+Set">EMG Physical Action Data Set</a>: The Physical Action Data Set includes 10 normal and 10 aggressive physical actions that measure the human activity. The data have been collected by 4 subjects using the Delsys EMG wireless apparatus. 163. <a href="datasets/Energy+efficiency">Energy efficiency</a>: This study looked into assessing the heating load and cooling load requirements of buildings (that is, energy efficiency) as a function of building parameters. 164. <a href="datasets/Entree+Chicago+Recommendation+Data">Entree Chicago Recommendation Data</a>: This data contains a record of user interactions with the Entree Chicago restaurant recommendation system. 165. <a href="datasets/Epileptic+Seizure+Recognition">Epileptic Seizure Recognition</a>: This dataset is a pre-processed and re-structured/reshaped version of a very commonly used dataset featuring epileptic seizure detection. 166. <a href="datasets/extention+of+Z-Alizadeh+sani+dataset">extention of Z-Alizadeh sani dataset</a>: It was collected for CAD diagnosis. 167. <a href="datasets/Facebook+Comment+Volume+Dataset">Facebook Comment Volume Dataset</a>: Instances in this dataset contain features extracted from facebook posts. The task associated with the data is to predict how many comments the post will receive. 168. <a href="datasets/Facebook+metrics">Facebook metrics</a>: Facebook performance metrics of a renowned cosmetic's brand Facebook page. 169. <a href="datasets/Farm+Ads">Farm Ads</a>: This data was collected from text ads found on twelve websites that deal with various farm animal related topics. The binary labels are based on whether or not the content owner approves of the ad. 170. <a href="dataset

ased on whether or not the content owner approves of the ad. 170. <a href="datasets/Fertility">Fertility</a>: 100 volunteers provide a semen sample analyzed according to the WHO 2010 criteria. Sperm concentration are related to socio-demographic data, environmental factors, health status, and life habits 171. <a href="datasets/Firm-Teacher_Clave-Direction_Classification">Firm-Teacher_Clave-Direction_Classification</a>: The data are binary attack-point vectors and their clave-direction class(es) according to the partido-alto-based paradigm. 172. <a href="datasets/First-order+theorem+proving">First-order theorem proving</a>: Given a theorem, predict which of five heuristics will give the fastest proof when used by a first-order prover. A sixth prediction declines to attempt a proof, should the theorem be too difficult. 173. <a href="datasets/Flags">Flags</a>: From Collins Gem Guide to Flags, 1986 174. <a href="datasets/FMA%3A+A+Dataset+For+Music+Analysis">FMA: A Dataset For Music Analysis</a>: FMA features 106,574 tracks and includes song title, album, artist, genres; play counts, favorites, comments; description, biography, tags; together with audio (343 days, 917 GiB) and features. 175. <a href="datasets/Folio">Folio</a>: 20 photos of leaves for each of 32 different species. 176. <a href="datasets/Forest+Fires">Forest Fires</a>: This is a difficult regression task, where the aim is to predict the burned area of forest fires, in the northeast region of Portugal, by using meteorological and other data (see details at: http://www.dsi.uminho.pt/~pcortez/forestfires). 177. <a href="datasets/Forest+type+mapping">Forest type mapping</a>: Multi-temporal remote sensing data of a forested area in Japan. The goal is to map different forest types using spectral data. 178. <a href="datasets/Function+Finding">Function Finding</a>: Cases collected mostly from investigations in physical science; intention is to evaluate function-finding algorithms 179. <a href="datasets/Gas+Sensor+Array+Drift+Dataset">Gas Sensor Array Drift Dataset</a>: This archive contains 13910 measurements from 16 chemical sensors utilized in simulations for drift compensation in a discrimination task of 6 gases at various levels of concentrations. 180. <a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations">Gas Sensor Array Drift Dataset at Different Concentrations</a>: This archive contains 13910 measurements from 16 chemical sensors exposed to 6 different gases at various concentration levels. 181. <a href="datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures">Gas sensor array exposed to turbulent gas mixtures</a>: A chemical detection platform composed of 8 chemoresistive gas sensors was exposed to turbulent gas mixtures generated naturally in a wind tunnel. The acquired time series of the sensors are provided. 182. <a href="datasets/Gas+sensor+array+under+dynamic+gas+mixtures">Gas sensor array under dynamic gas mixtures</a>: The data set contains the recordings of 16 chemical sensors exposed to two dynamic gas mixtures at varying concentrations. For each mixture, signals were acquired continuously during 12 hours. 183. <a href="datasets/Gas+sensor+array+under+flow+modulation">Gas sensor array under flow modulation</a>: The data set contains 58 time series acquired from 16 chemical sensors under gas flow modulation conditions. The sensors were exposed to different gaseous binary mixtures of acetone and ethanol. 184. <a href="datasets/Gas+sensor+arrays+in+open+sampling+settings">Gas sensor arrays in open sampling settings</a>: The dataset contains 18000 time-series recordings from a chemical detection platform at six different locations in a wind tunnel facility in response to ten high-priority chemical gaseous substances 185. <a href="datasets/Gas+sensors+for+home+activity+monitoring">Gas sensors for home activity monitoring</a>: 100 recordings of a sensor array under different conditions in a home setting: background, wine and banana presentations. The array includes 8 MOX gas sensors, and humidity and temperature sensors.
186. <a href="datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy">Gastrointestinal Lesions in Regular Colonoscopy</a>: This dataset contains features extracted from colonoscopy videos used to detect gastrointestinal lesions. It contains 76 lesions: 15 serrated adenomas, 21 hyperplastic lesions and 40 adenoma. 187. <a href="datasets/gene+expression+cancer+RNA-Seq">gene expression cancer RNA-Seq</a>: This collection of data is part of the RNA-Seq (HiSeq) PANCAN data set, it is a random extraction of gene expressions of patients having different types of tumor: BRCA, KIRC, COAD, LUAD and PRAD. 188. <a href="datasets/Geo-Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone">Geo-Magnetic field and WLAN dataset for indoor localisation from wristband and smartphone</a>: A multisource and multivariate dataset for indoor localisation methods based on WLAN and Geo-Magnetic field fingerprinting 189. <a href="datasets/Geographical+Original+of+Music">Geographical Original of Music</a>: Instances in this dataset contain audio features extracted from 1059 wave files. The task associated with the data is to predict the geographical origin of music.
190. <a href="datasets/Gesture+Phase+Segmentation">Gesture Phase Segmentation</a>: The dataset is composed by features extracted from 7 videos with people gesticulating, aiming at studying Gesture Phase Segmentation. It contains 50 attributes divided into two files for each video. 191. <a href="datasets/Gisette">Gisette</a>: GISETTE is a handwritten digit recognition problem. The problem is to separate the highly confusible digits '4' and '9'. This dataset is one of five datasets of the NIPS 2003 feature selection challenge.
192. <a href="datasets/Glass+Identification">Glass Identification</a>: From US A Forensic Science Service; 6 types of glass; defined in terms of their oxide content (i.e. Na, Fe, K, etc) <p

A Forensic Science Service; 6 types of glass; defined in terms of their oxide content (i.e. Na, Fe, K, etc) 193. <a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data">GNFUV Unmanned Surface Vehicles Sensor Data</a>: The data-set contains four (4) sets of mobile sensor readings data (humidity, temperature) corresponding to a swarm of four (4) Unmanned Surface Vehicles (USVs) in a test-bed in Athens (Greece). 194. <a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2">GNFUV Unmanned Surface Vehicles Sensor Data Set 2</a>: The data-set contains eight (2x4) data-sets of mobile sensor readings data (humidity, temperature) corresponding to a swarm of four Unmanned Surface Vehicles (USVs) in a test-bed, Athens, Greece. 195. <a href="datasets/GPS+Trajectories">GPS Trajectories</a>: The dataset has been feed by Android app called Go !Track. It is available at Goolge Play Store(https://play.google.com/store/apps/details?id=com.go.router). 196. <a href="datasets/Grammatical+Facial+Expressions">Grammatical Facial Expressions</a>: This dataset supports the development of models that make possible to interpret Grammatical Facial Expressions from Brazilian Sign Language (Libras). 197. <a href="datasets/Greenhouse+Gas+Observing+Network">Greenhouse Gas Observing Network</a>: Design an observing network to monitor emissions of a greenhouse gas (GHG) in California given time series of synthetic observations and tracers from weather model simulations.
198. <a href="datasets/Haberman%27s+Survival">Haberman's Survival</a>: Dataset contains cases from study conducted on the survival of patients who had undergone surgery for breast cancer 199. <a href="datasets/Hayes-Roth">Hayes-Roth</a>: Topic: human subjects study 200. <a href="datasets/HCC+Survival">HCC Survival</a>: Hepatocellular Carcinoma dataset (HCC dataset) was collected at a University Hospital in Portugal. It contains real clinical data of 165 patients diagnosed with HCC. 201. <a href="datasets/Health+News+in+Twitter">Health News in Twitter</a>: The data was collected in 2015 using Twitter API. This dataset contains health news from more than 15 major health news agencies such as BBC, CNN, and NYT. 202. <a href="datasets/Heart+Disease">Heart Disease</a>: 4 databases: Cleveland, Hungary, Switzerland, and the VA Long Beach 203. <a href="datasets/Hepatitis">Hepatitis</a>: From G.Gong: CMU; Mostly Boolean or numeric-valued attribute types; Includes cost data (donated by Peter Turney) 204. <a href="datasets/HEPMASS">HEPMASS</a>: The search for exotic particles requires sorting through a large number of collisions to find the events of interest. This data set challenges one to detect a new particle of unknown mass. 205. <a href="datasets/Heterogeneity+Activity+Recognition">Heterogeneity Activity Recognition</a>: The Heterogeneity Human Activity Recognition (HHAR) dataset from Smartphones and Smartwatches is a dataset devised to benchmark human activity recognition algorithms (classification, automatic data segmentation, sensor fusion, feature extraction, etc.) in real-world contexts; specifically, the dataset is gathered with a variety of different device models and use-scenarios, in order to reflect sensing heterogeneities to be expected in real deployments. 206. <a href="datasets/HIGGS">HIGGS</a>: This is a classification problem to distinguish between a signal process which produces Higgs bosons and a background process which does not. 207. <a href="datasets/Hill-Valley">Hill-Valley</a>: Each record represents 100 points on a two-dimensional graph. When plotted in order (from 1 through 100) as the Y co-ordinate, the points will create either a Hill (a &#65533;bump&#65533; in the terrain) or a Valley (a &#65533;dip&#65533; in the terrain). 208. <a href="datasets/HIV-1+protease+cleavage">HIV-1 protease cleavage</a>: The data contains lists of octamers (8 amino acids) and a flag (-1 or 1) depending on whether HIV-1 protease will cleave in the central position (between amino acids 4 and 5). 209. <a href="datasets/Horse+Colic">Horse Colic</a>: Well documented attributes; 368 instances with 28 attributes (continuous, discrete, and nominal); 30% missing values 210. <a href="datasets/HTRU2">HTRU2</a>: Pulsar candidates collected during the HTRU survey. Pulsars are a type of star, of considerable scientific interest. Candidates must be classified into pulsar and non-pulsar classes to aid discovery. 211. <a href="datasets/Human+Activity+Recognition+Using+Smartphones">Human Activity Recognition Using Smartphones</a>: Human Activity Recognition database built from the recordings of 30 subjects performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors. 212. <a href="datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer">Hybrid Indoor Positioning Dataset from WiFi RSSI, Bluetooth and magnetometer</a>: The dataset was created for the comparison and evaluation of hybrid indoor positioning methods. The dataset presented contains data from W-LAN and Bluetooth interfaces, and Magnetometer. 213. <a href="datasets/ICMLA+2014+Accepted+Papers+Data+Set">ICMLA 2014 Accepted Papers Data Set</a>: This data set compromises the metadata for the 2014 ICMLA conference's accepted papers, including ID, paper titles, author's keywords, abstracts and sessions in which they were exposed. 214. <a href="datasets/ICU">ICU</a>: Data set prepared for the use of participants for the 1994 AAAI Spring Symposium on Artificial Intelligence in Medicine. 215. <a href="datasets/IDA2016Challenge">IDA2016Challenge</a>: The dataset consists of data collected from heavy Scania trucks in everyday usage. 216. <a href="datasets/ILPD+%28Indian+Liver+Patient+Dataset%29">ILPD (Indian Liver Patient Dataset)</a>: This data set contains 10 variables that are age, gender, total Bilirubin, direct Bilirubin, total prot

eins, albumin, A/G ratio, SGPT, SGOT and Alkphos. 217. <a href="datasets/Image+Segmentation">Image Segmentation</a>: Image data described by high-level numeric-valued attributes, 7 classes 218. <a href="datasets/Immunotherapy+Dataset">Immunotherapy Dataset</a>: This dataset contains information about wart treatment results of 90 patients using immunotherapy. 219. <a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Parkinson%E2%80%99s+Disease">Improved Spiral Test Using Digitized Graphics Tablet for Monitoring Parkinson's Disease</a>: Handwriting database consists of 25 PWP(People with Parkinson) and 15 healthy individuals.Three types of recordings (Static Spiral Test, Dynamic Spiral Test and Stability Test) are taken. 220. <a href="datasets/Individual+household+electric+power+consumption">Individual household electric power consumption</a>: Measurements of electric power consumption in one household with a one-minute sampling rate over a period of almost 4 years. Different electrical quantities and some sub-metering values are available. 221. <a href="datasets/Indoor+User+Movement+Prediction+from+RSS+data">Indoor User Movement Prediction from RSS data</a>: This dataset contains temporal data from a Wireless Sensor Network deployed in real-world office environments. The task is intended as real-life benchmark in the area of Ambient Assisted Living. 222. <a href="datasets/Insurance+Company+Benchmark+%28COIL+2000%29">Insurance Company Benchmark (COIL 2000)</a>: This data set used in the CoIL 2000 Challenge contains information on customers of an insurance company. The data consists of 86 variables and includes product usage data and socio-demographic data 223. <a href="datasets/Internet+Advertisements">Internet Advertisements</a>: This dataset represents a set of possible advertisements on Internet pages. 224. <a href="datasets/Internet+Usage+Data">Internet Usage Data</a>: This data contains general demographic information on internet users in 1997. 225. <a href="datasets/Ionosphere">Ionosphere</a>: Classification of radar returns from the ionosphere 226. <a href="datasets/IPUMS+Census+Database">IPUMS Census Database</a>: This data set contains unweighted PUMS census data from the Los Angeles and Long Beach areas for the years 1970, 1980, and 1990. 227. <a href="datasets/Iris">Iris</a>: Famous database; from Fisher, 1936 228. <a href="datasets/ISOLET">ISOLET</a>: Goal: Predict which letter-name was spoken--a simple classification task. 229. <a href="datasets/ISTANBUL+STOCK+EXCHANGE">ISTANBUL STOCK EXCHANGE</a>: Data sets includes returns of Istanbul Stock Exchange with seven other international index; SP, DAX, FTSE, NIKKEI, BOVESPA, MSCE_EU, MSCI_EM from Jun 5, 2009 to Feb 22, 2011. 230. <a href="datasets/Japanese+Credit+Screening">Japanese Credit Screening</a>: Includes domain theory (generated by talking to Japanese domain experts); data in Lisp 231. <a href="datasets/Japanese+Vowels">Japanese Vowels</a>: This dataset records 640 time series of 12 LPC cepstrum coefficients taken from nine male speakers. 232. <a href="datasets/KASANDR">KASANDR</a>: KASANDR is a novel, publicly available collection for recommendation systems that records the behavior of customers of the European leader in e-Commerce advertising, Kelkoo. 233. <a href="datasets/KDC-4007+dataset+Collection">KDC-4007 dataset Collection</a>: KDC-4007 dataset Collection is the Kurdish Documents Classification text used in categories regarding Kurdish Sorani news and articles. 234. <a href="datasets/KDD+Cup+1998+Data">KDD Cup 1998 Data</a>: This is the data set used for The Second International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-98 235. <a href="datasets/KDD+Cup+1999+Data">KDD Cup 1999 Data</a>: This is the data set used for The Third International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-99 236. <a href="datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29">KEGG Metabolic Reaction Network (Undirected)</a>: KEGG Metabolic pathways modeled as un-directed reaction network. Variety of graphical features presented. 237. <a href="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29">KEGG Metabolic Relation Network (Directed)</a>: KEGG Metabolic pathways modeled as directed relation network. Variety of graphical features presented. 238. <a href="datasets/Kinship">Kinship</a>: Relational dataset 239. <a href="datasets/Labor+Relations">Labor Relations</a>: From Collective Bargaining Review 240. <a href="datasets/Las+Vegas+Strip">Las Vegas Strip</a>: This dataset includes quantitative and categorical features from online reviews from 21 hotels located in Las Vegas Strip, extracted from TripAdvisor (http://www.tripadvisor.com). 241. <a href="datasets/Leaf">Leaf</a>: This dataset consists in a collection of shape and texture features extracted from digital images of leaf specimens originating from a total of 40 different plant species. 242. <a href="datasets/LED+Display+Domain">LED Display Domain</a>: From Classification and Regression Trees book; We provide here 2 C programs for generating sample databases 243. <a href="datasets/Legal+Case+Reports">Legal Case Reports</a>: A textual corpus of 4000 legal cases for automatic summarization and citation analysis. For each document we collect catchphrases, citations sentences, citation catchphrases and citation classes. 244. <a href="datasets/Lenses">Lenses</a>: Database for fitting contact lenses 245. <a href="datasets/Letter+Recognition">Letter Recognition</a>: Database of character image features; try to identify the letter 246. <a href="dataset

s/Libras+Movement">Libras Movement</a>: The data set contains 15 classes of 24 instances each. Each class references to a hand movement type in LIBRAS (Portuguese
name 'L&#65533;ngua BRAsileira de Sinais', oficial brazilian signal language). 247. <a href="datasets/Liver+Disorders">Liver Disorders</a>: BUPA Medical Research Ltd. database donated by Richard S. Forsyth 248. <a href="datasets/Localization+Data+for+Person+Activity">Localization Data for Person Activity</a>: Data contains recordings of five people performing different activities. Each person wore four sensors (tags) while performing the same scenario five times. 249. <a href="datasets/Logic+Theorist">Logic Theorist</a>: All code for Logic Theorist 250. <a href="datasets/Low+Resolution+Spectrometer">Low Resolution Spectrometer</a>: From IRAS data -- NASA Ames Research Center 251. <a href="datasets/LSVT+Voice+Rehabilitation">LSVT Voice Rehabilitation</a>: 126 samples from 14 participants, 309 features. Aim: assess whether voice rehabilitation treatment lead to phonations considered 'acceptable' or 'unacceptable' (binary class classification problem). 252. <a href="datasets/Lung+Cancer">Lung Cancer</a>: Lung cancer data; no attribute definitions 253. <a href="datasets/Lymphography">Lymphography</a>: This lymphography domain was obtained from the University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia. (Restricted access) 254. <a href="datasets/M.+Tuberculosis+Genes">M. Tuberculosis Genes</a>: Data giving characteristics of each ORF (potential gene) in the M. tuberculosis bacterium. Sequence, homology (similarity to other genes) and structural information, and function (if known) are provided 255. <a href="datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014">Machine Learning based ZZAlpha Ltd. Stock Recommendations 2012-2014</a>: The data here are the ZZAlpha® machine learning recommendations made for various US traded stock portfolios the morning of each day during the 3 year period Jan 1, 2012 - Dec 31, 2014. 256. <a href="datasets/Madelon">Madelon</a>: MADELON is an artificial dataset, which was part of the NIPS 2003 feature selection challenge. This is a two-class classification problem with continuous input variables. The difficulty is that the problem is multivariate and highly non-linear. 257. <a href="datasets/MAGIC+Gamma+Telescope">MAGIC Gamma Telescope</a>: Data are MC generated to simulate registration of high energy gamma particles in an atmospheric Cherenkov telescope 258. <a href="datasets/Mammographic+Mass">Mammographic Mass</a>: Discrimination of benign and malignant mammographic masses based on BI-RADS attributes and the patient's age. 259. <a href="datasets/Mechanical+Analysis">Mechanical Analysis</a>: Fault diagnosis problem of electromechanical devices; also PUMPS DATA SET is newer version with domain theory and results 260. <a href="datasets/Mesothelioma%E2%80%99s+disease+data+set+">Mesothelioma’s disease data set </a>: Mesothelioma’s disease data set were prepared at Dicle University Faculty of Medicine in Turkey.
Three hundred and twenty-four Mesothelioma patient data. In the dataset, all samples have 34 features. 261. <a href="datasets/Meta-data">Meta-data</a>: Meta-Data was used in order to give advice about which classification method is appropriate for a particular dataset (taken from results of Statlog project). 262. <a href="datasets/MEU-Mobile+KSD">MEU-Mobile KSD</a>: This dataset contains keystroke dynamics data collected on a touch mobile device (Nexus 7). The dataset contains 2856 records, 51 records per subject for 56 subjects. 263. <a href="datasets/MHEALTH+Dataset">MHEALTH Dataset</a>: The MHEALTH (Mobile Health) dataset is devised to benchmark techniques dealing with human behavior analysis based on multimodal body sensing. 264. <a href="datasets/Mice+Protein+Expression">Mice Protein Expression</a>: Expression levels of 77 proteins measured in the cerebral cortex of 8 classes of control and Down syndrome mice exposed to context fear conditioning, a task used to assess associative learning. 265. <a href="datasets/microblogPCU">microblogPCU</a>: MicroblogPCU data is crawled from sina weibo microblog[http://weibo.com/]. This data can be used to study machine learning methods as well as do some social network research. 266. <a href="datasets/MicroMass">MicroMass</a>: A dataset to explore machine learning approaches for the identification of microorganisms from mass-spectrometry data. 267. <a href="datasets/MiniBooNE+particle+identification">MiniBooNE particle identification</a>: This dataset is taken from the MiniBooNE experiment and is used to distinguish electron neutrinos (signal) from muon neutrinos (background). 268. <a href="datasets/Miskolc+IIS+Hybrid+IPS">Miskolc IIS Hybrid IPS</a>: The dataset was created for the comparison and evaluation of hybrid indoor positioning methods. The dataset presented contains data from W-LAN and Bluetooth interfaces, and Magnetometer. 269. <a href="datasets/Mobile+Robots">Mobile Robots</a>: Learning concepts from sensor data of a mobile robot; set of data sets 270. <a href="datasets/MoCap+Hand+Postures">MoCap Hand Postures</a>: 5 types of hand postures from 12 users were recorded using unlabeled markers attached to fingers of a glove in a motion capture environment. Due to resolution and occlusion, missing values are common. 271. <a href="datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29">Molecular Biology (Promoter Gene Sequences)</a>: E. Coli promoter gene sequences (DNA) with partial domain theory 272. <a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular Biology (Protein Secondary Structure)</a>: From C

MU connectionist bench repository; Classifies secondary structure of certain globular proteins 273. <a href="datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29">Molecular Biology (Splice-junction Gene Sequences)</a>: Primate splice-junction gene sequences (DNA) with associated imperfect domain theory 274. <a href="datasets/MONK%27s+Problems">MONK's Problems</a>: A set of three artificial domains over the same attribute space; Used to test a wide range of induction algorithms 275. <a href="datasets/Moral+Reasoner">Moral Reasoner</a>: Horn-clause model that qualitatively simulates moral reasoning; Theory includes negated literals 276. <a href="datasets/Motion+Capture+Hand+Postures">Motion Capture Hand Postures</a>: 5 types of hand postures from 12 users were recorded using unlabeled markers on fingers of a glove in a motion capture environment. Due to resolution and occlusion, missing values are common. 277. <a href="datasets/Movie">Movie</a>: This data set contains a list of over 10000 films including many older, odd, and cult films. There is information on actors, casts, directors, producers, studios, etc. 278. <a href="datasets/MSNBC.com+Anonymous+Web+Data">MSNBC.com Anonymous Web Data</a>: This data describes the page visits of users who visited msnbc.com on September 28, 1999. Visits are recorded at the level of URL category (see description) and are recorded in time order. 279. <a href="datasets/Mturk+User-Perceived+Clusters+over+Images">Mturk User-Perceived Clusters over Images</a>: This dataset was collected by Shan-Hung Wu and DataLab members at NTHU, Taiwan. There're 325 user-perceived clusters from 100 users and their corresponding descriptions. 280. <a href="datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing">Multimodal Damage Identification for Humanitarian Computing</a>: 5879 captioned images (image and text) from social media related to damage during natural disasters/wars, and belong to 6 classes: Fires, Floods, Natural landscape, Infrastructural, Human, Non-damage. 281. <a href="datasets/Multiple+Features">Multiple Features</a>: This dataset consists of features of handwritten numerals (`0'--`9') extracted from a collection of Dutch utility maps 282. <a href="datasets/Mushroom">Mushroom</a>: From Audobon Society Field Guide; mushrooms described in terms of physical characteristics; classification: poisonous or edible 283. <a href="datasets/Musk+%28Version+1%29">Musk (Version 1)</a>: The goal is to learn to predict whether new molecules will be musks or non-musks 284. <a href="datasets/Musk+%28Version+2%29">Musk (Version 2)</a>: The goal is to learn to predict whether new molecules will be musks or non-musks 285. <a href="datasets/News+Aggregator">News Aggregator</a>: References to news pages collected from an web aggregator in the period from 10-March-2014 to 10-August-2014. The resources are grouped into clusters that represent pages discussing the same story. 286. <a href="datasets/News+Popularity+in+Multiple+Social+Media+Platforms">News Popularity in Multiple Social Media Platforms</a>: Large data set of news items and their respective social feedback on multiple platforms: Facebook, Google+ and LinkedIn. 287. <a href="datasets/Newspaper+and+magazine+images+segmentation+dataset">Newspaper and magazine images segmentation dataset</a>: Dataset is well suited for segmentation tasks. It contains 101 scanned pages from different newspapers and magazines in Russian with ground truth pixel-based masks. 288. <a href="datasets/NIPS+Conference+Papers+1987-2015">NIPS Conference Papers 1987-2015</a>: This data set contains the distribution of words in the full text of the NIPS conference papers published from 1987 to 2015. 289. <a href="datasets/NoisyOffice">NoisyOffice</a>: Corpus intended to do cleaning (or binarization) and enhancement of noisy grayscale printed text images using supervised learning methods. Noisy images and their corresponding ground truth provided. 290. <a href="datasets/Nomao">Nomao</a>: Nomao collects data about places (name, phone, localization...) from many sources.
Deduplication consists in detecting what data refer to the same place.
Instances in the dataset compare 2 spots. 291. <a href="datasets/Northix">Northix</a>: Northix is designed to be a schema matching benchmark problem for data integration of two entity relationship databases. 292. <a href="datasets/NSF+Research+Award+Abstracts+1990-2003">NSF Research Award Abstracts 1990-2003</a>: This data set consists of (a) 129,000 abstracts describing NSF awards for basic research, (b) bag-of-word data files extracted from the abstracts, (c) a list of words used for indexing the bag-of-word 293. <a href="datasets/Nursery">Nursery</a>: Nursery Database was derived from a hierarchical decision model originally developed to rank applications for nursery schools. 294. <a href="datasets/NYSK">NYSK</a>: NYSK (New York v. Strauss-Kahn) is a collection of English news articles about the case relating to allegations of sexual assault against the former IMF director Dominique Strauss-Kahn (May 2011). 295. <a href="datasets/Occupancy+Detection+">Occupancy Detection </a>: Experimental data used for binary classification (room occupancy) from Temperature,Humidity,Light and CO2. Ground-truth occupancy was obtained from time stamped pictures that were taken every minute. 296. <a href="datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes">OCT data & Color Fundus Images of Left & Right Eyes</a>: This dataset contains OCT data (in mat format) and color fundus data (in jpg format) of left & right eyes of 50 healthy persons. 297. <a href="datasets/One-hundred+plant+species+leaves+data+set">One-hundred plant species leaves data set</a>: Sixteen samples of leaf ea

ch of one-hundred plant species. For each sample, a shape descriptor, fine scale margin and texture histogram are given. 298. <a href="datasets/Online+Handwritten+Assamese+Characters+Data set">Online Handwritten Assamese Characters Dataset</a>: This is a dataset of 8235 online handwritten a ssamese characters. The “online” process involves capturing of data as text is written on a digitizing tablet with an electronic pen. 299. <a href="datasets/Online+News+Popularity">Online News Po pularity</a>: This dataset summarizes a heterogeneous set of features about articles published by Mashab le in a period of two years. The goal is to predict the number of shares in social networks (popularity). <p cl ass="normal">300. <a href="datasets/Online+Retail">Online Retail</a>: This is a transnational data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registere d non-store online retail. 301. <a href="datasets/Online+Shoppers+Purchasing+Inten tion+Dataset">Online Shoppers Purchasing Intention Dataset</a>: Of the 12,330 sessions in the dataset, 84.5% (10,422) were negative class samples that did not
end with shopping, and the rest (1908) were positive class
samples ending with shopping. 302. <a href="datasets/Online+Video+Characteristics +and+Transcoding+Time+Dataset">Online Video Characteristics and Transcoding Time Dataset</a>: The dataset contains a million randomly sampled video instances listing 10 fundamental video characteristics along with the YouTube video ID. 303. <a href="datasets/Open+University+Learning+Anal ytics+dataset">Open University Learning Analytics dataset</a>: Open University Learning Analytics Datase t contains data about courses, students and their interactions with Virtual Learning Environment for seven selec ted courses and more than 30000 students. 304. <a href="datasets/Opinosis+Opinio n+%26frasl%3B+Review">Opinosis Opinion ⁄ Review</a>: This dataset contains sentences extracted from user reviews on a given topic. Example topics are “performance of Toyota Camry” and “sound quality of ipod n ano”. 305. <a href="datasets/OpinRank+Review+Dataset">OpinRank Review Datas et</a>: This data set contains user reviews of cars and and hotels collected from Tripadvisor (~259,000 reviews) and Edmunds (~42,230 reviews). 306. <a href="datasets/OPPORTUNITY +Activity+Recognition">OPPORTUNITY Activity Recognition</a>: The OPPORTUNITY Dataset for Human Activity Recognition from Wearable, Object, and Ambient Sensors is a dataset devised to benchmark human ac tivity recognition algorithms (classification, automatic data segmentation, sensor fusion, feature extraction, etc). 307. <a href="datasets/Optical+Interconnection+Network+">Optical Interconnection Network </a>: This dataset contains 640 performance measurements from a simulation of 2-Dimensional Multiprocessor Optical Interconnection Network. 308. <a href="datasets/Optical+Re cognition+of+Handwritten+Digits">Optical Recognition of Handwritten Digits</a>: Two versions of this datab ase available; see folder 309. <a href="datasets/Othello+Domain+Theory">Othello D omain Theory</a>: Used in research to generate features for an inductive learning system 310. <a href="datasets/Ozone+Level+Detection">Ozone Level Detection</a>: Two ground ozon e level data sets are included in this collection. One is the eight hour peak set (eighthr.data), the other is the on e hour peak set (onehr.data). Those data were collected from 1998 to 2004 at the Houston, Galveston and Bra zoria area. 311. <a href="datasets/p53+Mutants">p53 Mutants</a>: The goal is to model mutant p53 transcriptional activity (active vs inactive) based on data extracted from biophysical simula tions.
312. <a href="datasets/Page+Blocks+Classification">Page Blocks Classification</a> : The problem consists of classifying all the blocks of the page layout of a document that has been detecte d by a segmentation process. 313. <a href="datasets/PAMAP2+Physical+Activity+Mo nitoring">PAMAP2 Physical Activity Monitoring</a>: The PAMAP2 Physical Activity Monitoring dataset conta ins data of 18 different physical activities, performed by 9 subjects wearing 3 inertial measurement units and a heart rate monitor. 314. <a href="datasets/PANDOR">PANDOR</a>: PANDOR i s a novel and publicly available dataset for online recommendation provided by Purch (http://www.purch.com/). 315. <a href="datasets/Paper+Reviews">Paper Reviews</a>: This sentiment a nalysis data set contains scientific paper reviews from an international conference on computing and informatic s. The task is to predict the orientation or the evaluation of a review. 316. <a href="d atasets/Parking+Birmingham">Parking Birmingham</a>: Data collected from car parks in Birmingham that are operated by NCP from
Birmingham City Council. UK Open Government Licence (OGL).
https://data.birmingham.gov.uk/dataset/birmingham-parking 317. <a href="datasets/ Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet">Parkinson Disease Spiral Drawings U sing Digitized Graphics Tablet</a>: Handwriting database consists of 62 PWP(People with Parkinson) and 15 healthy individuals. Three types of recordings (Static Spiral Test, Dynamic Spiral Test and Stability Test) are taken. 318. <a href="datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of +Sound+Recordings">Parkinson Speech Dataset with Multiple Types of Sound Recordings</a>: The traini ng data belongs to 20 Parkinson's Disease (PD) patients and 20 healthy subjects. From all subjects, multiple ty pes of sound recordings (26) are taken. 319. <a href="datasets/Parkinson%27s+Dis ease+Classification">Parkinson's Disease Classification</a>: The data used in this study were gathered fro

m 188 patients with PD (107 men and 81 women) with ages ranging from 33 to 87 (65.1±10.9). 320. <a href="datasets/Parkinsons">Parkinsons</a>: Oxford Parkinson's Disease Detection Dat aset 321. <a href="datasets/Parkinsons+Telemonitoring">Parkinsons Telemonitoring </a>: Oxford Parkinson's Disease Telemonitoring Dataset 322. <a href="dataset s/PEMS-SF">PEMS-SF</a>: 15 months worth of daily data (440 daily records) that describes the occupanc y rate, between 0 and 1, of different car lanes of the San Francisco bay area freeways across time. 323. <a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits">Pen-Based Recognition of Handwritten Digits</a>: Digit database of 250 samples from 44 writers 324. < a href="datasets/Perfume+Data">Perfume Data</a>: This data consists of odors of 20 different perfumes. Data was obtained by using a handheld odor meter (OMX-GR sensor) per second for 28 seconds period. < p class="normal">325. <a href="datasets/Phishing+Websites">Phishing Websites</a>: This dataset coll ected mainly from: PhishTank archive, MillerSmiles archive, Google’s searching operators. 326. <a href="datasets/Physical+Unclonable+Functions">Physical Unclonable Functions</a>: The da taset is generated from Physical Unclonable Functions (PUFs) simulation, specifically XOR Arbiter PUFs. PUFs are used for authentication purposes. For more info, refer to our paper below. 327. <a href="datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure">Physicochemical Properties of Protein Tertiary Structure</a>: This is a data set of Physicochemical Properties of Protein Tertiary Structur e. The data set is taken from CASP 5-9. There are 45730 decoys and size varying from 0 to 21 armstrong. 328. <a href="datasets/Pioneer-1+Mobile+Robot+Data">Pioneer-1 Mobile Robot Data</ a>: This dataset contains time series sensor readings of the Pioneer-1 mobile robot. The data is broken int o "experiences" in which the robot takes action for some period of time and experiences a control 329. <a href="datasets/Pittsburgh+Bridges">Pittsburgh Bridges</a>: Bridges database that h as original and numeric-discretized datasets 330. <a href="datasets/Planning+Relax" >Planning Relax</a>: The dataset concerns with the classification of two mental stages from recorded EEG signals: Planning (during imagination of motor act) and Relax state. 331. <a href="d atasets/Plants">Plants</a>: Data has been extracted from the USDA plants database. It contains all plants (species and genera) in the database and the states of USA and Canada where they occur. 332. <a href="datasets/PM2.5+Data+of+Five+Chinese+Cities">PM2.5 Data of Five Chinese Cities</a></ b>: This hourly data set contains the PM2.5 data in Beijing, Shanghai, Guangzhou, Chengdu and Shenyang. M eanwhile, meteorological data for each city are also included. 333. <a href="datasets /PMU-UD">PMU-UD</a>: The handwritten dataset was collected from 170 participants with a total of 5,180 numeral patterns. The dataset is named Prince Mohammad Bin Fahd University - Urdu/Arabic Database (PMU-UD). 334. <a href="datasets/Poker+Hand">Poker Hand</a>: Purpose is to predi ct poker hands 335. <a href="datasets/Polish+companies+bankruptcy+data">Polish companies bankruptcy data</a>: The dataset is about bankruptcy prediction of Polish companies.The bank rupt companies were analyzed in the period 2000-2012, while the still operating companies were evaluated fro m 2007 to 2013. 336. <a href="datasets/Post-Operative+Patient">Post-Operative Pa tient</a>: Dataset of patient features 337. <a href="datasets/Predict+keywords+ activities+in+a+online+social+media">Predict keywords activities in a online social media</a>: The data fro m Twitter was collected during 360 consecutive days. It was done by querying 1497 English keywords sampled from Wikipedia. This dataset is proposed in a Learning to rank setting. 338. <a href= "datasets/Primary+Tumor">Primary Tumor</a>: From Ljubljana Oncology Institute 3 39. <a href="datasets/Prodigy">Prodigy</a>: Assorted domains like blocksworld, eightpuzzle, and sche dworld. 340. <a href="datasets/Protein+Data">Protein Data</a>: Undocumented 341. <a href="datasets/Pseudo+Periodic+Synthetic+Time+Series">Pseudo Periodic Synthetic Time Series</a>: This data set is designed for testing indexing schemes in time series databases . The data appears highly periodic, but never exactly repeats itself. 342. <a href="dat asets/PubChem+Bioassay+Data">PubChem Bioassay Data</a>: These highly imbalanced bioassay datase ts are from the differing types of screening that can be performed using HTS technology. 21 datasets were crea ted from 12 bioassays. 343. <a href="datasets/QSAR+biodegradation">QSAR biode gradation</a>: Data set containing values for 41 attributes (molecular descriptors) used to classify 1055 ch emicals into 2 classes (ready and not ready biodegradable). 344. <a href="datasets/ QtyT40I10D100K">QtyT40I10D100K</a>: Since there is no numerical sequential data stream available in s tandard data sets, this data set is generated from the original T40I10D100K data set 34 5. <a href="datasets/Quadruped+Mammals">Quadruped Mammals</a>: The file animals.c is a data g enerator of structured instances representing quadruped animals 346. <a href="data sets/Qualitative+Structure+Activity+Relationships">Qualitative Structure Activity Relationships</a>: Two set s of datasets are given: pyrimidines and triazines 347. <a href="datasets/Qualitative _Bankruptcy">Qualitative_Bankruptcy</a>: Predict the Bankruptcy from Qualitative parameters from expert s. 348. <a href="datasets/Quality+Assessment+of+Digital+Colposcopies">Quality As sessment of Digital Colposcopies</a>: This dataset explores the subjective quality assessment of digital col poscopies. 349. <a href="datasets/Real+estate+valuation+data+set">Real estate val

uation data set</a>: The “real estate valuation” is a regression problem. The market historical data set of r eal estate valuation are collected from Sindian Dist., New Taipei City, Taiwan. 350. <a href="datasets/REALDISP+Activity+Recognition+Dataset">REALDISP Activity Recognition Dataset</a>: The REALDISP dataset is devised to evaluate techniques dealing with the effects of sensor displacement in we arable activity recognition as well as to benchmark general activity recognition algorithms 351. <a href="datasets/Record+Linkage+Comparison+Patterns">Record Linkage Comparison Patterns</ a>: Element-wise comparison of records with personal data from a record linkage setting. The task is to de cide from a comparison pattern whether the underlying records belong to one person. 3 52. <a href="datasets/Relative+location+of+CT+slices+on+axial+axis">Relative location of CT slices on axial axis</a>: The dataset consists of 384 features extracted from CT images. The class variable is numeric an d denotes the relative location of the CT slice on the axial axis of the human body. 353. <a href="datasets/Repeat+Consumption+Matrices">Repeat Consumption Matrices</a>: The dataset co ntains 7 datasets of User - Item matrices, where each entry represents how many times a user consumed an it em. Item is used as an umbrella term for various categories. 354. <a href="datasets/ Residential+Building+Data+Set">Residential Building Data Set</a>: Data set includes construction cost, sal e prices, project variables, and economic variables corresponding to real estate single-family residential apartm ents in Tehran, Iran. 355. <a href="datasets/Restaurant+%26+consumer+data">Re staurant & consumer data</a>: The dataset was obtained from a recommender system prototype. Th e task was to generate a top-n list of restaurants according to the consumer preferences. 356. <a href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+coll ection">Reuters RCV1 RCV2 Multilingual, Multiview Text Categorization Test collection</a>: This test collec tion contains feature characteristics of documents originally written in five different languages and their translati ons, over a common set of 6 categories. 357. <a href="datasets/Reuters+Transcrib ed+Subset">Reuters Transcribed Subset</a>: This dataset is created by reading out 200 files from the 10 l argest Reuters
classes and using an Automatic Speech Recognition system to create
corresponding transcriptions. 358. <a href="datasets/Reuters-21578+Text+Categori zation+Collection">Reuters-21578 Text Categorization Collection</a>: This is a collection of documents tha t appeared on Reuters newswire in 1987. The documents were assembled and indexed with categories. 359. <a href="datasets/Reuter_50_50">Reuter_50_50</a>: The dataset is used for aut horship identification in online Writeprint which is a new research field of pattern recognition. 360. <a href="datasets/Rice+Leaf+Diseases">Rice Leaf Diseases</a>: There are three classes/di seases: Bacterial leaf blight, Brown spot, and Leaf smut, each having 40 images. The format of all images is jpg . 361. <a href="datasets/Robot+Execution+Failures">Robot Execution Failures</a> : This dataset contains force and torque measurements on a robot after failure detection. Each failure is ch aracterized by 15 force/torque samples collected at regular time intervals 362. <a hr ef="datasets/Roman+Urdu+Data+Set">Roman Urdu Data Set</a>: Roman Urdu (the scripting style for Urd u language) is one of the limited resource languages.A data corpus comprising of more than 20000 records wa s collected. 363. <a href="datasets/Sales_Transactions_Dataset_Weekly">Sales_Tr ansactions_Dataset_Weekly</a>: Contains weekly purchased quantities of 800 over products over 52 wee ks. Normalised values are provided too. 364. <a href="datasets/SCADI">SCADI</a> : First self-care activities dataset based on ICF-CY. 365. <a href="datasets/SEC OM">SECOM</a>: Data from a semi-conductor manufacturing process 366. <a href="datasets/seeds">seeds</a>: Measurements of geometrical properties of kernels belonging to three d ifferent varieties of wheat. A soft X-ray technique and GRAINS package were used to construct all seven, real-v alued attributes. 367. <a href="datasets/seismic-bumps">seismic-bumps</a>: T he data describe the problem of high energy (higher than 10^4 J) seismic bumps forecasting in a coal
mine. Data come from two of longwalls located in a Polish coal mine. 368. <a href="d atasets/Semeion+Handwritten+Digit">Semeion Handwritten Digit</a>: 1593 handwritten digits from around 80 persons were scanned, stretched in a rectangular box 16x16 in a gray scale of 256 values. 369. <a href="datasets/sEMG+for+Basic+Hand+movements">sEMG for Basic Hand movements</a> : The “sEMG for Basic Hand movements” includes 2 databases of surface electromyographic signals of 6 h and movements using Delsys' EMG System. Healthy subjects conducted six daily life grasps. 370. <a href="datasets/Sentence+Classification">Sentence Classification</a>: Contains sentence s from the abstract and introduction of 30 articles annotated with a modified Argumentative Zones annotation s cheme. These articles come from biology, machine learning and psychology. 371. <a href="datasets/Sentiment+Labelled+Sentences">Sentiment Labelled Sentences</a>: The dataset contain s sentences labelled with positive or negative sentiment. 372. <a href="datasets/ser+ Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machines%29">ser Kno wledge Modeling Data (Students' Knowledge Levels on DC Electrical Machines)</a>: The dataset is about t he users' learning activities and knowledge levels on subjects of DC Electrical Machines. The dataset had been

obtained from online web-courses" and reported in my Ph.D. Thesis. 373. <a href="datasets/Servo">Servo</a>: Data was from a simulation of a servo system 374. <a href="datasets/SGEMM+GPU+kernel+performance">SGEMM GPU kernel performance</a>: Running times for multiplying two 2048 x 2048 matrices using a GPU OpenCL SGEMM kernel with varying parameters (using the library 'CLTune'). 375. <a href="datasets/Shuttle+Landing+Control">Shuttle Landing Control</a>: Tiny database; all nominal values 376. <a href="datasets/SIFT10M">SIFT10M</a>: In SIFT10M, each data point is a SIFT feature which is extracted from Caltech-256 by the open source VLFeat library. The corresponding patches of the SIFT features are provided. 377. <a href="datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set">Simulated Falls and Daily Living Activities Data Set</a>: 20 falls and 16 daily living activities were performed by 17 volunteers with 5 repetitions while wearing 6 sensors (3.060 instances) that attached to their head, chest, waist, wrist, thigh and ankle. 378. <a href="datasets/SkillCraft1+Master+Table+Dataset">SkillCraft1 Master Table Dataset</a>: This data was used in Thompson et al. (2013). A list of possible game actions is discussed in Thompson, Blair, Chen, & Henrey (2013). 379. <a href="datasets/Skin+Segmentation">Skin Segmentation</a>: The Skin Segmentation dataset is constructed over B, G, R color space. Skin and Nonskin dataset is generated using skin textures from face images of diversity of age, gender, and race people. 380. <a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Living+%28AAL%29">Smartphone Dataset for Human Actvity Recognition (HAR) in Ambient Assisted Living (AAL)</a>: This data is an addition to an existing dataset on UCI. We collected more data to improve the accuracy of our human activity recognition algorithms applied in the domain of Ambient Assisted Living. 381. <a href="datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions">Smartphone-Based Recognition of Human Activities and Postural Transitions</a>: Activity recognition data set built from the recordings of 30 subjects performing basic activities and postural transitions while carrying a waist-mounted smartphone with embedded inertial sensors.
382. <a href="datasets/SML2010">SML2010</a>: This dataset is collected from a monitor system mounted in a domotic house. It corresponds to approximately 40 days of monitoring data. 383. <a href="datasets/SMS+Spam+Collection">SMS Spam Collection</a>: The SMS Spam Collection is a public set of SMS labeled messages that have been collected for mobile phone spam research. 384. <a href="datasets/Solar+Flare">Solar Flare</a>: Each class attribute counts the number of solar flares of a certain class that occur in a 24 hour period 385. <a href="datasets/Somerville+Happiness+Survey">Somerville Happiness Survey</a>: A data extract of a non-federal dataset posted here https://catalog.data.gov/dataset/somerville-happiness-survey-responses-2011-2013-2015 386. <a href="datasets/Soybean+%28Large%29">Soybean (Large)</a>: Michalski's famous soybean disease database 387. <a href="datasets/Soybean+%28Small%29">Soybean (Small)</a>: Michalski's famous soybean disease database 388. <a href="datasets/Spambase">Spambase</a>: Classifying Email as Spam or Non-Spam 389. <a href="datasets/SPECT+Heart">SPECT Heart</a>: Data on cardiac Single Proton Emission Computed Tomography (SPECT) images. Each patient classified into two categories: normal and abnormal. 390. <a href="datasets/SPECTF+Heart">SPECTF Heart</a>: Data on cardiac Single Proton Emission Computed Tomography (SPECT) images. Each patient classified into two categories: normal and abnormal. 391. <a href="datasets/Spoken+Arabic+Digit">Spoken Arabic Digit</a>: This dataset contains timeseries of mel-frequency cepstrum coefficients (MFCCs) corresponding to spoken Arabic digits. Includes data from 44 male and 44 female native Arabic speakers. 392. <a href="datasets/Sponge">Sponge</a>: Data on sponges; Attributes in spanish 393. <a href="datasets/Sports+articles+for+objectivity+analysis">Sports articles for objectivity analysis</a>: 1000 sports articles were labeled using Amazon Mechanical Turk as objective or subjective. The raw texts, extracted features, and the URLs from which the articles were retrieved are provided. 394. <a href="datasets/Statlog+%28Australian+Credit+Approval%29">Statlog (Australian Credit Approval)</a>: This file concerns credit card applications. This database exists elsewhere in the repository (Credit Screening Database) in a slightly different form 395. <a href="datasets/Statlog+%28German+Credit+Data%29">Statlog (German Credit Data)</a>: This dataset classifies people described by a set of attributes as good or bad credit risks. Comes in two formats (one all numeric). Also comes with a cost matrix 396. <a href="datasets/Statlog+%28Heart%29">Statlog (Heart)</a>: This dataset is a heart disease database similar to a database already present in the repository (Heart Disease databases) but in a slightly different form 397. <a href="datasets/Statlog+%28Image+Segmentation%29">Statlog (Image Segmentation)</a>: This dataset is an image segmentation database similar to a database already present in the repository (Image segmentation database) but in a slightly different form. 398. <a href="datasets/Statlog+%28Landsat+Satellite%29">Statlog (Landsat Satellite)</a>: Multi-spectral values of pixels in 3x3 neighbourhoods in a satellite image, and the classification associated with the central pixel in each neighbourhood 399. <a href="datasets/Statlog+%28Shuttle%29">Statlog (Shuttle)</a>: The shuttle dataset contains 9 attributes all of which are numerical. Approximately 80% of the data belongs to class 1. 400. <a href="d

ch are numerical. Approximately 80% of the data belongs to class 1 400. <a href="datasets/Statlog+%28Vehicle+Silhouettes%29">Statlog (Vehicle Silhouettes)</a>: 3D objects within a 2D image by application of an ensemble of shape feature extractors to the 2D silhouettes of the objects. 401. <a href="datasets/Statlog+Project">Statlog Project</a>: Various Databases: Vehicle silhouttes, Landsat Sattelite, Shuttle, Australian Credit Approval, Heart Disease, Image Segmentation, German Credit 402. <a href="datasets/Steel+Plates+Faults">Steel Plates Faults</a>: A dataset of steel plates' faults, classified into 7 different types.
The goal was to train machine learning for automatic pattern recognition.
403. <a href="datasets/Stock+portfolio+performance">Stock portfolio performance</a>: The data set of performances of weighted scoring stock portfolios are obtained with mixture design from the US stock market historical database. 404. <a href="datasets/StoneFlakes">StoneFlakes</a>: Stone flakes are waste products of the stone tool production in
the prehistoric era. The variables are means of geometric and
stylistic features of the flakes contained in different inventories. 405. <a href="datasets/Student+Academics+Performance">Student Academics Performance</a>: The dataset tried to find the end semester percentage prediction based on different social, economic and academic attributes. 406. <a href="datasets/Student+Loan+Relational">Student Loan Relational</a>: Student Loan Relational Domain 407. <a href="datasets/Student+Performance">Student Performance</a>: Predict student performance in secondary education (high school). 408. <a href="datasets/Superconductivty+Data">Superconductivty Data</a>: Two file s contain data on 21 263 superconductors and their relevant features. 409. <a href="datasets/SUSY">SUSY</a>: This is a classification problem to distinguish between a signal process which produces supersymmetric particles and a background process which does not. 410. <a href="datasets/Synthetic+Control+Chart+Time+Series">Synthetic Control Chart Time Series</a>: This data consists of synthetically generated control charts. 411. <a href="datasets/Syskill+and+Webert+Web+Page+Ratings">Syskill and Webert Web Page Ratings</a>: This database contains HTML source of web pages plus the ratings of a single user on these web pages. Web pages are on four seperate subjects (Bands-recording artists; Goats; Sheep; and BioMedical) 412. <a href="datasets/Tamilnadu+Electricity+Board+Hourly+Readings">Tamilnadu Electricity Board Hourly Readings</a>: This data can be effectively produced the result to fewer parameter of the Load profile can be reduced in the Database 413. <a href="datasets/Tarvel+Review+Ratings">Tarvel Review Ratings</a>: Google reviews on attractions from 24 categories across Europe are considered. Google user rating ranges from 1 to 5 and average user rating per category is calculated. 414. <a href="datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015">Taxi Service Trajectory - Prediction Challenge, ECML PKDD 2015</a>: An accurate dataset describing trajectories performed by all the 442 taxis running in the city of Porto, in Portugal.
415. <a href="datasets/Teaching+Assistant+Evaluation">Teaching Assistant Evaluation</a>: The data consist of evaluations of teaching performance; scores are "low", "medium", or "high" 416. <a href="datasets/Tennis+Major+Tournament+Match+Statistics">Tennis Major Tournament Match Statistics</a>: This is a collection of 8 files containing the match statistics for both women and men at the four major tennis tournaments of the year 2013. Each file has 42 columns and a minimum of 76 rows. 417. <a href="datasets/Thoracic+Surgery+Data">Thoracic Surgery Data</a>: The data is dedicated to classification problem related to the post-operative life expectancy in the lung cancer patients: class 1 - death within one year after surgery, class 2 - survival. 418. <a href="datasets/Thyroid+Disease">Thyroid Disease</a>: 10 separate databases from Garavan Institute 419. <a href="datasets/Tic-Tac-Toe+Endgame">Tic-Tac-Toe Endgame</a>: Binary classification task on possible configurations of tic-tac-toe game 420. <a href="datasets/Trains">Trains</a>: 2 data formats (structured, one-instance-per-line) 421. <a href="datasets/Travel+Reviews">Travel Reviews</a>: Reviews on destinations in 10 categories mentioned across East Asia. Each traveler rating is mapped as Excellent(4), Very Good(3), Average(2), Poor(1), and Terrible(0) and average rating is used. 422. <a href="datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization">TTC-3600: Benchmark dataset for Turkish text categorization</a>: The TTC-3600 data set is a collection of Turkish news and articles including categorized 3,600 documents from 6 well-known portals in Turkey. It has 4 different forms in ARFF Weka format. 423. <a href="datasets/Turkiye+Student+Evaluation">Turkiye Student Evaluation</a>: This data set contains a total 5820 evaluation scores provided by students from Gazi University in Ankara (Turkey). There is a total of 28 course specific questions and additional 5 attributes. 424. <a href="datasets/TV+News+Channel+Commercial+Detection+Dataset">TV News Channel Commercial Detection Dataset</a>: TV Commercials data set consists of standard audio-visual features of video shots extracted from 150 hours of TV news broadcast of 3 Indian and 2 international news channels (30 Hours each). 425. <a href="datasets/Twenty+Newsgroups">Twenty Newsgroups</a>: This data set consists of 20000 messages taken from 20 newsgroups. 426. <a href="datasets/Twin+gas+sensor+a

rrays">Twin gas sensor arrays</a>: 5 replicates of an 8-MOX gas sensor array were exposed to different g as conditions (4 volatiles at 10 concentration levels each). 427. <a href="datasets/T witter+Data+set+for+Arabic+Sentiment+Analysis">Twitter Data set for Arabic Sentiment Analysis</a>: This problem of Sentiment Analysis (SA) has been studied well on the English language but not Arabic one. Two mai n approaches have been devised: corpus-based and lexicon-based. 428. <a href=" datasets/UbiqLog+%28smartphone+lifelogging%29">UbiqLog (smartphone lifelogging)</a>: UbiqLog is the smartphone lifelogging tool that runs on the smartphone of 35 users for about 2 months.
429. <a href="datasets/UJI+Pen+Characters">UJI Pen Characters</a>: Data co nsists of written characters in a UNIPEN-like format 430. <a href="datasets/UJI+Pen +Characters+%28Version+2%29">UJI Pen Characters (Version 2)</a>: A pen-based database with more t han 11k isolated handwritten characters 431. <a href="datasets/UJIIndoorLoc">UJII ndoorLoc</a>: The UJIIndoorLoc is a Multi-Building Multi-Floor indoor localization database to test Indoor Positioning System that rely on WLAN/WiFi fingerprint. 432. <a href="datasets/UJIIn doorLoc-Mag">UJIIndoorLoc-Mag</a>: The UJIIndoorLoc-Mag is an indoor localization database to test In door Positioning System that rely on Earth's magnetic field variations. 433. <a href=" datasets/Ultrasonic+flowmeter+diagnostics">Ultrasonic flowmeter diagnostics</a>: Fault diagnosis of four li quid ultrasonic flowmeters 434. <a href="datasets/Undocumented">Undocumented< /a>: Various datasets without documentation (feel free to explore!) 435. <a href= "datasets/University">University</a>: Data in original (LISP-readable) form 436. <a href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29">University of Tehran Question Dataset 2016 (UTQD.2016)</a>: Persian questions gathered from a jeopardy game broadcasted on Iranian national television. 437. <a href="datasets/UNIX+User+Data">UNIX User Data</a>: This file contains 9 sets of sanitized user data drawn from the command histories of 8 UNIX com puter users at Purdue over the course of up to 2 years. 438. <a href="datasets/Urba n+Land+Cover">Urban Land Cover</a>: Classification of urban land cover using high resolution aerial ima gery. Intended to assist sustainable urban planning efforts. 439. <a href="datasets/U RL+Reputation">URL Reputation</a>: Anonymized 120-day subset of the ICML-09 URL data containing 2. 4 million examples and 3.2 million features. 440. <a href="datasets/US+Census+Dat a+%281990%29">US Census Data (1990)</a>: The USCensus1990raw data set contains a one percent s ample of the Public Use Microdata Samples (PUMS) person records drawn from the full 1990 census sample.< /p> 441. <a href="datasets/User+Identification+From+Walking+Activity">User Identificatio n From Walking Activity</a>: The dataset collects data from an Android smartphone positioned in the chest pocket from 22 participants walking in the wild over a predefined path.
442. <a href="datasets/User+Knowledge+Modeling">User Knowledge Modeling</a> : It is the real dataset about the students' knowledge status about the subject of Electrical DC Machines. Th e dataset had been obtained from Ph.D. Thesis. 443. <a href="datasets/USPTO+Alg orithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Problem%3A+Pat">USP TO Algorithm Challenge, run by NASA-Harvard Tournament Lab and TopCoder Problem: Pat</a>: Data used for USPTO Algorithm Competition. Contains drawing pages from US patents with manually labeled figure and part labels. 444. <a href="datasets/Vertebral+Column">Vertebral Column</a></ b>: Data set containing values for six biomechanical features used to classify orthopaedic patients into 3 classe s (normal, disk hernia or spondilolysthesis) or 2 classes (normal or abnormal). 445. <a href="datasets/Vicon+Physical+Action+Data+Set">Vicon Physical Action Data Set</a>: The Physical Acti on Data Set includes 10 normal and 10 aggressive physical actions that measure the human activity. The data have been collected by 10 subjects using the Vicon 3D tracker. 446. <a href="datase ts/Victorian+Era+Authorship+Attribution">Victorian Era Authorship Attribution</a>: To create the largest aut horship attribution dataset, we extracted works of 50 well-known authors. To have a non-exhaustive learning, in training there are 45 authors whereas, in the testing, it's 50 447. <a href="datasets/V olcanoes+on+Venus+-+JARtool+experiment">Volcanoes on Venus - JARtool experiment</a>: The JARtool project was a pioneering effort to develop an automatic system for cataloging small volcanoes in the large set o f Venus images returned by the Magellan spacecraft. 448. <a href="datasets/Wall-F ollowing+Robot+Navigation+Data">Wall-Following Robot Navigation Data</a>: The data were collected as t he SCITOS G5 robot navigates through the room following the wall in a clockwise direction, for 4 rounds, using 24 ultrasound sensors arranged circularly around its 'waist'. 449. <a href="datasets/ Water+Treatment+Plant">Water Treatment Plant</a>: Multiple classes predict plant state 450. <a href="datasets/Waveform+Database+Generator+%28Version+1%29">Waveform Database G enerator (Version 1)</a>: CART book's waveform domains 451. <a href="datas ets/Waveform+Database+Generator+%28Version+2%29">Waveform Database Generator (Version 2)</a>: CART book's waveform domains 452. <a href="datasets/Wearable+Computing%3A+ Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29">Wearable Computing: Classification of Body Postures and Movements (PUC-Rio)</a>: A dataset with 5 classes (sitting-down, standing-up, standin g, walking, and sitting) collected on 8 hours of activities of 4 healthy subjects. We also established a baseline pe

rformance index. 453. <a href="datasets/Website+Phishing">Website Phishing</a>:

454. <a href="datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units">Weight Lifting Exercises monitored with Inertial Measurement Units</a>: Six young health subjects were asked to perform 5 variations of the biceps curl weight lifting exercise. One of the variations is the one predicted by the health professional. 455. <a href="datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29">WESAD (Wearable Stress and Affect Detection)</a>: WESAD (Wearable Stress and Affect Detection) contains data of 15 subjects during a stress-affect lab study, while wearing physiological and motion sensors. 456. <a href="datasets/Wholesale+customers">Wholesale customers</a>: The data set refers to clients of a wholesale distributor. It includes the annual spending in monetary units (m.u.) on diverse product categories 457. <a href="datasets/wiki4HE">wiki4HE</a>: Survey of faculty members from two Spanish universities on teaching uses of Wikipedia 458. <a href="datasets/Wilt">Wilt</a>: High-resolution Remote Sensing data set (Quickbird). Small number of training samples of diseased trees, large number for other land cover. Testing data set from stratified random sample of image. 459. <a href="datasets/Wine">Wine</a>: Using chemical analysis determine the origin of wines 460. <a href="datasets/Wine+Quality">Wine Quality</a>: Two datasets are included, related to red and white vinho verde wine samples, from the north of Portugal. The goal is to model wine quality based on physicochemical tests (see [Cortez et al., 2009], http://www3.dsi.uminho.pt/pcortez/wine/). 461. <a href="datasets/Wireless+Indoor+Localization">Wireless Indoor Localization</a>: Collected in indoor space by observing signal strengths of seven WiFi signals visible on a smartphone. The decision variable is one of the four rooms. 462. <a href="datasets/Yacht+Hydrodynamics">Yacht Hydrodynamics</a>: Delft data set, used to predict the hydodynamic performance of sailing yachts from dimensions and velocity. 463. <a href="datasets/YearPredictionMSD">YearPredictionMSD</a>: Prediction of the release year of a song from audio features. Songs are mostly western, commercial tracks ranging from 1922 to 2011, with a peak in the year 2000s. 464. <a href="datasets/Yeast">Yeast</a>: Predicting the Cellular Localization Sites of Proteins 465. <a href="datasets/YouTube+Comedy+Slam+Preference+Data">YouTube Comedy Slam Preference Data</a>: This dataset provides user vote data on which video from a pair of videos is funnier collected on YouTube Comedy Slam. The task is to automatically predict this preference based on video metadata. 466. <a href="datasets/YouTube+Multiview+Video+Games+Dataset">YouTube Multiview Video Games Dataset</a>: This dataset contains about 120k instances, each described by 13 feature types, with class information, specially useful for exploring multiview topics (cotraining, ensembles, clustering,..). 467. <a href="datasets/YouTube+Spam+Collection">YouTube Spam Collection</a>: It is a public set of comments collected for spam research. It has five datasets composed by 1,956 real messages extracted from five videos that were among the 10 most viewed on the collection period. 468. <a href="datasets/Z-Alizadeh+Sani">Z-Alizadeh Sani</a>: It was collected for CAD diagnosis. 469. <a href="datasets/Zoo">Zoo</a>: Artificial, 7 classes of animals </td></tr></table><hr/>
</td>
</tr>
</table>
<table align="center" cellpadding="5"><tr valign="center">
<td> Supported By: </td>
<td><img height="60" src="assets/nsfe.gif"/> </td>
<td> In Collaboration With: </td>
<td><img src="assets/rexaSmall.jpg"/></td>
</tr></table>

<a href="about.html">About</a> ||
<a href="citation_policy.html">Citation Policy</a> ||
<a href="donation_policy.html">Donation Policy</a> ||
<a href="contact.html">Contact</a> ||
<a href="http://cml.ics.uci.edu">CML</a>

</body></html>

In [179]:

```
try_1 = re.findall("b\>: (\w+.)",html_2)
```

try_1

Out[179]:

```
['Measurement ',
 '3D ',
 'This ',
 'This ',
 'Predict ',
 'The ',
 'The ',
 'This ',
 'The ',
 'This ',
 'Sequential ',
 'The ',
 'Predict ',
 'Contains ',
 'NASA ',
 "Amazon'",
 'The ',
 'Steel ',
 'Log ',
 'Acoustic ',
 'Experimental ',
 'The ',
 "ARCENE'",
 'Distinguish ',
 'Dataset ',
 'Nominal ',
 'Standardized ',
 'Exhaustive ',
 'This ',
 'This ',
 'Autistic ',
 'Autistic ',
 'Children ',
 'Revised ',
 'From ',
 'AutoUniv ',
 'The ',
 'The ',
 'Time-',
 'Badges ',
 'This ',
 'Balance ',
 'Data ',
 'The ',
 'Data ',
 'BAUM-',
 'A ',
 'The ',
 'This ',
 'This ',
 'This ',
 'Instances ',
 'In ',
 'Data ',
 'Breast ',
 'Clinical ',
 'Diagnostic ',
 'Original ',
 'Prognostic ',
```

'Prognostic ',
'Dataset ',
'User ',
'One ',
'This ',
'This ',
'This ',
'Derived ',
'This ',
'The ',
'Sanitized ',
'Predict ',
'This ',
'This ',
'Task:',
'Character ',
'Multiple,',
'6 ',
'Chess ',
'Knight ',
'King+',
'The ',
'ChIP-',
'This ',
'Given ',
'Little ',
'This ',
'This ',
'This ',
'The ',
'Communities ',
'Communities ',
'Relative ',
'Concrete ',
'Concrete ',
'Data ',
'The ',
'1984 ',
'Contains ',
'The ',
'The ',
'Speaker ',
'A ',
'Dataset ',
'This ',
'Forest ',
'This ',
'Crowdsourced ',
'This ',
'12 ',
'This ',
'Used ',
'The ',
'The ',
'This ',
'Data ',
'Recordings ',
'Features ',
'It ',
'This ',
'This ',
'Marine ',
'Aim ',

'Aim ',
'I ',
'This ',
'This ',
'DEXTER ',
'Generates ',
'This ',
'This ',
'This ',
'Discrete ',
'The ',
'Five ',
'Loop ',
'DOROTHEA ',
'Dota ',
'This ',
'This ',
'The ',
'Classify ',
'The ',
'The ',
'This ',
'This ',
'Data ',
'Predict ',
'Assorted ',
'Data ',
'This ',
'This ',
'Domain ',
'Educational ',
'This ',
'The ',
'This ',
'The ',
'The ',
'This ',
'These ',
'3 ',
'The ',
'This ',
'This ',
'This ',
'It ',
'Instances ',
'Facebook ',
'This ',
'100 ',
'The ',
'Given ',
'From ',
'FMA ',
'20 ',
'This ',
'Multi-',
'Cases ',
'This ',
'This ',
'A ',
'The ',
'The ',
'The ',

'100 ',
'This ',
'This ',
'A ',
'Instances ',
'The ',
'GISETTE ',
'From ',
'The ',
'The ',
'The ',
'This ',
'Design ',
'Dataset ',
'Topic:',
'Hepatocellular ',
'The ',
'4 ',
'From ',
'The ',
'The ',
'This ',
'Each ',
'The ',
'Well ',
'Pulsar ',
'Human ',
'The ',
'This ',
'Data ',
'The ',
'This ',
'Image ',
'This ',
'Handwriting ',
'Measurements ',
'This ',
'This ',
'This ',
'This ',
'Classification ',
'This ',
'Famous ',
'Data ',
'Includes ',
'This ',
'KASANDR ',
'KDC-',
'This ',
'This ',
'KEGG ',
'KEGG ',
'Relational ',
'From ',
'This ',
'This ',
'From ',
'A ',
'Database ',
'Database ',
'The ',
'BUPA ',

'Data ',
'All ',
'From ',
'126 ',
'Lung ',
'This ',
'The ',
'MADELON ',
'Data ',
'Discrimination ',
'Fault ',
'Mesothelioma”,
'Meta-',
'This ',
'The ',
'Expression ',
'MicroblogPCU ',
'A ',
'This ',
'The ',
'Learning ',
'5 ',
'E.',
'From ',
'Primate ',
'A ',
'Horn-',
'5 ',
'This ',
'This ',
'This ',
'5879 ',
'This ',
'From ',
'The ',
'The ',
'References ',
'Large ',
'Dataset ',
'This ',
'Corpus ',
'Nomao ',
'Northix ',
'This ',
'NYSK ',
'Experimental ',
'This ',
'Sixteen ',
'This ',
'This ',
'This ',
'Of ',
'The ',
'Open ',
'This ',
'This ',
'The ',
'This ',
'Two ',
'Used ',
'Two ',
'The ',

'The ',
'The ',
'PANDOR ',
'This ',
'Data ',
'Handwriting ',
'The ',
'The ',
'Oxford ',
'Oxford ',
'15 ',
'Digit ',
'This ',
'This ',
'The ',
'This ',
'This ',
'Bridges ',
'The ',
'Data ',
'This ',
'The ',
'Purpose ',
'The ',
'Dataset ',
'The ',
'From ',
'Assorted ',
'Undocumented<',
'This ',
'These ',
'Data ',
'Since ',
'Two ',
'Predict ',
'This ',
'The ',
'The ',
'Element-',
'The ',
'The ',
'Data ',
'The ',
'This ',
'This ',
'This ',
'The ',
'There ',
'This ',
'Roman ',
'Contains ',
'First ',
'Data ',
'Measurements ',
'The ',
'1593 ',
'The ',
'Contains ',
'The ',
'The ',
'Data ',
'Running ',

'Tiny ',
'In ',
'20 ',
'This ',
'The ',
'This ',
'Activity ',
'This ',
'The ',
'Each ',
'A ',
"Michalski'",
"Michalski'",
'Classifying ',
'Data ',
'Data ',
'This ',
'Data ',
'1000 ',
'This ',
'This ',
'This ',
'This ',
'Multi-',
'The ',
'3D ',
'Various ',
'A ',
'The ',
'Stone ',
'The ',
'Student ',
'Predict ',
'Two ',
'This ',
'This ',
'This ',
'This ',
'Google ',
'An ',
'The ',
'This ',
'The ',
'10 ',
'Binary ',
'2 ',
'Reviews ',
'The ',
'This ',
'TV ',
'This ',
'5 ',
'This ',
'UbiqLog ',
'Data ',
'A ',
'The ',
'The ',
'Fault ',
'Various ',
'Data ',
'Persian ',

'This ',
'Classification ',
'Anonymized ',
'The ',
'The ',
'It ',
'Data ',
'Data ',
'The ',
'To ',
'The ',
'The ',
'Multiple ',
'CART ',
'CART ',
'A ',
'Six ',
'WESAD ',
'The ',
'Survey ',
'High-',
'Using ',
'Two ',
'Collected ',
'Delft ',
'Prediction ',
'Predicting ',
'This ',
'This ',
'It ',
'It ',
'Artificial,']