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 7 5% 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 righest woman alive when the diverge is finalized."



```
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)
  a\tilde{n}o = 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 Al. 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 Al. 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 Il 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 Micro 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): Comput ational Microeconomics. COMPSCI 570 (Fall 2017): Artificial Intelligence. COMPSCI 590.3 (Fall 2017): Ethics a nd Al. COMPSCI 590.2 (Spring 2017): Computation, Information, and Learning in Market Design, COMPSCI 59 0.4 (Spring 2016): Computational Microeconomics: Game Theory, Social Choice, and Mechanism Design. CO MPSCI 290.4/590.4 (Spring 2015): Crowdsourcing Societal Tradeoffs. COMPSCI 570 (Fall 2014): Artificial Intelli gence. COMPSCI 590.4 (Spring 2014): Computational Microeconomics: Game Theory, Social Choice, and Mec hanism Design. COMPSCI 590.1 (Fall 2012): Linear and Integer Programming. COMPSCI 173 (Spring 2012): C omputational Microeconomics. COMPSCI 296.1 (Fall 2011): Computational Microeconomics: Game Theory, So cial 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 Microec onomics: Game Theory, Social Choice, and Mechanism Design. COMPSCI 170 (Spring 2009): Introduction to A rtificial 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 2 96.3 (Spring 2007): Topics in Computational Economics. COMPSCI 296.2 (Fall 2006): Computational Game Th eory 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', '590', '590', '173', '296', '173', '196', '170', '270', '196', '196', '296', '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]:

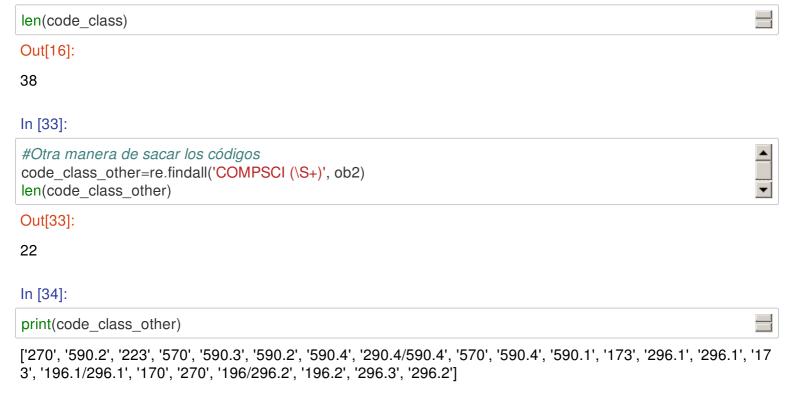
3

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', '590', '590', '173', '296', '296', '173', '196', '170', '270', '196', '196', '296', '296', '296', '590.4', '296.1', '296.2']

In [16]:



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, Mer hout, F, and Volfovsky, A. "Exposure to opposing views on social media can increase political polarization." P roceedings 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 e xamine 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." Am erican Sociological Review 82.6 (December 1, 2017): 1188-1213. Full Text. Bail, CA. "Taming Big Data: Usin g App Technology to Study Organizational Behavior on Social Media." Sociological Methods and Research 4 6.2 (March 1, 2017): 189-217. Full Text. McDonnell, TE, Bail, CA, and Tavory, I. "A Theory of Resonance." S ociological 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." Proc eedings 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 Auti sm 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), 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=&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



import pandas as nd

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 Mac hine Learning Repository: Data Sets</title>\n\n<!-- Stylesheet link -->\nlink 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 t erms.");\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// re turn 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(radio Obj.checked)\n\t\treturn radioObj.value;\n\t\telse\n\t\treturn "";\n\tfor(var i = 0; i < radioLength; i++) {\n\t\tif(radioLength) and in the content of the content n<body>\n\n<!-- SITE HEADER (INCLUDES LOGO AND SEARCH BOX) -->\n\n<!-- SITE HEADER (INCLUDE S LOGO AND SEARCH BOX) -->\n\n\n\n\t\n\t\t
br> p; Center for M achine Learning and Intelligent Systems\n\t\n\t<td width=100% valign=top align="right" >\n\t\\n\t\About \n\t\Cit ation Policy \n\t\tDonate a Data Set \n\t\tContact\n\t\t\n\n\t\t
\n\t\t<!-- Search Google -->\n\n\t\t<FORM method=GET acti on=http://www.google.com/custom onsubmit="return checkform(this);">\n\t\t<INPUT TYPE=text name=q size=3 0 maxlength=255 value="">\n\t\t<INPUT type=submit name=sa VALUE="Search">\n\t\t<INPUT type=hidden na me=cof VALUE="AH:center;LH:130;L:http://archive.ics.uci.edu/assets/logo.gif;LW:384;AWFID:869c0b2eaa8d51 8e;">\n\t\t<input type=hidden name=domains value="ics.uci.edu">\n\t\t
\n\t\t<input type=radio name=sitesea rch value="ics.uci.edu" checked> Repository\n\t\t<input t ype=radio name=sitesearch value=""> Web \n\t\t & nbsp; \n\t\<IM G SRC=http://www.google.com/logos/Logo_25blk.gif border=0 ALT=Google align=middle height=27>\n\t\t< br>\n\t\t</FORM>\n\t\t<!-- Search Google -->\n\n\n\t\tView ALL Data Sets\n\t\t
\n\t\n\n r />\n\n\n\t\n\t\t\n\n\t\t\tBr owse Through:\n\n\n\r\n\t\t<td bgcolor="#" 003366">Default Task\t\n\t\t\n\t<\tr>\n\t<p class= "normal">Classification (350)
Regression (96)
color=red>(96)
font>
color=red>(96)
font>
color=red>(96)
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
font>
 href=\'datasets.php?format=&task=clu&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table\'>Clus tering (84)
Other < font color=red>(55)\t/n\t$ > Attribute Type \t\n\t\n\t\n\t\n\t<td val ign=top>Categorical (38)
Numerical (30 7)
Data Type \t\n\t\t\n\t\n\t\t<a hre f=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=mvar&sort=nameUp&view=table\'>Multiv ariate (357)
Univariate (23)
<a href=\'datasets.

```
php?format=&task=&att=&area=&numAtt=&numIns=&type=seq&sort=nameUp&view=table\'>Sequential</a> <fo
nt color=red>(47)</font><br><a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=ts&s
ort=nameUp&view=table\'>Time-Series</a> <font color=red>(91)</font><br/>a href=\'datasets.php?format=&ta
sk=&att=&area=&numAtt=&numIns=&type=text&sort=nameUp&view=table\'>Text</a> <font color=red>(53)</fon
t><br><a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=dt&sort=nameUp&view=tab
le\'>Domain-Theory</a> <font color=red>(23)</font><br><a href=\'datasets.php?format=&task=&att=&area=&n
umAtt=&numIns=&type=other&sort=nameUp&view=table\'>Other</a> <font color=red>(21)</font><br>\t\n\t\
tn<</tr>/n<</tr>/n<</tr>/n<</tr>/n<</tr>/n/n/n/n/n/n/n/n/n/n/n/n/n/n/n/n/n/n/n/n/n/n/n/n/n//
t\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><br/>or><a href=\'datasets.php?f
ormat=&task=&att=&area=phys&numAtt=&numIns=&type=&sort=nameUp&view=table\'>Physical Sciences</a>
<font color=red>(49)</font><br><a href=\'datasets.php?format=&task=&att=&area=comp&numAtt=&numIns=&ty
pe=&sort=nameUp&view=table\'>CS / Engineering</a> <font color=red>(170)</font><br/>or> <a href=\'datasets.php
?format=&task=&att=&area=soc&numAtt=&numIns=&type=&sort=nameUp&view=table\'>Social Sciences</a> <f
ont color=red>(26)</font><br><a href=\'datasets.php?format=&task=&att=&area=bus&numAtt=&numIns=&type=
&sort=nameUp&view=table\'>Business</a> <font color=red>(29)</font><br><a href=\'datasets.php?format=&tas
k=&att=&area=game&numAtt=&numIns=&type=&sort=nameUp&view=table\'>Game</a> <font color=red>(10)</f
ont><br><a href=\'datasets.php?format=&task=&att=&area=other&numAtt=&numIns=&type=&sort=nameUp&vie
w=table\'>Other</a> <font color=red>(73)</font>\t\n\t\t\n\t<p class
="whitetext"><b># Attributes</b>\r\\n\t\\n\t\n\t\n\t\t<a href=\'
datasets.php?format=&task=&att=&area=&numAtt=less10&numIns=&type=&sort=nameUp&view=table\'>Less th
an 10</a> <font color=red>(113)</font><br><a href=\'datasets.php?format=&task=&att=&area=&numAtt=10to1
00&numIns=&type=&sort=nameUp&view=table\'>10 to 100</a> <font color=red>(210)</font><br><a href=\'data
sets.php?format=&task=&att=&area=&numAtt=greater100&numIns=&type=&sort=nameUp&view=table\'>Greate
r than 100</a> <font color=red>(84)</font>\t\n\t\t\n\t<\tr>\n\t(84)</font>\t
etext"><b>\# Instances</b>\n\t\t\n\n\t\t\n\t\t<a href='d to the following of 
atasets.php?format=&task=&att=&area=&numAtt=&numIns=less100&type=&sort=nameUp&view=table\'>Less th
an 100</a> <font color=red>(27)</font><br><a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numI
ns=100to1000&type=&sort=nameUp&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>\t\n\t\t\n\t<\tr><p cl
ass="whitetext"><b>Format Type</b>\t\n\t\\n\t\n\t\t<a h
ref=\'datasets.php?format=mat&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table\'>Matri
x</a> <font color=red>(324)</font><br><a href=\'datasets.php?format=nonmat&task=&att=&area=&numAtt=&n
umIns=&type=&sort=nameUp&view=table\'>Non-Matrix</a> <font color=red>(145)</font>\t\n\t\
> \n\t \n\n\t< d valign=top>\n\n\n\t\t \n\t\t< d>>< class="big">< b>469</b> Da
ta Sets\n\t\t<font color=gray>Table View</font>&nbsp;&nbsp;<a
href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list\'>List View
</a>\n\t\\n\t\n\t\t<td class="nor
mal, whitetext"><a href=\'datasets.php?format=&task=&att=&area=&numAtt=&nu
mIns=&type=&sort=nameDown&view=table\'><b>Name</b></a>\n\t\t<!-- <td><p class="normal, whitet
ext"><b>Abstract</b> -->\n\t\t<a href=\'datasets.php?format=&task=
&att=&area=&numAtt=&numIns=&type=&sort=typeUp&view=table\'><b>Data Types</b></a>\n\t\t<
p class="normal, whitetext"><a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort
=taskUp&view=table\'><b>Default Task</b></a>\n\t\t<a href=\'datas
ets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=attTypeUp&view=table\'><b>Attribute Typ
es</b></a>\n\t\t<a href=\'datasets.php?format=&task=&att=&area=&
numAtt=&numIns=&type=&sort=instUp&view=table\'><b># Instances</b></a>\n\t\t<p class="norm"
al, whitetext"><a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=attUp&view=t
able\'><b># Attributes</b></a>\n\t\t<a href=\'datasets.php?format=&
task=&att=&area=&numAtt=&numIns=&type=&sort=dateUp&view=table\'><b>Year</b></a>\n\n\t\<!-- <
td><b>Area</b> -->\n\n\t\t\n\t\t\t<a href="
datasets/Abalone"><img src="assets/MLimages/SmallLarge1.jpg" border=1 /></a>&nbsp;<p class="no
rmal"><b><a href="datasets/Abalone">Abalone</a></b>\n\t\t\t<!-- <td><p class="nor
mal">Predict the age of abalone from physical measurements  -->\n\t\t
Multivariate \n\t\tClassification \n\t\t\t
mal">Categorical, Integer, Real \n\t\t4177 \n\t\t\t<p
class="normal">8\ \n\t\t\t1995\ \n\t\t\t<!-- <td>1995\&nbsp;
ormal">Life  -->\n\t\t\\n\t\t\t<a href="datasets/A"
dult"><img src="assets/MLimages/SmallLarge2.jpg" border=1 /></a>&nbsp;<b><a
href="datasets/Adult">Adult</a></b>\n\t\t\t<!-- <td>Predict wheth
er income exceeds $50K/yr based on census data. Also known as "Census Income" dataset.
```

```
->\n\t\t<\td>Multivariate&nbsp;\n\t\t<\td>Classification&nbsp;
p>\n\t\tCategorical, Integer \n\t\t48842&
nbsp;   n\t t  p class = "normal" > 14 & nbsp;   (n t t  p class = "normal" > 1996 & nbsp;   (n t t  p class = "normal" > 1996 & nbsp;  
td > \ln t/t < -- < td > p class = "normal" > Social & nbsp;  -/ td > -- > \ln t/t / t  -table >  -td > < a href = td > \n/t / t/t  \n/t / t/t  -- > \n/t / t/t < \n/t < \n/
"datasets/Annealing"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p cl
ass="normal"><b><a href="datasets/Annealing">Annealing</a></b>\n\t\t\t<!-- <td><p
class="normal">Steel annealing data  -->\n\t\t\tMultivariate </t
d>\n\t\t\Categorical, Integer,
Real \n\t\t\t798 \n\t\t\t38 
p>\n\t\t\t \n\t\t<!-- <td>Physical&nbsp;
> -->\n\t\t\n\t\t<a href="datasets/Anonymous+Microsoft+Web"> -->\n\t\t</a>
+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="normal"
><b><a href="datasets/Anonymous+Microsoft+Web+Data">Anonymous Microsoft Web Data</a></b>
/tr>\n\t\t<!-- <td>Log of anonymous users of www.microsoft.com; predict area
s of the web site a user visited based on data on other areas the user visited.  -->\n\t\t\t<p c
>Categorical \n\t\t37711 \n\t\t>
294 \n\t\t\t1998 \n\t\t\t<!-- <td>1998&nbsp;
ss="normal">Computer  -->\n\t\t\z/tr>\n\t\t\t<a href="datasets/Arrhythmi">ctr><a 
a"><img src="assets/MLimages/SmallLarge5.jpg" border=1 /></a>&nbsp;<b><a hr
ef="datasets/Arrhythmia">Arrhythmia</a></b>\n\t\t<!-- <td>Disti
nguish between the presence and absence of cardiac arrhythmia and classify it in one of the 16 groups.  
on \n\t\tCategorical, Integer, Real \n\t\t<p class="
normal">452\ \n\t\t\t279\ \n\t\t\t19
n\t\t\t<a href="datasets/Artificial+Characters"><img src="assets/MLimages/SmallLarge6.jpg"
"border=1 /></a>&nbsp;<b><a href="datasets/Artificial+Characters">Artificial Char
acters</a></b>\n\t\t\t<!-- <td>Dataset artificially generated by usi
ng first order theory which describes structure of ten capital letters of English alphabet  -->\n\t\t\
Multivariate \n\t\tClassification \
n\t\t\tCategorical, Integer, Real \n\t\t6000&nbs
p;\n\t\t7 \n\t\t1992 \n
\t\t\t<!-- <td>Computer&nbsp; -->\n\t\t\t\n\t\t\t<a href="
datasets/Audiology+%28Original%29"><img src="assets/MLimages/SmallLarge7.jpg" border=1 /></a>&nbsp;</t
d><b><a href="datasets/Audiology+%28Original%29">Audiology (Original)</a></b>
\n\t\t<!-- <td>c/p class="normal">Nominal audiology dataset from Baylor&nbsp;
-->\n\t\t\tMultivariate \n\t\t\tClassification <
/p>\n\t\tCategorical \n\t\t\t226 </
Life  -->\n\t\t\\n\t\t\t<a
href="datasets/Audiology+%28Standardized%29"><img src="assets/MLimages/SmallLarge7.jpg" border=1 /></
a> <b><a href="datasets/Audiology+%28Standardized%29">Audiology (Stan
dardized)</a></b>\n\t\t\t<!-- <td>Standardized version of the original control of
nal audiology database  -->\n\t\t\tclass="normal">Multivariate \n\t\t\t
Classification \n\t\tCategorical \n\t\t\
tp class="normal">226 \n\t\tp class="normal">69 \n\t\tp class="normal">69 
="normal">1992 \n\t\t<!-- <td>Life&nbsp; -->\n\t\t\
d><a href="datasets/Auto+MPG"><img src="assets/MLimages/SmallLarge9.jpg" border=1 /></a
> <b><a href="datasets/Auto+MPG">Auto MPG</a></b>
e>\n\t\t\<!-- <td>Revised from CMU StatLib library, data concerns city-cycle fuel consu
mption  -->\n\t\t\tMultivariate \n\t\t\t<p class="normal"
">Regression \n\t\tCategorical, Real \n\t\t<p class
= "normal" > 398 \& nbsp;   \n\t  p class = "normal" > 8 \& nbsp;   \n\t  p class = "normal" > 19
93 \n\t\t\<!-- <td>Other&nbsp; -->\n\t\t\<tr bgcolor="DDEEFF"
>\n\t\t<a href="datasets/Automobile"><img src="assets/MLimages/SmallLarge10.jpg" bord
er=1 /></a>&nbsp;<b><a href="datasets/Automobile">Automobile</a></b>
>\n\t\t<!-- <td>From 1985 Ward\'s Automotive Yearbook&nbsp;
-->\n\t\t\ctd>Multivariate \n\t\t\ctd>Regression </p
>\n\t\t\ctd>Categorical, Integer, Real \n\t\t\ctd>205
 \n\t\t26 <math>\n\t\t\t1987&nbsp;<math>
/n\t/t<--< td><n class="normal">Other&nbsp: -->\n\t\t
```

```
="datasets/Badges"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p cla
ss="normal"><b><a href="datasets/Badges">Badges</a></b>\n\t\t\t<!-- <td><p class
="normal">Badges labeled with a "+" or "-" as a function of a person\'s name  -->\n\t\t\t<p cl
ass="normal">Univariate, Text \n\t\tClassification \n\t\t\t
class="normal"> \n\t\t\tclass="normal">294 \n\t\t\tp class="n
ormal">1 \n\t\t1994 \n\t\t\t<!-- <td>
Other  -->\n\t\t\t\n\t\t\t<a href="datasets/Balanc">datasets/Balanc</a>
e+Scale"><img src="assets/MLimages/SmallLarge12.jpg" border=1 /></a>&nbsp;<
b><a href="datasets/Balance+Scale">Balance Scale</a></b>\n\t\t\t<!-- <td><p class
="normal">Balance scale weight & distance database  -->\n\t\tMultivari
ate \n\t\t\tClassification \n\t\t\tCat
egorical \n\t\t625 <math>\n\t\t4&nbsp
;\n\t\t1994 \n\t\t<!-- <td>Social&nbsp;</p
> -->\n\t\t\n\t\t\t<dt><a href="datasets/Balloons"><img src="assets/MLimages/Sm">
allLarge13.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Balloons">Balloons
a></b>\table>\n\t\t<!-- <td>Data previously used in cognitive psycholog
y experiment; 4 data sets represent different conditions of an experiment  -->\n\t\t<p class
="normal">Multivariate \n\t\tClassification \n\t\t\t<p
class="normal">Categorical \n\t\t\t16 \n\t\t\t<p class
= "normal">4\  \n\t\t\t  \n\t\t\t<!-- <td>So
cial  -->\n\t\t\t bgcolor="DDEEFF">\n\t\t\t<a href="datasets/Breast+C">td><a href=
ancer"><img src="assets/MLimages/SmallLarge14.jpg" border=1 /></a>&nbsp;<b>
<a href="datasets/Breast+Cancer">Breast Cancer</a></b>\n\t\t<!-- <td><p class="
normal">Breast Cancer Data (Restricted Access)  -->\n\t\t\tMultivariate&
nbsp;\n\t\t\Classification \n\t\t\tCategor
ical\nbsp;  \n\t\t286   \n\t\t9  
/n\t/t<<d>1988&nbsp;/n\t/t<!-- <td>Life&nbsp;-
->\n\t\t\n\t\t<a href="datasets/Breast+Cancer+Wisconsin+%28Original%29"><im
g src="assets/MLimages/SmallLarge14.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="da
tasets/Breast+Cancer+Wisconsin+%28Original%29">Breast Cancer Wisconsin (Original)</a></b>
\n\t\t\<!-- <td>Original Wisconsin Breast Cancer Database&nbsp;
n\t\tMultivariate \n\t\tClassification 
/n\t\tfnteger /n\t\t699 /n\t\t699 /n\t\t699 
t/tp class="normal">10 \n/t/t/tp class="normal">1992 \n/t/t/t<!-- <td>= = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = 
p class="normal">Life  -->\n\t\t\\n\t\t\<a href="
datasets/Breast+Cancer+Wisconsin+%28Prognostic%29"><img src="assets/MLimages/SmallLarge14.jpg" bord
er=1 /></a>&nbsp;<b><a href="datasets/Breast+Cancer+Wisconsin+%28Prognosti
c%29">Breast Cancer Wisconsin (Prognostic)</a></b>\n\t\t<!-- <td><p class="norm"
al">Prognostic Wisconsin Breast Cancer Database  -->\n\t\tMultivariate
 \n\t\tClassification, Regression \n\t\tClassification, Regression 
rmal">Real \n\t\t198 \n\t\t34&
nbsp;   nt/t < d > p class = "normal" > 1995 & nbsp;   nt/t < !-- < td > p class = "normal" > Life & nbsp;  
/p> -->\n\t\t\\n\t\t\t<a href="datasets/Breast+Cancer+Wisconsin+%28Diagnost"
ic%29"><img src="assets/MLimages/SmallLarge14.jpg" border=1 /></a>&nbsp;<b>
<a href="datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29">Breast Cancer Wisconsin (Diagnostic)</a>
b>b>h\t\t\<!-- <td>Diagnostic Wisconsin Breast Cancer Database&
nbsp;-->\n\t\t\tclass="normal">Multivariate \n\t\t\tclass="normal">Classi
fication \n\t\t\tReal \n\t\t\t569&nb
sp;n/t/tp class="normal">32 n/t/t/tp class="normal">1995 
\ \n\t\t\t<!-- <td>Life&nbsp; -->\n\t\t\t bgcolor="DDEEFF">\n\t\t\t<
tr><a href="datasets/Pittsburgh+Bridges"><img src="assets/MLimages/SmallLarge18.jpg" border=1 /></a>
 <b><a href="datasets/Pittsburgh+Bridges">Pittsburgh Bridges</a></b>
\n\t\t<!-- <td>class="normal">Bridges database that has original and numeric-discreti
zed datasets  -->\n\t\tclass="normal">Multivariate \n\t\tclass="
normal">Classification \n\t\tCategorical, Integer \n\t\t\t<
td>108 \n\t\t\t13 \n\t\t\t<p class="
d><a href="datasets/Car+Evaluation"><img src="assets/MLimages/SmallLarge19.jpg" border=1
/></a>&nbsp;<b><a href="datasets/Car+Evaluation">Car Evaluation</a></b></
td>\n\t\t<!-- <td>Derived from simple hierarchical decision model, this dat
abase may be useful for testing constructive induction and structure discovery methods. 
-->\n\t\
t/t/td>n class="normal"-Multivariate&nhen:/n>//td\n/t/t/td>n class="normal"-Classification&nhen:/n>//td
```

```
mornial situativa latour bop, vps vias in the text vp olabo-
>\n\t\t\tCategorical \n\t\t\t1728 \n
\t\t\6 \n\t\t1997 \n\t\t<!-- <td>
Other  -->\n\t\t\t\n\t\t\t<a hre
f="datasets/Census+Income"><img src="assets/MLimages/SmallLarge2.jpg" border=1 /></a>&nbsp;
p class="normal"><b><a href="datasets/Census+Income">Census Income</a></b>
n\t\t<!-- <td>Predict whether income exceeds $50K/yr based on census data. Also known
as "Adult" dataset.  -->\n\t\t\tMultivariate \n\t\t\t<p cl
ass="normal">Classification \n\t\tCategorical, Integer \
n\t\t48842\nbsp;
<math display="block">n\t\t14\nbsp;
<math display="block">n\t\t14\nbsp;

p\ class="normal">1996\ \n\t\t\<!--<td><p\ class="normal">Social\&nbsp;\-->\n\t\t\<p\ class="normal">Social&nbsp;
r>\n\t\t<a href="datasets/Chess+%28King-Rook+vs.+King-Knight%29"><img src="assets/"
MLimages/SmallLarge24.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Chess"
+%28King-Rook+vs.+King-Knight%29">Chess (King-Rook vs. King-Knight)</a></b>
n\t\t\t<!-- <td>Knight Pin Chess End-Game Database Creator&nbsp; -->\n\t\t\t<
p class="normal">Multivariate, Data-Generator \n\t\tClassification&nbs
p;\n\t\t\tCategorical, Integer \n\t\t\t&nbs
p;\n\t\t22 \n\t\t1988 
n\t\t<!-- <td>Game&nbsp; -->\n\t\t\t bgcolor="DDEEFF">\n\t\t\t
<a href="datasets/Chess+%28King-Rook+vs.+King-Pawn%29"><img src="assets/MLimages/SmallLarg"><a href="datasets/Chess+%28King-Rook+vs.+King-Pawn%29"><img src="assets/MLimages/SmallLarg"><a href="datasets/Chess+%28King-Rook+vs.+King-Pawn%29"><img src="assets/MLimages/SmallLarg"><a href="datasets/Chess+%28King-Rook+vs.+King-Pawn%29"><a href="datasets/Chess+%28King-Rook-vs.+King-Pawn%29"><a href="datasets/Chess+%28King-Rook-vs.+King-Pawn%29"><a href="datasets/Chess+%28King-Rook-vs.+King-Pawn%29"><a href="datasets/Chess+%28King-Rook-vs.+King-Pawn%29"><a href="datasets/Chess+%28King-Rook-vs.+King-Pawn%29"><a href="datasets/Chess+%28King-Rook-vs.+King-Pawn%29"><a href="datasets/Chess+%28King-Rook-vs.+King-Pawn%29"><a href="datasets/Chess+%28King-Rook-vs.+King-Pawn%29"><a href="datasets/Chess+%28King-Pawn"><a href="datasets/Chess+%28King-Pawn"><a href="datasets/Chess+%28King-Pawn"><a href="datasets/Chess+%28King-Pawn"><a href="datasets/Chess+%28King-Pawn"><a href="datasets/Chess+%28King-Pawn"><a href="data
e24.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Chess+%28King-Rook+vs.+"
King-Pawn%29">Chess (King-Rook vs. King-Pawn)</a></b>\n\t\t\<!-- <td><p class=
"normal">King+Rook versus King+Pawn on a7 (usually abbreviated KRKPA7).  -->\n\t\t\t<p
class="normal">Multivariate \n\t\t\tClassification \n\t\t\t<t
d>Categorical \n\t\t\t3196 \n\t\t\t
<p class="normal">36 </p></td>\n\t\t<td><p class="normal">1989 </p></td>\n\t\t<!-- <td><p class="normal">1989 </p></td>\n\t\t<!-- <td><p class="normal">1989 </p></td>\n\t\t<!-- <td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td><td>
s="normal">Game\  --> \n\t\t \n\t\t< a href="datasets/Chess+%28Ki">tt>< a hr
ss="normal"><b><a href="datasets/Chess+%28King-Rook+vs.+King%29">Chess (King-Rook vs. King)</a></b>
\n\t\t\<!-- <td>Chess Endgame Database for White King and Roo
k against Black King (KRK). -->\n\t\tMultivariate \n\t\t
Classification \n\t\tCategorical, Integer </
p  n t < d < p class = "normal" > 28056 & nbsp;   n t  class = "normal" > 6 & nbsp;   \n t  class = "normal" > 6 & nbsp;   \n t  class = "normal" > 6 & nbsp;   \n t  \n t  class = "normal" > 6 & nbsp;   \n t  \n t < \n t < td > \n t < \n \\ t < \n t < \n
\t1994 \n\t\t\t<!-- <td>Game&nbsp; -->\n\t\t\
t\n\t\t\t<a href="datasets/Chess+%28Domain+Theories%29"><i
mg src="assets/MLimages/SmallLarge24.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="
datasets/Chess+%28Domain+Theories%29">Chess (Domain Theories)</a></b>
t\t<!-- <td>6 different domain theories for generating legal moves of chess&nbsp;
->\n\t\tDomain-Theory \n\t\t \n
\t < d > p class = "normal" > \nbsp;   \n  p class = "normal" > \nbsp;  
ormal"> \n\t\t\t \n\t\t\t<!-- <td>Game&
nbsp; -->\n\t\t\t\n\t\t\t<a href="datasets/Bach+Chorales"><img src="assets"
/MLimages/SmallLarge25.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Bach+
Chorales">Bach Chorales</a></b>\n\t\t\t<!-- <td>class="normal">Time-series dat
a based on chorales; challenge is to learn generative grammar; data in Lisp  -->\n\t\t\t<p cl
ass="normal">Univariate, Time-Series \n\t\t \n\t\t
Categorical, Integer \n\t\t\t100 \n\t\t\t
6 & nbsp; \n\t\t & nbsp; \n\t\t\t<!-- < td> & nbsp; 
"normal">Other  -->\n\t\t\\n\t\t\<a href="datase"
ts/Connect-4"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="n
ormal"><b><a href="datasets/Connect-4">Connect-4</a></b>\n\t\t<!-- <td><p class
="normal">Contains connect-4 positions  -->\n\t\tMultivariate, Spatial&
nbsp;\n\t\tClassification \n\t\t\tCategor
ical\ \n\t\t\t67557 \n\t\t\t42 
\n\t\t1995 \n\t\t<!-- <td>Game&nbsp;
-->\n\t\t\n\t\t<a href="datasets/Credit+Approval"><img src="assets/MLimag"><img src="assets/MLimag"></img src="assets/MLimag"></assets/MLimag</assets/MLimag</assets/MLimag</assets/MLimag</assets/MLimag</assets/MLimag</assets/MLimag</assets/MLimag</assets/MLimag</assets/MLimag</assets/MLimag</assets/MLimag</assets/MLimag</assets/MLimag</assets/MLimag</assets/MLimag</assets/MLimag</assets/MLimag</assets/MLimag</assets/MLimag</assets/ML
es/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Credit+App"><b><a href="datasets/Credit+App"><b><a href="datasets/Credit+App"><b><a href="datasets/Credit+App"><b style="datasets/Credit+App"><b style="d
roval">Credit Approval</a></b>\n\t\t\t<!-- <td>This data concerns
credit card applications; good mix of attributes  -->\n\t\tMultivariate&nb
sp;\n\t\tClassification \n\t\t\tCategoric
al, Integer, Real \n\t\t690 \n\t\t
15\    \\ lass="normal">\    \\ lass="normal">Financial\    \\ lass="normal">Financial\    \\ lass="normal">Financial     
                                - \n\t\t\ ./rr +r bacolor "DDEEEE" \n\t\t\ .td +table +trble +td +c brof "detector \la
```

```
5p,
+Screening"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="no
rmal"><b><a href="datasets/Japanese+Credit+Screening">Japanese Credit Screening</a></b>
table>\n\t\t\t<!-- <td>Includes domain theory (generated by talking to Japanese domain
experts); data in Lisp  -->\n\t\t\tMultivariate, Domain-Theory 
/td>\n\t\tClassification \n\t\t\tCategorical, Real, I
nteger \n\t\t\t125 \n\t\t\t 
p  n /t 1992 nbsp;  n /t /t < !-- < td < p class="normal">Financial nbsp;  n /t /t < !-- < td < p class="normal">Financial nbsp;  n /t /t < !-- < td < p class="normal">Financial nbsp;  n /t /t < !-- < td < p class="normal">Financial nbsp;  n /t /t < !-- < td < p class="normal">Financial nbsp;  n /t /t < !-- < td < p class="normal">Financial nbsp;  n /t /t < !-- < td < p class="normal">Financial nbsp;  n /t /t < !-- < td < p class="normal">Financial nbsp;  n /t /t < !-- < td < p class="normal">Financial nbsp; Financial nbsp; 
>-->\n\t\t\d>\n\t\t\t\n\t\t\t<dd><a href="datasets/Computer+Hardware"><img src="assets/M
Limages/SmallLarge29.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Compute"><b><a href="datasets/Compute"><b><a href="datasets/Compute"><b><a href="datasets/Compute"><b ><a href="datasets/Compute"><b ><a href="datasets/Compute"><b ><a href="datasets/Compute"><b ><a href="datasets/Compute"><a href="datasets/Compute</a></a></a></a></a>
r+Hardware">Computer Hardware</a></b>\n\t\t\<!-- <td>Relativ
e CPU Performance Data, described in terms of its cycle time, memory size, etc.  -->\n\t\t<
p class="normal">Multivariate \n\t\tRegression \n\t\t<t
d>Integer \n\t\t\t209 \n\t\t\t209 
ss="normal">9 \n\t\t\1987 \n\t\t\<!-- <td>1987&nbsp;
mal">Computer  -->\n\t\t\\n\t\t\<a href="datase">ctr><a hr
ts/Contraceptive+Method+Choice"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
<b><a href="datasets/Contraceptive+Method+Choice">Contraceptive Method Choi
ce</a></b>\n\t\t\t<!-- <td>Dataset is a subset of the 1987 Nationa
I Indonesia Contraceptive Prevalence Survey.  -->\n\t\t\tMultivariate&nbs
p;/n\t\tClassification /n\t\tCategorical (and the context of the context
, Integer \n\t\t1473 \n\t\t\t9&nbs
p;\n\t\t1997 \n\t\t<!-- <td>Life&nbsp;
-->\n\t\t\\n\t\t\t<a href="datasets/Covertype"><img src="assets/MLimages/Sm"
allLarge31.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Covertype">Covertype
e</a></b>\n\t\t<!-- <td>Forest CoverType dataset&nbsp;</t
d> --> \n\t\t\t class="normal"> Multivariate & nbsp;   \n\t\t\t class="normal"> Classification & nbs < (a) < (b) < (b) < (b) < (c) < (c
p;\n\t\tCategorical, Integer \n\t\t\t5810
12\    \\ lass = "normal" > 54\    \\ lass = "normal" > 1998\   
p  n t < --> n t <
able><a href="datasets/Cylinder+Bands"><img src="assets/MLimages/SmallLargedefault.jpg" border=1
/></a>&nbsp;<b><a href="datasets/Cylinder+Bands">Cylinder Bands</a></b>
fr>/rd>/rd>fr>/rd>
elays known as "cylinder bands" in rotogravure printing  -->\n\t\t\tMultiva
riate \n\t\t\Classification \n\t\t\tC
ategorical, Integer, Real \n\t\t512 \n\t\t<p class="
normal">39\ \n\t\t\1995\ \n\t\t\<!-- <td>
>Physical  -->\n\t\t\\r\\t\t\<a href="datasets/Dermatology"><img src
="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="dat
asets/Dermatology">Dermatology</a></b>\n\t\t<!-- <td>Aim for
this dataset is to determine the type of Eryhemato-Squamous Disease.   -->\n\t\t\t<p class="
normal">Multivariate \n\t\t\tClassification \n\t\t\t<p cl
ass="normal">Categorical, Integer \n\t\t\t366 \n\t\t\t
33 \n\t\t\t1998 \n\t\t\t<!-- <td>1998&nbsp;
s="normal">Life  -->\n\t\t\t\n\t\t\t<a href="datas"
ets/Diabetes"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="n
ormal"><b><a href="datasets/Diabetes">Diabetes</a></b>\n\t\t\t<!-- <td><p class="n
ormal">This diabetes dataset is from AIM \'94  -->\n\t\t\tMultivariate, Tim
e-Series \n\t\t \n\t\tCategorical
, Integer\  \n\t\t  \n\t\t20  
n\t\t\<a href="datasets/DGP2+-+The+Second+Data+Generation+Program">
<img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a
href="datasets/DGP2+-+The+Second+Data+Generation+Program">DGP2 - The Second Data Generation Program
am</a></b>\n\t\t<!-- <td>Generates application domains based
on specific parameters, number of features, and proportion of positive to negative examples  --
\ \n\t\t\t \n\t\t\t \n\t\t\t \n\t\t\t \n\t\t\t 
t/t/Real \n/t/t/ \n/t/t/td> 
s="normal">\ \n\t\t\t \n\t\t\t<!-- <td>Ot
her  -->\n\t\t\n\t\t\t<a href="datasets/Documen">datasets/Documen</a>
t+Understanding"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class
="normal"><b><a href="datasets/Document+Understanding">Document Understanding</a></b>
```

```
/table></tu>\frac{1}{\table}</tu>\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1}{\table}<\frac{1
-->\n\t\t \n\t\t\t \n\t\t\t 
 \n\t\t\ \n\t\t\t
 \n\t\t1994 <math>\n\t\t<!-- <td>Other&nb
sp; -->\n\t\t\\n\t\t\<a href="datasets/EBL+Domain+Theories"><img src="
assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="data"
sets/EBL+Domain+Theories">EBL Domain Theories</a></b>\n\t\t\t<!-- <td>class
="normal">Assorted small-scale domain theories  -->\n\t\t </
ss="normal">\ \n\t\t\t \n\t\t\t 
;\n\t\t<!-- <td>Computer&nbsp; -->\n\t\t bgcolor="DDEEFF">\n\t\
t\t<a href="datasets/Echocardiogram"><img src="assets/MLimages/SmallLarge38.jpg" bord
er=1 /></a>&nbsp;<b><a href="datasets/Echocardiogram">Echocardiogram</a></
b>\n\t\t\<!-- <td>Data for classifying if patients will survive for at I
east one year after a heart attack -->\n\t\t\tMultivariate \
n\t\t\tClassification \n\t\t\tCategorical, Integer, R
eal \n\t\t\ctd>132 \n\t\t\ctd>12 </p
>nt/t<<d>1989&nbsp;\n/t/t<!-- <td>Life&nbsp;
-->\n\t\t\n\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\t\n\t\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\
0.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Ecoli">Ecoli</a></b><
/tr>\n\t\t<!-- <td>This data contains protein localization sites&nbsp;---
>\n\t\tMultivariate \n\t\tClassification </p
>\n\t\tReal \n\t\t\t336 \n\t\t\t336 
t8 \n\t\t\t1996 \n\t\t\t<!-- < td>1996&nbsp;
lass="normal">Life  -->\n\t\t\t\n\t\t\t<a href="dat
asets/Flags"><img src="assets/MLimages/SmallLarge40.jpg" border=1 /></a>&nbsp;<p class="normal
"><b><a href="datasets/Flags">Flags</a></b>\n\t\t<!-- <td>Fro
m Collins Gem Guide to Flags, 1986 
-->\n\t\tMultivariate </t
d>\n\t\tClassification \n\t\t\tCategorical, Integer
 \n\t\t194 \n\t\t\t30 
td > \ln t/t  p class = "normal" > 1990 \& nbsp;   \ln t/t < !-- < td > p class = "normal" > Other \& nbsp;   --  p class = "normal" > Other \& nbsp;   --  p class = "normal" > Other \& nbsp;  
->\n\t\t\n\t\t<a href="datasets/Function+Finding"><img src="assets/MLimages/S
mallLarge41.jpg" border=1 /></a>&nbsp;<pcclass="normal"><b><a href="datasets/Function+Finding">
Function Finding</a></b>\n\t\t\t<!-- <td>Cases collected mostly fr
om investigations in physical science; intention is to evaluate function-finding algorithms  -->\n\t\
t\t \n\t\tFunction-Learning \n\t\t\
tp class="normal">Real \n\t\tp class="normal">352 \n\t\tp class="normal">352 
ss="normal">\ \n\t\t1990\ \n\t\t\t<!-- <td>1990&nbsp;
al">Physical  -->\n\t\t\n\t\t\t<a href="datasets/"
Glass+Identification"><img src="assets/MLimages/SmallLarge42.jpg" border=1 /></a>&nbsp;<p class=
"normal"><b><a href="datasets/Glass+Identification">Glass Identification</a></b>
t/t/t<!-- <td>From USA Forensic Science Service; 6 types of glass; defined in terms of their o
xide content (i.e. Na, Fe, K, etc) -->\n\t\t\tMultivariate \n
\t\tClassification \n\t\tReal \n\t\
t\t214 \n\t\t\t10 \n\t\t\t10 
s="normal">1987\ \n\t\t<!-- <td>Physical\&nbsp; -->\n\t\t\  \n \to \n \t\t
\t\t<a href="datasets/Haberman%27s+Survival"><img src="assets/MLimages/SmallLarged"
efault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Haberman%27s+Survival"
>Haberman\'s Survival</a></b>\n\t\t\t<!-- <td>Dataset contains c
ases from study conducted on the survival of patients who had undergone surgery for breast cancer 
-->\n\t\tMultivariate \n\t\tClassification&n
bsp;\n\t\t\t306 
/n\t/t3 /n/t/t1999 /n/t/t<!-
- class="normal">Life  -->\n\t\t\\n\t\t\<a
href="datasets/Hayes-Roth"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
d><b><a href="datasets/Hayes-Roth">Hayes-Roth</a></b>
<!-- <td>Topic: human subjects study&nbsp; -->\n\t\t\tMulti
variate \n\t\t\tClassification \n\t\t\t<p class="normal"
>Categorical \n\t\t\ctd>160 \n\t\t\ctd>5&
nbsp;   nt/t < d > p class = "normal" > 1989 & nbsp;   nt/t < !-- < td > p class = "normal" > Social & nbsp;  
p;-->\n\t\t\\n\t\t\<a href="datasets/Heart+Disease"><img src="assets/ML"
images/SmallLarge45.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Heart+Dis
```

```
ease">Heart Disease</a></b>\n\t\t\t<!-- <td>4 databases: Clevel
and, Hungary, Switzerland, and the VA Long Beach  -->\n\t\t\tMultivariat
e \n\t\tClassification \n\t\t\tCate
gorical, Integer, Real \n\t\t\t303 \n\t\t\t
mal">75 \n\t\t\t1988 \n\t\t\t<!-- <td>Lif
e  -->\n\t\t\t\n\t\t\t<a href="datasets/Hepatitis"><
img src="assets/MLimages/SmallLarge46.jpg" border=1 /></a>&nbsp;<b><a href="
datasets/Hepatitis">Hepatitis</a></b>\n\t\t\t<!-- <td>From G.Gon
g: CMU; Mostly Boolean or numeric-valued attribute types; Includes cost data (donated by Peter Turney) 
-->\n\t\tMultivariate \n\t\t\tClassificati
on \n\t\tCategorical, Integer, Real \n\t\t<p class="
normal">155 \\ normal">155 \\ normal">108 \\ normal">198 \\ normal">108 \\ 
8\ -->\\n\t\t<-ctable>
<a href="datasets/Horse+Colic"><img src="assets/MLimages/SmallLarge47.jpg" border=1 /></a>&nbsp;</t
d><b><a href="datasets/Horse+Colic">Horse Colic</a></b>
n\t\t\t<!-- <td>Well documented attributes; 368 instances with 28 attributes (continuous, discr
ete, and nominal); 30% missing values  -->\n\t\t\tMultivariate 
\n\t\tClassification \n\t\tCategorical, Integ
er, Real \n\t\t27&nbs
p;\n\t\t1989 \n\t\t<!-- <td>Life&nbsp;
-->\n\t\t\ bgcolor="DDEEFF">\n\t\t\<a href="datasets/ICU"><img src="assets/">
MLimages/SmallLarge49.jpg" border=1 /></a>&nbsp;<b><a href="datasets/ICU">I
CU</a></b>
pants for the 1994 AAAI Spring Symposium on Artificial Intelligence in Medicine.  -->\n\t\t\t<
p class="normal">Multivariate, Time-Series \n\t\t\t \n\t\t\t
Real \n\t\t\t \n\t\t\t<p class="
normal">\ \n\t\t\t \n\t\t\t<!-- <td>Life&n &nbsp;
bsp; -->\n\t\t\t\n\t\t\t<a href="datasets/Image+Segmentation"><img src="a"><img src="a"><img src="a">
ssets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datas"
ets/Image+Segmentation">Image Segmentation</a></b>\n\t\t<!-- <td><p class="no
rmal">Image data described by high-level numeric-valued attributes, 7 classes  -->\n\t\t\t<p
class="normal">Multivariate \n\t\tClassification \n\t\t\t<t
d>Real \n\t\t\t2310 \n\t\t\t2310 
s="normal">19\ \n\t\t1990\ \n\t\t<!-- <td>1900&nbsp;
mal">Other  -->\n\t\t\ bgcolor="DDEEFF">\n\t\t\t<a href="datasets/In"
ternet+Advertisements"><img src="assets/MLimages/SmallLarge51.jpg" border=1 /></a>&nbsp;<p cla
ss="normal"><b><a href="datasets/Internet+Advertisements">Internet Advertisements</a></b>
able>\n\t\t\<!-- <td>This dataset represents a set of possible advertisements on Intern
et pages.  -->\n\t\t\Multivariate \n\t\t\<p class="normal"
mal">Classification \n\t\tCategorical, Integer, Real \n\t\t
tp class="normal">3279 \n\t\tp class="normal">1558 \n\t\tp
class="normal">1998 \n\t\t\<!-- <td>Computer&nbsp; -->\n\t\t\t
\n\t\t<a href="datasets/lonosphere"><img src="assets/MLimages/SmallLarge52.jpg" b
order=1 /></a>&nbsp;<b><a href="datasets/lonosphere">lonosphere</a></b>
\n\t\t<!-- <td>Classification of radar returns from the ionosphere&nb
sp;-->\n\t\t\tMultivariate \n\t\t\tClassific
ation \n\t\tInteger, Real \n\t\t3
51\nbsp;\n\t\t34\nbsp;\n\t\t\t1989\nbsp;
d><a href="datasets/Iris"><img src="assets/MLimages/SmallLarge53.jpg" border=1 /></a>&nbs
p:<b><a href="datasets/lris">|ris</a></b>\n\t\t\t<!-- <td>|ris</a>
Famous database; from Fisher, 1936  -->\n\t\tMulti
variate \n\t\t\tClassification \n\t\t\t<p class="normal"
>Real \n\t\t150 \n\t\t\t4 
p  n /t  1988 & nbsp;   n /t /t < !-- < td > class = "normal" > Life & nbsp;   n /t /t < !-- < td > class = "normal" > Life & nbsp;   n /t /t < !-- < td > class = "normal" > Life & nbsp;   n /t /t < !-- < td > class = "normal" > Life & nbsp;   n /t /t < !-- < td > class = "normal" > Life & nbsp;   n /t /t < !-- < td > class = "normal" > Life & nbsp;   n /t /t < !-- < td > class = "normal" > Life & nbsp;   n /t /t < !-- < td > class = "normal" > Life & nbsp;   n /t /t < !-- < td > class = "normal" > Life & nbsp;   n /t /t < !-- < td > class = "normal" > Life & nbsp;   n /t < !-- < td > class = "normal" > Life & nbsp;   n /t < !-- < td > class = "normal" > Life & nbsp;   n /t < !-- < td > class = "normal" > Life & nbsp;   n /t < !-- < td > class = "normal" > Life & nbsp;  
> -->\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\
gedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/ISOLET">ISOLET</a>
</b>/td>\n\t\t\t<!-- <td>oal: Predict which letter-name was spoken--
a simple classification task.  -->\n\t\t\tMultivariate \n\t\t\t<
td>Classification \n\t\t\tReal \n\t\t\t<td
>7797 \n\t\t\617 \n\t\t\<p class=
"normal">1994 \n\t\t\<!-- <td>Computer&nbsp; -->\n\t\t\<tr bg
```

```
color="DDEEFF">\n\t\t\t<a href="datasets/Kinship"><img src="assets/MLimages/SmallLarg
e55.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Kinship">Kinship</a></b></p
>\n\t\t\t<!-- <td>Relational dataset&nbsp; -->\n\t\t\t<p cl
ass="normal">Relational\ \n\t\tRelational-Learning\ \n\t\t\tRelational-Learning 
tCategorical \n\t\t\t104 \n\t\t\t<td
>cp class="normal">12 \n\t\t\tclass="normal">1990 \n\t\t\t<!-- <td>class="normal">1990&nbsp;
ss="normal">Social  -->\n\t\t\n\t\t<a href="datasets/Labor+Relating">to
ons"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<
b><a href="datasets/Labor+Relations">Labor Relations</a></b>\n\t\t<!-- <td><p cla
ss="normal">From Collective Bargaining Review  -->\n\t\t\tMultivariate&
nbsp;\n\t\t\ \n\t\t\tCategorical, Integer,
Real\   \\ lass="normal">57\   \\ lass="normal">16\   \\ lass="norm
>nt/tp class="normal">1988 \n/t/t<!-- <td>p class="normal">Social&nbsp;
d> -->\n\t\t\ d> -->\n\t\t\t d> -->\n\t\t\t<\t\t<\t\t<\t\t<\t\t<\t\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\t<\t\
mg src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a hr
ef="datasets/LED+Display+Domain">LED Display Domain</a></b>\n\t\t<!-- <td><p
class="normal">From Classification and Regression Trees book; We provide here 2 C programs for generating
sample databases  -->\n\t\tMultivariate, Data-Generator 
>\n\t\tClassification \n\t\tCategorical </p
>n/t/tp class="normal"> n/t/t/p class="normal">7 n/t/t/t
p class="normal">1988 \n\t\t\<!-- <td>Computer&nbsp; -->\n\t\t\</t
r>\n\t\ttor>r>r>r>r>r>r>r>r>r>r>r>r>r>rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr<t
border=1 /></a>&nbsp;<b><a href="datasets/Lenses">Lenses</a></b>
r>\n\t\t\<!-- <td>Database for fitting contact lenses&nbsp; -->\n\t\t\t<td
>Multivariate \n\t\tClassification \n\t\
t/tCategorical \n/t/t24 \n/t/t<td
> class="normal">4 \n\t\t\t1990 \n\t\t\t<!-- <td>1990&nbsp;
sets/Letter+Recognition"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
<b><a href="datasets/Letter+Recognition">Letter Recognition</a></b>
>\n\t\t<!-- <td>Database of character image features; try to identify the letter&nbsp;</p
> -->\n\t\t\tMultivariate \n\t\t\tClassification&
nbsp;\n\t\tInteger \n\t\t20000 
\n\t\t16 \n\t\t1991 \n\t\t\t1991 
<b><a href="datasets/Liver+Disorders">Liver Disorders</a></b>
n\t\t\t<!-- <td>BUPA Medical Research Ltd. database donated by Richard S. Forsyth&nbsp;<
/p> -->\n\t\t\tMultivariate \n\t\t\t <
/td>\n\t\tCategorical, Integer, Real \n\t\t\t345&n
bsp;\n\t\t7 \n\t\t\t1990 
\ \n\t\t\<!-- <td>Life&nbsp; -->\n\t\t\ bgcolor="DDEEFF">\n\t\t\<
tr><a href="datasets/Logic+Theorist"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>
 <b><a href="datasets/Logic+Theorist">Logic Theorist</a></b>
>\n\t\t<!-- <td>All code for Logic Theorist&nbsp; -->\n\t\t\t<p clas
s="normal">Domain-Theory \n\t\t\t \n\t\t\t<p class="
normal"> \n\t\t\t 
>\n\t\t\ \n\t\t\<!-- <td>Computer&nbsp;
> -->\n\t\t\n\t\t\t<a href="datasets/Lung+Cancer"><img src="assets/MLimages/Sm
allLarge62.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Lung+Cancer">Lung+Cancer">Lung+Cancer">Lung+Cancer">Lung+Cancer">Lung+Cancer">Lung+Cancer">Lung+Cancer">Lung+Cancer">Lung+Cancer">Lung+Cancer">Lung+Cancer">Lung+Cancer">Lung+Cancer">Lung+Cancer">Lung+Cancer">Lung+Cancer">Lung+Cancer">Lung+Cancer</br>
Cancer</a></b>\n\t\t<!-- <td>Lung cancer data; no attribute defi
nitions  -->\n\t\t\tMultivariate \n\t\t\t<p class="normal"
">Classification \n\t\tInteger \n\t\t<p class="normal"
l">32 \n\t\t\t56 \n\t\t\t1992 
\n\t\t<!-- <td>Life&nbsp; -->\n\t\t\t\n\t\t\t
<a href="datasets/Lymphography"><img src="assets/MLimages/SmallLargedefault.jpg" border=
1 /></a>&nbsp;<b><a href="datasets/Lymphography">Lymphography</a></b>
\n\t\t<!-- <td>This lymphography domain was obtained from the Univ
ersity Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia. (Restricted access)  -->\n\t\t
\tMultivariate \n\t\tClassification </td
>\n\t\tCategorical \n\t\t148 \n\t
tp class="normal">18 \n\t\tp class="normal">1988 \n\t\t<!-- <td>p class="normal">1988&nbsp;
```

```
Life  -->\n\t\t\</r>\n\t\t\td><a href="datasets/Mechanic">Life&nbsp;
al+Analysis"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="nor
mal"><b><a href="datasets/Mechanical+Analysis">Mechanical Analysis</a></b>
\t<!-- <td>Fault diagnosis problem of electromechanical devices; also PUMPS DATA SET is
newer version with domain theory and results  -->\n\t\t\tMultivariate&nbs
p;\n\t\tClassification \n\t\tCategorical
, Integer, Real \n\t\t209 \n\t\t8
 \n\t\t1990 \n\t\t\t<!-- <td>Computer
  -->\n\t\t\\n\t\t\<a href="datasets/Meta-data">
<img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a
href="datasets/Meta-data">Meta-data</a></b>\n\t\t<!-- <td>Met
a-Data was used in order to give advice about which classification method is appropriate for a particular dataset
(taken from results of Statlog project).  -->\n\t\tMultivariate 
/td>\n\t\tClassification \n\t\t\tCategorical, Intege
r, Real \n\t\t\t528 \n\t\t\t22&nbsp
;\n\t\t1996 \n\t\t<!-- <td>Other&nbsp;</p
> -->\n\t\t\t\n\t\t\t<a href="datasets/Mobile+Robots"><img src="assets/MLimag">
es/SmallLarge66.jpg" border=1 /></a>&nbsp;<pcclass="normal"><b><a href="datasets/Mobile+Robots"
">Mobile Robots</a></b>\r\t\t<!-- <td>Learning concepts from s
ensor data of a mobile robot; set of data sets  -->\n\t\t\tDomain-Theory&
nbsp;\n\t\t\t \n\t\t\tCategorical, Integer,
Real \n\t\t \n\t\t\t 
>\n\t\t1995 \n\t\t\t<!-- <td>Computer&nbsp;
> -->\n\t\t\c/tr>\n\t\t\c/ta><a href="datasets/Molecular+Biology+%28Pro">+%28Pro
moter+Gene+Sequences%29"><img src="assets/MLimages/SmallLarge67.jpg" border=1 /></a>&nbsp;
><b><a href="datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29">Molecula
r Biology (Promoter Gene Sequences)</a></b>\n\t\t\t<!-- <td>E.
Coli promoter gene sequences (DNA) with partial domain theory  -->\n\t\t\t<p class="normal
">Sequential, Domain-Theory \n\t\t\tClassification \n\t\t\t
Categorical \n\t\t106 \n\t\t
<p class="normal">58 </p></td>\n\t\t<p class="normal">1990 </p></td>\n\t\t\t<!-- <td><p class="normal">
s="normal">Life  -->\n\t\t\n\t\t\t<a href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog">href="datasets/Molecular+Biolog</d>href="datasets/Molecular+Biolog</d>href="datasets/"
y+%28Protein+Secondary+Structure%29"><img src="assets/MLimages/SmallLarge67.jpg" border=1 /></a>&nb
sp;<b><a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%2"><b><a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%2"><b><a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%2"><b><a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%2"><b><a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%2"><b><a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%2"><b href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%2"><b href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%2"><b href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%2"><b href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%2"><b href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%2"><b href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%2"><b href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%2"><b href="datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets
9">Molecular Biology (Protein Secondary Structure)</a></b>\n\t\t\c!-- <p class=
"normal">From CMU connectionist bench repository; Classifies secondary structure of certain globular proteins
ification \n\t\t\tCategorical \n\t\t\t
128\    \\ lass="normal">     
\ \n\t\t\t<!-- <td>Life&nbsp; -->\n\t\t\t bgcolor="DDEEFF">\n\t\t\t<
tr><a href="datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29"><img src="assets/MLi
mages/SmallLarge67.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Molecular+"
Biology+%28Splice-junction+Gene+Sequences%29">Molecular Biology (Splice-junction Gene Sequences)</a>
/b>//table>\n\t\t<!-- <td>Primate splice-junction gene sequences (DNA)
with associated imperfect domain theory  -->\n\t\tSequential, Domain-T
heory \n\t\t\tClassification \n\t\t\t
Categorical \n\t\t3190 \n\t\t\t61
\alpha = \frac{1}{2} - 
 -->\n\t\t\t\n\t\t\t<a href="datasets/MONK%27s+Problems"><img src="ass"
ets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets"
/MONK%27s+Problems">MONK\'s Problems</a></b>\/table>\/td\/\t\t\<!-- <td><p class="norm"
al">A set of three artificial domains over the same attribute space; Used to test a wide range of induction algorit
hms -->\n\t\tMultivariate \n\t\t
Classification \n\t\t\tCategorical \n\t\t\tCategorical 
mal">432\ \n\t\t7\ \n\t\t\t1992\ 
sp;--> \n\t<<!--<td>Other&nbsp;---> \n\t<</tr>
<tmages/SmallLargedefault.jpg" b
order=1 /></a>&nbsp;<b><a href="datasets/Moral+Reasoner">Moral Reasoner</a
></b>
s moral reasoning; Theory includes negated literals  -->\n\t\t\tDomain-T
heory \    \\ \n\t\t    \\ \n\t\t    \\
d>n\t\tp class="normal">202 \n\t\t\tp class="normal"> \n\t\t\tp class="normal">
```

```
class = "normal" > 1994 \& nbsp;   (n)t < !-- < td > p class = "normal" > Computer \& nbsp;   --> (n)t < !-- < td > p class = "normal" > Computer \& nbsp;   --> (n)t < !-- < td > p class = "normal" > Computer \& nbsp; 
\n\t\t\t<a href="datasets/Multiple+Features"><img src="assets/MLimages/SmallLargede"
fault.jpg" border=1 /></a>&nbsp;<pcclass="normal"><b><a href="datasets/Multiple+Features">Multiple
Features</a></b>\n\t\t<!-- <td>This dataset consists of features
of handwritten numerals (`0\'--`9\') extracted from a collection of Dutch utility maps  -->\n\t\t\t<td
>Multivariate \n\t\tClassification \n\t\
t\tInteger, Real \n\t\t\t2000 \n\t\t\
t< d>p class="normal">649&nbsp;\n\t\tp class="normal">&nbsp;\n\t\t<!-- <td>p class="normal">&nbsp;\n\t\t\t<!-- <td>p class="normal">&nbsp;
ss="normal">Computer  -->\n\t\t\t\n\t\t\t<a href=
"datasets/Mushroom"><img src="assets/MLimages/SmallLarge73.jpg" border=1 /></a>&nbsp;<p class
="normal"><b><a href="datasets/Mushroom">Mushroom</a></b>\n\t\t\t<!-- <td>c
lass="normal">From Audobon Society Field Guide; mushrooms described in terms of physical characteristics; cl
assification: poisonous or edible -->\n\t\tMultivariate \n
\t\t\ctd>Classification \n\t\t\ctd>Categorical </
td>/n\tp class="normal">8124\nbsp;| n\t\t\tp class="normal">22\nbsp;| n\t\t\tp class="normal">22\nbsp;| n\t\t\tp class="normal">22\nbsp;| n\t\t\tp class="normal">22\nbsp;
> class="normal">1987 \n\t\t<!-- <td>p class="normal">Life&nbsp; -->\n\t\t
r>\n\t\t<a href="datasets/Musk+%28Version+1%29"><img src="assets/MLimages/SmallLa
rgedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Musk+%28Version+1"><b><a href="datasets/Musk+%28Version+1"><b><a href="datasets/Musk+%28Version+1"><b><a href="datasets/Musk+%28Version+1"><b><a href="datasets/Musk+%28Version+1"><b><a href="datasets/Musk+%28Version+1"><b><a href="datasets/Musk+%28Version+1"><b><a href="datasets/Musk+%28Version+1"><b><a href="datasets/Musk+%28Version+1"><b href="datasets/Musk+%28Versi
n to predict whether new molecules will be musks or non-musks  -->\n\t\t<p class="normal
">Multivariate \n\t\tClassification \n\t\t<p class="n
ormal">Integer\ \n\t\t\t476 \n\t\t\t
168\    \\ lass="normal">1994\    \\ lass="normal">Physic + lass="normal">1994\    \\ lass="normal">1994\   
al  -->\n\t\t\t\n\t\t\t<a href="datasets/Musk+%28"
normal"><b><a href="datasets/Musk+%28Version+2%29">Musk (Version 2)</a></b>
>\n\t\t<!-- <td>The goal is to learn to predict whether new molecules will be musks or non-
musks -->\n\t\t\tMultivariate \n\t\t\t<p class="normal"
">Classification \n\t\tInteger \n\t\t<p class="normal"
I">6598\ \n\t\t\t168\ \n\t\t\t1994 
bsp;  n/t/t < !-- < td > Physical & nbsp;  --> \n/t/t  \n/t/t < \n/
<a href="datasets/Nursery"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;</t
d><b><a href="datasets/Nursery">Nursery</a></b>\n\t\t\t<!--
 Nursery Database was derived from a hierarchical decision model originally developed
to rank applications for nursery schools. anbsp; -->\n\t\tMultivariate </p
>\n\t\t\tClassification \n\t\t\tCategorical&nbs
p</b>n/t/tp class="normal">12960&nbspn/t/tp class="normal">8&nbsp
n\t\tp class="normal">1997\t\t>n\t\t<t--<td>p class="normal">Social\t\t>n>c/p>
n\t\t\n\t\t<a href="datasets/Othello+Domain+Theory"><img s
rc="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="d
atasets/Othello+Domain+Theory">Othello Domain Theory</a></b>
class="normal">Used in research to generate features for an inductive learning system  -->\n\t\t
\tDomain-Theory \n\t\t \n\t\t\t<t
d><p\ class="normal">\&nbsp;\n\t\t\t<p\ class="normal">&nbsp;\n\t\t\t<p\ class="normal">
">\ \n\t\t\t1991\ \n\t\t\t<!-- <td>Game&
nbsp;   --> \\ lt                
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="
datasets/Page+Blocks+Classification">Page Blocks Classification</a></b>
The problem consists of classifying all the blocks of the page layout of a document that
has been detected by a segmentation process.  -->\n\t\t\tMultivariate&n
bsp;\n\t\tClassification \n\t\tInteger,
Real\ \n\t\t5473 \n\t\t\t10 
\n\t\t1995 \n\t\t\t<!-- <td>Computer&nbsp;
-->\n\t\t\/tr>\n\t\t\t<a href="datasets/Optical+Recognition">+Recognition</a>
+of+Handwritten+Digits"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
p class="normal"><b><a href="datasets/Optical+Recognition+of+Handwritten+Digits">Optical Recognition of Ha
ndwritten Digits</a></b>\n\t\t<!-- <td>Two versions of this datab
ase available; see folder  -->\n\t\t\tMultivariate \n\t\t\t<td
>Classification \n\t\t\tInteger \n\t\t\t<t
d>5620 \n\t\t\t64 \n\t\t\t64 
"normal">1998\ </n/>
tht<dd><a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits"><img src="assets/M"
```

```
Limages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Pen-
Based+Recognition+of+Handwritten+Digits">Pen-Based Recognition of Handwritten Digits</a></b>
>\n\t\t<!-- <td>Digit database of 250 samples from 44 writers&nbsp;
-->\n\t\t\tMultivariate \n\t\t\tClassification 
/p>nt/tp class="normal">Integer nt/t/tp class="normal">10992 
d>/n/t/t16 /n/t/t/t1998 /n/t/t/t<!--
Computer  -->\n\t\t\t\n\t\t\t<t
d><a href="datasets/Post-Operative+Patient"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /><
/a> <b><a href="datasets/Post-Operative+Patient">Post-Operative Patient
a></b>
-->\n\t\t\ctd>Multivariate \n\t\t\ctd>Classification <
/p>\n\t\tCategorical, Integer \n\t\t\t90&nbs
p;\n\t\t8 \n\t\t\t1993 \n
ts/Primary+Tumor"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p cla
ss="normal"><b><a href="datasets/Primary+Tumor">Primary Tumor</a></b>
!-- From Ljubljana Oncology Institute  -->\n\t\t\tM
ultivariate \n\t\t\tClassification \n\t\t\t<p class="normal"
al">Categorical \n\t\t\ctd>339 \n\t\t\ctd>
17\    \\ lass="normal">1988\    \\ lass="normal">Life\    \\ lass="normal">Life\    \\ lass="normal">Life\    \\ lass="normal">Life     
sp; -->\n\t\t\n\t\t\t<a href="datasets/Prodigy"><img sr
c="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;c="assets/mclimages/SmallLargedefault.jpg" border=1 /></a>
atasets/Prodigy">Prodigy</a></b>\n\t\t<!-- <td>Assorted domai
ns like blocksworld, eightpuzzle, and schedworld.  -->\n\t\t\tDomain-The
ory \n\t\t \n\t\t 
\n\t\t \n\t\t\t \n\t\t\t<p class=
table><a href="datasets/Qualitative+Structure+Activity+Relationships"><img src="assets/MLimages/Sm
allLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Qualitative+Struc"><b><a href="datasets/Qualitative+Struc"><b><a href="datasets/Qualitative+Struc"><b><a href="datasets/Qualitative+Struc"><b><a href="datasets/Qualitative+Struc"><b><a href="datasets/Qualitative+Struc"><b><a href="datasets/Qualitative+Struc"><b><a href="datasets/Qualitative+Struc"><b href="datasets/Qualitative
ture+Activity+Relationships">Qualitative Structure Activity Relationships</a></b>
\t<!-- <td>Two sets of datasets are given: pyrimidines and triazines&nbsp; -->\n\t\t\
Domain-Theory \n\t\t\t \n\t\t\t<td
>class="normal"> \n\t\t\t \n\t\t\t<p class="normal"
> nbsp;\n\t\t\t hbsp;\n\t\t\t<!-- <td> Physical&nbs
p;-->\n\t\t\bgcolor="DDEEFF">\n\t\t\<a href="datasets/Quadruped+Mam">+Mam</a>
mals"><img src="assets/MLimages/SmallLarge86.jpg" border=1 /></a>&nbsp;<b><
a href="datasets/Quadruped+Mammals">Quadruped Mammals</a></b></to>\n\t\t<!-- <t
d> The file animals.c is a data generator of structured instances representing quadruped ani
mals  -->\n\t\tMultivariate, Data-Generator \n\t\t\t<
p class="normal">Classification \n\t\t\tReal \n\t\t\t<p
class="normal"> \n\t\t72 \n\t\t\t
1992\  n/t/t/<!-- <td>p class="normal">Life&nbsp; -->/n/t/t//n/t/t/
<a href="datasets/Servo"><img src="assets/MLimages/SmallLarge87.jpg" border=1 /></a>&nbsp;
<b><a href="datasets/Servo">Servo</a></b>\rt<!-- <td><
p class="normal">Data was from a simulation of a servo system  -->\n\t\t\t<p class="normal"
">Multivariate \n\t\tRegression \n\t\t\t<p class="normal"
mal">Categorical, Integer \n\t\t\167 \n\t\t\<p class=
"normal">4\ \n\t\t\t1993\ \n\t\t\t<!-- <td>1993&nbsp;
>Computer  -->\n\t\t\\n\t\t\<a href="datasets/S"
ass="normal"><b><a href="datasets/Shuttle+Landing+Control">Shuttle Landing Control</a></b>
/table>\n\t\t<!-- <td>Tiny database; all nominal values&nbsp; -->\n\t\t\t<p
class="normal">Multivariate \n\t\t\tClassification \n\t\t\t<t
d>Categorical \n\t\t\t15 \n\t\t\t<p
class="normal">6\ \n\t\t\t1988\ \n\t\t\t<!-- <td>1988&nbsp;
ormal">Physical  -->\n\t\t\n\t\t<a href="datasets/Solar+Flare"><i
mg src="assets/MLimages/SmallLarge89.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="
datasets/Solar+Flare">Solar Flare</a></b>\n\t\t<!-- <td>Each cl
ass attribute counts the number of solar flares of a certain class that occur in a 24 hour period 
-->\n\t\t\ctd>Multivariate \n\t\t\ctd>Regression </p
>\n\t\tCategorical \n\t\t\t1389 </
td>n/t/tn class="normal">10&nhsn:</n>
10&nhsn:</n>
10&nhsn:
10&nhsn:
10&nhsn:
10&nhsn:
10&nhsn:
10&nhsn:
10&nhsn:
10&nhsn:
10&nhsn:
10&n
```

```
-<\!td><\!p\ class="normal">Physical\&nbsp;<\!/p><\!/td> -->\!n\t\t<\!/tr><tr\ bgcolor="DDEEFF">\n\t\t\t<\!td><\!table><\!tr>
d><a href="datasets/Soybean+%28Large%29"><img src="assets/MLimages/SmallLarge90.jpg" border=1 /></a
> <b><a href="datasets/Soybean+%28Large%29">Soybean (Large)</a></b
>\n\t\t<!-- <td>Michalski\'s famous soybean disease database&
nbsp; -->\n\t\t\tMultivariate \n\t\t\tClassi
fication \n\t\t\tCategorical \n\t\t\t3
07\ /n/t/tp class="normal">35\ /n/t/t/tp class="normal">1988\ 
p  n/t/t < -- < td > p class = "normal" > Life & nbsp;  -/td > -- > n/t/t / t  -table >  td > < a hree life & nbsp; </p > -/td > -- > \n/t/t / t  -- > 
f="datasets/Soybean+%28Small%29"><img src="assets/MLimages/SmallLarge90.jpg" border=1 /></a>&nbsp;</
td><b><a href="datasets/Soybean+%28Small%29">Soybean (Small)</a></b>
\n\t\t<!-- <td>Michalski\'s famous soybean disease database&nbsp;</
td>-->\n\t\t\tMultivariate \n\t\t\tClassification&nb
sp;\n\t\tCategorical \n\t\t47 </
p  n/t/t  p class = "normal" > 35 & nbsp;   n/t/t  p class = "normal" > 1987 & nbsp;   \n/t/t  p class = "normal" > 1987 & nbsp;   \n/t/t  p class = "normal" > 1987 & nbsp;   \n/t/t  p class = "normal" > 1987 & nbsp;   \n/t/t  p class = "normal" > 1987 & nbsp;   \n/t/t  p class = "normal" > 1987 & nbsp;   \n/t/t  p class = "normal" > 1987 & nbsp;   \n/t/t  p class = "normal" > 1987 & nbsp;   \n/t/t  p class = "normal" > 1987 & nbsp;   \n/t/t  p class = "normal" > 1987 & nbsp;   \n/t/t  p class = "normal" > 1987 & nbsp;   \n/t/t  p class = "normal" > 1987 & nbsp;   \n/t/t  p class = "normal" > 1987 & nbsp;   \n/t/t  p class = "normal" > 1987 & nbsp;  
\t<!-- <td>class="normal">Life&nbsp; -->\n\t\t\t\n\t\t\t<td
><a href="datasets/Challenger+USA+Space+Shuttle+O-Ring"><img src="assets/MLimages/SmallLarge92.jpg" b
order=1 /></a>&nbsp;<b><a href="datasets/Challenger+USA+Space+Shuttle+O-Ri
ng">Challenger USA Space Shuttle O-Ring</a></b>\/tt>\/td\\\\t\t\<!-- <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  -->\n\t\t\tMultivariate \n\t\t
\tRegression \n\t\tInteger \n\t\t
<\!td><\!p\ class="normal">\!23\&nbsp;<\!/p><\!/td>\n\t\t<\!td><\!p\ class="normal">\!4\&nbsp;<\!/p><\!/td>\n\t\t<\!td><\!p\ class="normal">\!4\&nbsp;<\!/p><\!/td>
ormal">1993\ \n\t\t\<!-- <td>Physical&nbsp; --> \n\t\t\\n\t\t\</r>
edefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Low+Resolution+Spectr"><br/>
ometer">Low Resolution Spectrometer</a></b>\n\t\t\t<!-- <td>Fr
om IRAS data -- NASA Ames Research Center  -->\n\t\t\tMultivariate&nb
sp;\n\t\tClassification \n\t\t\tInteger, R
eal \n\t\t531 \n\t\t102 
p  n /t /t  p class = "normal" > 1988 & nbsp;  n /t /t /t < !-- < td > p class = "normal" > Physical & nbsp;  n /t /t /t < !-- < td > p class = "normal" > Physical & nbsp;  n /t /t /t < !-- < td > p class = "normal" > Physical & nbsp;  n /t /t /t < !-- < td > p class = "normal" > Physical & nbsp;  n /t /t /t < !-- < td > p class = "normal" > Physical & nbsp;  n /t /t /t < !-- < td > p class = "normal" > Physical & nbsp;  n /t /t /t < !-- < td > p class = "normal" > Physical & nbsp;  n /t /t /t < !-- < td > p class = "normal" > Physical & nbsp;  n /t /t /t < !-- < td > p class = "normal" > Physical & nbsp;  n /t /t /t < !-- < td > p class = "normal" > Physical & nbsp;  n /t /t /t < !-- < td > p class = "normal" > Physical & nbsp;  n /t /t /t < !-- < td > p class = "normal" > Physical & nbsp;  n /t /t /t < !-- < td > p class = "normal" > Physical & nbsp;  n /t /t /t < !-- < td > p class = "normal" > Physical & nbsp;  n /t /t /t < !-- < td > p class = "normal" > Physical & nbsp;  n /t /t /t < !-- < td > p class = "normal" > Physical & nbsp;  n /t /t /t < !-- < td > p class = "normal" > Physical & nbsp;  n /t /t /t < !-- < td > p class = "normal" > Physical & nbsp;  n /t /t /t < !-- < td > p class = "normal" > Physical & nbsp;  n /t /t < |td > n /t /t < |td > n /t < |
> -->\n\t\t\t\n\t\t\t<a href="datasets/Spambase"><img src="
assets/MLimages/SmallLarge94.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/"
Spambase">Spambase</a></b>\n\t\t<!-- <td>Classifying Email
as Spam or Non-Spam  -->\n\t\tMultivariate \n\t\t
Classification \n\t\tInteger, Real \n\t\
t/tp class="normal">4601 \n/t/tp class="normal">57 \n/t/tp class="normal">57 
ass="normal">1999\ \n\t\t\<!--<td>Computer\&nbsp;\-->\n\t\t\
tr>\n\t\t\t<a href="datasets/SPECT+Heart"><img src="assets/MLimages/SmallLarge45.jpg"
border=1 /></a>&nbsp;<b><a href="datasets/SPECT+Heart">SPECT Heart</a></b
>\n\t\t<!-- <td>class="normal">Data on cardiac Single Proton Emission Compute
d Tomography (SPECT) images. Each patient classified into two categories: normal and abnormal. 
/td> -->\n\t\tMultivariate \n\t\t\tClassification&nb
sp;\n\t\tCategorical \n\t\t267 
/p  n t 22 & nbsp;  n t 2001 & nbsp;  n t  n t 2001 & nbsp;   n t  n t 2001 & nbsp;   n t 2001 & nbsp;  
t/t<!--< td>Life&nbsp; --> \n/t/t/  bgcolor="DDEEFF">\n/t/t/t  Life&nbsp; --> \n/t/t/t  \n/t/t/t  \n/t/t/t  \n/t/t/t  \n/t/t/t  \n/t/t/t  \n/t/t/t  \n/t/t/t  \n/t/t/t \n/t/t/t \n/t/t/t \n/t/t/t \n/t/t/t \n/t/t/t \n/t/t/t \n/t/t/t
d><a href="datasets/SPECTF+Heart"><img src="assets/MLimages/SmallLarge45.jpg" border=1 /></a>&nbsp;</
td><b><a href="datasets/SPECTF+Heart">SPECTF Heart</a></b>
>\n\t\t<!-- <td>Data on cardiac Single Proton Emission Computed Tomography (SPEC
T) images. Each patient classified into two categories: normal and abnormal.  -->\n\t\t\t<p cl
ass="normal">Multivariate \n\t\tClassification \n\t\t\t<td
>Integer \n\t\t\ctd>267 \n\t\t\ctd><p clas
s="normal">44\ \n\t\t\t2001\ \n\t\t\t<!-- <td>2001&nbsp;
mal">Life  -->\n\t\t\t\n\t\t\t<a href="datasets/Sponge"><img src="as
sets/MLimages/SmallLarge97.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Sp
onge">Sponge</a></b>\n\t\t\t<!-- <td>Data on sponges; Attribute
s in spanish  -->\n\t\t\tMultivariate \n\t\t\t<p class="n
ormal">Clustering \n\t\t\tCategorical, Integer \n\t\t\t<
p class="normal">76 \n\t\t\t45 \n\t\t\t45 
al"> \n\t\t<!-- <td>Life&nbsp; -->\n\t\t\
\n\t\t<a href="datasets/Statlog+Project"><img src="assets/MLimages/SmallLargedefault.jp"
g" border=1 /></a>&nbsp;<b><a href="datasets/Statlog+Project">Statlog Project</
```

```
dsat Sattelite, Shuttle, Australian Credit Approval, Heart Disease, Image Segmentation, German Credit </
p>--> \ln t  p> --> \lambda \
d> \n\t\t\t \n\t\t\t<p class="normal"
bsp; -->\n\t\t\\r\t\t\t<a href="datasets/Student+Loan+Relational"><img sr
c="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="d
atasets/Student+Loan+Relational">Student Loan Relational</a></b>\n\t\t\<!-- <td><
p class="normal">Student Loan Relational Domain  -->\n\t\t\tDomain-Th
eory \    \\ lass="normal"> \    \\ lass="normal">        lass="normal">       lass="normal">       lass="normal">       lass="normal">       lass="normal">       lass="normal">       lass="normal">       lass="normal">        lass="normal">       lass="normal">       lass="normal">       lass="normal">       lass="normal">       lass="normal">       lass="normal">       lass="normal">       lass="normal">      lass="normal">      lass="normal"> &nbs
\/\ \n\t\t\t\n\t\t\t\&n\t\t\t\&n\t\t\t<p
class="normal">1993 \n\t\t\<!-- <td>Social&nbsp; -->\n\t\t\
bgcolor="DDEEFF">\n\t\t\t<a href="datasets/Teaching+Assistant+Evaluation"><img src="as
sets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="dataset"
s/Teaching+Assistant+Evaluation">Teaching Assistant Evaluation</a></b>
The data consist of evaluations of teaching performance; scores are "low", "medium", o
r "high"  -->\n\t\t\tMultivariate \n\t\t\t<p class="normal"
I">Classification \n\t\tCategorical, Integer \n\t\t<p
class="normal">151 \n\t\t\t5 \n\t\t\t<p class="normal"
">1997 \n\t\t\t<!-- <td>Other&nbsp; -->\n\t\t\t
le><a href="datasets/Tic-Tac-Toe+Endgame"><img src="assets/MLimages/SmallLarge101.jpg" border
=1 /></a>&nbsp;<b><a href="datasets/Tic-Tac-Toe+Endgame">Tic-Tac-Toe Endg
ame</a></b>
configurations of tic-tac-toe game  -->\n\t\t\tMultivariate 
n\t\t\tClassification \n\t\t\tCategorical 
/n\t\t98 /n\t\t9 /n\t\t
> class="normal">1991 \n\t\t\t<!-- <td>Game&nbsp; -->\n\t\t\t
>\n\t\t<a href="datasets/Thyroid+Disease"><img src="assets/MLim">
ages/SmallLargedefault.jpg" border=1 /></a>&nbsp;</d><b><a href="datasets/Thyroid"
+Disease">Thyroid Disease</a></b>\n\t\t\t<!-- <td>class="normal">10 separate d
atabases from Garavan Institute  -->\n\t\t\tMultivariate, Domain-Theory&
nbsp;\n\t\tClassification \n\t\tCategor
ical, Real \n\t\t7200 \n\t\t21&n
bsp;\n\t\t\t1987 \n\t\t\t<!-- <td>Life&nbsp;</
p> -->\n\t\t\t\n\t\t\t<a href="datasets/Trains"><img src="assets/MLimages/Sma">
IlLarge103.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Trains">Trains</a></
b>\n\t\t\t<!-- <td>2 data formats (structured, one-instance-per-lin
e)  -->\n\t\t\tMultivariate \n\t\t\tCl
assification \n\t\tCategorical \n\t\tCategorical 
al">10 \n\t\t32 <math>\n\t\t\t1994&nbs
p</d>n/t/t<--<d>p class="normal">Other&nbsp-->\n/t/t
td><a href="datasets/University"><img src="assets/MLimages/SmallLarge104.jpg" border=1 /><
/a> <b><a href="datasets/University">University</a></b>
e>\n\t\t<!-- <td>Data in original (LISP-readable) form&nbsp; -->\n\t\t\t<p
class="normal">Multivariate \n\t\tClassification \n\t\t\t<t
d>Categorical, Integer \n\t\t285 \n\t\
p class="normal">Other  -->\n\t\t\t\n\t\t\t<a href="datasets/Congres">datasets/Congres</a>
sional+Voting+Records"><img src="assets/MLimages/SmallLarge105.jpg" border=1 /></a>&nbsp;<p cl
ass="normal"><b><a href="datasets/Congressional+Voting+Records">Congressional Voting Records</a></b></
p>\n\t\t\t<!-- <td>1984 United Stated Congressional Voting Records;
Classify as Republican or Democrat  -->\n\t\tMultivariate </t
d>\n\t\t\Classification \n\t\t\tCategorical </
\n\t\t<a href="datasets/Water+Treatment+Plant"><img src="a</td>
ssets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datas"
ets/Water+Treatment+Plant">Water Treatment Plant</a></b>\n\t\t<!-- <td><p class
="normal">Multiple classes predict plant state  -->\n\t\t\tMultivariate&nbs
p;\n\t\tClustering \n\t\tInteger, Real&
nbsp; /nt/t/t 527  /nt/t/t 38  </td
d>\n\t\t< d>p class="normal">1993\nbsp;
```

```
1%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="normal"
><b><a href="datasets/Waveform+Database+Generator+%28Version+1%29">Waveform Database Generator (
Version 1)</a></b>\n\t\t\t<!-- <td>CART book\'s waveform domai
ns  -->\n\t\t\tMultivariate, Data-Generator \n\t\t\t<p c
lass="normal">Classification \n\t\tReal \n\t\tReal 
ss="normal">5000\ \n\t\t\t21\ \n\t\t\t21 
>1988 % nbsp; \n\t\t<!-- <td>Physical & nbsp; <math>--- \n\t\t\t
EEFF">\n\t\t<a href="datasets/Waveform+Database+Generator+%28Version+2%29"><im
g src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href
="datasets/Waveform+Database+Generator+%28Version+2%29">Waveform Database Generator (Version 2)</
>-->\n\t\t\tMultivariate, Data-Generator \n\t\t\t<p class="norma"
l">Classification \n\t\t\tReal \n\t\t\t<p class="normal"
>5000%nbsp;\n\t\t40%nbsp;\n\t\t\t1988%nbs
d><a href="datasets/Wine"><img src="assets/MLimages/SmallLarge109.jpg" border=1 /></a>&nbsp;<
p class="normal"><b><a href="datasets/Wine">Wine</a></b>\n\t\t<!-- <td><p class
="normal">Using chemical analysis determine the origin of wines  -->\n\t\t\t<p class="norma
l">Multivariate \n\t\tClassification \n\t\t\t<p class="n
ormal">Integer, Real \n\t\t178 \n\t\t178 
mal">13 \n\t\t1991 \n\t\t<!-- <td>P
hysical  -->\n\t\t\\n\t\t\t<a href="datasets/Yeast"
><img src="assets/MLimages/SmallLarge110.jpg" border=1 /></a>&nbsp;<b><a hr
ef="datasets/Yeast">Yeast</a></b>\n\t\t<!-- <td>Predicting the
Cellular Localization Sites of Proteins  -->\n\t\tMultivariate </
td>\n\t\tClassification \n\t\tReal </td
\ \n\t\t\t1484 \n\t\t\t8 \n\t\t\t<
p\ class="normal">1996\ \n\t\t\<!--<td><p\ class="normal">Life&nbsp;\--->\n\t\t\
n\t\t\t<a href="datasets/Zoo"><img src="assets/MLimages/SmallLarge111.jpg" border=1 />
</a>&nbsp;<b><a href="datasets/Zoo">Zoo</a></b>
\t\t<!-- <td>Artificial, 7 classes of animals&nbsp; -->\n\t\t\tM
ultivariate \n\t\t\tClassification \n\t\t\t<p class="normal"
al">Categorical, Integer \n\t\t\t101 \n\t\t\t<p class="n
ormal">17 \n\t\t1990 \n\t\t\t<!-- <td>
Life  -->\n\t\t\ bgcolor="DDEEFF">\n\t\t\<a href="datasets/Undocum">t\t\t</a></a>
ented"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
><b><a href="datasets/Undocumented">Undocumented</a></b>\n\t\t\t<!-- <td><p cl
ass="normal">Various datasets without documentation (feel free to explore!)  -->\n\t\t\t<p cl
ass="normal"> \n\t\t <math>\n\t\t&nbsp;<math></td
p;\n\t\t \n\t\t\t \n\t\t\t 
d> \n\t\t<!-- <td>Other&nbsp; -->\n\t\t\t
>\n\t\t\t<a href="datasets/Twenty+Newsgroups"><img src="assets/MLimages/SmallLargede">
fault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/Twenty+Newsgroups">Twe</a>
nty Newsgroups</a></b>\n\t\t\t<!-- <td>This data set consists of
20000 messages taken from 20 newsgroups.  -->\n\t\t\tText 
/n\t/t /n\t/t /n/t/t\t /n/t/t\t /n/t/t\t /n/t/t\t /n/t/t\t /n/t/t\t /n/t/t\t /n/t/t\t /n/t/t\t /n/t\t 
lass="normal">20000\ \n\t\t\t \n\t\t\t
">1999\ \n\t\t<!--<td>Other\&nbsp; -->\n\t\t\t
EFF">\n\t\t\ctd><a href="datasets/Australian+Sign+Language+signs"><img src="assets/MLima"
ges/SmallLarge114.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Australian+S"
ign+Language+signs">Australian Sign Language signs</a></b>\n\t\t\t<!-- <td><p cla
ss="normal">This data consists of sample of Auslan (Australian Sign Language) signs. Examples of 95 signs w
ere collected from five signers with a total of 6650 sign samples.  -->\n\t\t\ctd><p class="normal
">Multivariate, Time-Series \n\t\tClassification \n\t\t<td
>Categorical, Real \n\t\t\t6650 \n\t\t\t
15 \n\t\t1999 \n\t\t\t<!-- < td>10
class="normal">Other  -->\n\t\t\n\t\t<a href="datasets/Australian+"
Sign+Language+signs+%28High+Quality%29"><img src="assets/MLimages/SmallLarge114.jpg" border=1 /></a
> <b><a href="datasets/Australian+Sign+Language+signs+%28High+Quality">
%29">Australian Sign Language signs (High Quality)</a></b>\n\t\t<!-- <td><p class
="normal">This data consists of sample of Auslan (Australian Sign Language) signs. 27 examples of each of 95
```

```
Ausian signs were captured from a native signer using high-quality position trackers&hosp;
-->\n\t\t\t\t<t
d>Multivariate, Time-Series \n\t\tClassification&nbs
p</b>n/t/tp class="normal">Real&nbsp;n/t/tp class="normal">2565&nbsp;
\ \n\t\t\t22 \n\t\t\t2002 \n\t\t\t<!--
Other  -->\n\t\t\\n\t\t\<
a href="datasets/US+Census+Data+%281990%29"><img src="assets/MLimages/SmallLarge2.jpg" border=1 />
</a>&nbsp;<b><a href="datasets/US+Census+Data+%281990%29">US Census D
et contains a one percent sample of the Public Use Microdata Samples (PUMS) person records drawn from the
full 1990 census sample.  -->\n\t\t\tMultivariate \n\t\t\t<td
>Clustering \n\t\t\tCategorical \n\t\t\t<
td><p\ class="normal">2458285\&nbsp;\n\t\t\t<p\ class="normal">68\&nbsp;\n\t\t\t<p\ class="normal">68&nbsp;
ass="normal">\ +--p class="normal">Social\ +-->\n\t\t\
2.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Census-Income+%28KDD%29"
">Census-Income (KDD)</a></b>>/table>\r\t\t<!-- <td>This data set con
tains weighted census data extracted from the 1994 and 1995 current population surveys conducted by the U.S
. Census Bureau.  -->\n\t\t\tMultivariate \n\t\t\t
ss="normal">Classification \n\t\tCategorical, Integer \n\
t/t/299285 \n/t/t/t40 \n/t/t/t<
p class="normal">2000 \n\t\t<!-- <td>Social&nbsp; -->\n\t\t\t
r bgcolor="DDEEFF">\n\t\t\t<a href="datasets/Coil+1999+Competition+Data"><img src="as
sets/MLimages/SmallLarge118.jpg" border=1 /></a>&nbsp;<b><a href="datasets/"
Coil+1999+Competition+Data">Coil 1999 Competition Data</a></b>
p class="normal">This data set is from the 1999 Computational Intelligence and Learning (COIL) competition. T
he data contains measurements of river chemical concentrations and algae densities. 
-->\n\t\t\t
Multivariate \n\t\t\t \n\t\t\t<p
class="normal">Categorical, Real \n\t\t\t340 \n\t\t\t<
p\ class="normal">17\ \n\t\t\t<p\ class="normal">1999\&nbsp;\n\t\t\t<!--<td><p\ class="normal">10
="normal">Physical  -->\n\t\t<\r\\\t\t<td><a href="datasets/Corel+Image">-->\n\t\t
+Features"><img src="assets/MLimages/SmallLarge119.jpg" border=1 /></a>&nbsp;<p class="normal
"><b><a href="datasets/Corel+Image+Features">Corel Image Features</a></b>
t\t<!-- <td>This dataset contains image features extracted from a Corel image collection. Fo
ur sets of features are available based on the color histogram, color histogram layout, color moments, and co-o
ccurence \    --> \n\t\tMultivariate \    \n\t\t\t \n\t\t\t
mal">\ \n\t\t\Real\ \n\t\t\68040&n (lass="normal")
d>/n/t/t<!--< td>Other&nbsp; -->/n/t/t/t/n/t/t/t
e><a href="datasets/E.+Coli+Genes"><img src="assets/MLimages/SmallLarge120.jpg" border=1 /></a>
 <b><a href="datasets/E.+Coli+Genes">E. Coli Genes</a></b>
>\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 kno
wn) are provided.  -->\n\t\t\tRelational \n\t\t\tRelational 
s="normal">\ \n\t\t\t \n\t\t\t 
\n\t\t \n\t\t2001 \n\t\t\t2001 
t<!-- <td>Life&nbsp; -->\n\t\t\t\n\t\t\t<a href="datasets/"
EEG+Database"><img src="assets/MLimages/SmallLarge121.jpg" border=1 /></a>&nbsp;<p class="n
ormal"><b><a href="datasets/EEG+Database">EEG Database</a></b>\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 </t
d> -->\n\t\t\tMultivariate, Time-Series \n\t\t\t&nbs
p;\n\t\tCategorical, Integer, Real \n\t\t\t<p class="normal"
>122%nbsp;\n\t\t\4%nbsp;\n\t\t\t1999%nbsp;<
able><a href="datasets/El+Nino"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>
 <b><a href="datasets/El+Nino">El Nino</a></b>
n\t\t\t<!-- <td>The data set contains oceanographic and surface meteorological readings tak
en from a series of buoys positioned throughout the equatorial Pacific.  -->\n\t\t\t<p class="
normal">Spatio-temporal \n\t\t \n\t\t \n\t\t 
rmal">12\ /nt/ttp class="normal">1999\ /nt/tt<!-- <td>p class="normal">P
hysical  -->\n\t\t\\n\t\t\<a href="datasets/Entree+Chicago+Recom">hysical&nbsp;
```

```
mendation+Data"><img src="assets/MLimages/SmallLarge123.jpg" border=1 /></a>&nbsp;<p class="
normal"><b><a href="datasets/Entree+Chicago+Recommendation+Data">Entree Chicago Recommendation D
ata</a></b>\n\t\t\t<!-- <td>This data contains a record of user int
eractions with the Entree Chicago restaurant recommendation system.  -->\n\t\t\t<p class="
normal">Transactional, Sequential \n\t\t\tRecommender-Systems 
\n\t\t50672 \n\t\t50672 
p  n t t < d < p class = "normal" > 4 n t t  2000 & n b sp;  n t t t  p class = "normal" > 2000 & n b sp;   n t t t  n t t t  p class = "normal" > 2000 & n b sp;   n t t t  n t t t  p class = "normal" > 2000 & n b sp;   n t t t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t 
!-- Other  -->\n\t\t\t\n\t\t\t<td
><a href="datasets/CMU+Face+Images"><img src="assets/MLimages/SmallLarge124.jpg" border=1 /></a>&nb
sp;<b><a href="datasets/CMU+Face+Images">CMU Face Images</a></b></t
d>\n\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  -->\n\t\t\tImage \n\t\t\t
Classification \n\t\tInteger \n\t\t<td
> class="normal">640 \n\t\t\t \n\t\t\t
mal">1999 \n\t\t\<!-- <td>Other&nbsp; -->\n\t\t\
<a href="datasets/Insurance+Company+Benchmark+%28COIL+2000%29"><img src="assets/M
Limages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Insur
ance+Company+Benchmark+%28COIL+2000%29">Insurance Company Benchmark (COIL 2000)</a></b>
2000 Challenge contain
s information on customers of an insurance company. The data consists of 86 variables and includes product u
sage data and socio-demographic data  -->\n\t\t\tMultivariate 
\n\t\tRegression, Description \n\t\t\tCatego
rical, Integer\  \n\t\t9000\  \n\t\t\t8
6 \n\t\t2000 \n\t\t\t<!-- <td>Social&n
bsp; -->\n\t\t\t\n\t\t\t<a href="datasets/Internet+Usage">tr><a href="datasets/Internet+Usage">tr>
+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="normal"
><b><a href="datasets/Internet+Usage+Data">Internet Usage Data</a></b>
-- This data contains general demographic information on internet users in 1997. 
-->\n\t\t\tMultivariate \n\t\t\t 
/n/t/t/Categorical, Integer /n/t/t/10104&nbs
p</b>n\t\t<d>p class="normal">72&nbsp;\n\t\tp class="normal">1999&nbsp;\n\t\tp class="normal">1999&nbsp;
n/t/t/<--< td>Computer&nbsp; -->/n/t/t//n/t/t/<a href="mailto:nref">tr></a></a></a></a></a></a></a></a></a></a></a></a></a></a>
"datasets/IPUMS+Census+Database"><img src="assets/MLimages/SmallLarge2.jpg" border=1 /></a>&nbsp;</t
d><b><a href="datasets/IPUMS+Census+Database">IPUMS Census Database</a></b>
\n\t\t\<!-- <td>This data set contains unweighted PUMS census d
ata from the Los Angeles and Long Beach areas for the years 1970, 1980, and 1990.  -->\n\t\t\
Multivariate \n\t\t\t \n\t\t\t<p
class="normal">Categorical, Integer \n\t\t\t256932 \n\t\t\t
<td>61 \n\t\t1999 \n\t\t\t<!-- <td><p
class="normal">Social  -->\n\t\t\t\n\t\t\t<a href="
d><b><a href="datasets/Japanese+Vowels">Japanese Vowels</a></b>
>\n\t\t<!-- <td>This dataset records 640 time series of 12 LPC cepstrum coefficients ta
ken from nine male speakers.  -->\n\t\tp class="normal">Multivariate, Time-Series 
\n\t\tClassification \n\t\tReal 
p >  n t t < d > p class = "normal" > 640 nbsp;   n t t t < d > p class = "normal" > 12 nbsp;   n t t t < d > p class = "normal" > 12 nbsp;   n t t t < d > p class = "normal" > 12 nbsp;   n t t t < d > p class = "normal" > 12 nbsp;   n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n < d > n t < d > n t < d > n t < d > n t < d > n t < d > n t < d > n < d > n < d > n < d > n < d > n < d > n < d > n < d > n < d > n < d > n < d > n < d > n < d > n < d > n < d > n < d > n < d >
class="normal"> \n\t\t\t<!-- <td>class="normal">Other&nbsp; -->\n\t\t\t
r>\n\t\t<a href="datasets/KDD+Cup+1998+Data"><img src="assets/MLimages/SmallLarge"></a>
default.jpg" border=1 /></a>&nbsp;<b><a href="datasets/KDD+Cup+1998+Data">
sed for The Second International Knowledge Discovery and Data Mining Tools Competition, which was held in c
onjunction with KDD-98 -->\n\t\t\tp class="normal">Multivariate \n\t\t\t
Regression \n\t\tCategorical, Integer 
\ \n\t\t\t191779 \n\t\t\t481 \n\t\t\t<
td>1998 \n\t\t\<!-- <td>Other&nbsp; -->\n\t\t\</t
r>\n\t\t<a href="datasets/KDD+Cup+1999+Data"><img src="assets"
/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/K"
DD+Cup+1999+Data">KDD Cup 1999 Data</a></b>
">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  -->\n\t\t\tMultivariate&nbsp
```

```
Integer\  \n\t\t4000000\  \n\t\t\t42
src="assets/MLimages/SmallLarge131.jpg" border=1 /></a>&nbsp;<b><a href="dat
asets/M.+Tuberculosis+Genes">M. Tuberculosis Genes</a></b>\n\t\t\t<!-- <td>cl
ass="normal"> Data giving characteristics of each ORF (potential gene) in the M. tuberculosis bacterium. Sequ
ence, homology (similarity to other genes) and structural information, and function (if known) are provided&nbs
p;-->\n\t\tRelational \n\t\t 
/n\t\t /n\t\t\t /n\t\t\t
lass="normal">\ \n\t\t\t2001 \n\t\t\t<!-- <td>2001&nbsp;
mal">Life  -->\n\t\t\\n\t\t\<a href="datasets/Mo"
vie"><img src="assets/MLimages/SmallLarge132.jpg" border=1 /></a>&nbsp;<b>
a href="datasets/Movie">Movie</a></b>\n\t\t\t<!-- <td>class="normal">This data
set contains a list of over 10000 films including many older, odd, and cult films. There is information on actors, c
asts, directors, producers, studios, etc.  -->\n\t\t\tMultivariate, Relational
 \n\t\t <math>\n\t\t&nbsp;<math>\n\t
tp class="normal">10000 \n\t\tp class="normal"> \n\t\tp class="normal">
ss="normal">1999\ \n\t\t<!-- <td>Other&nbsp; -->\n\t\t\t
\t\t<a href="datasets/MSNBC.com+Anonymous+Web+Data"><img src="assets/MLimages/S
mallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/MSNBC.com+A"
nonymous+Web+Data">MSNBC.com Anonymous Web Data</a></b>\n\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. 
d> --> \ln t \leq d> 
\t\tCategorical \n\t\t989818 \n\t\
t\t \n\t\t\t \n\t\t\t<!-- <td>&nbsp;
="normal">Computer  -->\n\t\t\t\n\t\t\t<a href="d
atasets/NSF+Research+Award+Abstracts+1990-2003"><img src="assets/MLimages/SmallLarge134.jpg" border
=1 /></a>&nbsp;<b><a href="datasets/NSF+Research+Award+Abstracts+1990-20"
03">NSF Research Award Abstracts 1990-2003</a></b>\n\t\t\t<!-- <td>\neq class="no"
rmal">This data set consists of (a) 129,000 abstracts describing NSF awards for basic research, (b) bag-of-wor
d data files extracted from the abstracts, (c) a list of words used for indexing the bag-of-word  --
>\n\t\t\t\text \n\t\t\t\text \n\t\t\t<p
class="normal"> 
$$/p>
$$/p>
$$/p>
$$/p>
$$/p>
$$/p>
$$/p>
mal">\ \n\t\t\t2003\ \n\t\t\t<!-- <td>Othe
r  -->\n\t\t\n\t\t<a href="datasets/Pioneer-1+Mobile+Robot+Data"
><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a
href="datasets/Pioneer-1+Mobile+Robot+Data">Pioneer-1 Mobile Robot Data</a></b>
d>\n\t\t\t<!-- <td>This dataset contains time series sensor readings of the Pioneer-1 mobile r
obot. The data is broken into "experiences" in which the robot takes action for some period of time and experie
nces a control  -->\n\t\tMultivariate, Time-Series \n\t\t\
 \n\t\t\tCategorical, Real \n\t\t\t<t
d> \n\t\t\t \n\t\t\t<p class="normal"
">1999 \n\t\t<!-- <td>Computer&nbsp; -->\n\t\t\<tr bgcolor="
DDEEFF">\n\t\t\t<a href="datasets/Pseudo+Periodic+Synthetic+Time+Series"><img src="as"
sets/MLimages/SmallLarge136.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/P"
seudo+Periodic+Synthetic+Time+Series">Pseudo Periodic Synthetic Time Series</a></b>
>\n\t\t<!-- <td>This data set is designed for testing indexing schemes in time series da
tabases. The data appears highly periodic, but never exactly repeats itself.  -->\n\t\t\t<p cla
ss="normal">Univariate, Time-Series\  \n\t\t\t  \n\t\t\t<<p class="normal">&nbsp; 
p class="normal">nt/t<0p class="normal">1000008nbsp;\n\t\t\tcp class="normal">1000008nbsp;
rmal">\ \n\t\t\t1999\ \n\t\t\t<!-- <td>Oth
er  -->\n\t\t\n\t\t<a href="datasets/Reuters-21578+Text+Categori
zation+Collection"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class
s="normal"><b><a href="datasets/Reuters-21578+Text+Categorization+Collection">Reuters-21578 Text Categorization+Collection">Reuters-21578 Text Categorization+Collection">Reuters-21578 Text Categorization+Collection
ocuments that appeared on Reuters newswire in 1987. The documents were assembled and indexed with cate
gories.  -->\n\t\t\tclass="normal">Text \n\t\t\tclass="normal">Clas
sification \n\t\tCategorical \n\t\t<p class="normal"
>21578%nbsp;\n\t\t5%nbsp;\n\t\t1997%nbs
p;\n\t\t\<!-- <td>Other&nbsp; -->\n\t\t\
td><a href="datasets/Robot+Execution+Failures"><img src="assets/MLimages/SmallLarge138.j"
```

```
pg" border=1 /></a>&nbsp;<b><a href="datasets/Robot+Execution+Failures">Rob
ot Execution Failures</a></b>\n\t\t<!-- <td>class="normal">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  -->\n\t\t\tMultivariate, Ti
me-Series \n\t\t\tClassification \n\t\t\tClassification 
al">Integer \n\t\t\463 \n\t\t\t90&
nbsp;   n\t  p\ class = "normal" > 1999 & nbsp;   n\t  p\ class = "normal" > Physical & nbsp;  
bsp; -->\n\t\t\\n\t\t\t<a href="datasets/Synthetic+Control+Chart+Time+Ser">+Ser</a>
ies"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b
><a href="datasets/Synthetic+Control+Chart+Time+Series">Synthetic Control Chart Time Series</a></b>
td>\n\t\t\<!-- <td>This data consists of synthetically generated control char
ts. -->\n\t\t\tC
lassification, Clustering \n\t\t\tReal \n\t\t\t<p class="n
ormal">600\ \n\t\t\t \n\t\t\t1999&n class="normal"> 
bsp; /n/t/t <!-- <td>Other&nbsp; -->/n/t/t/t 
\t<a href="datasets/Syskill+and+Webert+Web+Page+Ratings"><img src="assets/MLimages"
/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Syskill+and+"
Webert+Web+Page+Ratings">Syskill and Webert Web Page Ratings</a></b>
!-- 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) -->\n\t\t\tMultivariate, Text \n\t\t\t<p class
="normal">Classification \n\t\tCategorical \n\t\t<p
class="normal">332\ \n\t\t\t5\ \n\t\t\t5 
">1998 \n\t\t\</r>\n\t\t\t
<a href="datasets/UNIX+User+Data"><img src="assets/MLimages/SmallLarge141.jpg" border=
1 /></a>&nbsp;<b><a href="datasets/UNIX+User+Data">UNIX User Data</a></b>
/p>\n\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. 
-->\n\t\t\tText, Sequential \n\t\t\t 
/n\t\t /n\t\t\t /n\t\t\t 
lass="normal">\ \n\t\t\ \n\t\t\<!-- <td>
Computer  -->\n\t\t\\n\t\t\t<a href="datasets/Volume of the color by the col
canoes+on+Venus+-+JARtool+experiment"><img src="assets/MLimages/SmallLarge142.jpg" border=1 /></a>&
nbsp;<b><a href="datasets/Volcanoes+on+Venus+-+JARtool+experiment">Volcan
oes on Venus - JARtool experiment</a></b>\n\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.  -->\n\t\tI
mage \n\t\tClassification \n\t\t
 \n\t\t <math>\n\t\t&nbsp;<math>\n\t\t \n\t
t<0>c class="normal">&nbsp;\n\t\t<!-- <td>c class="normal">Physical&nbsp; -->\n\t\t\t<!-- <td>c class="normal">Physical&nbsp;
/tr>\n\t\t<a href="datasets/Statlog+%28Australian+Credit+Approval%29"><img src="as
sets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="dataset"
s/Statlog+%28Australian+Credit+Approval%29">Statlog (Australian Credit Approval)</a></b>
le>\n\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  -->\n\t\t\t<p cl
ass="normal">Multivariate \n\t\tClassification \n\t\t\t<td
>Categorical, Integer, Real \n\t\t\t690 
\ \n\t\t\t14 \n\t\t\t \n\t\t\t<!-- <td><
p class="normal">Financial  -->\n\t\t\t\n\t\t\t<a h
ref="datasets/Statlog+%28German+Credit+Data%29"><img src="assets/MLimages/SmallLargedefault.jpg" bord
er=1 /></a>&nbsp;<b><a href="datasets/Statlog+%28German+Credit+Data%29">S
tatlog (German Credit Data)</a></b>\/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  -->\n\t\t\tMultivariate 
>\n\t\tClassification \n\t\tCategorical, Integer&
nbsp;  nt/t  1000 & nbsp;  n/t/t  20 & nbsp; 
td>\n\t\t1994 \n\t\t\t<!-- <td>Financial&nbsp;</t
d> -->\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\n\t\t\t\n\t\t\n\t\t\t\n\t\t\t\n\t\t\n\t\t\t
mages/SmallLarge45.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Statlog+%"
28Heart%29">Statlog (Heart)</a></b>\n\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&nbsp: -->\n\t\t\tMultivariate&nbsp:\n\t\t\
```

```
tClassification \n\t\tCategorical, Real </p
>n\t\t270 n\t\t13 n\t\t\<
td> \n\t\t<!-- <td>Life&nbsp; -->\n\t\t
gcolor="DDEEFF">\n\t\t\t<a href="datasets/Statlog+%28Landsat+Satellite%29"><img src="
assets/MLimages/SmallLarge146.jpg" border=1 /></a>&nbsp;<b><a href="dataset"
s/Statlog+%28Landsat+Satellite%29">Statlog (Landsat Satellite)</a></b>
td>Multi-spectral values of pixels in 3x3 neighbourhoods in a satellite image, and the classifi
cation associated with the central pixel in each neighbourhood  -->\n\t\t\t
Multivariate \n\t\t\tClassification \n\t\t\t
mal">Integer\ \n\t\t\6435\ \n\t\t\3
6\ \n\t\t1993\ \n\t\t\t<!-- <td>Physical 
  -->\n\t\t\\n\t\t\<a href="datasets/Statlog+%28Image+Segmentatio">
<a href="datasets/Statlog+%28Im
n%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
><b><a href="datasets/Statlog+%28Image+Segmentation%29">Statlog (Image Segmentation)</a></b>
>\n\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.&nb
sp;-->\n\t\t\tMultivariate \n\t\tClassific
ation \n\t\tReal \n\t\t2310&nbs
p;n\t\t19 \n\t\t1990 \
n/t/t/t<!--< td>Other&nbsp; -->/n/t/t/t/n/t/t/t
tr><a href="datasets/Statlog+%28Shuttle%29"><img src="assets/MLimages/SmallLarge92.jpg" border=1 />
</a>&nbsp;<b><a href="datasets/Statlog+%28Shuttle%29">Statlog (Shuttle)</a></
b>\n\t\t\<!-- <td>The shuttle dataset contains 9 attributes all of w
hich are numerical. Approximately 80% of the data belongs to class 1  -->\n\t\t\t<p class="n
ormal">Multivariate \n\t\t\tClassification \n\t\t\tClassification 
ss="normal">lnteger \n\t\t\t58000 \n\t\t\t58000 
ormal">9\ \n\t\t\t \n\t\t\t<!-- <td>Physic 
al  -->\n\t\t\\r\\t\t<a href="datasets/Statlog+%28Vehicle+Silhouette">te+Silhouette</a>
s%29"><img src="assets/MLimages/SmallLarge149.jpg" border=1 /></a>&nbsp;<b
><a href="datasets/Statlog+%28Vehicle+Silhouettes%29">Statlog (Vehicle Silhouettes)</a></b>
table>\n\t\t\t<!-- <td>3D objects within a 2D image by application of an ensemble of sha
pe feature extractors to the 2D silhouettes of the objects.  -->\n\t\t\tMulti
variate \    \\ lass="normal"> Classification \    \\ lass="normal"> class="normal"> class="normal">
>Integer \n\t\t946 \n\t\t18&nbs
p;\n\t\t \n\t\t<!-- <td>Other&nbsp;
d> -->\n\t\t\z/tr>\n\t\t\z<a href="datasets/Connectionist+Bench+%28"
Nettalk+Corpus%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p cl
ass="normal"><b><a href="datasets/Connectionist+Bench+%28Nettalk+Corpus%29">Connectionist Bench (Net
talk Corpus)</a></b>\n\t\t\t<!-- <td>The file "nettalk.data" contain
s 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  -->\n\t\t\tMultivariate 
n\t\t\t \n\t\t\tCategorical \n\t\t\t<td
> class="normal">20008 \n\t\t\t> class="normal">4 \n\t\t\t> class="normal">4 
normal">\ \n\t\t\<!-- <td>Other&nbsp; --- \n\t\t\t
able><a href="datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29"><img src="asset"
s/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/"</pre>
Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29">Connectionist Bench (Sonar, Mines vs. Rocks)</
a></b>\n\t\t<!-- <td>The task is to train a network to discriminat
e between sonar signals bounced off a metal cylinder and those bounced off a roughly cylindrical rock. </
p> -->\n\t\t\Multivariate \n\t\t\Classification
 \n\t\tReal \n\t\t208 
/n\t/t60 /n\t/t\t /n\t/t\t<!-- <tr>
d>Physical  -->\n\t\t\t bgcolor="DDEEFF">\n\t\t\t<
a href="datasets/Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29"><img src="assets/MLi
mages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Conne
ctionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29">Connectionist Bench (Vowel Recognition - Det
erding Data)</a></b>\n\t\t\t<!-- <td>Speaker independent recogn
ition of the eleven steady state vowels of British English using a specified training set of lpc derived log area rati
os.  -->\n\t\t\t \n\t\t\tClassification
 \n\t\tReal \n\t\t\t528 
/n\t\t10 /n\t\t\t /n\t\t\t<!-- <td>&nbsp;/n\t\t\t<!-- <td>&nbsp;/n\t\t<!-- <td>&nbsp;/n\t\t<!-- <td>&nbsp;/n\t\t<!-- <td>&nbsp;/n\t\t<!-- <td>&nbsp;/n\
d><p_class="normal">Other&nbsp: -->\n\t\t
```

```
omic+Sanctions"><img src="assets/MLimages/SmallLarge153.jpg" border=1 /></a>&nbsp;<p class="n
ormal"><b><a href="datasets/Economic+Sanctions">Economic Sanctions</a></b>
\t\t\t<!-- <td>Domain Theory on Economic Sanctions; Undocumented&nbsp; -->\n\t\
t\tDomain-Theory \n\t\t \n\t\t\t<t
d> \n\t\t\t \n\t\t\t<p class="normal"
"> \n\t\t \n\t\t<!-- <td>Financial&nb
sp; -->\n\t\t\t\n\t\t\t<a href="datasets/Protein+Data"><i
mg src="assets/MLimages/SmallLarge154.jpg" border=1 /></a>&nbsp;class="normal"><b><a href=
"datasets/Protein+Data">Protein Data</a></b>\n\t\t<!-- <td>Und
ocumented  -->\n\t\t\t \n\t\t\t&nbs
p;n\t\t \n\t\t\t \n\t\t\t \n\t\t\t \n\t\t\t 
d> \n\t\t\t \n\t\t\t<!-- <td>&nbsp;
rmal">Life  -->\n\t\t\<a href="datasets/Cloud"><img src="ass"
ets/MLimages/SmallLarge155.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Cl
oud">Cloud</a></b>\n\t\t\t<!-- <td>class="normal">Little Documentation&nbsp;</
p> -->\n\t\t\Multivariate \n\t\t\t </
td>\n\t\tReal\ \n\t\t\t1024\ \n\t\t\t1024 
td><p\ class="normal">10\&nbsp;<p\ class="normal">1989\&nbsp;<p\ class="normal">10&nbsp;
ass="normal">Physical  -->\n\t\t\\n\t\t\t<a href=
"datasets/Callt2+Building+People+Counts"><img src="assets/MLimages/SmallLarge156.jpg" border=1 /></a>&n
bsp;class="normal"><b><a href="datasets/Callt2+Building+People+Counts">Callt2 Building People
Counts</a></b>\n\t\t<!-- <td>This data comes from the main do
or of the Callt2 building at UCI.  -->\n\t\t\tMultivariate, Time-Series&nbsp
;\n\t\t \n\t\t\tCategorical, Integer&nbsp
;\n\t\t\t10080 \n\t\t4 \n
t/tp class="normal">2006 \n\t\t<!-- <td>p class="normal">Other&nbsp; -->\n\t
\t\t\n\t\t\t<a href="datasets/Dodgers+Loop+Sensor"><img src="assets/MLimages/S
mallLarge157.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/Dodgers+Loop+S"
ensor">Dodgers Loop Sensor</a></b>\n\t\t\t<!-- <td>Loop senso
r data was collected for the Glendale on ramp for the 101 North freeway in Los Angeles  -->\n\t\
t\tMultivariate, Time-Series \n\t\t </t
d>\n\t\tCategorical, Integer \n\t\t50400 
\n\t\t3 \n\t\t\t2006 \n\t\t\t2006 
t\t<!-- <td>Other&nbsp; -->\n\t\t\t bgcolor="DDEEFF">\n\t\t\t
<a href="datasets/Poker+Hand"><img src="assets/MLimages/SmallLarge158.jpg" border=1 /></a>&nbsp;</
td><b><a href="datasets/Poker+Hand">Poker Hand</a></b>
\n\t\t<!-- <td>Purpose is to predict poker hands&nbsp; -->\n\t\t\t<p class="normal"
mal">Multivariate \n\t\t\tClassification \n\t\t\t<p class
="normal">Categorical, Integer \n\t\t1025010 \n\t\t<td
>11 \n\t\t\2007 \n\t\t\<!-- <td>2007&nbsp;
ss="normal">Game  -->\n\t\t\n\t\t<a href="datasets/MAGIC+Gam">datasets/MAGIC+Gam</a>
ma+Telescope"><img src="assets/MLimages/SmallLarge159.jpg" border=1 /></a>&nbsp;<p class="no
rmal"><b><a href="datasets/MAGIC+Gamma+Telescope">MAGIC Gamma Telescope</a></b>
able>\n\t\t<!-- <td>Data are MC generated to simulate registration of high energy gam
ma particles in an atmospheric Cherenkov telescope  -->\n\t\tMultivaria
te \n\t\tClassification \n\t\t\tReal
\alpha = \frac{10^{-10}}{h} / h t/t  p class="normal">19020 hbsp;   \n/t/t  p class="normal">11 hbsp; 
>\n\t\t\2007 \n\t\t\<!-- <td>Physical&nbsp;
-->\n\t\t\ datasets/UJI+Pen+Characters"><i</td>
mg src="assets/MLimages/SmallLarge160.jpg" border=1 /></a>&nbsp;class="normal"><b><a href=
"datasets/UJI+Pen+Characters">UJI Pen Characters</a></b>\n\t\t<!-- <td><p class
="normal">Data consists of written characters in a UNIPEN-like format  -->\n\t\t\t<p class="n
ormal">Multivariate, Sequential \n\t\tClassification \n\t\t
\tInteger \n\t\t1364 \n\t\t\t<
p\ class="normal">\ \n\t\t\t<p\ class="normal">2007\&nbsp;\n\t\t\t<!-- <td><p\ class="normal">2007\&nbsp;
ormal">Computer  -->\n\t\t\r\t\t<a href="datasets/Mammographi">datasets/Mammographi</a>
c+Mass"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="norma"
I"><b><a href="datasets/Mammographic+Mass">Mammographic Mass</a></b>
\t<!-- <td>Discrimination of benign and malignant mammographic masses based on BI-RAD
S attributes and the patient\'s age.  -->\n\t\t\tMultivariate </td
>\n\t\t\tClassification \n\t\t\tInteger </t
d_n/t/t/td>n class="normal">961&nbsn://n>//td\n\t\t\t-td>n class="normal">6&nbsn://n>//td\n\t\t\t-td>
```

```
mornial 200 randop, 4p2 4ta2 intitit ta2 4p diado— normal 20ambop, 4p2 4ta2 intitit ta
p class="normal">2007 \n\t\t\<!-- <td>Life&nbsp; -->\n\t\t\
gcolor="DDEEFF">\n\t\t\t<a href="datasets/Forest+Fires"><img src="assets/MLimages/Sm"
allLarge162.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/Forest+Fires">Fore
st Fires</a></b>\r\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).  -->\n\t\t\t<t
d>Multivariate \n\t\tRegression \n\t\t
\tReal \n\t\t\t517 \n\t\t\t<p cl
ass="normal">13\ \n\t\t\2008\ \n\t\t\<!-- <td>13&nbsp;
rmal">Physical  -->\n\t\t\<\r\>\n\t\t\t<a href="datasets/Reuters+Transcri
bed+Subset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="no
rmal"><b><a href="datasets/Reuters+Transcribed+Subset">Reuters Transcribed Subset</a></b>
\n\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. -->\n\t\t\tText \n\t\t\tClassi
fication \n\t\t\t \n\t\t\t200 
p  n t t < d > p class = "normal" > 2008 & nbsp;  n t t  p class = "normal" > 2008 & nbsp;  n t t  p class = "normal" > 2008 & nbsp;  n t  p class = "normal" > 2008 & nbsp;  n t  p class = "normal" > 2008 & nbsp;  n t  p class = "normal" > 2008 & nbsp;  n t  p class = "normal" > 2008 & nbsp;  n t  p class = "normal" > 2008 & nbsp;  n t  p class = "normal" > 2008 & nbsp;  n t  p class = "normal" > 2008 & nbsp;  n t  p class = "normal" > 2008 & nbsp;  n t  p class = "normal" > 2008 & nbsp;  n t  p class = "normal" > 2008 & nbsp; 
!-- Business  -->\n\t\t\z\tr>\n\t\t\z\ta>
<a href="datasets/Bag+of+Words"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&n
bsp;<b><a href="datasets/Bag+of+Words">Bag of Words</a></b>
able>\n\t\t<!-- <td>This data set contains five text collections in the form of bags-of-wo
rds.  -->\n\t\tText \n\t\tClusteri
/n/t/t<!--<td>Other&nbsp;-->/n/t/t/t/n/t/t/t<a
href="datasets/Concrete+Compressive+Strength"><img src="assets/MLimages/SmallLarge165.jpg" border=1 />
</a>&nbsp;<b><a href="datasets/Concrete+Compressive+Strength">Concrete Co
mpressive Strength</a></b>\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.  -->\n\t\t\tMultivariate \n\t\t\t<p class
="normal">Regression \n\t\t\tReal \n\t\t\t<p class="n
ormal">1030\ \n\t\t\t9\ \n\t\t\t2007
 \n\t\t<!-- <td>Physical&nbsp;<math>--- \n\t\t\t bgcolor="DDEEFF"
>\n\t\t\t<a href="datasets/Hill-Valley"><img src="assets/MLimages/SmallLarge166.jpg" bord
er=1 /></a>&nbsp;<b><a href="datasets/Hill-Valley">Hill-Valley</a></b></
tr>\n\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 $\text{$\text{$\text{$}}\text{bump}}$
� in the terrain) or a Valley (a �dip� in the terrain).  -->\n\t\t\tSequenti
al \n\t\tClassification \n\t\t\tReal
\alpha = \frac{\sqrt{y}}{\sqrt{t}} \sqrt{t} \cdot \frac{100 - t}{t} \cos \frac{y}{t} = \frac{y}{t} \sin \frac{y
n\t\tp class="normal">2008 n\t\t\t<!-- <td>p class="normal">Other&nbsp;
> -->\n\t\t\n\t\t\t<a href="datasets/Arcene"><img src="assets/MLimages/SmallLarg"
e167.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Arcene">Arcene</a></b></
p>\n\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.  -->\n\t\t\t<td
>Multivariate \n\t\tClassification \n\t\
t\tReal \n\t\t900 \n\t\t<p cl
ass="normal">10000\ 
$$ rormal">2008\ 
$$ rormal">2008\ 
$$ rormal">2008\ 
$$ rormal">2008\ 
="normal">Life  -->\n\t\t\t\n\t\t\t<a href="dataset"
s/Dexter"><img src="assets/MLimages/SmallLarge168.jpg" border=1 /></a>&nbsp;
<b><a href="datasets/Dexter">Dexter</a></b>\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  -->\n\t\tMultivariate \n\t\t<p class="no
rmal">Classification \n\t\t\tInteger \n\t\t\t<p class="n
ormal">2600\ \n\t\t\20000\ \n\t\t\
2008 \n\t\t\<!-- <td>Other&nbsp; -->\n\t\t
><a href="datasets/Dorothea"><img src="assets/MLimages/SmallLarge169.jpg" border=1 /></a>&nbsp;
<b><a href="datasets/Dorothea">Dorothea</a></b>
t\t<!-- <td>DOROTHEA is a drug discovery dataset. Chemical compounds represented by st
            alacular factures must be algorified as active (hinding to thrombin) or inactive. This is one of F
```

```
ructural molecular reatures must be classified as active (binding to tinombin) of mactive. This is one of 3 datase
ts of the NIPS 2003 feature selection challenge.  -->\n\t\t\tMultivariate&n
bsp;\n\t\tClassification \n\t\t\tInteger&
nbsp;   nt/t  p class = "normal" > 1950 & nbsp;   nt/t  p class = "normal" > 100000 & nbsp; 
> -->\n\t\t\n\t\t\t<a href="datasets/Gisette"><img src="assets/"
MLimages/SmallLarge170.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/Gisett"
e">Gisette</a></b>\n\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  -->\n\t\t\t
Multivariate \n\t\t\tClassification \n\t\t\t
mal">Integer\ \n\t\t\13500\ \n\t\t\
5000\ \n\t\t2008\ \n\t\t\t<!-- <td>Com
puter  -->\n\t\t\\r\t\t\</d><a href="datasets/Madelon"><img src="assets"
/MLimages/SmallLarge171.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Mad"
elon">Madelon</a></b>\n\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.  
-->\n\t\t\tMultivariate \n\t\t\tClassification&n
bsp;\n\t\t\t4400 \n\t\t4400 
td>nttp class="normal">500 \n\t\tp class="normal">2008 \n\t\t\t<
!-- Other  -->\n\t\t\\n\t\t\t<td
><a href="datasets/Ozone+Level+Detection"><img src="assets/MLimages/SmallLarge172.jpg" border=1 /></a>
 <b><a href="datasets/Ozone+Level+Detection">Ozone Level Detection</a>
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. 
->\n\t\tMultivariate, Sequential, Time-Series \n\t\t\t<p class="normal"
al">Classification \n\t\t\tReal \n\t\t<p class="normal"
">2536\ \n\t\t\t73\ \n\t\t\t2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">2008\&nbs="normal">20
p;/n\t\t<!-- <td>Physical&nbsp; -->\n\t\t\t<t
d><a href="datasets/Abscisic+Acid+Signaling+Network"><img src="assets/MLimages/SmallLargedefault.jpg" bo
rder=1 /></a>&nbsp;<b><a href="datasets/Abscisic+Acid+Signaling+Network">Abs
cisic Acid Signaling Network</a></b>\n\t\t\t<!-- <td>class="normal">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.  -->\n\t\t\tclass="normal">Multivariate \n\t\t\tclass="normal"
mal">Causal-Discovery \n\t\t\tInteger \n\t\t\t<p class
="normal">300\ \n\t\t\t43\ \n\t\t\t2
008 \n\t\t<!-- <td>Life&nbsp; -->\n\t\t\c/tr>
\n\t\t<ahref="datasets/Parkinsons"><img src="assets/MLimages/SmallLarge174.jpg" bord
er=1 /></a>&nbsp;<b><a href="datasets/Parkinsons">Parkinsons</a></b>
>\n\t\t<!-- <td>Oxford Parkinson\'s Disease Detection Dataset&nbsp;
-->\n\t\t\tMultivariate \n\t\t\tClassification&n
bsp;  \ln t  p class = "normal" > Real   \ln t  p class = "normal" > 197   \ln t  p class = "normal" > 197  
d>\n\t\t23 \n\t\t\t2008 \n\t\t\t<!--
Life  -->\n\t\t\z\tr>\n\t\t\z\table><a href="datasets/Char
acter+Trajectories"><img src="assets/MLimages/SmallLarge175.jpg" border=1 /></a>&nbsp;<p class=
"normal"><b><a href="datasets/Character+Trajectories">Character Trajectories</a></b>
\n\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.  -->\n\t\t\tTime-S
eries \n\t\tClassification, Clustering \n\t\t<p class=
"normal">Real\ \n\t\t2858\ \n\t\t
3 \n\t\t2008 \n\t\t\t<!-- <td>Compute
r\ -->\n\t\t bgcolor="DDEEFF">\n\t\t\t<a href="datasets/Blood+Tran">+Tran</a>
sfusion+Service+Center"><img src="assets/MLimages/SmallLarge176.jpg" border=1 /></a>&nbsp;<p
class="normal"><b><a href="datasets/Blood+Transfusion+Service+Center">Blood Transfusion Service Center</a>
a></b>/td>\n\t\t<!-- <td>Data taken from the Blood Transfusion Ser
vice Center in Hsin-Chu City in Taiwan -- this is a classification problem.   -->\n\t\t\t<p class
class="normal">Real\ 
$$ rormal">748\ 
$$ class="normal">748\ 
$$ rormal">Real\ 
$$ rormal">748\
```

```
mai >5&nosp;</tu>\n\t\\\<!d>2006&nosp;</tu>\n\t\\\\<!-- <tu>bu
siness  -->\n\t\t\<a href="datasets/UJI+Pen+Characters+%28">
Version+2%29"><img src="assets/MLimages/SmallLarge160.jpg" border=1 /></a>&nbsp;<p class="no
rmal"><b><a href="datasets/UJI+Pen+Characters+%28Version+2%29">UJI Pen Characters (Version 2)</a>
>ormal">A pen-based database with more than 11k isolat
ed handwritten characters -->\n\t\t\ctd>Multivariate, Sequential <
/td>\n\t\tClassification \n\t\t\tInteger 
/n\t\t11640\n\t\t /n\t\t\t 
d>2009 \n\t\t\t<!-- <td>Computer&nbsp; -->\n\t\t
\t\n\t\t\cd><a href="datasets/Semeion+Handwritten+Digit"><img sr
c="assets/MLimages/SmallLarge178.jpg" border=1 /></a>&nbsp;c class="normal"><b><a href="data"
sets/Semeion+Handwritten+Digit">Semeion Handwritten Digit</a></b>
1593 handwritten digits from around 80 persons were scanned, stretched in a rectangular b
ox 16x16 in a gray scale of 256 values.  -->\n\t\t\tMultivariate 
\n\t\tClassification \n\t\tInteger </p
>/n\t\t1593 /n\t\t256 /n\t\t
\t2008 \n\t\t<!-- <td>Computer&nbsp;-->\
n\t\t\n\t\t\t\n\t\t\t<dt><a href="datasets/SECOM"><img src="assets/MLimages/SmallLarge17" |
9.jpg" border=1 /></a>&nbsp;<b><a href="datasets/SECOM">SECOM</a></b></p
>\n\t\t<!-- <td>Data from a semi-conductor manufacturing process&
nbsp; -->\n\t\t\tMultivariate \n\t\t\tClassi
fication, Causal-Discovery \n\t\tReal \n\t\t<p class
="normal">1567\ \n\t\t\t591 \n\t\t\t
>2008 \n\t\t\t<!-- <td>Computer&nbsp; -->\n\t\t\t
DEEFF">\n\t\t\t<a href="datasets/Plants"><img src="assets/MLimages/SmallLarge180.jpg"
border=1 /></a>&nbsp;<b><a href="datasets/Plants">Plants</a></b>
\n\t\t<!-- <td>Data has been extracted from the USDA plants database. It cont
ains all plants (species and genera) in the database and the states of USA and Canada where they occur. &nbs
p;-->\n\t\tclass="normal">Multivariate \n\t\t\tcp class="normal">Clusterin
g \n\t\tCategorical \n\t\t\t22632
 \n\t\t70 \n\t\t\t2008 
/n\t/t<-<-td>Life&nbsp; -->\n\t\t
datasets/Libras+Movement"><img src="assets/MLimages/SmallLarge181.jpg" border=1 /></a>&nbsp;
<b><a href="datasets/Libras+Movement">Libras Movement</a></b>
/td>\n\t\t<!-- <td>The data set contains 15 classes of 24 instances each. Each class referen
ces to a hand movement type in LIBRAS (Portuguese\r\nname \'L\phingua BRAsileira de Sinais\', oficial brazilian
signal language).  -->\n\t\tMultivariate, Sequential \n\t\t
tClassification, Clustering \n\t\tReal 
/n\t\t360 /n\t\t91 /n\t\t\t91 /n\t\t\t
d><p\ class="normal">2009\&nbsp;\n\t\t\t<!--<td><p\ class="normal">Other&nbsp;\-->\n\t\t\t
>\n\t\t\t<a href="datasets/Concrete+Slump+Test"><img src="assets/"
MLimages/SmallLarge165.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/Concr
ete+Slump+Test">Concrete Slump Test</a></b>\n\t\t<!-- <td>class="normal">C
oncrete is a highly complex material. The slump flow of concrete is not only determined by the water content, b
ut that is also influenced by other concrete ingredients.  -->\n\t\t\tMultiva
riate \n\t\t\tRegression \n\t\tRe
al\   \\ lass = "normal" > 103\   \\ lass = "normal" > 10\   \\ lass = "nor
\n\t\t2009 \n\t\t\t<!-- <td>Computer&nbsp;
-->\n\t\t\<a href="datasets/Communities+and+Crime"><img src="asset">
s/MLimages/SmallLarge183.jpg" border=1 /></a>&nbsp;cd><b><a href="datasets/Co
mmunities+and+Crime">Communities and Crime</a></b>\n\t\t\t<!-- <td><p class="n
ormal">Communities within the United States. The data combines socio-economic data from the 1990 US Cens
us, law enforcement data from the 1990 US LEMAS survey, and crime data from the 1995 FBI UCR. 
-->\n\t\t\tMultivariate \n\t\t\tRegression&nb
sp;\n\t\tReal \n\t\t\t1994 
d>\n\t\t128\ \n\t\t\t2009\ \n\t\t\t<!--
<a href="datasets/Acute+Inflammations"><img src="assets/MLimages/SmallLarge184.jpg" border=1 /></a>&nb
sp;<b><a href="datasets/Acute+Inflammations">Acute Inflammations</a></b>
\n\t\t<!-- <td>The data was created by a medical expert as a data se
t to test the expert system, \r\nwhich will perform the presumptive diagnosis of two diseases of the urinary syste
m.\r\n\ -->\n\t\tclass="normal">Multivariate\ /\n\t\t\t
```

```
>Classification \n\t\t\ctd>Categorical, Integer \n\t\t\t<p cl
ass="normal">120 \n\t\t\t6 \n\t\t\t
2009\ -->\n\t\t<--<td>Life\&nbsp;-->\n\t\t<--<tr>
<a href="datasets/Wine+Quality"><img src="assets/MLimages/SmallLarge186.jpg" border=1 /></a>&nb
sp;<b><a href="datasets/Wine+Quality">Wine Quality</a></b>
e>\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 [Co
rtez et al., 2009], http://www3.dsi.uminho.pt/pcortez/wine/).  -->\n\t\tMu
ltivariate \n\t\tClassification, Regression \n\t\t\t<p cl
ass="normal">Real\ \n\t\t\t4898\ \n\t\t\t4898\ 
mal">12\ \n\t\t\t2009\ \n\t\t\t<!-- <td>B
usiness  -->\n\t\t\\n\t\t\t<a href="datasets/URL+"
Reputation"><img src="assets/MLimages/SmallLarge187.jpg" border=1 /></a>&nbsp;
I"><b><a href="datasets/URL+Reputation">URL Reputation</a></b>\n\t\t\<!-- <td><
p class="normal">Anonymized 120-day subset of the ICML-09 URL data containing 2.4 million examples and 3.
2 million features.  -->\n\t\t\Multivariate, Time-Series \n\
t\t\tClassification \n\t\tInteger, Real 
mallLarge188.jpg" border=1 /></a>&nbsp;<b><a href="datasets/p53+Mutants">p5
3 Mutants</a></b>\n\t\t\t<!-- <td>The goal is to model mutant p5
3 transcriptional activity (active vs inactive) based on data extracted from biophysical simulations.\r\n 
>-->\n\t\t\tMultivariate \n\t\t\tClassification&
nbsp;\n\t\tReal \n\t\t16772 </p
>/n\t\tp class="normal">5409 /n\t\tp class="normal">2010 /n\t
\t\t<!-- <td>class="normal">Life&nbsp; -->\n\t\t\n\t\t\t<t
d><a href="datasets/Parkinsons+Telemonitoring"><img src="assets/MLimages/SmallLarge174.jpg" border=1 />
</a>&nbsp;<b><a href="datasets/Parkinsons+Telemonitoring">Parkinsons Telemo
nitoring</a></b>\n\t\t\t<!-- <td>Oxford Parkinson\'s Disease Tele
monitoring Dataset  -->\n\t\t\tMultivariate \n\t\t\t<p cl
ass="normal">Regression \n\t\t\tInteger, Real \n\t\t\t
5875\ \n\t\t\t26\ \n\t\t\t26 
ormal">2009 \n\t\t\<!-- <td>Life&nbsp; -->\n\t\t\
<a href="datasets/Demospongiae"><img src="assets/MLimages/SmallLarge190.jpg" border=1 /
></a>&nbsp;<b><a href="datasets/Demospongiae">Demospongiae</a></b></
td>\n\t\t\<!-- <td>Marine sponges of the Demospongiae class classification
n domain.  -->\n\t\t\tMultivariate \n\t\t\t<p class="normal"
mal">Classification \n\t\t\tInteger \n\t\t\t<p class="no
rmal">503\ \n\t\t\ \n\t\t\2010\ 
sp; \\ life  --> \\ life  --> \\ life  --> \\ life  --> \\ life 
d><a href="datasets/Opinosis+Opinion+%26frasl%3B+Review"><img src="assets/MLimages/S
mallLarge191.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/Opinosis+Opinion"
+%26frasl%3B+Review">Opinosis Opinion ⁄ Review</a></b>\/tr>\/tr>
class="normal">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".   -->\n\t\t<p class="
normal">Text \n\t\t\t \n\t\t\t&nbsp
;\n\t\t51 \n\t\t \n\t\t\t
td>2010 \n\t\t<!-- <td>Computer&nbsp; -->\n\t\
t\t\n\t\t<a href="datasets/Breast+Tissue"><img src="assets/MLimages/SmallLarge"><a href="datasets/Breast+Tissue"><img src="assets/MLimages/SmallLarge"></a>
default.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Breast+Tissue">Breast Ti
ssue</a></b>\n\t\t\t<!-- <td>Dataset with electrical impedance m
easurements of freshly excised tissue samples from the breast.  -->\n\t\t\t<p class="normal"
>Multivariate \n\t\t\etd>Classification \n\t\t\etd><p class="no
rmal">Real\ \n\t\t\t106\ \n\t\t\t10&
nbsp;   nt/t < d > p class = "normal" > 2010 & nbsp;   nt/t < !-- < td > p class = "normal" > Life & nbsp;  
/p> -->\n\t\t\t\n\t\t\t<a href="datasets/Cardiotocography"><
img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a h
ref="datasets/Cardiotocography">Cardiotocography</a></b>\n\t\t\t<!-- <td><p class=
"normal">The dataset consists of measurements of fetal heart rate (FHR) and uterine contraction (UC) features
on cardiotocograms classified by expert obstetricians.  -->\n\t\tMultivari
ate \    \\ lass="normal"> Classification \    \\ lass="normal"> Remaining the last of t
```

```
al \n\t\t2126 \n\t\t23 </p
>n\t\tp class="normal">2010 n\t\t<!-- <td>p class="normal">Life&nbsp;
c="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="d
atasets/Wall-Following+Robot+Navigation+Data">Wall-Following Robot Navigation Data</a></b>
\n\t\t\<!-- <td>The data were collected as the SCITOS G5 robot navigates thro
ugh the room following the wall in a clockwise direction, for 4 rounds, using 24 ultrasound sensors arranged circ
ularly around its \'waist\'.  -->\n\t\t\tMultivariate, Sequential </t
d>\n\t\t\Classification \n\t\t\tReal 
\ \n\t\t\t24 \n\t\t\t24 \n\t\t\t
2010 \n\t\t<!-- <td>Computer&nbsp; -->\n\t\t\t<</pre>
/tr>\n\t\t\t<a href="datasets/Spoken+Arabic+Digit"><img src="assets"
/MLimages/SmallLarge195.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Spok"><a href="datasets/Spok
en+Arabic+Digit">Spoken Arabic Digit</a></b>\n\t\t\t<!-- <td>This
dataset contains timeseries of mel-frequency cepstrum coefficients (MFCCs) corresponding to spoken Arabic di
gits. Includes data from 44 male and 44 female native Arabic speakers.  -->\n\t\t\ctd><p class=
"normal">Multivariate, Time-Series \n\t\tClassification \
n\t\t\tReal \n\t\t\t8800 \n\t\t\t
<p class="normal">13 </p></td>\n\t\t<p class="normal">2010 </p></td>\n\t\t\t<!-- <td><p class="normal">
s="normal">Other  -->\n\t\t\\n\t\t\t<a href="datasets/Localization+D">->\n\t\t\
ata+for+Person+Activity"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
<b><a href="datasets/Localization+Data+for+Person+Activity">Localization Data for Person
Activity</a></b>\n\t\t<!-- <td>class="normal">Data contains recordings of five pe
ople performing different activities. Each person wore four sensors (tags) while performing the same scenario fi
ve times.   -->\n\t\t\ctd>Univariate, Sequential, Time-Series \
n\t\tClassification \n\t\tReal \n\
t/tp class="normal">164860 \n\t\t\tp class="normal">8 \n\t\t\tp
class="normal">2010\ \n\t\t\<!-- <td>Life&nbsp; --- \n\t\t\
color="DDEEFF">\n\t\t\t<a href="datasets/AutoUniv"><img src="assets/MLimages/SmallLar"
gedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/AutoUniv">AutoUniv</
a></b>
classifications tasks. The aim is to reflect the nuances and heterogeneity of real data. Data can be generated i
n.csv, ARFF \ or \ C4.5 \ formats. \    --> \\ lt/t Multivariate \    
t\tClassification \n\t\tCategorical, Integer, Real
 \n\t\t <math>\n\t\t&nbsp;<math>\n\t
t<d>2010&nbsp;\n\t\t<!-- <td>Other&nbsp; -->\n\t\t
\t\n\t\t<a href="datasets/Steel+Plates+Faults"><img src="assets/MLimages/Small"
Largedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Steel+Plates+Fault"><b><a href="datasets/Steel+Plates+Fault"><b><a href="datasets/Steel+Plates+Fault"><b><a href="datasets/Steel+Plates+Fault"><b><a href="datasets/Steel+Plates+Fault"><b><a href="datasets/Steel+Plates+Fault"><b><a href="datasets/Steel+Plates+Fault"><b><a href="datasets/Steel+Plates+Fault"><b><a href="datasets/Steel+Plates+Fault"><b><a href="datasets/Steel+Plates+Fault"><b href="dat
s">Steel Plates Faults</a></b>\n\t\t<!-- <td>A dataset of steel pl
ates' faults, classified into 7 different types. \r\nThe goal was to train machine learning for automatic pattern rec
ormal">Classification \n\t\t\tInteger, Real \n\t\t\t<p cl
ass="normal">1941\ \n\t\t\t27\ \n\t\t\t27\ 
I">2010\ \n\t\t<!-- <td>Physical&nbsp;--->\n\t\t\t
DEEFF">\n\t\t\ctd><a href="datasets/MiniBooNE+particle+identification"><img src="assets/MLi
mages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/MiniBo"
oNE+particle+identification">MiniBooNE particle identification</a></b>
This dataset is taken from the MiniBooNE experiment and is used to distinguish electron ne
utrinos (signal) from muon neutrinos (background).  -->\n\t\tMultivariat
e \n\t\tClassification \n\t\t\tReal
\alpha = \frac{\sqrt{y}}{\sqrt{t}\sqrt{t}\sqrt{t}}
p>\n\t\t2010 \n\t\t<!-- <td>Physical&nbsp;</p
>-->\n\t\t\d><a href="datasets/YearPredictionMSD"><img src="assets/MLi">
mages/SmallLarge203.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/YearPred"
ictionMSD">YearPredictionMSD</a></b>\n\t\t\t<!-- <td>class="normal">Prediction
of the release year of a song from audio features. Songs are mostly western, commercial tracks ranging from 1
922 to 2011, with a peak in the year 2000s.  -->\n\t\t\tMultivariate 
\n\t\tRegression \n\t\tReal </p
>\n\t\t515345 \n\t\t90 \n\t
\t\n\t\t\t<a href="datasets/PEMS-SF"><img src="assets/MLimag"><a href="datasets/PEMS-SF"><img src="assets/MLimag"><a href="datasets/PEMS-SF"><img src="assets/MLimag"><a href="datasets/PEMS-SF"><img src="assets/MLimag"><a href="datasets/PEMS-SF"><img src="assets/MLimag"><a href="datasets/PEMS-SF"><a href="da
es/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/PEMS-SF"
```

```
>PEMS-SF</a></b></tp>\n\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 Franci
sco bay area freeways across time.  -->\n\t\t\tMultivariate, Time-Series&
nbsp;\n\t\t\tReal&n
bsp; /n/t/t440  /n/t/t/t138672  
>\n\t\t2011 \n\t\t<!-- <td>Computer&nbsp;</p
>-->\n\t\t\n\t\t<a href="datasets/OpinRank+Review+Dataset"><img src="as"
sets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="dataset"
s/OpinRank+Review+Dataset">OpinRank Review Dataset</a></b>\n\t\t<!-- <td><p
class="normal">This data set contains user reviews of cars and and hotels collected from Tripadvisor (~259,00
0 \r\nreviews) and Edmunds (~42,230 reviews).   -->\n\t\tText 
p\ class="normal">\ \n\t\t\t<p\ class="normal">&nbsp;\n\t\t\t<p\ class="normal">2
011 \n\t\t<!-- <td>class="normal">Computer&nbsp; -->\n\t\t\t
EFF">\n\t\t<a href="datasets/Relative+location+of+CT+slices+on+axial+axis"><img src="a
ssets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datas"
ets/Relative+location+of+CT+slices+on+axial+axis">Relative location of CT slices on axial axis</a></b>
>\n\t\t\<!-- <td>The dataset consists of 384 features extracted from CT im
ages. The class variable is numeric and denotes the relative location of the CT slice on the axial axis of the hu
man body.  -->\n\t\t\tDomain-Theory \n\t\t\t<p class
="normal">Regression \n\t\t\tReal \n\t\t\t<p class="n
ormal">53500\ \n\t\t\t386\ \n\t\t\t2
011 \n\t\t<!-- <td>Computer&nbsp; -->\n\t\t
ble><a href="datasets/Online+Handwritten+Assamese+Characters+Dataset"><img src="assets/MLima"
ges/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/Online+H"><br/>
andwritten+Assamese+Characters+Dataset">Online Handwritten Assamese Characters Dataset</a></b>
td>\n\t\t\<!-- <td>This is a dataset of 8235 online handwritten assamese c
haracters. The "online" process involves capturing of data as text is written on a digitizing tablet with an electron
ic pen.  -->\n\t\tMultivariate, Sequential \n\t\t\t<p cl
ass="normal">Classification \n\t\t\Integer \n\t\t\t<p c
lass="normal">8235\ \n\t\t\t \n\t\t\t
>2011 \n\t\t<!-- <td>Computer&nbsp; -->\n\t\t\t
DEEFF">\n\t\t\t<a href="datasets/PubChem+Bioassay+Data"><img src="assets/MLimages/
SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/PubChem+Bi"><b><a href="datasets/PubChem+Bi"><b><a href="datasets/PubChem+Bi"><b><a href="datasets/PubChem+Bi"><b><a href="datasets/PubChem+Bi"><b><a href="datasets/PubChem+Bi"><b><a href="datasets/PubChem+Bi"><b><a href="datasets/PubChem+Bi"><b><a href="datasets/PubChem+Bi"><b><a href="datasets/PubChem+Bi"><b href="datasets/PubChem+Bi
oassay+Data">PubChem Bioassay Data</a></b>\n\t\t<!-- <td>T
hese highly imbalanced bioassay datasets are from the differing types of screening that can be performed usin
g HTS technology. 21 datasets were created from 12 bioassays.  -->\n\t\t\t<p class="normal
">Multivariate \n\t\tClassification \n\t\t<p class="n
ormal">Integer, Real \n\t\t\t \n\t\t\t<p class="normal"
p; -->\n\t\t\t<tt>\n\t\t\t<a href="datasets/Record+Linkage+Comparison+Patter">+Patter</a>
ns"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b
><a href="datasets/Record+Linkage+Comparison+Patterns">Record Linkage Comparison Patterns</a></b>
>\n\t\t<!-- <td>Element-wise comparison of records with personal da
ta from a record linkage setting. The task is to decide from a comparison pattern whether the underlying record
s belong to one person.  -->\n\t\t\tp class="normal">Multivariate \n\t\t\t
Classification \n\t\t\tReal \n\t\t\t<</p>
p class="normal">5749132 \n\t\t12 \n\t\t\t<p class=
"normal">2011\ \n\t\t\<!-- <td>Other\&nbsp; --> \n\t\t\
="DDEEFF">\n\t\t\ctd><a href="datasets/Communities+and+Crime+Unnormalized"><img src="
assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="data"
sets/Communities+and+Crime+Unnormalized">Communities and Crime Unnormalized</a></b>
table>\n\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  -->\n\t\t\tMultivariate 
\n\t\tRegression \n\t\tReal 
>class="normal">2011 \n\t\t\t<!-- <td>Social&nbsp; -->\n\t\t\t
>\n\t\t<a href="datasets/Vertebral+Column"><img src="assets/MLimages/SmallLarged"><a href="datasets/Vertebral+Column"><img src="assets/MLimages/SmallLarged"><a href="datasets/Vertebral+Column"><a href="datasets/Vertebral+Column"><a
efault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Vertebral+Column">Verteb
ral Column</a></b>\n\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).  -->\n\t\tMultivariate&nbsp
;\n\t\tClassification \n\t\t\tReal <
/p>\n\t\t310 \n\t\t\t6 \n\t\t\t
r bgcolor="DDEEFF">\n\t\t\t<a href="datasets/EMG+Physical+Action+Data+Set"><img src="
assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="data"
sets/EMG+Physical+Action+Data+Set">EMG Physical Action Data Set</a></b>
<!-- <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.  -->\n\t\tTime-Series \n\t\t\t<p class="n
ormal">Classification \n\t\t\tReal \n\t\t\tReal 
mal">10000\ \n\t\t8 \n\t\t2011&
nbsp;   nt/t/t <!-- <td>p class="normal">Physical&nbsp;   --> \n/t/t   \n/t/t   \n/t/t   \n/t/t   \n/t/t  
><a href="datasets/Vicon+Physical+Action+Data+Set"><img src="assets/MLimages/SmallLargedefault.jpg"
border=1 /></a>&nbsp;<b><a href="datasets/Vicon+Physical+Action+Data+Set">Vi
con Physical Action Data Set</a></b>\n\t\t\t<!-- <td>con Class="normal">The Physical
Action Data Set includes 10 normal and 10 aggressive physical actions that measure the human activity. The d
ata have been collected by 10 subjects using the Vicon 3D tracker.  -->\n\t\t\t<p class="nor
mal">Time-Series \n\t\tClassification \n\t\t\t<p clas
s="normal">Real\ \n\t\t3000\ \n\t\t\t4>\n\t\t\t5000 
">27\ \n\t\t2011\ \n\t\t\t<!-- <td>Physi
cal  -->\n\t\t\\n\t\t\t<a href="datasets/Amazon+"
Commerce+reviews+set"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
<b><a href="datasets/Amazon+Commerce+reviews+set">Amazon Commerce reviews set</
a></b>\n\t\t\t<!-- <td>The dataset is used for authorship identifica
tion in online Writeprint which is a new research field of pattern recognition.   -->\n\t\t\t<p cl
ass="normal">Multivariate, Text, Domain-Theory \n\t\tClassification&n
bsp;\n\t\t\tReal \n\t\t1500 
td>n\tp class="normal">10000 \n\t\tp class="normal">2011 \n\t\t
\t<!-- <td>Physical&nbsp; -->\n\t\t\r\t\t<a href="data"
sets/Amazon+Access+Samples"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;</t
d><b><a href="datasets/Amazon+Access+Samples">Amazon Access Samples</a></b>
\n\t\t\<!-- <td>Amazon\'s InfoSec is getting smarter about the wa
y Access data is leveraged. This is an anonymized sample of access provisioned within the company.  
>-->\n\t\t\tTime-Series, Domain-Theory \n\t\t\t<p class="normal"
al">Regression, Clustering, Causal-Discovery \n\t\t \n\t\
t/tp class="normal">30000 \n/t/tp class="normal">20000 \n/t/t\tp class="normal">20000 
2011 \n\t\t\t<!-- <td>Business&nbsp; -->\n\t\t\t
r>\n\t\t<a href="datasets/Reuter_50_50"><img src="assets/MLima"
ges/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Reuter_5"
0_50">Reuter_50_50</a></b>\n\t\t<!-- <td>class="normal">The dataset is used
for authorship identification in online Writeprint which is a new research field of pattern recognition.  
-->\n\t\t\tMultivariate, Text, Domain-Theory \n\t\t\t<p class="no
rmal">Classification, Clustering \n\t\t\tReal \n\t\t\t<p
class="normal">2500\ \n\t\t\t10000\ \n\t\t\t10000\ 
ormal">2011 \n\t\t<!-- <td>Computer&nbsp; -->\n\t\t\t
\t<a href="datasets/Farm+Ads"><img src="assets/MLimages/SmallLargedefault.jpg" border
=1 /></a>&nbsp;<b><a href="datasets/Farm+Ads">Farm Ads</a></b>
>\n\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 ow
ner approves of the ad.  -->\n\t\tText \n\t\t\t<p clas
s="normal">Classification \n\t\t\ \n\t\t\
mal">4143\ \n\t\t\54877\ \n\t\t\t20
11 \n\t\t\t<!-- <td>Business&nbsp; -->\n\t\t\t bgcolor="DDEE"
FF">\n\t\t\t<a href="datasets/DBWorld+e-mails"><img src="assets/MLimages/SmallLargede"
fault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails">DBWorld+e-mails</a>
d e-mails</a></b>\n\t\t\t<!-- <td>It contains 64 e-mails which I ha
ve manually collected from DBWorld mailing list. They are classified in: \'announces of conferences\' and \'every
thing else\'.  -->\n\t\t\tText \n\t\t\t
Classification \n\t\t\ctd> \n\t\t\ctd>64 
sp;/n/t/t4702 /n/t/t/t2011 
td>\n\t\t<!-- <td>Computer&nbsp: -->\n\t\t\
```

```
ref="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29"><img src="assets/MLimages/SmallLarge">
default.jpg" border=1 /></a>&nbsp;<b><a href="datasets/KEGG+Metabolic+Relatio"
n+Network+%28Directed%29">KEGG Metabolic Relation Network (Directed)</a></b>
>\n\t\t<!-- <td>KEGG Metabolic pathways modeled as directed relation network. Variety of
graphical features presented. anbsp; -->\n\t\t\tMultivariate, Univariate, Text&nb
sp;\n\t\t\tClassification, Regression, Clustering \n\t\t\t<p cl
ass="normal">Integer, Real \n\t\t53414 \n\t\t<p cl
ass="normal">24\ \n\t\t2011\ \n\t\t<!-- <td>2011&nbsp;
rmal">Life  -->\n\t\t bgcolor="DDEEFF">\n\t\t<a href="datasets/KE"></a>
GG+Metabolic+Reaction+Network+%28Undirected%29"><img src="assets/MLimages/SmallLargedefault.jpg" b
order=1 /></a>&nbsp;<b><a href="datasets/KEGG+Metabolic+Reaction+Network+"
%28Undirected%29">KEGG Metabolic Reaction Network (Undirected)</a></b>
t<!-- <td>KEGG Metabolic pathways modeled as un-directed reaction network. Variety of gra
phical features presented.  -->\n\t\t\tMultivariate, Univariate, Text 
\n\t\tClassification, Regression, Clustering \n\t\t\t<p class
="normal">Integer, Real \n\t\t65554 \n\t\t<p class
="normal">29 \n\t\t2011 <math>\n\t\t<!-- <td>2011&nbsp;<math>
al">Life  -->\n\t\t\r\t\t<a href="datasets/Bank+Marketing"><img s
rc="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="d
atasets/Bank+Marketing">Bank Marketing</a></b>\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).  -->\n\t\t\t<p
class="normal">Multivariate \n\t\tClassification \n\t\t\t<t
d><p\ class="normal">Real\&nbsp;\n\t\t\t<p\ class="normal">45211\&nbsp;\n\t\t\t<p\ class="normal">45211\&nbsp;
ss="normal">17 \n\t\t\t2012 \n\t\t\t<!-- <td>17&nbsp;
mal">Business  -->\n\t\t\t\n\t\t\t<a href="dataset">tr><a 
s/YouTube+Comedy+Slam+Preference+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 />
</a>&nbsp;<b><a href="datasets/YouTube+Comedy+Slam+Preference+Data">Yo
uTube Comedy Slam Preference Data</a></b>\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.  -->\n\t\t\t<t
d>Text \n\t\t\tClassification \n\t\t\t<td
>class="normal"> \n\t\t\t1138562 \n\t\t\t1138562 
"normal">3 \n\t\t\t2012 \n\t\t\t<!-- <td><p class="normal"
>Computer  -->\n\t\t\t\n\t\t\t<a href="datasets/Gas+Sensor+Array+">Computer&nbsp;
Drift+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="n
ormal"><b><a href="datasets/Gas+Sensor+Array+Drift+Dataset">Gas Sensor Array Drift Dataset</a></b>
/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. -->\n\t\tMultivariate \n\t\t\t<p clas
s="normal">Classification \n\t\t\tReal \n\t\t\t<p class
= "normal" > 13910 \& nbsp;   (nt/t t  p class = "normal" > 128 \& nbsp;   (nt/t t  p class = "normal" > 128 \& nbsp;   (nt/t t  p class = "normal" > 128 \& nbsp; 
">2012 \n\t\t<!-- <td>Computer&nbsp; -->\n\t\t<tr bgcolor="
DDEEFF">\n\t\t\ctd><a href="datasets/ILPD+%28Indian+Liver+Patient+Dataset%29"><img src
="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="dat
asets/ILPD+%28Indian+Liver+Patient+Dataset%29">ILPD (Indian Liver Patient Dataset)</a></b>
\n\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.  -->\n\t\t\t<t
d>Multivariate \n\t\tClassification \n\
t/t/Integer, Real \n/t/t/583 \n/t/t/
tp class="normal">10 \n\t\tp class="normal">2012 \n\t\t\t<!-- <td>p class="normal">2012&nbsp;\n\t\t\t<!-- <td>p class="normal">2012&nbsp;
class="normal">Life  -->\n\t\t
TY+Activity+Recognition"><img src="assets/MLimages/SmallLarge226.jpg" border=1 /></a>&nbsp;<p
class="normal"><b><a href="datasets/OPPORTUNITY+Activity+Recognition">OPPORTUNITY Activity Recognition">OPPORTUNITY Activity Recognition
on</a></b>
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).  -->\n\t\tMultivariate, Time-Series \n\t\t\t<p clas
s="normal">Classification \n\t\t\tReal \n\t\t\t<p class
="normal">2551\ \n\t\t\t242\ \n\t\t\t242 
>2012\ \n\t\t<!--<td>Computer\&nbsp; -->\n\t\t\t
DEFEE">\n\t\tztd>>tr><a href="datasets/Nomao"><img src="assets/MI images/Smalll argedefault i
```

```
pg" border=1 /></a>&nbsp;<b><a href="datasets/Nomao">Nomao</a></b>
d>\n\t\t\t<!-- <td>class="normal">Nomao collects data about places (name, phone, localiza
tion...) from many sources.\r\nDeduplication consists in detecting what data refer to the same place.\r\nInstance
s in the dataset compare 2 spots.  -->\n\t\t\tUnivariate \n
\t\t\tClassification \n\t\tReal \n\t\
t\t34465\nbsp;\n\t\t\t120\nbsp;\n\t\t\t120\nbsp;
class="normal">2012\ \n\t\t<!-- <td>Computer\&nbsp; -->\n\t\t\t
\n\t\t\t<a href="datasets/SMS+Spam+Collection"><img src="assets/MLimages/SmallLar">
gedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/SMS+Spam+Collection"
">SMS Spam Collection</a></b>\n\t\t<!-- <td>The SMS Spam C
ollection is a public set of SMS labeled messages that have been collected for mobile phone spam research.&n
bsp; -->\n\t\t\tMultivariate, Text, Domain-Theory \n\t\t\t<p
class="normal">Classification, Clustering \n\t\tReal \n\t\
t\t5574\n\t \n\t\t\t \n\t\t\t 
="normal">2012\ \n\t\t<!--< td>Computer&nbsp; --> \n\t\t\
gcolor="DDEEFF">\n\t\t\ctd><a href="datasets/Skin+Segmentation"><img src="assets/MLimag"
es/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/Skin+Seg"
mentation">Skin Segmentation</a></b>\n\t\t\t<!-- <td>The Skin S
egmentation 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.  -->\n\t\t<p class="
normal">Univariate \n\t\t\tClassification \n\t\t\t<p cla
ss="normal">Real\ \n\t\t\245057 \n\t\t\245057 
mal">4\ \n\t\t\t2012\ \n\t\t\t<!-- <td>Continue (normal) < normal </p>
mputer  -->\n\t\t\n\t\t<a href="datasets/Planning+Relax"><img sr
c="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;c class="normal"><b><a href="d
atasets/Planning+Relax">Planning Relax</a></b>\n\t\t\t<!-- <td>
The dataset concerns with the classification of two mental stages from recorded EEG signals: Planning (during i
magination of motor act) and Relax state.   -->\n\t\t\tUnivariate </
p>\n\t\t\tClassification \n\t\t\tReal </p
>\n\t\t182 \n\t\t\t13 \n\t\t\t<
td>2012 \n\t\t<!-- <td>Computer&nbsp; -->\n\t\
t\t\n\t\t<a href="datasets/PAMAP2+Physical+Activity+Monitori"
ng"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b
><a href="datasets/PAMAP2+Physical+Activity+Monitoring">PAMAP2 Physical Activity Monitoring</a></b>
/td>\n\t\t<!-- <td>The PAMAP2 Physical Activity Monitoring dataset contains
ns data of 18 different physical activities, performed by 9 subjects wearing 3 inertial measurement units and a h
eart rate monitor.  -->\n\t\t\tMultivariate, Time-Series \n\t
\t\tClassification \n\t\tReal \n\t\t\
t3850505 \n\t\t\t52 \n\t\t\t52 
class="normal">2012 \n\t\t\<!-- <td>Computer&nbsp; -->\n\t\t\t
\n\t\ttr>\n\t\tscing src="assets/MLimage">scing scing sci
s/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Restaurant+
%26+consumer+data">Restaurant & consumer data</a></b>\n\t\t\t<!-- <td><p class
="normal">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.   -->\n\t\tMul
tivariate    \\ lass = "normal" >    \\ lass = "normal" 
/n\t/t<<d>138&nbsp;/n/t/t<<td>47&nbsp;/n/t/t/t<<tr>
d>2012 \n\t\t\t<!-- <td>Computer&nbsp; -->\n\t\t
\t\n\t\t\t<a href="datasets/CNAE-9"><img src="assets/MLimage"
s/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/CNAE-9">C
documents of free text business descriptions of Brazilian companies categorized into a\r\nsubset of 9 categorie
s  -->\n\t\t\tMultivariate, Text \n\t\t\t<p class="normal"
">Classification \n\t\tInteger \n\t\t<p class="normal"
l">1080\ \n\t\t857\ \n\t\t2012\&n class="normal">2012&n class
bsp;   + t/t < --> 
><a href="datasets/Individual+household+electric+power+consumption"><img src="assets/MLimages/Small"></a>
Largedefault.jpg" border=1 /></a>&nbsp;<pcclass="normal"><b><a href="datasets/Individual+househo"><br/>the datasets/Individual+househo
ld+electric+power+consumption">Individual household electric power consumption</a></b>
e>\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 2 phon; //n> //td> >\n\t\t\t +td> -n class="normal"> Multivariate. Time Series 2 phon; //n> //td>
```

```
values are available.aribsp,~/p>~/ta>
                                           \n\t\tRegression, Clustering \n\t\tReal </
p>/n\t\t2075259\ /n\t\t9\ /n
t/tp class="normal">2012 \n\t\t<!-- <td>p class="normal">Physical&nbsp; -->\
n/t/t/t\n/t/t/<a href="datasets/seeds"><img src="assets/MLimag"
es/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/seeds">se
eds</a></b>li>-- p class="normal">Measurements of geometrical properti
es 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.   -->\n\t\t\tclass="normal">Multivariate
 \n\t\tClassification, Clustering \n\t\tClassification, Clustering 
mal">Real\ \n\t\t210\ \n\t\t\t7 
sp:\n\t\t2012&nbsp:\n\t\t<!-- <td>Life&nbsp:
> -->\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t</
|Largedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Northix">Northix">Northix
a></b>
g benchmark problem for data integration of two entity relationship databases.   -->\n\t\t\t<p
class="normal">Multivariate, Univariate, Text \n\t\tclass="normal">Classification 
\n\t\t115 </
p>/n/t/t200 /n/t/t/t2012 /n/t/t/t2012 /n/t/t/t2012 /n/t/t/t2012 /n/t/t/t2012 /n/t/t/t2012 /n/t/t/t2012 /n/t/t2012 /n/t
\t\t<!-- <td>Computer&nbsp; -->\n\t\t\t\n\t\t\t<table
><a href="datasets/QtyT40I10D100K"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 />
</a>&nbsp;<b><a href="datasets/QtyT40I10D100K">QtyT40I10D100K</a></b></p
>\n\t\t<!-- <td>Since there is no numerical sequential data stream av
ailable in standard data sets, this data set is generated from the original T40I10D100K data set 
-->\n\t\t\ctd>Sequential \n\t\t \n\t\t\
Integer \n\t\t3960456 \n\t\t<td
> class="normal">4 \n\t\t\t2012 \n\t\t\t<!-- <td>2012&nbsp;
s="normal">  -->\n\t\t\n\t\t<a href="datasets/Legal+Case+Report">+Report</a>
s"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b>
<a href="datasets/Legal+Case+Reports">Legal Case Reports</a></b>\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. 
-->\n\t\t\tText \n\t\t\tClassification&nbs
p;/nt/t /n/t/t /n/t/t/t /n/t/t/t /n/t/t/t /n/t/t/t /n/t/t/t /n/t/t/t /n/t/t/t /n/t/t/t /n/t/t/t /n/t/t /n/t/t /n/t/t /n/t/t /n/t/t /n/t /n/t /n/t /n/t /n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t<
d><p\ class="normal">&nbsp;\n\t\t\t<p\ class="normal">2012&nbsp;\n\t\t\t<!--<td><p\ class="normal">2012&nbsp;
s="normal">Other  -->\n\t\t\t\n\t\t\t<a href="data"
sets/Human+Activity+Recognition+Using+Smartphones"><img src="assets/MLimages/SmallLargedefault.jpg" b
order=1 /></a>&nbsp;<b><a href="datasets/Human+Activity+Recognition+Using+S"
martphones">Human Activity Recognition Using Smartphones</a></b>\n\t\t\<!-- <td
>Human Activity Recognition database built from the recordings of 30 subjects performing a
ctivities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors. & nbsp;
 -->\n\t\tMultivariate, Time-Series \n\t\t<p class="normal"
">Classification, Clustering \n\t\t\t \n\t\t\t<p class="n
ormal">10299\ \n\t\t\t561\ \n\t\t\t2
012 \n\t\t<!-- <td>Computer&nbsp; -->\n\t\t
ble><a href="datasets/One-hundred+plant+species+leaves+data+set"><img src="assets/MLimages/Sm
allLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/One-hundred+pl"><b><a href="datasets/One-hundred+pl"><b hre
ant+species+leaves+data+set">One-hundred plant species leaves data set</a></b>
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.  -->\n\t\t\t<p class="
normal"> \n\t\t\tClassification \n\t\t\t<p class="normal"
I">Real\ \n\t\t\t1600\ \n\t\t\t64\ 
p</b>n\t\t<<d>p class="normal">2012&nbsp;n\t\t\<!-- <td>p class="normal">Life&nbsp;
-->\n\t\t\n\t\t\t<a href="datasets/Energy+efficiency"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="
datasets/Energy+efficiency">Energy efficiency</a></b></tp>\n\t\t<!-- <td><p class="nor
mal">This study looked into assessing the heating load and cooling load requirements of buildings (that is, ener
gy efficiency) as a function of building parameters.  -->\n\t\t\tMultivariate
 \n\t\tClassification, Regression \n\t\t\t<p class="no
rmal">Integer, Real \n\t\t\t768 \n\t\t\t<p class="normal"
al">8\nbsp;\n\t\t\t2012\nbsp;\n\t\t\t<!-- <td>Com
puter  -->\n\t\t\\n\t\t\<a href="datasets/Yacht+Hydrodynamics"><i
```

```
ef="datasets/Yacht+Hydrodynamics">Yacht Hydrodynamics</a></b>\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.  -->\n\t\t\tMultivariate \n\t\t\t<p class=
"normal">Regression \n\t\tReal \n\t\t<p class="no
rmal">308\ \n\t\t7\ \n\t\t\t2013&n the class="normal">13 
\t\t\ctd><a href="datasets/Fertility"><img src="assets/MLimages/SmallLargedefault.jpg" border
=1 /></a>&nbsp;<b><a href="datasets/Fertility">Fertility</a></b>
ble>\n\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  -->\n\t\t\tMultivariate \n\t\t\t<p
class="normal">Classification, Regression \n\t\t\tReal \n\
t/t/ctd><p\ class="normal">100&nbsp;<p\ class="normal">10&nbsp;<p\ class="normal">10&nbsp;
ass="normal">2013\ \n\t\t<!--<td>Life&nbsp; -->\n\t\t\n\t\
t\t<a href="datasets/Daphnet+Freezing+of+Gait"><img src="assets/MLimages/SmallLarged"><a href="datasets/Daphnet+Freezing+of+Gait"><a href="datasets/Daphnet-Freezing+of+Gait"><a href="
efault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/Daphnet+Freezing+of+Gai
t">Daphnet Freezing of Gait</a></b>\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  -->\n\t\t\t
Multivariate, Time-Series \n\t\t\tClassification \n\t\t\t
Real \n\t\t237 \n\t\t<p class="
normal">9\ \n\t\t\t2013\ \n\t\t\t<!-- <td>
Life  -->\n\t\t\\n\t\t\t<a href="datasets/3D+Road">tr><a href="datasets/3D+Road
+Network+%28North+Jutland%2C+Denmark%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=
1 /></a>&nbsp;<b><a href="datasets/3D+Road+Network+%28North+Jutland%2C+"
Denmark%29">3D Road Network (North Jutland, Denmark)</a></b>\n\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.  -->\n\t\t\tSequen
tial, Text \n\t\tRegression, Clustering \n\t\t\t<p clas
mal">4\ \n\t\t\t2013\ \n\t\t\t<!-- <td>Continue (normal) < normal </p>
mputer  -->\n\t\t\n\t\t</d><ahref="datasets/ISTANBUL+STOCK+EXC">mputer&nbsp;
HANGE"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="norma"
I"><b><a href="datasets/ISTANBUL+STOCK+EXCHANGE">ISTANBUL STOCK EXCHANGE</a></b></
/tr>\n\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.  -->\n\t\tclass="normal">Multivariate, Univariate, Time-Series 
/td>\n\t\t\tClassification, Regression \n\t\t\tReal&
nbsp;   nt/t  p class = "normal" > 536 & nbsp;   (nt/t)  p class = "normal" > 8 & nbsp;   (nt/t)  p class = "normal" > 8 & nbsp;  
\ \n\t\t\t2013 \n\t\t\t<!-- <td>Business&nbsp;
-->\n\t\t\ bgcolor="DDEEFF">\n\t\t\<a href="datasets/Buzz+in+social+media+"><im
g src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href
="datasets/Buzz+in+social+media+">Buzz in social media </a></b>\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\tTime-Series, Multivariate \n\t\tReg
ression, Classification \n\t\t\tInteger, Real \n\t\t\t<p c
lass="normal">140000\ 
$$ \n\t\t$77\ 
$$ \n\t\t$77\ 
mal">2013\ \n\t\t<!-- <td>Computer\&nbsp; --> \n\t\t\
efault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/First-order+theorem+provi
ng">First-order theorem proving</a></b>\n\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.  -->\n\t\t\t<p class="no
rmal">Multivariate \n\t\tClassification \n\t\t\t<p clas
s="normal">Real \n\t\t\t6118 \n\t\t\t<p class="normal"
">51 \n\t\t2013 \n\t\t<!-- <td>Comp
uter  -->\n\t\t\c/tr>\n\t\t\ctd><a href="datasets/Wearable">tr><a h
e+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29"><img src="assets/M
Limages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><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></b></fd>
A dataset with 5 classes (sitting-down, standing-up, standing, walking, and sitting) collec
ted on 8 hours of activities of 4 healthy subjects. We also established a baseline performance index. 
-->\n\t\t\tSequential \n\t\t\tClassification&nb
sp;\n\t\tInteger, Real \n\t\t165632&n
bsp; /n/t/t18  /n/t/t/t2013  /n/t/t2013  /n/t/t2013  /n/t/t2013  /n/t2013  /n/t2013  /n/t/n/t2013  /n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t/n/t
d>/n/t/t<!--< td>Computer&nbsp; -->/n/t/t/t/n/t/t/t<a hr
ef="datasets/Gas+sensor+arrays+in+open+sampling+settings"><img src="assets/MLimages/SmallLargedefault."
jpg" border=1 /></a>&nbsp;<b><a href="datasets/Gas+sensor+arrays+in+open+sa"
mpling+settings">Gas sensor arrays in open sampling settings</a></b>
>The dataset contains 18000 time-series recordings from a chemical detection platform at si
x different locations in a wind tunnel facility in response to ten high-priority chemical gaseous substances 
 -->\n\t\tMultivariate, Time-Series \n\t\t<p class="normal"
">Classification \n\t\tReal \n\t\t
18000 \n\t\t1950000 \n\t\t\t201
3 \n\t\t<!-- <td>Computer&nbsp; -->\n\t\t\
FF">\n\t\t\ctd><a href="datasets/Climate+Model+Simulation+Crashes"><img src="assets/MLim
ages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/Climate"
+Model+Simulation+Crashes">Climate Model Simulation Crashes</a></b>\r\t\t\t<!--
Given Latin hypercube samples of 18 climate model input parameter values, predict cli
mate model simulation crashes and determine the parameter value combinations that cause the failures.  
 -->\n\t\t\Multivariate \n\t\t\Classificati
on\ \n\t\tFeal\ \n\t\t\t540 
p  n t < d < p class = "normal" > 18 & nbsp;  n t  2013 & nbsp;   n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t
\t<!-- <td>Physical&nbsp; -->\n\t\t\n\t\t\t<a href="data"
sets/MicroMass"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class
="normal"><b><a href="datasets/MicroMass">MicroMass</a></b>\n\t\t<!-- <td><p
class="normal">A dataset to explore machine learning approaches for the identification of microorganisms from
mass-spectrometry data.  -->\n\t\t\t>p class="normal">Multivariate \n\t\t\t<td
>Classification \n\t\t\tReal \n\t\t\t
931\ \n\t\t\t1300\ \n\t\t\t1300 
normal">2013\ \n\t\t\<!-- <td>Life&nbsp;\-->\n\t\t\
DDEEFF">\n\t\t\ctd><a href="datasets/QSAR+biodegradation"><img src="assets/MLimages/S
mallLargedefault.jpg" border=1 /></a>&nbsp;<pcclass="normal"><b><a href="datasets/QSAR+biodegr"
adation">QSAR biodegradation</a></b>\n\t\t\t<!-- <td>Data set c
ontaining values for 41 attributes (molecular descriptors) used to classify 1055 chemicals into 2 classes (ready
and not ready biodegradable). -->\n\t\tMultivariate \n\t\
t\tClassification \n\t\t\tInteger, Real </t
>2013 \n\t\t\<!-- <td>Other&nbsp; -->\n\t\t\
\n\t\t\t<a href="datasets/BLOGGER"><img src="assets/MLimages/SmallLargedefault.jp">
g" border=1 /></a>&nbsp;<b><a href="datasets/BLOGGER">BLOGGER</a></b></
p>\n\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  -->\n\t\t\t<p
class="normal">Multivariate \n\t\t\tClassification \n\t\t\t<t
d> \n\t\t\t100 \n\t\t\t<p class="no
rmal">6\ \n\t\t2013\ \n\t\t\t<!-- <td>Continue (normal) < normal </p>
mputer  -->\n\t\t\ dolor="DDEEFF">\n\t\t\<a href="datasets/Daily+"> dolor="DDEEFF">\n\t\t\<a href="datasets/Daily+"> dolor="DDEEFF">\n\t\t\<a href="datasets/Daily+"> dolor="DDEEFF">\n\t\t\<a href="datasets/Daily+"> dolor="DDEEFF">\n\t\t\t</a>
and+Sports+Activities"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p
class="normal"><b><a href="datasets/Daily+and+Sports+Activities">Daily and Sports Activities</a></b>
>\n\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 o
n the torso, arms, and legs.\r\n  -->\n\t\t\tMultivariate, Time-Series&nbsp
;\n\t\tClassification, Clustering \n\t\tR
eal\ \n\t\t\9120 \n\t\t\t5625&nbsp
\n\t\t2013 \n\t\t<!-- <td>Computer&nbsp;
 -->\n\t\t\\r\t\t\<a href="datasets/User+Knowledge+Modeling"><img src=
"assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="data"
sets/User+Knowledge+Modeling">User Knowledge Modeling</a></b>\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.  -->\n\t\t\t
Multivariate \n\t\tClassification, Clustering \n\t\t<p
```

```
class="normal">Integer \n\t\t\t403 \n\t\t\t<p class="n
ormal">5\ \n\t\t2013\ \n\t\t\t<!-- <td>C
omputer  -->\n\t\t\\n\t\t\<a href="datasets/Reut"><a href="datasets/Reut"><ta
ers+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection"><img src="assets/MLimag"
es/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Reuters+R
CV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection">Reuters RCV1 RCV2 Multilingua
I, Multiview Text Categorization Test collection</a></b>\r\t\t<!-- <td><p class="nor
mal">This test collection contains feature characteristics of documents originally written in five different languag
es and their translations, over a common set of 6 categories.   -->\n\t\t\t
Multivariate \n\t\t\tClassification \n\t\t\tClassification 
mal">Real\ \n\t\t\t111740\ \n\t\t\t&
nbsp;\n\t\t\t2013 \n\t\t\t<!-- <td>Life&nbsp;
/p> -->\n\t\t\\n\t\t\t<dd><a href="datasets/NYSK"><img src="assets/MLimages/Sma"
on of English news articles about the case relating to allegations of sexual assault against the former IMF direct
or Dominique Strauss-Kahn (May 2011).  -->\n\t\t\tMultivariate, Sequenti
al, Text \n\t\tClustering \n\t\t&
nbsp;   nbt < class="normal" > 10421 & nbsp;   nbt < class="normal" > 7 & nbsp;  
td > \ln t < - td > \ln t < - td > - td
->\n\t\t\ bgcolor="DDEEFF">\n\t\t\<a href="datasets/Turkiye+Student+Evaluation">
<img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a
href="datasets/Turkiye+Student+Evaluation">Turkiye Student Evaluation</a></b></fo>
\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.&
nbsp; -->\n\t\t\tMultivariate \n\t\t\tClassi
fication, Clustering \n\t\t \n\t\t5
820\   \\ lass="normal">33\   \\ lass="normal">2013\   \\ lass="normal">2013\   \\ lass="normal">2013\   \\ lass="normal">2013   \\ lass="n
/p>/n/t/t<!--<td>Other&nbsp;-->/n/t/t/
href="datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Mach
ines%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="norm"
al"><b><a href="datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Elect
rical+Machines%29">ser Knowledge Modeling Data (Students\' Knowledge Levels on DC Electrical Machines)</
a></b>\n\t\t\<!-- <td>The dataset is about the users\' learning ac
tivities and knowledge levels on subjects of DC Electrical Machines. The dataset had been obtained from onlin
e web-courses and reported in my Ph.D. Thesis.  -->\n\t\t\tMultivariate&
nbsp;\n\t\t\Real&n
bsp;/n/t/t403 /n/t/t/t5 /
n\t\tp class="normal">2013 

-->\n\t\t\ bgcolor="DDEEFF">\n\t\t\<a href="datasets/EEG+Eye+State"><img src="a
ssets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datas"
ets/EEG+Eye+State">EEG Eye State</a></b>\n\t\t\t<!-- <td>The
data set consists of 14 EEG values and a value indicating the eye state.  -->\n\t\t\t<p class=
"normal">Multivariate, Sequential, Time-Series \n\t\tClassification&nbs
p;\n\t\t14980&nbs
p;\n\t\t15 \n\t\t2013 \
n\t\t\t<!-- <td>class="normal">Life&nbsp; -->\n\t\t\tclass="normal">Life&nbsp;
ets/Physicochemical+Properties+of+Protein+Tertiary+Structure"><img src="assets/MLimages/SmallLargedefaul"
t.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Physicochemical+Properties+of"
+Protein+Tertiary+Structure">Physicochemical Properties of Protein Tertiary Structure</a></b>
able>\n\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 armstro
egression \n\t\t\tReal \n\t\t4573
0\ \n\t\t9\ \n\t\t\t2013\ 
\n\t\t\<!-- <td>Life&nbsp; -->\n\t\t bgcolor="DDEEFF">\n\t\t\t<table
e><a href="datasets/seismic-bumps"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /><
/a> <b><a href="datasets/seismic-bumps">seismic-bumps</a></b>
\n\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. \   --> \n\t\tMultivariate \   \n\t\t\t
>Classification \n\t\tReal \n\t\t\t
```

```
2584%nbsp;\n\t\t\t19%nbsp;\n\t\t\t2013%nbsp;
\n\t\t<!--<td>Other&nbsp;-->\n\t\t\t\n\t\t\t<a
href="datasets/banknote+authentication"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&
nbsp;<b><a href="datasets/banknote+authentication">banknote authentication</a
></b>\n\t\t\t<!-- <td>Data were extracted from images that were t
aken for the evaluation of an authentication procedure for banknotes.  -->\n\t\t\t<p class="n
ormal">Multivariate \n\t\t\tClassification \n\t\t\tClassification 
ss="normal">Real \n\t\t\t1372 \n\t\t\t<p class="normal"
al">5 \n\t\t\t2013 \n\t\t\t<!-- <td>2013
puter  -->\n\t\t\\n\t\t\<a href="datasets/USPTO">tt><a href="da
+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Problem%3A+Pat">
<img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<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
oblem: Pat</a></b>
m Competition. Contains drawing pages from US patents with manually labeled figure and part labels.  </
p> -->\n\t\t\tDomain-Theory \n\t\t\tClassifica
tion \n\t\t\tInteger \n\t\t\t306&nbs
p;/n\t\t5 /n\t\t\t2013 /n
sets/YouTube+Multiview+Video+Games+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=
1 /></a>&nbsp;<b><a href="datasets/YouTube+Multiview+Video+Games+Dataset""
>YouTube Multiview Video Games Dataset</a></b>\n\t\t\<!-- <td><p class="normal"
>This dataset contains about 120k instances, each described by 13 feature types, with class information, speci
ally useful for exploring multiview topics (cotraining, ensembles, clustering,..).  -->\n\t\t<p c
lass="normal">Multivariate, Text \n\t\tClassification, Clustering </
p>\n\t\t\t120000 
/n\t\t1000000 /n\t\t2013 
/td>\n\t\t<!-- <td>Computer&nbsp; -->\n\t\t bgcolor="DDEEFF">\n\t\t\t
<a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations"><img src="a
ssets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datas"
ets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations">Gas Sensor Array Drift Dataset at Different
Concentrations</a></b>\n\t\t<!-- <td>This archive contains 139
10 measurements from 16 chemical sensors exposed to 6 different gases at various concentration levels. &nbs
p;-->\n\t\tMultivariate, Time-Series \n\t\t<p class="norm
al">Classification, Regression, Clustering, Causa \n\t\tReal 
/n\t\t13910 /n\t\t\t129 /n\t\t
\t2013 \n\t\t\t<!-- < td>Computer&nbsp;-->\
n\t\t\c/tr>\n\t\t\t\n\t\t\t<a href="datasets/Activities+of+Daily+Living+%28ADLs%29+Recognitio">Activities+of+Daily+Living+%28ADLs%29+Recognitio</a>
n+Using+Binary+Sensors"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
><b><a href="datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary"><b>
+Sensors">Activities of Daily Living (ADLs) Recognition Using Binary Sensors</a></b></t
d>\n\t\t\t<!-- <td>This dataset comprises information regarding the ADLs performed by two u
sers on a daily basis in their \r\nown homes.   -->\n\t\t\tMultivariate, Seq
uential, Time-Series \n\t\tClassification, Clustering \n\t\t
tp class="normal"> \n\t\t\tp class="normal">2747 \n\t\t\tp class="normal">2747 
= "normal" >     \\ \n\t\t\t  p class = "normal" > 2013     \\ \n\t\t\t < !-- < td > p class = "normal" > 2013 &nbsp;   \\ \n\t\t\t < !-- < td > p class = "normal" > 2013 &nbsp;  
>Computer  -->\n\t\t\\n\t\t\<a href="datasets/S"
killCraft1+Master+Table+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<
/td><b><a href="datasets/SkillCraft1+Master+Table+Dataset">SkillCraft1 Master Table
Dataset</a></b>\/tr>\/td>\/tr\/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). </p
>-->\n\t\t\tMultivariate \n\t\t\tRegression&n
bsp;\n\t\t\tlnteger, Real \n\t\t\t3395&nb
sp;\n\t\t20 \n\t\t\t2013 
\ \\n\t\t<!-- <td>Game&nbsp; -->\n\t\t\t<a href="d
atasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units"><img src="assets/MLimages/S
mallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Weight+Lifting+"
Exercises+monitored+with+Inertial+Measurement+Units">Weight Lifting Exercises monitored with Inertial Meas
urement Units</a></b>\n\t\t\t<!-- <td>Six young health subjects w
ere asked to perform 5 variations of the biceps curl weight lifting exercise. One of the variations is the one predi
cted by the health professional.  -->\n\t\t\tMultivariate \n\
```

```
t\t\tClassification \n\t\tReal \n\t\t
tp class="normal">39242 \n\t\tp class="normal">152 \n\t\tp
class="normal">2013 \n\t\t<!-- <td>Physical&nbsp; -->\n\t\t\t
r bgcolor="DDEEFF">\n\t\t<a href="datasets/SML2010"><img src="assets/MLimages/Sma
IlLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/SML2010">SML20
10</a></b></fo>/td>/tr>\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. </t
d> -->\n\t\tMultivariate, Sequential, Time-Series, Text \n\t\t<p cla
ss="normal">Regression \n\t\t\tReal \n\t\t\t<p class=
"normal">4137 \n\t\t\t24 \n\t\t\t2
014\    + td>Computer    + --> + td>Computer     < ---> + td>Computer&nbsp;  Computer&nbsp;  Computer&
ble><a href="datasets/Bike+Sharing+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" bor
der=1 /></a>&nbsp;<b><a href="datasets/Bike+Sharing+Dataset">Bike Sharing Da
taset</a></b>\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 w
eather and seasonal information.  -->\n\t\t\tUnivariate \n\
t/t/Regression \n/t/tInteger, Real </t
d>n\t\tp class="normal">17389\nbsp;\n\t\t\t16\nbsp;\n\t\t\t16\nbsp;
r>\n\t\t\t<a href="datasets/Predict+keywords+activities+in+a+online"><a href="datasets/Predict+keywords+activities+in+a+on
+social+media"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="
normal"><b><a href="datasets/Predict+keywords+activities+in+a+online+social+media">Predict keywords activit
ies in a online social media</a></b>\n\t\t\t<!-- <td>The data from
Twitter was collected during 360 consecutive days. It was done by querying 1497 English keywords sampled fro
m Wikipedia. This dataset is proposed in a Learning to rank setting.  -->\n\t\t\t<p class="nor
mal">Multivariate, Seguential, Time-Series \n\t\t \n\t\t\t
Integer, Real \n\t\t\t51 \n\t\t\t<td
>cp class="normal">35 \n\t\t\tclass="normal">2013 \n\t\t\t<!-- <td>class="normal">2013&nbsp;
+Surgery+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class=
"normal"><b><a href="datasets/Thoracic+Surgery+Data">Thoracic Surgery Data</a></b>
\n\t\t\<!-- <td>The data is dedicated to classification problem related to the post-operati
ve life expectancy in the lung cancer patients: class 1 - death within one year after surgery, class 2 - survival.&n
bsp; -->\n\t\t\tMultivariate \n\t\t\tClassifi
cation \n\t\t\tInteger, Real \n\t\t\t
470\  \n\t\t17\  \n\t\t\t2013\  
\label{linear_posterior} $$ \p>\n\t\t<!-- <td>p class="normal">Life&nbsp;--->\n\t\t\t bgcolor="DDEEFF">\n\t\t\t<t
able><a href="datasets/EMG+dataset+in+Lower+Limb"><img src="assets/MLimages/SmallLargedefaul"><
t.jpg" border=1 /></a>&nbsp;<b><a href="datasets/EMG+dataset+in+Lower+Limb"
>EMG dataset in Lower Limb</a></b>\n\t\t<!-- <td>3 different e
xercises: sitting, standing and walking in the muscles: biceps femoris, vastus medialis, rectus femoris and semit
endinosus addition to goniometry in the exercises.   -->\n\t\tMultivariate
, Time-Series \n\t\t\t \n\t\t\tReal&
nbsp;   \\ n\t  132 & nbsp;   \\ n\t  5 & nbsp;   </pr>
>\n\t\t\t2014 \n\t\t\t<!-- <td>Computer&nbsp;
>-->\n\t\t\\n\t\t\</d>><ahref="datasets/SUSY"><imq src="assets/MLimages/SmallLarge">-->\n\t\t\</a>
default.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/SUSY">SUSY</a></b></
p>\n\t\t\<!-- <td>This is a classification problem to distinguish betwee
n a signal process which produces supersymmetric particles and a background process which does not. 
-->\n\t\t \n\t\t\tClassification </p
>n\t\tReal n\t\t\t5000000 
\label{local-control} $$ \ln t < d>\ \rho \ class = "normal">18 \ nbsp;   \ n\ t \ \rho \ class = "normal">2014 \ nbsp;   \ n\ t < t< < t < td>\ o \ class = "normal">2014 \ nbsp;  
d>Physical  -->\n\t\t\ bgcolor="DDEEFF">\n\t\t\t<
a href="datasets/HIGGS"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
<b><a href="datasets/HIGGS">HIGGS</a></b>\n\t\t\t<!-- <td><p</p>
class="normal">This is a classification problem to distinguish between a signal process which produces Higgs b
osons and a background process which does not.   -->\n\t\t 
\n\t\tClassification \n\t\t\tReal </
td>/n/t/tp class="normal">11000000 /n/t/tp class="normal">28 /n/t/
t/tp class="normal">2014 \n/t/t<!-- <td>p class="normal">Physical&nbsp; -->\n/t
t\t\t\n\t\t<a href="datasets/Qualitative_Bankruptcy"><img src="assets/MLimages/S"
mallLargedefault.ipg" border=1 /></a>&nbsp:class="normal"><b><a href="datasets/Qualitative" Ba</pre>
```

```
nkruptcy">Qualitative\_Bankruptcy</a></b>\n\t\t<!-- <td>Predict t
he Bankruptcy from Qualitative parameters from experts.  -->\n\t\t\tMulti
variate \n\t\t\Classification \n\t\t\t<p class="normal"
td>\n\t\t2014 \n\t\t\t<!-- <td>Computer&nbsp;</t
d> -->\n\t\t\t\n\t\t\t<a href="datasets/LSVT+Voice+Rehabilitation">to the control of the control
"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><
a href="datasets/LSVT+Voice+Rehabilitation">LSVT Voice Rehabilitation</a></b>
\t\t<!-- <td>126 samples from 14 participants, 309 features. Aim: assess whether voice reha
bilitation treatment lead to phonations considered \'acceptable\' or \'unacceptable\' (binary class classification pr
oblem). \   --> \\ lt/tMultivariate    /n/t/tMultivariate   
al">Classification \n\t\t\tReal \n\t\t\t<p class="normal"
">126\ \n\t\t309\ \n\t\t2014&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&nbs=126&n
pn/t/t<--<td>p class="normal">Life&nbsp;-->\n/t/t
href="datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer"><img src="assets/MLimages/S
mallLargedefault.jpg" border=1 /></a>&nbsp;<pcclass="normal"><b><a href="datasets/Dataset+for+A"
DL+Recognition+with+Wrist-worn+Accelerometer">Dataset for ADL Recognition with Wrist-worn Accelerometer
</a></b>
4 Activities of Daily Living (ADL) while carrying a single wrist-worn tri-axial accelerometer.  -->\n\
t\t\tMultivariate, Time-Series \n\t\tClassification,
Clustering \n\t\t \n\t\t 
<!-- <td>Computer&nbsp; -->\n\t\t\t\n\t\t\t<t
r><a href="datasets/Wilt"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
<b><a href="datasets/Wilt">Wilt</a></b>\n\t\t\t<!-- <td><b>
ss="normal">High-resolution Remote Sensing data set (Quickbird). Small number of training samples of diseas
ed trees, large number for other land cover. Testing data set from stratified random sample of image. </p
>-->\n\t\tMultivariate \n\t\tClassification&
nbsp;   n\t t  p \ class = "normal" > 4889 & nbsp;   (nt) t  p \ class = "normal" > 4889 & nbsp;   (nt) t  p \ class = "normal" > 4889 & nbsp;  
\ \\n\t\t6 \\n\t\t2014 \\n\t\t\t<!-- <t
d>Life  -->\n\t\t\\n\t\t\<a href="datasets/User+I">t</a>
dentification+From+Walking+Activity"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbs
p;<b><a href="datasets/User+Identification+From+Walking+Activity">User Identific
ation From Walking Activity</a></b>\n\t\t<!-- <td>The dataset c
ollects data from an Android smartphone positioned in the chest pocket from 22 participants walking in the wild
over a predefined path. \r\n  -->\n\t\tUnivariate, Sequential, Time-Serie
s \n\t\tClassification, Clustering \n\t\t<p class="no
rmal">Real \n\t\t\t \n\t\t\t 
/p>nt/tp class="normal">2014 \n/t/t<!-- <td>p class="normal">Other&nbsp;
/td> -->\n\t\t\ bgcolor="DDEEFF">\n\t\t<a href="datasets/Activity+Recognition+from">tobs/>tr><a href="datasets/Activity+Recognition+from">tobs/>tracentare</a></a></a></a></a></a></a></a>
+Single+Chest-Mounted+Accelerometer"><img src="assets/MLimages/SmallLargedefault.ipg" border=1 /></a>
ccelerometer">Activity Recognition from Single Chest-Mounted Accelerometer</a></b>
d>\n\t\t\t<!-- <td>The dataset collects data from a wearable accelerometer mounted on the c
hest. The dataset is intended for Activity Recognition research purposes.  -->\n\t\t\t<p class
="normal">Univariate, Sequential, Time-Series \n\t\t\tClassification, Clus
tering \n\t\t\tReal \n\t\t\t 
!-- Other  -->\n\t\t\\n\t\t\t<a href="datasets/"
Leaf"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
<b><a href="datasets/Leaf">Leaf</a></b>\n\t\t<!-- <td>This dat
aset consists in a collection of shape and texture features extracted from digital images of leaf specimens origin
ating from a total of 40 different plant species.    -->\n\t\t\tMultivariate&nb
sp;\n\t\t\Real&nbs
p;n\t\t340 \n\t\t16 \n
\t\t2014 \n\t\t<!-- <td>Computer&nbsp; --
>\n\t\t\n\t\t<a href="datasets/Dresses_Attribute_Sales"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="
datasets/Dresses_Attribute_Sales">Dresses_Attribute_Sales</a></b>\n\t\t<!-- <td>
This dataset contain Attributes of dresses and their recommendations according to their sal
es.Sales are monitor on the basis of alternate days.   -->\n\t\t\tclass="normal">Text&nbs
p:\n\t\to class="normal">Classification Clustering&nbsp:\n\t\to class="normal">
```

```
 \n\t\t501 \n\t\t\t13 
td>\n\t\t2014 \n\t\t<!-- <td>Computer&nbsp;</t
d> -->\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\t\n\t\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t<t
s"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b>
<a href="datasets/Tamilnadu+Electricity+Board+Hourly+Readings">Tamilnadu Electricity Board Hourly Reading
s</a></b>\n\t\t<!-- <td>This data can be effectively produced the
e result to fewer parameter of the Load profile can be reduced in the Database   -->\n\t\t\t<
p class="normal">Multivariate \n\t\tClassification, Regression, Clusterin
g\ \n\t\tReal\ \n\t\t\t45781 
\n\t\t5 \n\t\t\t2013 \n\t\t\t2013 
t\t<!-- <td>Life&nbsp; -->\n\t\t\t\n\t\t\t<t
d><a href="datasets/Airfoil+Self-Noise"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&n
bsp;class="normal"><b><a href="datasets/Airfoil+Self-Noise">Airfoil Self-Noise</a></b>
tr>\n\t\t<!-- <td>NASA data set, obtained from a series of aerodynamic and ac
oustic tests of two and three-dimensional airfoil blade sections conducted in an anechoic wind tunnel. </p
>-->\n\t\t\tMultivariate \n\t\t\tRegression&n
bsp;  lass = "normal" > Real   lass = "normal" > 1503  
\label{local-control} $$ \frac{ht}{td>nt}t< d>p class="normal">6&nbsp;\n\t\t\<d>p class="normal">2014&nbsp;\n\t\t\<!--
Physical  -->\n\t\t\\n\t\t\t<a href="datasets/"
Wholesale+customers"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p
class="normal"><b><a href="datasets/Wholesale+customers">Wholesale customers</a></b>
ble>\n\t\t<!-- <td>The data set refers to clients of a wholesale distributor. It includes th
e annual spending in monetary units (m.u.) on diverse product categories  -->\n\t\t\t<p clas
s="normal">Multivariate \n\t\t\tClassification, Clustering \
n\t\tInteger \n\t\t\t440 \n\t\t\t<td
> class="normal">8 \n\t\t\t2014 \n\t\t\t<!-- <td>2014&nbsp;
s="normal">Business  -->\n\t\t\\n\t\t\t<a href="d
atasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis"><img src="assets/MLimages/SmallLargedefault.jpg"
ent+Analysis">Twitter Data set for Arabic Sentiment Analysis</a></b>\n\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.  
->\n\t\t\tText \n\t\t\tClassification 
\nttt< d>&nbsp;\n\t\t\t< d>2000&nbsp;\n\t\t\t
lass="normal">2\ \n\t\t\t2014\ \n\t\t\t<!-- <td>2014&nbsp;
mal">Social  -->\n\t\t\t<a href="datasets/Combined+Cycle+Po">htt\t<a href="datasets/Combined+Cycle+Po">httl<a href="datasets/Combined+Cycle+Po">httl</a>
wer+Plant"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
mal"><b><a href="datasets/Combined+Cycle+Power+Plant">Combined Cycle Power Plant</a></b>
r>\n\t\t\t<!-- <td>The dataset contains 9568 data points collected from a Combi
ned Cycle Power Plant over 6 years (2006-2011), when the plant was set to work with full load.  
>-->\n\t\t\tMultivariate \n\t\tRegression <
a href="datasets/Urban+Land+Cover"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nb
sp;<b><a href="datasets/Urban+Land+Cover">Urban Land Cover</a></b>
>\n\t\t\<!-- <td>Classification of urban land cover using high resolution aeri
al imagery. Intended to assist sustainable urban planning efforts.  -->\n\t\t\t<p class="norma"
l">Multivariate \n\t\t\tClassification \n\t\t\t<p class="n
ormal">\ \n\t\t168\ \n\t\t148\ 
sp;\n\t\t2014 \n\t\t\t<!-- <td>Physical&nbsp
; -->\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\n\t\t\\\n\t\t\\\n\t\t\\\n\t\t\\\n\t\t\\\n\t\t\\\n\t\t\\\n\t\t\\\\n\t\t\\\n\t\t\\\n\t\t\\\n\t\t\\\n\t\t\\\n\t\t\\\n\t\t\\\n\t\t\\\n\t\t\\\\n\t\t\\\\n\t\t\\\\n\t\t\\\\n\t\t\\\\n\t\t\\\\n\t\\\\n\t\\\\n\t\\\\n\t\\\\n\t\\\\n\t\\\\n\t\\\\n\t\\\\n\t\\\n\t\\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\\n\t\\
+1999-2008"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
ormal"><b><a href="datasets/Diabetes+130-US+hospitals+for+years+1999-2008">Diabetes 130-US hospitals f
or years 1999-2008</a></b>\n\t\t\t<!-- <td>This data has been pr
epared to analyze factors related to readmission as well as other \r\n\r\noutcomes pertaining to patients with di
abetes. -->\n\t\t\tMultivariate \n\t\t\t<p class="normal"
al">Classification, Clustering \n\t\tInteger \n\t\t<p
class="normal">100000 \n\t\t\t55 \n\t\t\t<p class="no
rmal">2014 \n\t\t<!-- <td>Life&nbsp; -->\n\t\t\t bgcolor="DD"
EEFF">\n\t\t\t<a href="datasets/Bach+Choral+Harmony"><img src="assets/MLimages/Smal
Il argedefault ing" horder-1 /-/a-&nhsn://td->td->n class-"normal"->h->a href-"datasets/Rach+Choral+Har
```

```
argodoraditijpg bordor- i />
mony">Bach Choral Harmony</a></b>\n\t\t\t<!-- <td>class="normal">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.  -->\n\t\t\t<p class="
normal">Sequential \n\t\tClassification \n\t\t\t
ss="normal">\ \n\t\t5665\ \n\t\t1
7\n\t< d>\p/n\t<<d>\p>/n\t\t<d>\p>/n\t\t\t<d>\p>/n\t\t\t<d>\p>/n\t\t\t<d>\p>
bsp; -->\n\t\t\t\n\t\t\t<a href="datasets/StoneFlakes"><img src="assets/ML">
images/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/Stone"
Flakes">StoneFlakes</a></b>\n\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.  -->\n\t\t<p class="normal"
>Multivariate \n\t\tClassification, Clustering, Causal-Discovery </
p>n\t\t<d>79&nbsp;\n\t\t\t<d>79&nbsp;\n\t\t\t<d>79&nbsp;
tp class="normal">8 \n\t\t\tp class="normal">2014 \n\t\t\t<!-- <td>p c
lass="normal">Other  -->\n\t\t\t\n\t\t\t<a href="d
atasets/Tennis+Major+Tournament+Match+Statistics"><img src="assets/MLimages/SmallLargedefault.jpg" bord
er=1 /></a>&nbsp;<b><a href="datasets/Tennis+Major+Tournament+Match+Statis"
tics">Tennis Major Tournament Match Statistics</a></b>\n\t\t<!-- <td><p class="no
rmal">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. 
t\tMultivariate \n\t\tClassification, Regression,
Clustering \n\t\tInteger, Real \n\t\t\t
al">127\ \n\t\t\42\ \n\t\t\2014\ 
sp:n/t/t<!--<td>Other&nbsp:-->\n/t/t\c/tr>\n/t/t\tOther&nbsp:-->\n/t/t\c/tr><ta
><a href="datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings"><img src="asse
ts/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/"
Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings">Parkinson Speech Dataset with M
ultiple Types of Sound Recordings</a></b>\n\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.  -->\n\t\t\tMultivariate </
p>\n\t\t\tClassification, Regression \n\t\t\tInt
eger, Real \n\t\t\t1040 <math>\n\t\t\t26&
nbsp;   nt/t < d > p class = "normal" > 2014 & nbsp;   nt/t < !-- < td > p class = "normal" > Life & nbsp;  
/p> -->\n\t\t bgcolor="DDEEFF">\n\t\t<a href="datasets/Gesture+Phase+Seg">
mentation"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="nor
mal"><b><a href="datasets/Gesture+Phase+Segmentation">Gesture Phase Segmentation</a></b>
r>\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.  -->\n\t\tp class="normal">Multivariate, Sequential, Time-Series&
nbsp;\n\t\tClassification, Clustering \n\t\t\tClassification, Clustering 
al">Real\ \n\t\t9900\ \n\t\t\t50&n | lass="normal">50&n | lass=
bsp;\n\t\t\t2014 \n\t\t\t<!-- <td>Other&nbsp;
 -->\n\t\t\<\r\n\t\t\<td><a href="datasets/Perfume+Data"><img src="assets/MLim
ages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Perfume"
+Data">Perfume Data</a></b>\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.  -->\n\t\t\tUnivariate, Domain-Theory 
/p>\n\t\tClassification, Clustering \n\t\t\tInt
eger\ \n\t\t560 \n\t\t2 
>\n\t\t2014 \n\t\t<!-- <td>Computer&nbsp;</p
>-->\n\t\t\ bgcolor="DDEEFF">\n\t\t\t<a href="datasets/BlogFeedback"><img s
rc="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="d
atasets/BlogFeedback">BlogFeedback</a></b>\n\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.  -->\n\t\t\tMultivariate 
/p>\n\t\tRegression \n\t\tInteger, Real&n
bsp;   \\ lass = "normal" > 60021 \\ lass = "normal" > 281 \\ lass = "
\n\t\t2014 \n\t\t\t<!-- <td>Social&nbsp;
> -->\n\t\t\n\t\t<a href="datasets/REALDISP+Activity+Recognition+Dataset"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="
datasets/REALDISP+Activity+Recognition+Dataset">REALDISP Activity Recognition Dataset</a></b>
                                the an along "normal". The DEAL DICD detect is deviced to evaluate tech
```

```
/ti>
aling with the effects of sensor displacement in wearable activity recognition as well as to benchmark general a
ctivity recognition algorithms   -->\n\t\t\tMultivariate, Time-Series 
>/n\t\t1419 /n\t\t120 /n\t\t
\t2014 \n\t\t<!-- <td>Computer&nbsp; -->\
n/t/t/t\n/t/t/<a href="datasets/Newspaper+and+magazine+imag">tr><a href="datasets/Newspaper+and+magazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagazine+imagaz
es+segmentation+dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
><b><a href="datasets/Newspaper+and+magazine+images+segmentation+dataset">Newsp
aper and magazine images segmentation dataset</a></b>\n\t\t\c!-- <p class="n
ormal">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.  -->\n\t\t\t<p class="normal
"> \n\t\tClassification \n\t\t&nb
sp;/n\t101 /n\t\t /n\t
\t\t2014 \n\t\t<!-- <td>Computer&nbsp; -->
Limages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/AAAI"
+2014+Accepted+Papers">AAAI 2014 Accepted Papers</a></b>\n\t\t\c!-- <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.   -->\n\t\t\t<p class
="normal">Multivariate \n\t\t\tClustering \n\t\t<p cla
ss="normal">\ \n\t\t\399 \n\t\t\6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6&normal">6
nbsp;  nt/t < d > p class = "normal" > 2014 & nbsp;  n/t/t < !-- < td > p class = "normal" > Computer & last = "normal" > Computer & last
nbsp; -->\n\t\t\t\n\t\t\t<a href="datasets/Gas+sensor+a">+a</a>
rray+under+flow+modulation"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
<b><a href="datasets/Gas+sensor+array+under+flow+modulation">Gas sensor array u
nder flow modulation</a></b>\n\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.  -->\n\t\t\t<p class="no
rmal">Multivariate, Time-Series \n\t\t\ctd>Classification, Regression 
/p>\n\t\tReal \n\t\t58 \n\t\t
td>Computer  -->\n\t\t\t\n\t\t\t<a href="datasets/"
Gas+sensor+array+exposed+to+turbulent+gas+mixtures"><img src="assets/MLimages/SmallLargedefault.jpg"
border=1 /></a>&nbsp;<b><a href="datasets/Gas+sensor+array+exposed+to+turb"><b><a href="datasets/Gas+sensor+array+exposed+to+turb"><b href="datasets/Gas+sensor+array+expose
ulent+gas+mixtures">Gas sensor array exposed to turbulent gas mixtures</a></b>
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.  -->\n\t\t\tMultivariate, Time-Series \n\t\t\t<
td>Classification, Regression \n\t\t\tReal <
td>nttp class="normal">180 \n\t\tp class="normal">150000 \n\t\t
\t2014 \n\t\t<!-- < td>Computer&nbsp;-->\t
n\t\t\t\n\t\t\t<a href="datasets/UJIIndoorLoc"><img src="assets/"
MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/UJ"
IIndoorLoc">UJIIndoorLoc</a></b>\n\t\t\t<!-- <td>class="normal">The UJIIndoorL
oc is a Multi-Building Multi-Floor indoor localization database to test Indoor Positioning System that rely on WLA
N/WiFi fingerprint.  -->\n\t\t\tMultivariate \n\t\t\t<p cla
ss="normal">Classification, Regression \n\t\tInteger, Real </
td>n/t/tp class="normal">21048 \n/t/tp class="normal">529 \n/t/t\t
class="normal">2014 \n\t\t<!-- <td>class="normal">Computer&nbsp;-->\n\
SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/Sentence+Cla
ssification">Sentence Classification</a></b>\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.  -->\n\
t\t\tText \n\t\tClassification \n\t\t
\tInteger \n\t\t\t \n\t\t\t<p cla
ss="normal">\ \n\t\t2014 \n\t\t\t<!-- <td>2014&nbsp;
al">Other  -->\n\t\t\t\n\t\t\t<a href="datasets/Do"
w+Jones+Index"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class
="normal"><b><a href="datasets/Dow+Jones+Index">Dow Jones Index</a></b>
\t<!-- <td>This dataset contains weekly data for the Dow Jones Industrial Index. It has been
```

```
used in computational investing research.&hbsp; -->\h\t\t\t\clus<<p class= normal > i ine-series&hbsp;<
/p>\n\t\tClassification, Clustering \n\t\t\tInt
eger, Real \n\t\t750 <math>\n\t\t16&n class="normal">16&n class="normal">16&n
bsp; /n/t/tp class="normal">2014  /n/t/t/t<!-- <td>p class="normal">Business&nb
sp; -->\n\t\t\d>\n\t\t\t<a href="datasets/sEMG+for+Basic+Hand+movements"
"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><
a href="datasets/sEMG+for+Basic+Hand+movements">sEMG for Basic Hand movements</a></b>
>\n\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 cond
ucted six daily life grasps.  -->\n\t\tTime-Series \n\t\t<t
d>Classification \n\t\t\tReal \n\t\t\t<td
>class="normal">3000 \n\t\t\tclass="normal">2500 \n\t\t\tclass="normal">2500 
="normal">2014 \n\t\t<!-- <td>class="normal">Life&nbsp; -->\n\t\t
="DDEEFF">\n\t\t<a href="datasets/AAAI+2013+Accepted+Papers"><img src="assets/MLi
mages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/AAAI+
2013+Accepted+Papers">AAAI 2013 Accepted Papers</a></b>\n\t\t\t<!-- <td>cla
ss="normal">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.   -->\n\t\t\t<
p class="normal">Multivariate \n\t\t\tClustering \n\t\t\t<td
> class="normal"> \n\t\t150 \n\t\t
mal">5\ \n\t\t\t2014\ \n\t\t\t<!-- <td>Continue (normal) < normal </p>
mputer  -->\n\t\t\\n\t\t\<a href="datasets/Geographical+Original+of">+Original+of</a>
+Music"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="normal
"><b><a href="datasets/Geographical+Original+of+Music">Geographical Original of Music</a></b>
>\n\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 
-->\n\t\tMultivariate \n\t\tClassification, R
egression \n\t\t\tReal \n\t\t\t1059
 \n\t\t68 <math>\n\t\t\t2014&nbsp;<math>
\n\t\t\<!-- <td>Other&nbsp; -->\n\t\t\n\t\t\c\d><ta
ble><a href="datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants"><img src="assets/"
MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Co
ndition+Based+Maintenance+of+Naval+Propulsion+Plants">Condition Based Maintenance of Naval Propulsion
Plants</a></b>\n\t\t\t<!-- <td>Data have been generated from a s
ophisticated simulator of a Gas Turbines (GT), mounted on a Frigate characterized by a COmbined Diesel eLe
ctric And Gas (CODLAG) propulsion plant type.  -->\n\t\t\ctd>Multivariate&n
bsp;\n\t\t\tRegression \n\t\t\tReal&nbsp
;\n\t\t11934 \n\t\t16 
n\t\t\t2014 \n\t\t\t<!-- <td>Computer&nbsp;
-->\n\t\t\\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t</t
ets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets"
/Grammatical+Facial+Expressions">Grammatical Facial Expressions</a></b>
!-- This dataset supports the development of models that make possible to interpret Gra
mmatical Facial Expressions from Brazilian Sign Language (Libras).  -->\n\t\t\t<p class="nor
mal">Multivariate, Sequential \n\t\tClassification, Clustering 
n\t\tReal n\t\t27965 n\t\t27965 
t\t100 \n\t\t2014 \n\t\t\t<!-- <td>2014&nbsp;
Computer  -->\n\t\t\t\n\t\t\t<</pre>
a href="datasets/NoisyOffice"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
<b><a href="datasets/NoisyOffice">NoisyOffice</a></b>\n\t\t
\t<!-- <td>Corpus intended to do cleaning (or binarization) and enhancement of noisy graysc
ale printed text images using supervised learning methods. Noisy images and their corresponding ground truth
provided.  -->\n\t\t\Multivariate \n\t\t\
mal">Classification, Regression \n\t\t\tReal \n\t\t\t<p
class="normal">216\  
<math display="block">class="normal">216\&nbsp; 
mal">2015\ \n\t\t\<!-- <td>Computer\&nbsp; --> \n\t\t\
<tm, sets/MLimages/SmallLargedefault.jpg
border=1 /></a>&nbsp;<b><a href="datasets/MHEALTH+Dataset">MHEALTH Da
taset</a></b>\n\t\t<!-- <td>The MHEALTH (Mobile Health) data
set is devised to benchmark techniques dealing with human behavior analysis based on multimodal body sensi
ng.  -->\n\t\t\tMultivariate, Time-Series \n\t\t\tMultivariate, Time-Series 
s="normal">Classification \n\t\t\tReal \n\t\t\t<p class
```

```
="normal">120 \n\t\t\t23 \n\t\t\t2
014 \n\t\t<!-- <td>Computer&nbsp; -->\n\t\t\t
EFF">\n\t\t\<a href="datasets/Student+Performance"><img src="assets/MLimages/SmallLa"
rgedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/Student+Performance"
">Student Performance</a></b>\n\t\t<!-- <td>Predict student pe
rformance in secondary education (high school).   -->\n\t\t\tMultivariate&
nbsp;\n\t\tClassification, Regression \n\t\t\tClassification, Regression 
mal">Integer\ \n\t\t\t649\ \n\t\t\t33
 \n\t\t\2014 <math>\n\t\t\<!-- <td>Social&nb
sp; -->\n\t\t\t\n\t\t\t<a href="datasets/ElectricityLoadDiagrams20112014">
<img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a
href="datasets/ElectricityLoadDiagrams20112014">ElectricityLoadDiagrams20112014</a></b>
able>\n\t\t<!-- <td>This data set contains electricity consumption of 370 points/clients.\
r\n  -->\n\t\t\tTime-Series \n\t\t\t
Regression, Clustering \n\t\t\tReal \n\t\t\tReal 
ormal">370%nbsp;\n\t\t<math>140256%nbsp;\n\t\t<math>140256%nbsp;
2015\    + td>Computer\    --> + td>Computer\    --> + td>Computer\   
EEFF">\n\t\t\t<a href="datasets/Gas+sensor+array+under+dynamic+gas+mixtures"><img s
rc="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="d
atasets/Gas+sensor+array+under+dynamic+gas+mixtures">Gas sensor array under dynamic gas mixtures</a>
</b>
emical sensors exposed to two dynamic gas mixtures at varying concentrations. For each mixture, signals were
acquired continuously during 12 hours.  -->\n\t\t\tMultivariate, Time-Seri
es \n\t\tClassification, Regression \n\t\t<p class="
normal">Real \n\t\t\t4178504 \n\t\t\t<p class="normal"
al">19 \n\t\t2015 <math>\n\t\t<!-- <td>Co
="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="dat
asets/microblogPCU">microblogPCU</a></b>\n\t\t\t<!-- <td>Micr
oblogPCU data is crawled from sina weibo microblog[http://weibo.com/]. This data can be used to study machin
e learning methods as well as do some social network research.   -->\n\t\t\t<p class="norm
al">Multivariate, Univariate, Sequential, Text \n\t\t\tClassification, Causal
-Discovery \n\t\tInteger, Real \n\t\t\t
al">221579\ \n\t\t\t20\ \n\t\t\t2015
 \n\t\t<!-- <td>Computer&nbsp; -->\n\t\t\t
F">\n\t\t\t<a href="datasets/Firm-Teacher Clave-Direction Classification"><img src="asset"
s/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/F
irm-Teacher_Clave-Direction_Classification">Firm-Teacher_Clave-Direction_Classification</a></b>
r>\n\t\t<!-- <td>The data are binary attack-point vectors and their clave-directi
on class(es) according to the partido-alto-based paradigm.  -->\n\t\t\tMu
ltivariate \n\t\tClassification \n\t\t\t<p class="normal"
"> \n\t\t10800 \n\t\t20 </
p  n t t < d > p class = "normal" > 2015 & nbsp;  n t t < d > p class = "normal" > Other & nbsp;  n t t < d > p class = "normal" > Other & nbsp;  n t t < d > p class = "normal" > Other & nbsp;   n t < d > p class = "normal" > Other & nbsp;   n t < d > p class = "normal" > Other & nbsp;   n t < d > p class = "normal" > Other & nbsp;   n t < d > p class = "normal" > Other & nbsp;   n t < d > p class = "normal" > Other & nbsp;   n t < d > p class = "normal" > Other & nbsp;   n t < d > p class = "normal" > Other & nbsp;   n t < d > p class = "normal" > Other & nbsp;   n t < d > p class = "normal" > Other & nbsp;   n t < d > p class = "normal" > Other & nbsp;   n t < d > p class = "normal" > Other & nbsp;   n t < d > p class = "normal" > Other & nbsp;   n t < d > p class = "normal" > Other & nbsp;   n t < d > p class = "normal" > Other & nbsp;   n t < d > p class = "normal" > Other & nbsp;   n t < d > p class = "normal" > Other & nbsp;   n t < d > p class = "normal" > Other & nbsp;   n t < d > p class = "normal" > Other & nbsp;   n t < d > p class = "normal" > Other & nbsp;  < d > n t < d > p class = "normal" > Other & nbsp;  n t < d > p class = "normal" > Other & nbsp;  n t < d > p class = "normal" > Other & nbsp;  n t < d > p class = "normal" > Other & nbsp;  n t < d > p class = "normal" > Other & nbsp;  n t < d > p class = "normal" > Other & nbsp;  n t < d > p class = "normal" > Other & nbsp;  n t < d > p class = "normal" > Other & nbsp;  n t < d > p class = "normal" > Other & nbsp;  n t < d > p class = "normal" > Other & nbsp;  n t < d > p class = "normal" > Othe
td>-->\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\t\n\t\t\t\t\t\n\t\t\t\t\n\t\t\t\t\t\n\t\t\t\t\t\n\t\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\t\n\t\t\t\t\t\n\t\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t
mg src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a hr
ef="datasets/Dataset+for+Sensorless+Drive+Diagnosis">Dataset for Sensorless Drive Diagnosis</a></b></
td>\n\t\t\<!-- <td>Features are extracted from motor current. The motor h
as intact and defective components. This results in 11 different classes with different conditions.  
> -->\n\t\t\ctd>Multivariate \n\t\t\ctd>Classification&nbsp
\n\t\tp class="normal">Real \n\t\tp class="normal">58509 
\ \n\t\t\t49 \n\t\t\t2015 \n\t\t\t<!--
Computer  -->\n\t\t\t\n\t\t\t<t
d><a href="datasets/TV+News+Channel+Commercial+Detection+Dataset"><img src="assets/MLimages/SmallL"
argedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/TV+News+Channel+
Commercial+Detection+Dataset">TV News Channel Commercial Detection Dataset</a></b>
e>\n\t\t\<!-- <td>TV Commercials data set consists of standard audio-visual features o
f video shots extracted from 150 hours of TV news broadcast of 3 Indian and 2 international news channels (3
0 Hours each).   -->\n\t\t\tMultivariate \n\t\t\tMultivariate 
s="normal">Classification, Clustering \n\t\t\tReal \n\t\t\t<t
d><p\ class="normal">129685\&nbsp;\n\t\t\t<p\ class="normal">12\&nbsp;\n\t\t\t<p\ class="normal">12&nbsp;
s="normal">2015\ \n\t\t<!--<td>Computer&nbsp; -->\n\t\t\t
\n\t\t<a href="datasets/Phishing+Websites"><img src="assets/MLimages/SmallLargedefau">
```

```
lt.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Phishing+Websites">Phishing
Websites</a></b>\n\t\t\t<!-- <td>This dataset collected mainly fro
m: PhishTank archive, MillerSmiles archive, Google's searching operators.  -->\n\t\t\t<p clas
s="normal"> \n\t\tClassification \n\t\t\t<p class="normal"
mal">Integer\ \n\t\t\t2456\ \n\t\t\t3
0\ \n\t\t2015\ \n\t\t\t<!-- <td>Compute
r Security  -->\n\t\t\\n\t\t\<a href="datasets/Gre
enhouse+Gas+Observing+Network"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbs
p;<b><a href="datasets/Greenhouse+Gas+Observing+Network">Greenhouse Gas
Observing Network</a></b>
etwork to monitor emissions of a greenhouse gas (GHG) in California given time series of synthetic observation
s and tracers from weather model simulations.\r\n  -->\n\t\t\tMultivariate,
Time-Series \n\t\tRegression \n\t\tRegression 
al">Real \n\t\t5232
nbsp; -->\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\
al"><b><a href="datasets/Diabetic+Retinopathy+Debrecen+Data+Set">Diabetic Retinopathy Debrecen Data Set
rom the Messidor image set to predict whether an image contains signs of diabetic retinopathy or not.  </
p  --> /n /t /t  class = "normal" > Multivariate & nbsp;   /n /t /t  class = "normal" > Classification
 \n\t\tInteger, Real \n\t\t\t1151&
nbsp;  \\ lass = "normal" > 20 & nbsp;  \\ lass = "normal" > 2014 & nbsp; 
td>\n\t\t<!-- <td>Life&nbsp; -->\n\t\t\t bgcolor="DDEEFF">\n\t\t\t
><a href="datasets/HIV-1+protease+cleavage"><img src="assets/MLimages/SmallLargedefault.jpg" bor
der=1 /></a>&nbsp;<b><a href="datasets/HIV-1+protease+cleavage">HIV-1 prote
ase cleavage</a></b>\n\t\t<!-- <td>class="normal">The data contains lists of oct
amers (8 amino acids) and a flag (-1 or 1) depending on whether HIV-1 protease will cleave in the central positi
on (between amino acids 4 and 5).  -->\n\t\tMultivariate 
>\n\t\t\tClassification \n\t\t\tCategorical </p
>n/t/tp class="normal">6590 \n/t/tp class="normal">1 \n/t/t/t<
td><p\ class="normal">2015\&nbsp;\n\t\t\<!--< td><p\ class="normal">Life\&nbsp;\-->\n\t\t\
\n\t\t\t<a href="datasets/Sentiment+Labelled+Sentences"><img src="assets/MLimages/"
SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/Sentiment+La"
belled+Sentences">Sentiment Labelled Sentences</a></b>\n\t\t\t<!-- <td>class="
normal">The dataset contains sentences labelled with positive or negative sentiment.  -->\n\t\t\
Text \n\t\tClassification \n\t\t\t<t
d><p\ class="normal">&nbsp;\n\t\t\t<p\ class="normal">3000&nbsp;\n\t\t\t<p\ class="normal">0.000&nbsp;
ormal">\ \n\t\t\t2015\ \n\t\t\t<!-- <td>Ot
ews+Popularity"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class=
"normal"><b><a href="datasets/Online+News+Popularity">Online News Popularity</a></b>
>\n\t\t\t<!-- <td>This dataset summarizes a heterogeneous set of features about article
s published by Mashable in a period of two years. The goal is to predict the number of shares in social network
s (popularity).  -->\n\t\tMultivariate \n\t\t<p class="
normal">Classification, Regression \n\t\tInteger, Real \
nttt39797 \nttt61 \nttt<<p class="normal">61&nbsp;\nttt<<p class="normal">61&nbsp;\nttt<<p class="normal">61&nbsp;\nttt<<p class="normal">61&nbsp;\nttt<<p class="normal">61&nbsp;\nttt<<p class="normal">61&nbsp;\nttt<<p class="normal">61&nbsp;
p class="normal">2015 \n\t\t\t<!-- <td>Business&nbsp; -->\n\t\t\t</tr
>\n\t\t<a href="datasets/Forest+type+mapping"><img src="assets/MLimages/SmallLar">
gedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Forest+type+mapping"
>Forest type mapping</a></b>\n\t\t\<!-- <td>Multi-temporal rem
ote sensing data of a forested area in Japan. The goal is to map different forest types using spectral data.&nbs
p;-->\n\t\t\tMultivariate \n\t\t\tClassifica
tion \n\t\t\t \n\t\t326 
/n\t\t27 /n\t\t2015 /n\t\t\t<
!-- Life  -->\n\t\t\\n\t\t\t<
a href="datasets/wiki4HE"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
><b><a href="datasets/wiki4HE">wiki4HE</a></b>\n\t\t<!-- <td>
Survey of faculty members from two Spanish universities on teaching uses of Wikipedia&nbs
p;-->\n\t\t\tMultivariate \n\t\tRegressi
on, Clustering, Causal-Discovery \n\t\t\t \n\t\t\t<p cla
ss="normal">913\ \n\t\t53\ \n\t\t\t
```

```
2015\ \n\t\t\<!-- <td>Social\&nbsp; -->\n\t\t\
><a href="datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset"><img src="assets/"
MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/On"
line+Video+Characteristics+and+Transcoding+Time+Dataset">Online Video Characteristics and Transcoding Ti
me Dataset</a></b>\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
.   -->\n\t\t\tMultivariate \n\t\t\tRe
gression \n\t\tInteger, Real \n\t\t<p class="normal"
nbsp;\n\t\t\<!-- <td>Computer&nbsp; -->\n\t\t\
>\n\t\t<a href="datasets/Chronic Kidney Disease"><img src="assets/MLimages/SmallLar"></a>
gedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/Chronic_Kidney_Disea"
se">Chronic_Kidney_Disease</a></b>\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 p
eriod.  -->\n\t\t\tMultivariate \n\t\t\t<p class="normal"
>Classification \n\t\t\tReal \n\t\t\t
400\    \\ lass = "normal" > 25\    \\ lass = "normal" > 2015\    \\ lass = "normal" > 2015\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ | 400\   \\ 
/p>n/t/t<!--<td>p class="normal">Other&nbsp;-->\n/t/t\
href="datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014"><img src="ass
ets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets"
/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014">Machine Learning based ZZ
Alpha Ltd. Stock Recommendations 2012-2014</a></b>\n\t\t\c!-- <p class="nor
mal">The data here are the ZZAlpha® machine learning recommendations made for various US traded stock p
ortfolios the morning of each day during the 3 year period Jan 1, 2012 - Dec 31, 2014.  
t\tSequential, Time-Series \n\t\tClassification&n
bsp;\n\t\t\t314080 
>\n\t\t0 \n\t\t2015 \n\t\t\t<
!-- Business  -->\n\t\t\z\tr>\n\t\t\z\ta>
<a href="datasets/Folio"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
td><b><a href="datasets/Folio">Folio</a></b>\n\t\t<!-- <td><p cl
ass="normal">20 photos of leaves for each of 32 different species.  -->\n\t\t\t2
mal">Multivariate \n\t\tClassification, Clustering \n\t\t\<
td> \n\t\t\t637 \n\t\t\t<p class="no
rmal">20\ \n\t\t\t2015\ \n\t\t\t<!-- <td>2015\&nbsp;
Other  -->\n\t\t\n\t\t<a href="datasets/Taxi+Service+Trajectory+-">+-</a>
+Prediction+Challenge%2C+ECML+PKDD+2015"><img src="assets/MLimages/SmallLargedefault.jpg" border=
1 /></a>&nbsp;class="normal"><b><a href="datasets/Taxi+Service+Trajectory+-+Prediction+Challe"></a>
nge%2C+ECML+PKDD+2015">Taxi Service Trajectory - Prediction Challenge, ECML PKDD 2015</a></b>
\n\t\t<!-- <td>An accurate dataset describing trajectories performed
by all the 442 taxis running in the city of Porto, in Portugal.\r\n  -->\n\t\t
Multivariate, Sequential, Time-Series, Domain-Theory \n\t\t\tClustering,
Causal-Discovery \n\t\t\tReal \n\t\t\t<p class="norma
l">1710671 & p; /n/t/tp class="normal">9 & p; /n/t/tp class="normal">2015 & p; /n/t/tp class="normal">2015 & p; 
nbsp;\n\t\t\<!-- <td>Computer&nbsp; -->\n\t\t\
>\n\t\t\t<a href="datasets/Cuff-Less+Blood+Pressure+Estimation"><img src="assets/MLima"><a href="datasets/Cuff-Less+Blood+Pressure+Estimation"><img src="assets/MLima"><a href="datasets/Cuff-Less+Blood+Pressure+Estimation"><img src="assets/MLima"></a></a></a>
ges/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/Cuff-Less"
+Blood+Pressure+Estimation">Cuff-Less Blood Pressure Estimation</a></b>
-- This Data set provides preprocessed and cleaned vital signals which can be used in d
esigning algorithms for cuff-less estimation of the blood pressure.  -->\n\t\t\t<p class="norm
al">Multivariate \n\t\t\tClassification, Regression \n\t\t\t<t
d>Real \n\t\t\t12000 \n\t\t\t12000 
ss="normal">3 \n\t\t\2015 \n\t\t\<!-- <td>2015&nbsp;
mal">Life  -->\n\t\t\n\t\t</d><ahref="datasets/Smartphone-Based+Re"
cognition+of+Human+Activities+and+Postural+Transitions"><img src="assets/MLimages/SmallLargedefault.jpg"
border=1 /></a>&nbsp;<b><a href="datasets/Smartphone-Based+Recognition+of+">
Human+Activities+and+Postural+Transitions">Smartphone-Based Recognition of Human Activities and Postural
Transitions</a></b>\n\t\t<!-- <td>Activity recognition data set bu
ilt from the recordings of 30 subjects performing basic activities and postural transitions while carrying a waist-
mounted smartphone with embedded inertial sensors.\r\n  -->\n\t\t\tMulti
variate, Time-Series \n\t\t\ctd>Classification \n\t\t\ctd><p cl
ass="normal">Real\ \n\t\t\t10929\ \n\t\t\t10929\ 
mal">561\ \n\t\t\t2015\ \n\t\t\t<!-- <td>
```

```
Life  -->\n\t\t\\n\t\t\t<a href="datasets/Mice+Pro">tr><a 
tein+Expression"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class
="normal"><b><a href="datasets/Mice+Protein+Expression">Mice Protein Expression</a></b>
able>\n\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 asso
ciative learning.  -->\n\t\t\tMultivariate \n\t\t\t<p class
="normal">Classification, Clustering \n\t\tReal \n\t\t<td
>1080 \n\t\t\t82 \n\t\t\t<p class="
normal">2015\ \n\t\t\<!-- <td>Life&nbsp; --> \n\t\t\
><a href="datasets/UJIIndoorLoc-Mag"><img src="assets/MLimages/SmallLargedefault.jpg" bor
der=1 /></a>&nbsp;<b><a href="datasets/UJIIndoorLoc-Mag">UJIIndoorLoc-Mag">UJIIndoorLoc-Mag<
/a></b>\n\t\t<!-- <td>The UJIIndoorLoc-Mag is an indoor localiz
ation database to test Indoor Positioning System that rely on Earth\'s magnetic field variations. 
-->\n\t\tMultivariate, Sequential, Time-Series \n\t\t<p class="norm
al">Classification, Regression, Clustering \n\t\tInteger, Real 
/n\t\t40000 /n\t\t\t13 /n\t\t\t13 
2015 \n\t\t<!-- <td>Computer&nbsp; -->\n\
t\t\t\n\t\t<a href="datasets/Heterogeneity+Activity+Recognition"
"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><
a href="datasets/Heterogeneity+Activity+Recognition">Heterogeneity Activity Recognition</a></b>
>\n\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; speci
fically, the dataset is gathered with a variety of different device models and use-scenarios, in order to reflect se
nsing heterogeneities to be expected in real deployments.  -->\n\t\t\tMult
ivariate, Time-Series \n\t\tClassification, Clustering \n\t\
t/tp class="normal">Real \n/t/tp class="normal">43930257 \n/t/t
>cp class="normal">16 \n\t\t\t2015 \n\t\t\t<!-- <td>2015&nbsp;
ss="normal">Computer  -->\n\t\t\\n\t\t\<a href="datasets/Educatio">catasets/Educatio</a>
nal+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set"><img src="assets/MLimages/SmallL
argedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Educational+Process"
+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set">Educational Process Mining (EPM): A Learning
Analytics Data Set</a></b>\n\t\t\t<!-- <td>Educational Process Mi
ning data set is built from the recordings of 115 subjects\' activities through a logging application while learning
with an educational simulator.  -->\n\t\tMultivariate, Sequential, Time-S
eries \n\t\tClassification, Regression, Clustering \n\t\t\t<
td>Integer \n\t\t\t230318 \n\t\t\t<
p class="normal">13 \n\t\t\t2015 <math>\n\t\t\t<!-- <td>2015&nbsp;<math>
="normal">Computer  -->\n\t\t\t\n\t\t\t<a href="d
atasets/HEPMASS"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p cla
ss="normal"><b><a href="datasets/HEPMASS">HEPMASS</a></b>\n\t\t\t<!-- <td>
class="normal">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.  -->\
n\t\tMultivariate \n\t\tClassification 
/n\t/tReal /n/t/t10500000 
d>Physical  -->\n\t\t\c/tr>\n\t\t\c/td><cd><a href="datasets/In"
door+User+Movement+Prediction+from+RSS+data"><img src="assets/MLimages/SmallLargedefault.jpg" borde
r=1 /></a>&nbsp;<b><a href="datasets/Indoor+User+Movement+Prediction+from+"
RSS+data">Indoor User Movement Prediction from RSS data</a></b>\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. <
/p> -->\n\t\t\tMultivariate, Sequential, Time-Series \n\t\t\t<p cla
ss="normal">Classification \n\t\t\tReal \n\t\t\tReal 
s="normal">13197\ 
$$/p>
$$/nt/t<<d>$<pclass="normal">4&nbsp;
$$/nt/t<<d>$<pclass="normal">4&nbsp;
2016 \n\t\t<!-- <td>class="normal">Computer&nbsp; -->\n\t\t
EEFF">\n\t\t<a href="datasets/Open+University+Learning+Analytics+dataset"><img src="a
ssets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datas"
ets/Open+University+Learning+Analytics+dataset">Open University Learning Analytics dataset</a></b>
>\n\t\t\<!-- <td>Open University Learning Analytics Dataset contains data
about courses, students and their interactions with Virtual Learning Environment for seven selected courses an
d more than 30000 students.&nbsp:
```

```
ries \n\t\tClassification, Regression, Clustering \n\t\t\t<t
d>Integer \n\t\t\t \n\t\t\t<p class=
"normal"> \n\t\t2015 \n\t\t<!-- <td>
Computer  -->\n\t\t\n\t\t\t<a href="datasets/default+of+credit+card">+credit+card</a>
+clients"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="normal"
"><b><a href="datasets/default+of+credit+card+clients">default of credit card clients</a></b>
ble>\n\t\t<!-- <td>This research aimed at the case of customers' default payments in T
aiwan and compares the predictive accuracy of probability of default among six data mining methods. 
>-->\n\t\tMultivariate \n\t\tClassification&
nbsp;   \\ lass = "normal" > Integer, Real \    \\ lass = "normal" > 30000 \& lass = "normal" > 10000 \& lass = "normal" > 10000
nbsp;   nt/t  p class = "normal" > 24 & nbsp;   (nt/t  p class = "normal" > 2016 & nbsp;  
td>n/t/t<--< td>p class="normal">Business&nbsp; -->\n/t/t/
table><a href="datasets/Mesothelioma%E2%80%99s+disease+data+set+"><img src="assets/MLimage"
s/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/Mesothelio"
ma%E2%80%99s+disease+data+set+">Mesothelioma's disease data set </a></b>
\t\t<!-- <td>Mesothelioma's disease data set were prepared at Dicle University Faculty of M
edicine in Turkey.\r\nThree hundred and twenty-four Mesothelioma patient data. In the dataset, all samples hav
e 34 features. \   --> \\ lt/tMultivariate \   /n/t/t/tMultivariate   
"normal">Classification \n\t\tReal \n\t\t<p class="n
ormal">324\ \n\t\t\t34\ \n\t\t\t2016
 \n\t\t\<!-- <td>Computer&nbsp;<math>-->\n\t\t\
><a href="datasets/Online+Retail"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a
> <b><a href="datasets/Online+Retail">Online Retail</a></b>
\n\t\t<!-- <td>This is a transnational data set which contains all the transaction
s occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail. <
/p> -->\n\t\t\tMultivariate, Sequential, Time-Series \n\t\t\t<p cla
ss="normal">Classification, Clustering \n\t\t\tInteger, Real </t
d>\n\t\t541909\ \n\t\t\t8\ \n\t\t\t8 
d><p class="normal">2015 \n\t\t\c!-- Business  -->\n\t\t\t
\n\t\t\t<a href="datasets/SIFT10M"><img src="assets/MLimage"><a href="datasets/SIFT10M"><img src="assets/MLimage"><a href="datasets/SIFT10M"><timg src="assets/MLimage"><a href="datasets/SIFT10M"><ta><a href="da
s/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/SIFT10M">
a SIFT feature which is extracted from Caltech-256 by the open source VLFeat library. The corresponding patc
hes of the SIFT features are provided. -->\n\t\t\tMultivariate 
\n\t\tCausal-Discovery \n\t\tInteger&nbsp
\n\t\t11164866 \n\t\t128 
\n\t\t\t2016 \n\t\t\t<!-- <td>Computer&nbsp;
-->\n\t\t\\n\t\t\t<a href="datasets/GPS+Trajectories"><img src="assets/MLima">
ges/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/GPS+Traj"
ectories">GPS Trajectories</a></b>\n\t\t\t<!-- <td>The dataset h
as 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).   -->\n\t\t\tMultivariate \
n\t\tClassification, Regression \n\t\tReal&nbsp
\n\t\t163 \n\t\t15 \n\t
\t\t2016 \n\t\t<!-- <td>Computer&nbsp; -->
\n\t\t\n\t\t\ctd><a href="datasets/Detect+Malacious+Executable%"
28AntiVirus%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class
="normal"><b><a href="datasets/Detect+Malacious+Executable%28AntiVirus%29">Detect Malacious Executable
e(AntiVirus)</a></b>\n\t\t\t<!-- <td>p class="normal">I extract features from malaci
ous and non-malacious and create and training dataset to teach svm classifier. Dataset made of unknown exec
utable to detect if it is virus or normal safe executable.  -->\n\t\t\tMultivar
iate \n\t\t\tClassification \n\t\t\tRe
al\    \\ lass = "normal" > 373\    \\ lass = "normal" > 513\    \\ lass = "normal" > 513\   
>\n\t\t2016 \n\t\t<!-- <td>Computer&nbsp;</p
>-->\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t</
MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Oc
cupancy+Detection+">Occupancy Detection </a></b>\n\t\t\t<!-- <td>class="norma"
I">Experimental data used for binary classification (room occupancy) from Temperature, Humidity, Light and CO
2. Ground-truth occupancy was obtained from time stamped pictures that were taken every minute. 
-->\n\t\t\tMultivariate, Time-Series \n\t\t\tCl
assification \n\t\t\tReal \n\t\t\t205
60%nbsn:</n>\n\t\t\t< n class="normal">7%nbsn:</n>\n\t\t\t< n class="normal">2016%nbsn:</n
```

```
>\n\t\t<!--<td>Computer&nbsp; -->\n\t\t\t bgcolor="DDEEFF">\n\t\t\t<t
d><a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+">d><a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+">d><a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+">d><a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+">d><a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+">d><a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+">d</a></a>
Parkinson%E2%80%99s+Disease"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
r+Monitoring+Parkinson%E2%80%99s+Disease">Improved Spiral Test Using Digitized Graphics Tablet for Mon
itoring Parkinson's Disease</a></b>\n\t\t\t<!-- <td>Handwriting d
atabase consists of 25 PWP(People with Parkinson) and 15 healthy individuals. Three types of recordings (Stati
c Spiral Test, Dynamic Spiral Test and Stability Test) are taken. -->\n\t\t\t<p class="normal"
>Multivariate \n\t\tClassification, Regression, Clustering 
\ \n\t\tReal \n\t\t40 \n\t\t<
p\ class="normal">7\ \n\t\t\t<p\ class="normal">2016\&nbsp;\n\t\t\t<!--<td><p\ class="normal">2016&nbsp;
normal">Computer  -->\n\t\t\t\n\t\t\t<a href="datasets/News+Aggreg">normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">Normal">
ator"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<
b><a href="datasets/News+Aggregator">News Aggregator</a></b>\n\t\t\t<!-- <td><p
class="normal">References to news pages collected from an web aggregator in the period from 10-March-201
4 to 10-August-2014. The resources are grouped into clusters that represent pages discussing the same story.
 --> \n\t\tp class="normal">Multivariate <math>/n\t\tp class="normal">Clas
sification, Clustering \n\t\t \n\t\t<p class="normal"
>422937%nbsp;\n\t\t\t5%nbsp;\n\t\t\t2016%nb
sp; \\ large sp; --> \\ large sp;
<tmg src="assets/MLimages/SmallLargedefault.jpg" border=
1 /></a>&nbsp;</d></b></a></b></d>
>\n\t\t<!-- <td>Contains the responses of a gas multisensor device deployed o
n the field in an Italian city. Hourly responses averages are recorded along with gas concentrations references f
rom a certified analyzer.  -->\n\t\tMultivariate, Time-Series 
\n\t\tRegression \n\t\tReal </td
\ \n\t\t\t9358 \n\t\t\t15 \n\t\t\t
2016 \n\t\t<!-- <td>Computer&nbsp; -->\n\t\t\t<
/tr>\n\t\t<a href="datasets/Twin+gas+sensor+arrays"><img src="assets/MLimages/Sm
allLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Twin+gas+senso"
r+arrays">Twin gas sensor arrays</a></b>\n\t\t<!-- <td>5 replic
ates of an 8-MOX gas sensor array were exposed to different gas conditions (4 volatiles at 10 concentration lev
els each).  -->\n\t\t\tMultivariate, Time-Series, Domain-Theory </p
>\n\t\tClassification, Regression \n\t\tRea
l\ /n\t\t640\ /n\t\t480000\ 
\n\t\t\t2016 \n\t\t\t<!-- <td>Computer&nbsp;
 -->\n\t\t bgcolor="DDEEFF">\n\t\t<a href="datasets/Gas+sensors+for+h">
ome+activity+monitoring"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
<b><a href="datasets/Gas+sensors+for+home+activity+monitoring">Gas sensors for home
activity monitoring</a></b>\n\t\t<!-- <td>100 recordings of a sen
sor array under different conditions in a home setting: background, wine and banana presentations. The array i
ncludes 8 MOX gas sensors, and humidity and temperature sensors.\r\n  -->\n\t\t<p class
="normal">Multivariate, Time-Series \n\t\t\tClassification 
\n\t<d>Real \n\t\t919438 \n\t\t\t919438 
d><p\ class="normal">11&nbsp;\n\t\t\t<p\ class="normal">2016&nbsp;\n\t\t\t<!--<td><p\ class="normal">2016&nbsp;
ass="normal">Computer  -->\n\t\t\<a href="datasets/Facebo">tr><a hre
ok+Comment+Volume+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;</t
d><b><a href="datasets/Facebook+Comment+Volume+Dataset">Facebook Comment
Volume Dataset</a></b>\n\t\t<!-- <td>Instances in this dataset c
ontain features extracted from facebook posts. The task associated with the data is to predict how many comm
ents the post will receive.     -->\n\t\tMultivariate    \n\t\t<td
>Regression \n\t\tInteger, Real \n\t\
t\t40949 \n\t\t\t54 \n\t\t\t54 
lass="normal">2016\ 
<math display="block">lass="normal">Other\&nbsp;
gcolor="DDEEFF">\n\t\t\ctd><a href="datasets/Smartphone+Dataset+for+Human+Activity+Rec
ognition+%28HAR%29+in+Ambient+Assisted+Living+%28AAL%29"><img src="assets/MLimages/SmallLargedef"
ault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Smartphone+Dataset+for+H">
uman+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Living+%28AAL%29">Smartphone Dataset for
Human Activity Recognition (HAR) in Ambient Assisted Living (AAL)</a></b>
-- This data is an addition to an existing dataset on UCI. We collected more data to impr
ove the accuracy of our human activity recognition algorithms applied in the domain of Ambient Assisted Living.
9 phon:
```

```
ssification \  \n\t\t\Real \  \n\t\t\5744
 \n\t\t561 \n\t\t2016 
>\n\t\t<!-- <td>Computer&nbsp; -->\n\t\t\t\n\t\t\t<
a href="datasets/Polish+companies+bankruptcy+data"><img src="assets/MLimages/SmallLargedefault.jpg" bor
der=1 /></a>&nbsp;<b><a href="datasets/Polish+companies+bankruptcy+data">P
olish companies bankruptcy data</a></b>\n\t\t\t<!-- <td>The data
set is about bankruptcy prediction of Polish companies. The bankrupt companies were analyzed in the period 20
00-2012, while the still operating companies were evaluated from 2007 to 2013.  -->\n\t\t<
p class="normal">Multivariate \n\t\t\tClassification \n\t\t\t
class="normal">Real \n\t\t\tclass="normal">10503 \n\t\t\tc
lass="normal">64\ \n\t\t\t2016\ \n\t\t\t<!-- <td>2016&nbsp;
ormal">Business  -->\n\t\t\t>table><a href="datas">\n\t\t\t<a href="datas"><a hr
ets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29"><img src="assets/MLim"
ages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Activity+"
Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29">Activity Recognition system based o
n Multisensor data fusion (AReM)</a></b>\n\t\t\t<!-- <td>This dat
aset contains temporal data from a Wireless Sensor Network worn by an actor performing the activities: bendin
g, cycling, lying down, sitting, standing, walking.  -->\n\t\t\tMultivariate, S
equential, Time-Series \n\t\t\tClassification \n\t\t\t<p
class="normal">Real\ \n\t\t\t42240\ \n\t\t\t42240\ 
ormal">6\ \n\t\t\t2016\ \n\t\t\t<!-- <td>C
omputer  -->\n\t\t\t\n\t\t\t<a href="datasets/Dota2+Games+Results"
><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a
href="datasets/Dota2+Games+Results">Dota2 Games Results</a></b>\r\t\t\t<!-- <td
>Dota 2 is a popular computer game with two teams of 5 players. At the start of the game e
ach player chooses a unique hero with different strengths and weaknesses.  -->\n\t\t<p cla
ss="normal">Multivariate \n\t\tClassification \n\t\t\t
 \n\t\t\102944 \n\t\t\<p class="n
ormal">116\ \n\t\t2016\ \n\t\t<!-- <td>
>Game  -->\n\t\t\\n\t\t\<a href="datasets/Face"><a href="datasets/
book+metrics"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="
normal"><b><a href="datasets/Facebook+metrics">Facebook metrics</a></b>
<!-- <td>Facebook performance metrics of a renowned cosmetic\'s brand Facebook page.&
nbsp; -->\n\t\t\tMultivariate \n\t\t\tRegre
ssion \n\t\tInteger \n\t\t\t500&n
bsp;  \ln t  p class = "normal" > 19 & nbsp;  \ln t  p class = "normal" > 2016 & nbsp; 
d>n\t\t<!--< td>Business&nbsp; -->\n\t\t\t\n\t\t\t<a hr
ef="datasets/UbiqLog+%28smartphone+lifelogging%29"><img src="assets/MLimages/SmallLargedefault.jpg" b
order=1 /></a>&nbsp;<b><a href="datasets/UbiqLog+%28smartphone+lifelogging"><b><a href="datasets/UbiqLog+%28smartphone+lifelogging"><b><a href="datasets/UbiqLog+%28smartphone+lifelogging"><b><a href="datasets/UbiqLog+%28smartphone+lifelogging"><b><a href="datasets/UbiqLog+%28smartphone+lifelogging"><b><a href="datasets/UbiqLog+%28smartphone+lifelogging"><b><a href="datasets/UbiqLog+%28smartphone+lifelogging"><b><a href="datasets/UbiqLog+%28smartphone+lifelogging"><b><a href="datasets/UbiqLog+%28smartphone+lifelogging"><b><a href="datasets/UbiqLog+%28smartphone+lifelogging"><a href="datasets/UbiqLog+%28smartphone+lifelogging</a><a hre
%29">UbiqLog (smartphone lifelogging)</a></b>\n\t\t\t<!-- <td>U
biqLog is the smartphone lifelogging tool that runs on the smartphone of 35 users for about 2 months. \r\n&nbs
p;-->\n\t\tclass="normal">Multivariate \n\t\tclass="normal">Causal-Di
scovery \    \\ lass="normal"> \    \\ lass="normal"> 9782222 \&n lass="normal"> 8782222 \&n lass="normal"> 87822222 \&n lass="normal"> 8782222 \&n lass="normal"> 8782222 \&n lass
bsp;/nt/tt /nt/tt/td>2016 /
le><a href="datasets/NIPS+Conference+Papers+1987-2015"><img src="assets/MLimages/SmallLarge"><a href="datasets/NIPS+Conference+Papers+1987-2015"><img src="assets/MLimages/SmallLarge"><a href="datasets/NIPS+Conference+Papers+1987-2015"><a href="datasets/NIPS+2015"><a href="datasets/NIPS+2015">
default.jpg" border=1 /></a>&nbsp;<b><a href="datasets/NIPS+Conference+Paper"
s+1987-2015">NIPS Conference Papers 1987-2015</a></b>\n\t\t\t<!-- <td><p class
="normal">This data set contains the distribution of words in the full text of the NIPS conference papers publish
ed from 1987 to 2015. anbsp;  --> \ln t < d>= "normal">Text & nbsp; <math> \ln t < d>= class = "normal">Text & nbsp; <math>
="normal">Clustering \n\t\tInteger \n\t\t<p class="
normal">11463\ \n\t\t5812\ \n\t\t\t
>2016 \n\t\t<!-- <td>Computer&nbsp; -->\n\t\t\t
<a href="datasets/HTRU2"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a
> <b><a href="datasets/HTRU2">HTRU2</a></b>
>\n\t\t\<!-- <td>Pulsar candidates collected during the HTRU survey. Pulsars are a type of st
ar, of considerable scientific interest. Candidates must be classified in to pulsar and non-pulsar classes to aid di
scovery.  -->\n\t\t\tMultivariate \n\t\t\t<p class="normal"
al">Classification, Clustering \n\t\t\tReal \n\t\t\tReal 
ss="normal">17898\ \n\t\t\t9\ \n\t\t\t
>2017%nbsp;\n\t\t<!--<td>Physical%nbsp;---\n\t\t\ bgcolor="DD">2017
```

```
EEFF >\II\\\\\ella>\table><\ti>\ella>\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\table\tabl
MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Dr
ug+consumption+%28quantified%29">Drug consumption (quantified)</a></b>
<!-- <td>Classify type of drug consumer by personality data&nbsp; -->\n\t\t\t<p
class="normal">Multivariate \n\t\t\tClassification \n\t\t\t<t
d>Real \n\t\t1885 \n\t\t\t1885 
s="normal">32\ \n\t\t\t2016\ \n\t\t\t<!-- <td>2016&nbsp;
mal">Social  -->\n\t\t\\etr>\n\t\t\<a href="datasets/Appliances+energy+"
prediction"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="nor
mal"><b><a href="datasets/Appliances+energy+prediction">Appliances energy prediction</a></b>
>\n\t\t<!-- <td>Experimental data used to create regression models of applian
ces energy use in a low energy building. -->\n\t\t\tMultivariate, Time-Se
ries \n\t\t\tRegression \n\t\t\tReal
>\n\t\t2017 \n\t\t<!-- <td>Computer&nbsp;</p
>-->\n\t\t\ bgcolor="DDEEFF">\n\t\t\<a href="datasets/Miskolc+IIS+Hybrid+IP">
S"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b>
<a href="datasets/Miskolc+IIS+Hybrid+IPS">Miskolc IIS Hybrid IPS</a></b>
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. &nbs
p;-->\n\t\tText \n\t\tClassification, Cl
ustering, Causal-Discovery \n\t\tInteger \n\t\t<p cl
ass="normal">1540\ \n\t\t\t67\ \n\t\t\t67\ 
I">2016\ \n\t\t<!-- <td>op class="normal">Computer&nbsp; -->\n\t\t\t
><a href="datasets/KDC-4007+dataset+Collection"><img src="assets/MLimages/SmallLargedef"><a href="datasets/KDC-4007+dataset+Collection"><img src="assets/MLimages/SmallLargedef"><a href="datasets/KDC-4007+dataset+Collection"><a href="datasets/KDC-4007+dataset+Collection"><a href="datasets/KDC-4007+dataset+Collection"><a href="datasets/KDC-4007+dataset+Collection"><a href="datasets/KDC-4007+dataset+Collection"><a href="datasets/KDC-4007+dataset+Collection"><a href="datasets/KDC-4007+dataset+Collection"><a href="datasets/KDC-4007+dataset+Collection"><a href="datasets/KDC-4007+dataset+Collection">
ault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/KDC-4007+dataset+Collecti
on">KDC-4007 dataset Collection</a></b>\n\t\t<!-- <td>class="normal">KDC-40
07 dataset Collection is the Kurdish Documents Classification text used in categories regarding Kurdish Sorani
news\ and\ articles. \   --> \\ l't <p\ class="normal">Multivariate,\ Text &nbsp;  
Classification, Regression \n\t\tInteger </
td>n\tp class="normal">4007 \n\t\tp class="normal"> \n\t\t\tp class="normal"> 
p class="normal">2017 \n\t\t\t<!-- <td>Computer&nbsp; -->\n\t\t\t</t
r>\n\t\t<a href="datasets/Geo-Magnetic+field+and+WLAN+dataset"><a href="datasets/Geo-Magnetic+field+and+wLAN+datasets/"><a href="datasets/Geo-Magnetic+field+and+wLAN+datasets/"><a href="datasets/Geo-Magnetic+field+and+wLAN+datasets/"><a href="datasets/"><a href="datasets/
+for+indoor+localisation+from+wristband+and+smartphone"><img src="assets/MLimages/SmallLargedefault.jp"
g" border=1 /></a>&nbsp;<b><a href="datasets/Geo-Magnetic+field+and+WLAN+d"
ataset+for+indoor+localisation+from+wristband+and+smartphone">Geo-Magnetic field and WLAN dataset for in
door localisation from wristband and smartphone</a></b>\/\table>\/\table>
ormal">A multisource and multivariate dataset for indoor localisation methods based on WLAN and Geo-Magne
tic field fingerprinting  -->\n\t\t\t>Multivariate, Sequential, Time-Series&nb
sp;\n\t\t\tClassification, Regression, Clustering \n\t\t\t<p cl
ass="normal">Integer, Real \n\t\t\t153540 \n\t\t\t<p
class="normal">25 \n\t\t\2017 \n\t\t\<!-- <td><p class="
normal">Computer  -->\n\t\t\n\t\t\t<a href="datasets/DrivFace"><i
mg src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a hr
ef="datasets/DrivFace">DrivFace</a></b>\n\t\t\t<!-- <td>The Driv
Face contains images sequences of subjects while driving in real scenarios. It is composed of 606 samples of 6
40×480, acquired over different days from 4 drivers with several facial features.  -->\n\t\t\ctd><
p class="normal">Multivariate \n\t\tClassification, Regression, Clusterin
g\nbsp;\n\t\tReal\nbsp;\n\t\t606\nbsp;
>n/t/tp class="normal">6400 \n/t/tp class="normal">2016 \n/t
\t\t<!-- <td>Computer&nbsp; -->\n\t\t\c/tr>\n\t\t\c/ta><table
><a href="datasets/Website+Phishing"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /
></a>&nbsp;<b><a href="datasets/Website+Phishing">Website Phishing</a></b></
p>/nt/tt<!-- <td>\r\n\r\n&nbsp;-->\n\t\t\t\r\n\r\n&nbsp;
al">Multivariate \n\t\t\tClassification \n\t\t\t<p class="
normal">Integer\ \n\t\t1353 \n\t\t\t1353 
">10\ \n\t\t\2016\ \n\t\t\<!-- <td>Comp
uter  -->\n\t\t\n\t\t<a href="datasets/YouTube+Spam+Collection"
><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a
href="datasets/YouTube+Spam+Collection">YouTube Spam Collection</a></b>
\t<!-- <td>It is a public set of comments collected for spam research. It has five datasets co
mposed by 1,956 real messages extracted from five videos that were among the 10 most viewed on the collecti
```

```
on period.&nosp;</ta> -->\n\t\t\t<ta>Lext&nosp;</ta>\n\t\t\t<ta>C
lassification \n\t\t \n\t\t1956&n
bsp;nt/tp class="normal">5 nt/t/tp class="normal">2017 
\ \n\t\t\t<!-- <td>Computer&nbsp; -->\n\t\t\t bgcolor="DDEEFF">\n\t\t\t<t
able><a href="datasets/Beijing+PM2.5+Data"><img src="assets/MLimages/SmallLargedefault.jpg" bord
er=1 /></a>&nbsp;<b><a href="datasets/Beijing+PM2.5+Data">Beijing PM2.5 Data
ata of US Embassy in Beijing. Meanwhile, meteorological data from Beijing Capital International Airport are also
included.   -->\n\t\t\tMultivariate, Time-Series \n\t\t
Regression \n\t\t\Integer, Real \n\t\t\
tp class="normal">43824 \n\t\t\tp class="normal">13 \n\t\t\tp class="normal">13 
ass="normal">2017\ \n\t\t\<!-- <td>Physical\&nbsp; -->\n\t\t\
>\n\t\t<a href="datasets/Cargo+2000+Freight+Tracking+and+Tracing"><img src="assets/"
MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Ca
rgo+2000+Freight+Tracking+and+Tracing">Cargo 2000 Freight Tracking and Tracing</a></b>
able>\n\t\t<!-- <td>Sanitized and anonymized Cargo 2000 (C2K) airfreight tracking an
d tracing events, covering five months of business execution (3,942 process instances, 7,932 transport legs, 56
,082 activities).   -->\n\t\t\tMultivariate, Sequential \n\t\t\t
Classification, Regression \n\t\t\tInteger 
\t2016 \n\t\t<!-- <td>Business&nbsp; -->\n\
t\t\t\n\t\t\t<a href="datasets/Cervical+cancer+%28Risk+Factors">t\t\t<a href="datasets/Cervical+cancer+%28Risk+Factors">t\t\t</a>
%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
<b><a href="datasets/Cervical+cancer+%28Risk+Factors%29">Cervical cancer (Risk Factors)</a></b>
>\n\t\t\<!-- <td>This dataset focuses on the prediction of indicators/diagno
sis of cervical cancer. The features cover demographic information, habits, and historic medical records. 
 -->\n\t\t\Multivariate \n\t\t\Classificati
on \n\t\t\tInteger, Real \n\t\t\t858
 \n\t\t36 <math>\n\t\t\t2017&nbsp;<math>
\n\t\t\t<!-- <td>Life&nbsp; -->\n\t\t\n\t\t\t<a href=""">href=""</a>
datasets/Quality+Assessment+of+Digital+Colposcopies"><img src="assets/MLimages/SmallLargedefault.jpg" bo
rder=1 /></a>&nbsp;<b><a href="datasets/Quality+Assessment+of+Digital+Colpos"
copies">Quality Assessment of Digital Colposcopies</a></b>\n\t\t\t<!-- <td><p class=
"normal">This dataset explores the subjective quality assessment of digital colposcopies.  -->\n\
t\t\tMultivariate \n\t\tClassification </t
d>\n\t\tReal\ \n\t\t\t287\ \n\t\t\t287\ 
>class="normal">69 \n\t\t\tclass="normal">2017 \n\t\t\t<!-- <td>class="normal">2017&nbsp;
ss="normal">Life  -->\n\t\t\t\n\t\t\t<a href="datas"
ets/KASANDR"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="
normal"><b><a href="datasets/KASANDR">KASANDR</a></b>\n\t\t\t<!-- <td><p clas
s="normal">KASANDR is a novel, publicly available collection for recommendation systems that records the beh
avior of customers of the European leader in e-Commerce advertising, Kelkoo.   -->\n\t\t\t<
p class="normal">Multivariate \n\t\t\tCausal-Discovery \n
\tp class="normal">2158859 \n\t\t\tp class="normal">2017 \n\t\t\t<!--
<\!td><\!p\ class="normal">\!Life\&nbsp;<\!/p><\!/td> -->\n\t\t\<\!dr><\!h\t\t\t<\!td><\!table><\!tr><<\!a\ href="datasets/FMA">th\t\t\t<\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table><\table>
%3A+A+Dataset+For+Music+Analysis"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&n
bsp;<b><a href="datasets/FMA%3A+A+Dataset+For+Music+Analysis">FMA: A Dat
aset For Music Analysis</a></b>\n\t\t<!-- <td>class="normal">FMA features 106
,574 tracks and includes song title, album, artist, genres; play counts, favorites, comments; description, biograp
hy, tags; together with audio (343 days, 917 GiB) and features.  -->\n\t\t\t<p class="normal"
>Multivariate, Time-Series \n\t\tClassification, Clustering </t
d>\n\t\tReal\ \n\t\t\t106574\ \n\t\t\t106574\ 
t<p\ class="normal">518\&nbsp;\n\t\t\t<p\ class="normal">2017\&nbsp;\n\t\t\t<!--< td><p\ class="normal">2017&nbsp;
p class="normal">Computer  -->\n\t\t\\n\t\t\<a
href="datasets/Air+quality"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
><b><a href="datasets/Air+quality">Air quality</a></b>
 Contains the responses of a gas multisensor device deployed on the field in an Italian c
ity.   -->\n\t\tMultivariate, Time-Series \n\t\t<p clas
s="normal">Regression\ \n\t\t\tReal\ \n\t\t\tReal\ 
normal">9358\ \n\t\t\t15\ \n\t\t\t20
```

```
tr><a href="datasets/Epileptic+Seizure+Recognition"><img src="assets/MLimages/SmallLargedefault.jpg" b
order=1 /></a>&nbsp;<b><a href="datasets/Epileptic+Seizure+Recognition">Epilep
tic Seizure Recognition</a></b>\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.   -->\n\t\tMultivariate, Time-Series \n\t\t\t<
p class="normal">Classification, Clustering \n\t\t\tInteger, Real </p
>n/t/tp class="normal">11500 n/t/t/p class="normal">179 
tp class="normal">2017 \n\t\t<!-- <td>p class="normal">Life&nbsp; -->\n\t\t\t < !-- <td>p class="normal">Life&nbsp;
\n\t\t\t<a href="datasets/Devanagari+Handwritten+Character+D">thatasets/Devanagari+Handwritten+Character+D
ataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="normal"
><b><a href="datasets/Devanagari+Handwritten+Character+Dataset">Devanagari Handwritten Character Data
set</a></b>\n\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%).   -->\n\t\t \n\t
\t\tClassification \n\t\tInteger \n\
t/tc class="normal">92000 \n/t/tc class="normal"> \n/t/t/tc class="normal"> \n/t/t/t 
ass="normal">2016\ \n\t\t\<!-- <td>Computer\&nbsp; -->\n\t\t\
tr>\n\t\t\t<a href="datasets/Stock+portfolio+performance"><img src="assets/MLimages/Sma">
IlLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/Stock+portfolio+pe">href="datasets/Stock+portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Stock-portfolio+pe">href="datasets/Datasets/Datasets/Datasets/Datasets/Datasets/Datasets/Datasets/D
rformance">Stock portfolio performance</a></b>\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.  -->\n\t\t\tMultivariate 
n/t/t/tRegression \n/t/tReal \n/t/t
tp class="normal">315 \n\t\tp class="normal">12 \n\t\t\tp class="normal">12 
s="normal">2016 \n\t\t\<!-- <td>Business&nbsp; -->\n\t\t\
gcolor="DDEEFF">\n\t\t\t<a href="datasets/MoCap+Hand+Postures"><img src="assets/MLi
mages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/MoCa"
p+Hand+Postures">MoCap Hand Postures</a></b>\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. 
-->\n\t\t\tMultivariate \n\t\t\tClassification, Cluster
ing \n\t\t\t780
95\ \n\t\t38\ \n\t\t2016 
p>/n/t/t<!--< td>Computer&nbsp; -->/n/t/t/t
<a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="
datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech">Early biomark
ers of Parkinson s disease based on natural connected speech</a></b>
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;-->\n\t\tMultivariate \n\t\tClassifica
tion, Regression \n\t\t\tInteger, Real \n\t\t\t<p class=
"normal">130\ \n\t\t65\ \n\t\t20
17\    + td> Life\    --> + td> Life\   Life\  Life\  Life\  Life\  Life\  Life\  Life\  Life\  Life\  Life\  Life\  Life\  Life\  Life\  Life\  Life\  Life\  Life\  Life\  Life\  Life\  Life\  Life\  
n\t\t\t<a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Educ">n\t\t\t<a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Educ">n\t\t\t<a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Educ">n\t\t\t<\t<a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Educ">n\t\t\t<a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Educ">n\t\t<a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Educ">n\t\t<a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Educ">n\t\t<a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Educ">n\t\t<a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Educ">n\t\t<a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Educ">n\t\t<a href="datasets/Data=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=for+Bota=
ation+Setting"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="n
ormal"><b><a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting"
>Data for Software Engineering Teamwork Assessment in Education Setting</a></b>
>\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  -->\n\t\tSequential, Time-Series \n\t\t\t<
td>Classification \n\t\tInteger, Real \
n\t\tp class="normal">74\nbsp;|n\t\tp class="normal">102\nbsp;|n\t\tp class="normal">102\nbsp;
class="normal">2017 \n\t\t\<!-- <td>Computer&nbsp; -->\n\t\t\
\n\t\t\t<a href="datasets/PM2.5+Data+of+Five+Chinese+Cities"><img src="assets/MLim">
ages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><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, Chengd
u and Shenyang. Meanwhile, meteorological data for each city are also included. hbsp; -->\n\t\t\t
Multivariate, Time-Series \n\t\tRegression 
p>\n\t\t\t52854 \n\t\t\t52854
```

```
\t<!-- <td>Physical&nbsp; -->\n\t\t\n\t\t\t<tr
><a href="datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet"><img src="as
sets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="dataset"
s/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet">Parkinson Disease Spiral Drawings
Using Digitized Graphics Tablet</a></b>\n\t\t\t<!-- <td>Handwriti
ng 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.  -->\n\t\t\t<p class="nor
mal">Multivariate \n\t\tClassification, Regression, Clustering </p
>\n\t\t17 \n\t\t77 \n\t\t77 
t\t7 \n\t\t\t2017 \n\t\t\t<!-- <td><p
class="normal">Computer  -->\n\t\t\\r\\t\t\<a href="datasets/Sales"
Transactions_Dataset_Weekly"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;</td
><b><a href="datasets/Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly Weekly W
aset Weekly</a></b>\n\t\t<!-- <td>Contains weekly purchased
quantities of 800 over products over 52 weeks. Normalised values are provided too.  -->\n\t\t\t<t
d>Multivariate, Time-Series \n\t\tClustering 
/p>\n\t\tInteger, Real \n\t\t\t811 </p
>n/t/tp class="normal">53 n/t/t/t<d>p class="normal">2017&nbsp;\n/t/t/t
<!-- <td>&nbsp; -->\n\t\t\t\n\t\t\t<a
href="datasets/Las+Vegas+Strip"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;</
td><b><a href="datasets/Las+Vegas+Strip">Las Vegas Strip</a></b>
ble>\n\t\t\t<!-- <td>This dataset includes quantitative and categorical features from onli
ne reviews from 21 hotels located in Las Vegas Strip, extracted from TripAdvisor (http://www.tripadvisor.com).&
nbsp;-->\n\t\t \n\t\tClassification, Re
gression \n\t\t\tInteger \n\t\t\t504
 \n\t\t20 \n\t\t\t2017 
href="datasets/Eco-hotel"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
<b><a href="datasets/Eco-hotel">Eco-hotel</a></b>\d>\n\t\t\t<!-- <td</pre>
>This dataset includes Online Textual Reviews from both online (e.g., TripAdvisor) and offlin
e (e.g., Guests\' book) sources from the Areias do Seixo Eco-Resort.  -->\n\t\t\t<p class="no
rmal">Text \n\t\t\t \n\t\t\t 
p  n t t < d < p class = "normal" > 401 nbsp;  n t t  1 nbsp;  n t t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n
td>2017 \n\t\t\t<!-- <td>Business&nbsp; -->\n\t\t\
t bgcolor="DDEEFF">\n\t\t\t<a href="datasets/MEU-Mobile+KSD"><img src="assets/
MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/M"
EU-Mobile+KSD">MEU-Mobile KSD</a></b>\n\t\t\t<!-- <td>This d
ataset contains keystroke dynamics data collected on a touch mobile device (Nexus 7). The dataset contains 2
856 records, 51 records per subject for 56 subjects.   -->\n\t\t\tMultivaria
te \n\t\tClassification \n\t\t\tInte
ger, Real \n\t\t\t2856 \n\t\t71&n
bsp; /n/t/t2016  /n/t/t/t<!-- <td>Computer&n 
bsp; -->\n\t\t\\chip<a href="datasets/Crowdsourced+Mapping"><img src</td>
="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="dat
asets/Crowdsourced+Mapping">Crowdsourced Mapping</a></b>\n\t\t\t<!-- <td><p cl
ass="normal">Crowdsourced data from OpenStreetMap is used to automate the classification of satellite image
s into different land cover classes (impervious, farm, forest, grass, orchard, water).   -->\n\t\t\t<t
d>Multivariate \n\t\tClassification \n\
t/tp class="normal"> \n/t/tp class="normal">10546 \n/t/tp class="normal">
ass="normal">29\ \n\t\t2016\ \n\t\t<!-- <td>2016\&nbsp;
rmal">Physical  -->\n\t\t\t\n\t\t\t<a href="dataset"><a href=
s/gene+expression+cancer+RNA-Seq"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&n
bsp;<b><a href="datasets/gene+expression+cancer+RNA-Seq">gene expression c
ancer RNA-Seq</a></b>\n\t\t\t<!-- <td>This collection of data is p
art 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. 
-->\n\t\t\t<p class="normal"
>Multivariate \n\t\tClassification, Clustering \n\t\t<
p class="normal">Real \n\t\t\t801 \n\t\t\t<p class="n
ormal">20531\ \n\t\t\t2016\ \n\t\t\t<!-- <td>2016\&nbsp;
al">Life  -->\n\t\t\\r\t\t\<a href="datasets/Hybrid+Indoor+Positionin"
g+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer"><img src="assets/MLimages/SmallLargedefa
ult.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Hybrid+Indoor+Positioning+D"><b><a href="datasets/Hybrid+Indoor+Positioning+D"><b><a href="datasets/Hybrid+Indoor+Positioning+D"><b><a href="datasets/Hybrid+Indoor+Positioning+D"><b><a href="datasets/Hybrid+Indoor+Positioning+D"><b><a href="datasets/Hybrid+Indoor+Positioning+D"><b><a href="datasets/Hybrid+Indoor+Positioning+D"><b><a href="datasets/Hybrid+Indoor+Positioning+D"><b><a href="datasets/Hybrid+Indoor+Positioning+D"><b><a href="datasets/Hybrid+Indoor+Positioning+D"><b href="datasets/Hybrid+Indoor+P"><b href="datasets/Hybrid+Indoor+P"><b href="datasets/Hybrid+D"><b href="datasets/Hybri
```

```
ataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer">Hybrid Indoor Positioning Dataset from WiFi RSS
I, Bluetooth and magnetometer</a></b>\n\t\t\t<!-- <td>The datas
et 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.   -->\n\t\t\t<p class
="normal">Multivariate, Sequential, Time-Series \n\t\t\tClassification&nb
sp;\n\t\tReal \n\t\t\t1540 
d>\n\t\t65\ \n\t\t\t2016\ \n\t\t\t<!--
Computer  -->\n\t\t\t\n\t\t\t<t
d><a href="datasets/chestnut+%E2%80%93+LARVIC"><img src="assets/MLimages/SmallLargedefault.jpg" bor
der=1 /></a>&nbsp;<b><a href="datasets/chestnut+%E2%80%93+LARVIC">chest
nut – LARVIC</a></b></tp>\n\t\t\t<!-- <td>The research project prese
nts this database, shows the images of chestnuts that will be processed to determine the presence or absence
of defects  -->\n\t\t\t \n\t\t\tClassi
fication, Clustering \n\t\t\t \n\t\t\t1
451\   /td > lt/t  p class = "normal" > 3\   /td > lt/t  p class = "normal" > 2017\   /td > lt/t  p class = "normal" > 2017\   /td > lt/t  p class = "normal" > 2017\   /td > lt/t  p class = "normal" > 2017\   /td > lt/t  p class = "normal" > 2017\   /td > lt/t  p class = "normal" > 2017\   /td > lt/t  p class = "normal" > 2017\   /td > lt/t  p class = "normal" > 2017\   /td > lt/t  p class = "normal" > 2017\   /td > lt/t  p class = "normal" > 2017\  
<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>&nbsp;
normal"><b><a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switchi
ng+%28OBS%29+Network">Burst Header Packet (BHP) flooding attack on Optical Burst Switching (OBS) Netw
fying the risks of the Burst Header Packet (BHP) flood attacks in Optical Burst Switching networks (OBS) is the s
carcity of reliable historical data.  -->\n\t\t\tText \n\t\t\t<t
d>Classification \n\t\t\tInteger \n\t\t\t<
td><p\ class="normal">1075\&nbsp;\n\t\t\t<p\ class="normal">22\&nbsp;\n\t\t\t<p\ class="normal">22&nbsp;
="normal">2017 \n\t\t<!-- <td>Computer&nbsp; -->\n\t\t\t
gcolor="DDEEFF">\n\t\t\t<a href="datasets/Motion+Capture+Hand+Postures"><img src="as
sets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="dataset"
s/Motion+Capture+Hand+Postures">Motion Capture Hand Postures</a></b>
-- 5 types of hand postures from 12 users were recorded using unlabeled markers on fi
ngers of a glove in a motion capture environment. Due to resolution and occlusion, missing values are common
. \    --> \\ lt   Multivariate & nbsp;   \\ lt   Class = "normal"> Class = "norma
ssification, Clustering \n\t\t\tReal \n\t\t\tReal 
rmal">78095\ \n\t\t38\ \n\t\t\t201
7\ \n\t\t<!--<td>Computer\&nbsp; -->\n\t\t\t
e><a href="datasets/Anuran+Calls+%28MFCCs%29"><img src="assets/MLimages/SmallLargedefault.j"
pg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/Anuran+Calls+%28MFCCs%29">
Anuran Calls (MFCCs)</a></b>\n\t\t\t<!-- <td>Acoustic features e
xtracted from syllables of anuran (frogs) calls, including the family, the genus, and the species labels (multilabel
).   -->\n\t\t\tMultivariate \n\t\t\tCl
assification, Clustering \n\t\t\tReal \n\t\t\t<p class="n
ormal">7195\ \n\t\t22\ \n\t\t\t201
7\ + (--p class="normal">Life\ + --> (n)t/t/t+ (n)t/t/t+ (n)t/t/t
t/t/t<a href="datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization">tt/t<a href="datasets/TTC-3600%3A+Benchmark+dataset-for+Turkish+text+categorization">tt/t<a href="datasets/TTC-3600%3A+Benchmark+dataset-for+Turkish+text+categorization">tt/t<a href="datasets/TTC-3600%3A+Benchmark+dataset-for+Turkish+text+categorization">tt/t<a href="datasets/TTC-3600%3A+Benchmark+dataset-for+Turkish+text+categorization">tt/t<a href="datasets/TTC-3600%3A+Benchmark+dataset-for+Turkish+text+categorization">tt/t<a href="datasets/TTC-3600%3A+Benchmark+dataset-for+Turkish+text+categorization">tt/t<a href="datasets/TTC-3600%3A+Benchmark+dataset-for+Turkish+text+categorization">tt/t<a href="dataset-for+Turkish+text+categorization">tt/t<a href="dataset-for+Turkish+text+categorization">tt/t<a href="dataset-for+Turkish+text+categorization">tt/t<a href="dataset-for+Turkish+text+categorization">tt/t<a href="dataset-for+Turkish+text+categorization">tt/t<a href="dataset-for+Turkish+text+categor
"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><
a href="datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization">TTC-3600: Benchmark
dataset for Turkish text categorization</a></b>\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 w
ell-known portals in Turkey. It has 4 different forms in ARFF Weka format.  -->\n\t\t\t<p clas
s="normal">Text \n\t\tClassification, Clustering \n\t\t<t
d>Integer \n\t\t3600 \n\t\t\t<p cl
ass="normal">4814 \n\t\t\t2017 \n\t\t\t<!-- <td><p class="
normal">Computer  -->\n\t\t\\r\t\t<a href="datasets/Gastrointestin">rointestin</a>
al+Lesions+in+Regular+Colonoscopy"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nb
sp;<b><a href="datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy">Gas
trointestinal Lesions in Regular Colonoscopy</a></b>\n\t\t<!-- <td><p class="norma"
I">This dataset contains features extracted from colonoscopy videos used to detect gastrointestinal lesions. It c
ontains 76 lesions: 15 serrated adenomas, 21 hyperplastic lesions and 40 adenoma.   -->\n\t\t\
Multivariate \n\t\tClassification \
n\t\t\tReal \n\t\t\t76 \n\t\t\t<p
class="normal">698\ \n\t\t\t2016\ \n\t\t\t<!-- <td>2016&nbsp;
"normal">Computer&nbsp: -->\n\t\t\t\n\t\t\t<a href="da"
```

```
tasets/Daily+Demand+Forecasting+Orders"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></
a> <b><a href="datasets/Daily+Demand+Forecasting+Orders">Daily Deman
d Forecasting Orders</a></b>\n\t\t\t<!-- <td>The dataset was coll
ected during 60 days, this is a real database of a brazilian logistics company.   -->\n\t\t\t<p cl
ass="normal">Time-Series \n\t\tRegression \n\t\t
Integer \n\t\t\t60 \n\t\t\t
"normal">13\ \n\t\t2017\ \n\t\t\t<!-- <td>13&nbsp;
">Business  -->\n\t\t\\n\t\t\t<a href="datasets/Paper+Reviews"><im
g src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href
="datasets/Paper+Reviews">Paper Reviews</a></b>\n\t\t\t<!-- <td><p class="normal"
">This sentiment analysis data set contains scientific paper reviews from an international conference on comput
ing and informatics. The task is to predict the orientation or the evaluation of a review.  -->\n\t\t\t
Text \n\t\tClassification, Regression 
/n\t\t405 /n\t\t405 /n\t\t405 /n\t\t405 /n\t\t405 
p class="normal">Computer  -->\n\t\t\\n\t\t\t<a
href="datasets/extention+of+Z-Alizadeh+sani+dataset"><img src="assets/MLimages/SmallLargedefault.jpg" bor
der=1 /></a>&nbsp;<b><a href="datasets/extention+of+Z-Alizadeh+sani+dataset">
extention of Z-Alizadeh sani dataset</a></b>\n\t\t<!-- <td>It was
collected for CAD diagnosis.  -->\n\t\t\t \n\t\t\t<p clas
s="normal">Classification \n\t\t\Integer, Real \n\t\t\t
<p class="normal">303 </p></td>\n\t\t\t<p class="normal">59 </p></td>\n\t\t\t<p class="normal">50 </p></td>\n\t\t\t<p class="normal">50 </p></td>
rmal">2017 \n\t\t<!-- <td>Life&nbsp; -->\n\t\t
able><a href="datasets/Z-Alizadeh+Sani"><img src="assets/MLimages/SmallLargedefault.jpg" border=
1 /></a>&nbsp;<b><a href="datasets/Z-Alizadeh+Sani">Z-Alizadeh Sani</a></b></
p>\n\t\t\<!-- <td>It was collected for CAD diagnosis.&nbsp;
->\n\t\t \n\t\tClassification \n\t\t
\tInteger, Real \n\t\t\t303 \n\t\t\t<
td>56 \n\t\t2017 \n\t\t\t<!-- <td>2017&nbsp;
ass="normal">Life  -->\n\t\t bgcolor="DDEEFF">\n\t\t<a href="dat
asets/Dynamic+Features+of+VirusShare+Executables"><img src="assets/MLimages/SmallLargedefault.jpg" bor
der=1 /></a>&nbsp;<b><a href="datasets/Dynamic+Features+of+VirusShare+Exec
utables">Dynamic Features of VirusShare Executables</a></b>\n\t\t\t<!-- <td>cla
ss="normal">This dataset contains the dynamic features of 107,888 executables, collected by VirusShare from
Nov/2010 to Jul/2014. -->\n\t\tclass="normal">Multivariate, Time-Series 
>\n\t\tClassification, Regression \n\t\t\tInteger&
nbsp;   nt/t  p class = "normal" > 107888 & nbsp;   (nt/t  p class = "normal" > 482 & nbsp;  
p  n t < d < p class = "normal" > 2017 & nbsp;  n t  Computer & nbsp; 
p> -->\n\t\t\\n\t\t\t\n\t\t\t<a href="datasets/IDA2016Challenge"><img src="assets/ML">
images/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/IDA20"
16Challenge">IDA2016Challenge</a></b>\n\t\t<!-- <td>The dat
aset consists of data collected from heavy Scania trucks in everyday usage.   -->\n\t\t\t<p cl
>Integer \n\t\t\t76000 \n\t\t\t<p cl
ass="normal">171 \n\t\t\t2017 \n\t\t\t<!-- <td>2017&nbsp;
ormal">Computer  -->\n\t\t\\n\t\t\t<a href="data"
sets/DSRC+Vehicle+Communications"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&n
bsp;<b><a href="datasets/DSRC+Vehicle+Communications">DSRC Vehicle Comm
unications</a></b>h\t\t\t<!-- <td>This set Provides data regardin
g wireless communications between vehicles and road side units. two separate data sets are provided (normal
scenario) and in the presence of attacker (jammer).  -->\n\t\t\tSequential
, Text \n\t\tClustering \n\t\tRe
al\ /n\t\t10000\ /n\t\t5\ 
>\n\t\t\t2017 \n\t\t\t<!-- <td>Computer&nbsp;</p
>-->\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\t\n\
ges"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<
b><a href="datasets/Mturk+User-Perceived+Clusters+over+Images">Mturk User-Perceived Clusters over Imag
es</a></b>
ng Wu and DataLab members at NTHU, Taiwan. There\'re 325 user-perceived clusters from 100 users and thei
r corresponding descriptions.  -->\n\t\t\tMultivariate, Text 
>\n\t\tClustering \n\t\tInteger \
```

```
class="normal">2016\ \n\t\t<!--<td>Computer\&nbsp; -->\n\t\t\t
\n\t\t<a href="datasets/Character+Font+Images"><img src="asset"><a href="datasets/Character+Font+Images"><a href="datasets/Character+Font+Images"><a href="datasets/Character+Font+Images"><a href="datasets/Character+Font+Images"><a href="datasets/Character+Font+Images"><a href="datasets/Character+Font+Images"><a href="datasets/Character+Font+Images"><ta
s/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/"
Character+Font+Images">Character Font Images</a></b>\n\t\t\t<!-- <td><p class="
normal">Character images from scanned and computer generated fonts.  -->\n\t\t<p class
="normal">Multivariate \n\t\tClassification \n\t\t\t<p
class="normal">Integer, Real \n\t\t\t745000 \n\t\t\t<
p class="normal">411 \n\t\t\t2016 <math>\n\t\t\t<!--<td>2016&nbsp;<math>
s="normal">Computer  -->\n\t\t\r\t\t<a href="datasets/DeliciousM"
IL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Instance+Labels"><img src="assets/MLima"
ges/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Delicious"
MIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Instance+Labels">DeliciousMIL: A Data S
et for Multi-Label Multi-Instance Learning with Instance Labels</a></b>
>This dataset includes 1) 12234 documents (8251 training, 3983 test) extracted from Delicio
usT140 dataset, 2) class labels for all documents, 3) labels for a subset of sentences of the test documents.&n
bsp;-->\n\t\t\tText \n\t\t\tClassification&
nbsp;   n\t  p\ class = "normal" > lnteger & nbsp;   n\t  p\ class = "normal" > 12234 & nbsp;
/n\t\t8519 /n\t\t\t2016 
\n\t\t<!-- <td>Computer&nbsp; -->\n\t\t\t\n\t\t\t<ta
ble><a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++"><img src="assets"
/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/A"
utistic+Spectrum+Disorder+Screening+Data+for+Children++">Autistic Spectrum Disorder Screening Data for Ch
ildren </a></b>\n\t\t<!-- <td>Children screening data for autism
suitable for classification and predictive tasks   -->\n\t\t\tMultivariate&nbs
p;\n\t\tClassification \n\t\t\tInteger&nb
sp;\n\t\t292 \n\t\t21 \n\t\t21 
n\t\t\t2017 \n\t\t\t<!-- <td>Life&nbsp; -->\n\t\
t\t\n\t\t\t<a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adole">t\t<a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adole">t\t<a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adole">t\t<a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adole">t\t</a>
scent+++"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="norm"
al"><b><a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++">Autistic Spectrum
Disorder Screening Data for Adolescent </a></b>\r\t\t\t<!-- <td>
Autistic Spectrum Disorder Screening Data for Adolescent. This dataset is related to classification and predictiv
e tasks.  -->\n\t\t\tMultivariate \n\t\t\t<p class="normal"
al">Classification \n\t\t\Integer \n\t\t\t
mal">104\ \n\t\t\t21\ \n\t\t\t2017\&n line (line for the line fo
bsp;   + t < -- < td> -- > h < t < -- > h <
edefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/APS+Failure+at+Scania"><b>
+Trucks">APS Failure at Scania Trucks</a></b>\n\t\t<!-- <td>T
he datasets\' positive class consists of component failures for a specific component of the APS system. The neg
ative class consists of trucks with failures for components not related to the APS.  -->\n\t\t\t
Multivariate \n\t\tClassification \n\t\t\
tInteger, Real \n\t\t60000 \n\t\t\
p class="normal">Computer  -->\n\t\t\\n\t\t\t<a href="datasets/Wirel">this class="normal">Computer&nbsp;
<b><a href="datasets/Wireless+Indoor+Localization">Wireless Indoor Localization</a></b><
/p>\n\t\t\<!-- <td>Collected in indoor space by observing signal stren
gths of seven WiFi signals visible on a smartphone. The decision variable is one of the four rooms.  
-->\n\t\t\tMultivariate \n\t\t\tClassification&n
bsp;\n\t\t\tReal \n\t\t2000 
\label{thm:local_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_control_con
Computer  -->\n\t\t\t\n\t\t\t<t
d><a href="datasets/HCC+Survival"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp
;<b><a href="datasets/HCC+Survival">HCC Survival</a></b></table
>\n\t\t<!-- <td>Hepatocellular Carcinoma dataset (HCC dataset) was collected at a Uni
versity Hospital in Portugal. It contains real clinical data of 165 patients diagnosed with HCC. 
>\n\t\tMultivariate \n\t\t\tClassification </p
>\n\t\tInteger, Real \n\t\t\t165 <
\label{local-condition} $$ \frac{\t\class=\normal">49\nbsp;\n\t\t\class=\normal">2017\nbsp;\n\t\t\<<!--
- class="normal">Life  -->\n\t\t\t\n\t\t\t<a href="datasets/CS"
M_%28Conventional_and_Social_Media_Movies%29_Dataset_2014_and_2015"~ima_src="assets/MI images/
```

```
SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/CSM+%28Co"
nventional+and+Social+Media+Movies%29+Dataset+2014+and+2015">CSM (Conventional and Social Media M
ovies) Dataset 2014 and 2015</a></b>\n\t\t<!-- <td>12 features
categorized as conventional and social media features. Both conventional features, collected from movies data
bases on Web as well as social media features(YouTube,Twitter).  -->\n\t\t\t<p class="norm
al">Multivariate \n\t\t\tClassification, Regression \n\t\t\t<t
d>Integer \n\t\t\t217 \n\t\t\t217 
ss="normal">12 \n\t\t\t2017 \n\t\t\t<!-- <td>10
mal">Computer  -->\n\t\t\\n\t\t\<a href="datase"><a hr
ts/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29"><img src="assets/MLimages/SmallLar"
gedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/University+of+Tehran+">theran+" border=1 /></a>
Question+Dataset+2016+%28UTQD.2016%29">University of Tehran Question Dataset 2016 (UTQD.2016)</a>
ame broadcasted on Iranian national television.   -->\n\t\t\tText </
p>\n\t\t\tClassification \n\t\t\t 
\ \\n\t\t\t3 \\n\t\t\t3 \\n\t\t\t<
p class="normal">2017 \n\t\t\<!-- <td>Other&nbsp; -->\n\t\t\
r>\n\t\t<a href="datasets/Autism+Screening+Adult"><img src="assets/MLimages/SmallLar"><ntd><a href="datasets/Autism+Screening+Adult"><a href="datasets/Autism+Adult"><a href="datasets/Autism+Adult"><a href="datasets/Autism+Adult"><a href="datasets/Autism+Adult"><a href="datasets/Autism+Adult"><a href="datasets/Autism+Adult"><a href="datasets/Autism+Adult"><a href="datasets/Autism+Adult"><a href="datasets/Autism+Adult">datasets/Autism+Adult</a><a href="datasets/Autism+Adult">datasets/Autism+Adult</a><a href="dat
gedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/Autism+Screening+Ad">
ult">Autism Screening Adult</a></b>\n\t\t<!-- <td>Autistic Spectr
um Disorder Screening Data for Adult. This dataset is related to classification and predictive tasks. 
td> -->\n\t\t \n\t\t\tClassification 
\n\t< d>p class="normal">Integer&nbsp;\n\t\t< d>p class="normal">704&nbsp;\n\t\t\tp class="normal">704&nbsp;
>21 \n\t\t\t2017 \n\t\t\t<!-- <td>2017&nbsp;
ss="normal">Social  -->\n\t\t\t\n\t\t\t<a href="dat
asets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor"><img src="asset"
s/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/A"
ctivity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor">Activity recognition with
class="normal">Sequential motion data from 14 healthy older people aged 66 to 86 years old using a batteryles
s, wearable sensor on top of their clothing for the recognition of activities in clinical environments. 
d> -->\n\t\t\tSequential \n\t\t\tClassification&nbsp
;\n\t\t\tReal \n\t\t\t75128 </td
\ \\n\t\t9 \\n\t\t2016 \\n\t\t<!-- <t
d>Life  -->\n\t\t<\r\\n\t\t<td><a href="datasets/Immun">Life&nbsp;
otherapy+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p cla
ss="normal"><b><a href="datasets/Immunotherapy+Dataset">Immunotherapy Dataset</a></b>
table>\n\t\t<!-- <td>This dataset contains information about wart treatment results of 9
0 patients using immunotherapy.  -->\n\t\t\tUnivariate \n\
t/t/Classification \n/t/t/Integer, Real 
p class="normal">2018 \n\t\t\<!-- <td>Life&nbsp; -->\n\t\t\
gcolor="DDEEFF">\n\t\t\t<a href="datasets/Cryotherapy+Dataset+"><img src="assets/MLim
ages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Cryother"
apy+Dataset+">Cryotherapy Dataset </a></b>\n\t\t\t<!-- <td>This
dataset contains information about wart treatment results of 90 patients using cryotherapy. 
nttttUnivariate \ntttClassification </
td>\n\t\tInteger, Real \n\t\t90 \
n/t/tcp class="normal">7 \n/t/tcp class="normal">2018 \n/t/t/t<!-- <td>1&nbsp;\n/t/t<!-- <td>1&nbsp;\n/t/t<|-- <td>1&nbsp;
>Life  -->\n\t\t\\n\t\t\t<a href="datasets/OCT+da"
ta+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes"><img src="assets/MLimages/SmallLargedefault.jpg"
"border=1 /></a>&nbsp;<b><a href="datasets/OCT+data+%26+Color+Fundus+Im"
ages+of+Left+%26+Right+Eyes">OCT data & Color Fundus Images of Left & Right Eyes</a></b>
\n\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.  -->\n\t\t\t
Multivariate \n\t\t\tClassification \n\t\t\tClassification 
mal">Real \n\t\t50 \n\t\t\t2&nbs
p;\n\t\t2016 \n\t\t\t<!-- <td>Computer&nbs
p; -->\n\t\t\t\n\t\t\t<a href="datasets/Discrete+Tone+I">t
mage+Dataset"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class=
"normal"><b><a href="datasets/Discrete+Tone+Image+Dataset">Discrete Tone Image Dataset</a></b></t
                 the an alone "normal". Discrete Tana Imagos (DTI) are
                                                                                              available which a
```

```
u></ti><ti>ti></ti><ti>tiable></ti><ti>tiu></ti><ti>tiu></ti><ti>tiable willow heads to be
e analyzed in detail. Here, we created this dataset for those who do research in DTI. \r\n  -->\n\t
\t\tMultivariate \n\t\tClassification </t
d>n\t\t< d>p class="normal">&nbsp;\n\t\t\t< d>p class="normal">71&nbsp;\n\t\t\t< d>p c
ormal">Computer  -->\n\t\t\\n\t\t\<a href="datasets/News+Populari">Computer&nbsp;
ty+in+Multiple+Social+Media+Platforms"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&
nbsp;<b><a href="datasets/News+Popularity+in+Multiple+Social+Media+Platforms""
>News Popularity in Multiple Social Media Platforms</a></b>\n\t\t\t<!-- <td><p class=
"normal">Large data set of news items and their respective social feedback on multiple platforms: Facebook, G
oogle+ and LinkedIn.  -->\n\t\t\tMultivariate, Time-Series, Text </p
>\n\t\t\tRegression \n\t\t\tInteger, Real&nbs
p;n\t\t93239 n\t\t\t11 
>\n\t\t2018 \n\t\t<!-- <td>Computer&nbsp;
> -->\n\t\t bgcolor="DDEEFF">\n\t\t\t<a href="datasets/Ultrasonic+flowmeter+diagn"> -->\n\t\t\t</a>
ostics"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
><b><a href="datasets/Ultrasonic+flowmeter+diagnostics">Ultrasonic flowmeter diagnostics</a></b>
tr>\n\t\t<!-- <td>Fault diagnosis of four liquid ultrasonic flowmeters&nbsp;
-->\n\t\t\tMultivariate \n\t\t\tClassification&n
bsp;\n\t\t\t540 \n\t\t540 
d>n\t\tp class="normal">173 n\t\t\tp class="normal">2018 
- class="normal">Computer  -->\n\t\t\t\n\t\t\t<a href="datase"><a h
ts/ICMLA+2014+Accepted+Papers+Data+Set"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 />
</a>&nbsp;<b><a href="datasets/ICMLA+2014+Accepted+Papers+Data+Set">ICMLA+2014+Accepted+Papers+Data+Set">ICMLA+2014+Accepted+Papers+Data+Set">ICMLA+2014+Accepted+Papers+Data+Set">ICMLA+2014+Accepted+Papers+Data+Set">ICMLA+2014+Accepted+Papers+Data+Set">ICMLA+2014+Accepted+Papers+Data+Set">ICMLA+2014+Accepted+Papers+Data+Set">ICMLA+2014+Accepted+Papers+Data+Set</a>
LA 2014 Accepted Papers Data Set</a></b>\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.  -->\n\t\t\t<p clas
s="normal">Multivariate \n\t\tClassification, Clustering \
n/t/tc class="normal"> \n/t/tc class="normal">105 \n/t/tc class="normal">
ss="normal">5\ \n\t\t\2018\ \n\t\t\<!-- <td>2018\&nbsp;
mal">Other  -->\n\t\t\\n\t\t\<a href="datasets/B"
LE+RSSI+Dataset+for+Indoor+localization+and+Navigation"><img src="assets/MLimages/SmallLargedefault.jp"
g" border=1 /></a>&nbsp;<b><a href="datasets/BLE+RSSI+Dataset+for+Indoor+lo"
calization+and+Navigation">BLE RSSI Dataset for Indoor localization and Navigation</a></b>
ble>\n\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.  -->\n\t\t\tMultivariate, Sequential, Time-Series </p
>\n\t\tClassification, Clustering \n\t\tInteg
er\ \n\t\t6611\ \n\t\t\t15\ 
>\n\t\t\2018 \n\t\t\<!-- <td>Computer&nbsp;</p
>-->\n\t\t\d><a href="datasets/Container+Crane+Controller+Data+Set"><i
mg src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a hr
ef="datasets/Container+Crane+Controller+Data+Set">Container Crane Controller Data Set</a></b>
r>\n\t\t\t<!-- <td>A container crane has the function of transporting containers fr
om one point to another point.  -->\n\t\t\tUnivariate, Domain-Theory&nb
sp;\n\t\tClassification, Regression \n\t\t<p class="normal"
">Real \n\t\t\t15 \n\t\t\t3 </
p  n t < d < p class = "normal" > 2018 & nbsp;  n t  Computer & nbsp; 
p> -->\n\t\t\c/tr>\n\t\t\c/td><a href="datasets/Residential+Building+">+Building+</a>
Data+Set"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="norm"
al"><b><a href="datasets/Residential+Building+Data+Set">Residential Building Data Set</a></b>
\n\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; -->\n\t\t\tMultivariate \n\t\tRegress
ion \n\t\t\t372 \n\t\t\t372 
atasets/Health+News+in+Twitter"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;</
td><b><a href="datasets/Health+News+in+Twitter">Health News in Twitter</a></b>
\n\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;-->\n\t\t\tclass="normal">Text \n\t\t\tp class="normal">Clustering&nb
```

```
$p;\n\t\\\\>6000 
\t<!-- <td>Computer&nbsp; -->\n\t\t\t bgcolor="DDEEFF">\n\t\t\t
<a href="datasets/chipseq"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp
;<b><a href="datasets/chipseq">chipseq</a></b>
<!-- <td>ChIP-seq experiments characterize protein modifications or binding at\r\nspecific ge
nomic locations in specific samples. The machine learning\r\nproblem in these data is structured binary classific
ation. -->\n\t\t\Sequential \n\t\t\t
Classification \  \n\t\t\tInteger \  \n\t\t\t
>4960%nbsp;\n\t\t%nbsp;<math>\n\t\t2018%nbsp;
/p>\n/t/t<!--<td>Life&nbsp;-->\n/t/t
ef="datasets/SGEMM+GPU+kernel+performance"><img src="assets/MLimages/SmallLargedefault.jpg" border=
1 /></a>&nbsp;<b><a href="datasets/SGEMM+GPU+kernel+performance">SGEM
M GPU kernel performance</a></b>\n\t\t\t<!-- <td>Running times
for multiplying two 2048 x 2048 matrices using a GPU OpenCL SGEMM kernel with varying parameters (using t
he library \'CLTune\'). -->\n\t\t\tMultivariate \n\t\t\t<
p class="normal">Regression \n\t\tInteger \n\t\t<p
class="normal">241600 \n\t\t\t18 \n\t\t\t<p class="no
rmal">2018\ \n\t\t\<!-- <td>Computer&nbsp; --> \n\t\t\
or="DDEEFF">\n\t\t\<a href="datasets/Repeat+Consumption+Matrices"><img src="assets/
MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Re
peat+Consumption+Matrices">Repeat Consumption Matrices</a></b>\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. 
-->\n\t\tMultivariate \n\t\tClustering 
/n\t\tReal /n\t\t130000 /n\t\t
t/t/tclass="normal">21000 \n/t/t/tclass="normal">2018 \n/t/t/t<!--
ctd>class="normal">Computer  -->\n\t\t\t\n\t\t\tctable>ca href="datasets"
/detection_of_loT_botnet_attacks_N_BaloT"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></
a> <b><a href="datasets/detection_of_loT_botnet_attacks_N_BaloT">detec
tion_of_loT_botnet_attacks_N_BaloT</a></b>\n\t\t\t<!-- <td>This
dataset addresses the lack of public botnet datasets, especially for the IoT. It suggests *real* traffic data, gather
ed from 9 commercial IoT devices authentically infected by Mirai and BASHLITE.  -->\n\t\t\t
Multivariate, Sequential \n\t\tClassification, Clusteri
ng \n\t\tReal \n\t\t7062606&n
bsp;  \ln t  p class = "normal" > 115 & nbsp;  \ln t  p class = "normal" > 2018 & nbsp; 
td>\n\t\t<!-- <td>Computer&nbsp; -->\n\t\t\t\n\t\t\t
<a href="datasets/Absenteeism+at+work"><img src="assets/MLimages/SmallLargedefault.jpg"
border=1 /></a>&nbsp;<b><a href="datasets/Absenteeism+at+work">Absenteeism
at work</a></b>\n\t\t<!-- <td>The database was created with re
cords of absenteeism at work from July 2007 to July 2010 at a courier company in Brazil. 
-->\n\
t\t\tMultivariate, Time-Series \n\t\tClassification,
Clustering \n\t\tInteger, Real \n\t\t\t
al">740\ \n\t\t\t21\ \n\t\t\t2018 
sp;--> \ln t Business --> \ln t 
<a href="datasets/SCADI"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;</td
>class="normal"><b><a href="datasets/SCADI">SCADI</a></b>\n\t\t\t<!-- <td>!-- <td
>First self-care activities dataset based on ICF-CY.  -->\n\t\t<p class="
normal">Multivariate \n\t\tClassification, Clustering \n\t\
t\t \n\t\t\t70 \n\t\t\t70 
normal">206 \n\t\t\t2018 \n\t\t\t<!-- <td><p class="norma
l">Life  -->\n\t\t\t\n\t\t\t<a href="datasets/Conditi">to datasets/Conditi</a>
on+monitoring+of+hydraulic+systems"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nb
sp;<b><a href="datasets/Condition+monitoring+of+hydraulic+systems">Condition
monitoring of hydraulic systems</a></b>\n\t\t<!-- <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.  -->\n\t\t\t<p cl
ass="normal">Multivariate, Time-Series \n\t\tClassification, Regression
 \n\t\tReal \n\t\t2205 </p
>n/t/tp class="normal">43680 n/t/t/p class="normal">2018 n
\t\t\t<!-- <td>Computer&nbsp; -->\n\t\t\t\r\t\t\t<a href="
datasets/Carbon+Nanotubes"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
```

```
<b><a href="datasets/Garbon+Nanotubes">Garbon Nanotubes</a></b>
table>\n\t\t\t<!-- <td>This dataset contains 10721 initial and calculated atomic coordinat
es of carbon nanotubes.  -->\n\t\t\tUnivariate \n\t\t\t
Regression \n\t\t\tReal \n\t\t\t<p</p>
class="normal">10721\ 
<math display="block">class="normal">8\&nbsp;
<math display="block">class="normal">8\&nbsp;<
mal">2018\ \n\t\t<!--<td>p class="normal">Computer\&nbsp; -->\n\t\t\t
r="DDEEFF">\n\t\t\t<a href="datasets/Optical+Interconnection+Network+"><img src="asset"
s/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/"
Optical+Interconnection+Network+">Optical Interconnection Network </a></b>
<!-- <td>This dataset contains 640 performance measurements from a simulation of 2-Dime
nsional Multiprocessor Optical Interconnection Network.   -->\n\t\t\tMultiv
ariate \n\t\tClassification, Regression \n\t\t\t<p clas
s="normal">Integer, Real \n\t\t\t640 \n\t\t\t<p class=
"normal">10\ \n\t\t\t2018\ \n\t\t\t<!-- <td>10&nbsp;
">Computer  -->\n\t\t\n\t\t<a href="datasets/Sports+articles+for+"
objectivity+analysis"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p cl
ass="normal"><b><a href="datasets/Sports+articles+for+objectivity+analysis">Sports articles for objectivity anal
ysis</a></b>
g Amazon Mechanical Turk as objective or subjective. The raw texts, extracted features, and the URLs from wh
ich the articles were retrieved are provided.  -->\n\t\t\tMultivariate, Text&
nbsp;\n\t\tClassification \n\t\t\tInteger
\alpha = \frac{1000 \cdot 1000 \cdot 10000 \cdot 1000 \cdot 
\n\t\t2018 \n\t\t\t<!-- <td>Social&nbsp;
> -->\n\t\t\n\t\t<a href="datasets/Breast+Cancer+Coimbra"><
img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a h
ref="datasets/Breast+Cancer+Coimbra">Breast Cancer Coimbra</a></b>
td>Clinical features were observed or measured for 64 patients with breast cancer and 52 h
ealthy controls.   -->\n\t\t\tMultivariate \n\t\t\t<p class
="normal">Classification \n\t\t\tInteger \n\t\t\t<p clas
s="normal">116 \n\t\t10 \n\t\t2
018\    + td>Life\    --> + td>Life\   
r><a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data"><img src="assets/MLimages/S
mallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/GNFUV+Unma"
nned+Surface+Vehicles+Sensor+Data">GNFUV Unmanned Surface Vehicles Sensor Data</a></b>
r>\n\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).   -->\n\t\tMultivariate, Time-Series 
\n\t\tRegression \n\t\tReal 
\ \\n\t\t\t5 \\n\t\t\t5 \\n\t\t\t<
p class="normal">2018 \n\t\t<!-- <td>p class="normal">Computer&nbsp; -->\n\t\t\t</t
r>\n\t\t<a href="datasets/Dishonest+Internet+users+Dataset"><im
g src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href
="datasets/Dishonest+Internet+users+Dataset">Dishonest Internet users Dataset</a></b>
>\n\t\t<!-- <td>The dataset was used to test an architecture based on a trust model ca
pable to cope with the evaluation of the trustworthiness of users interacting in pervasive environments. </
p  --> /n /t /t  class = "normal" > Multivariate & nbsp;   /n /t /t  class = "normal" > Classification | figure | f
, Clustering \n\t\t\t \n\t\t\t322&nb
sp;\n\t\t5 \n\t\t\t2018 \n\t\t\t2018 \n\t\t\t2018 
n\t\t\t<!-- <td>class="normal">Computer&nbsp; -->\n\t\t\t\n\t\t\t<a href=
"datasets/Victorian+Era+Authorship+Attribution"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /
></a>&nbsp;<b><a href="datasets/Victorian+Era+Authorship+Attribution">Victorian
Era Authorship Attribution</a></b>\n\t\t\t<!-- <td>create the la
rgest authorship attribution dataset, we extracted works of 50 well-known authors. To have a non-exhaustive le
arning, in training there are 45 authors whereas, in the testing, it\'s 50  -->\n\t\t\t<p class="n
ormal">Text \n\t\tClassification \n\t\t\t
mal">\ /nt/tt93600 /nt/tt/1000 
bsp;\n\t\t\tclass="normal">2018 \n\t\t\t<!-- <td>class="normal">Computer&n
bsp; -->\n\t\t\\n\t\t\<a href="datasets/Simulated+Falls">\n\t\t</d>
+and+Daily+Living+Activities+Data+Set"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&
nbsp;<b><a href="datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set"><b</a>
">Simulated Falls and Daily Living Activities Data Set</a></b>\n\t\t<!-- <td><p class
="normal">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. 
n\t\t\tTime-Series \n\t\t\tClassification 
n\t\tInteger \n\t\t3060 \n\t\t3060 
>Life  -->\n\t\t\\n\t\t\t<a href="datasets/Multimo"><
dal+Damage+Identification+for+Humanitarian+Computing"><img src="assets/MLimages/SmallLargedefault.jpg"
border=1 /></a>&nbsp;<b><a href="datasets/Multimodal+Damage+Identification+f"><b><a href="datasets/Multimodal+Damage+Identification+f"><b href="datasets/Mul
or+Humanitarian+Computing">Multimodal Damage Identification for Humanitarian Computing</a></b>
>\n\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, Infr
astructural, Human, Non-damage. -->\n\t\t\tMultivariate, Text </p
>\n\t\tClassification \n\t\tInteger 
p  n t t < d < p class = "normal" > 5879 nbsp;  n t t  8 nbsp;   n t t  p class = "normal" > 8 nbsp;   n t \ t  p class = "normal" > 8 nbsp;   n t \ t  p class = "normal" > 8 nbsp;   n t \ t  p class = "normal" > 8 nbsp;   n t \ t  p class = "normal" > 8 nbsp;   n t \ t  p class = "normal" > 8 nbsp;   n t \ t  p class = "normal" > 8 nbsp;   n t \ t  p class = "normal" > 8 nbsp;   n t \ t  p class = "normal" > 8 nbsp;   n t \ t  p class = "normal" > 8 nbsp;   n t \ t  p class = "normal" > 8 nbsp;   n t \ t  p class = "normal" > 8 nbsp;   n t \ t  p class = "normal" > 8 nbsp;   n t \ t  p class = "normal" > 8 nbsp;   n t \ t  p class = "normal" > 8 nbsp;   n t \ t  n t \ t  p class = "normal" > 8 nbsp;  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t  n t 
td>2018 \n\t\t\<!-- <td>Social&nbsp; -->\n\t\t\</
tr>\n\t\t\t<a href="datasets/EEG+Steady-State+Visual+Evoked+Pote">tr><a href="datasets/EEG+State+Visual+Evoked+Pote">tr><a href="datasets/EEG+State+Visual+Evoked+Pote">tr><a href="datasets/EEG+State+Visual+Evoked+Pote">tr><a href="datasets/EEG+State+Visual+Evoked+Pote">tr><a href="datasets/EEG+State+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+Visual+Evoked+Pote+
ntial+Signals"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="n
ormal"><b><a href="datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals">EEG Steady-State Visual
Evoked Potential Signals</a></b>\n\t\t\t<!-- <td>This database co
nsists on 30 subjects performing Brain Computer Interface for Steady State Visual Evoked Potentials (BCI-SSV
EP).   -->\n\t\t\tMultivariate, Time-Series \n\t\t\t<p cl
ass="normal">Classification, Regression \n\t\tInteger \
n/t/tp class="normal">9200 \n/t/t/tp class="normal">16 \n/t/t/t16 
class="normal">2018\ \n\t\t\<!-- <td>Life&nbsp; -->\n\t\t\
t/t/t<a href="datasets/Roman+Urdu+Data+Set"><img src="assets/MLimages/SmallLargedef"><a href="datasets/Roman+Urdu+Data+Set"><img src="assets/MLimages/SmallLargedef"><a href="datasets/Roman+Urdu+Data+Set"><a href="data
ault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Roman+Urdu+Data+Set">R
oman Urdu Data Set</a></b>\n\t\t\t<!-- <td>Roman Urdu (the scr
ipting style for Urdu language) is one of the limited resource languages. A data corpus comprising of more than
20000 records was collected.  -->\n\t\t\tText \n\t\t\t<
p class="normal">Classification \n\t\t\t \n\t\t\t
s="normal">20000\ \n\t\t\t2\ \n\t\t\t
2018 \n\t\t<!-- <td>Computer&nbsp; -->\n\t\t<tr bgcolor="DD"
EEFF">\n\t\t<a href="datasets/Avila"><img src="assets/MLimages/SmallLargedefault.jpg"
border=1 /></a>&nbsp;<b><a href="datasets/Avila">Avila</a></b>
able>\n\t\t\<!-- <td>The Avila data set has been extracted from 800 images of the \'Avil
a Bible\', an XII century giant Latin copy of the Bible. The prediction task consists in associating each pattern to
a copyist.  -->\n\t\t\Multivariate \n\t\t\
mal">Classification \n\t\tReal \n\t\t\tReal 
al">20867\ \n\t\t\t10\ \n\t\t\t2018\ 
nbsp;   + nbsp;  
tr><a href="datasets/PANDOR"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp
;<b><a href="datasets/PANDOR">PANDOR</a></b>\n\t
\t\t<!-- <td>PANDOR is a novel and publicly available dataset for online recommendation pro
vided by Purch (http://www.purch.com/).   -->\n\t\t\tMultivariate </
p>\n\t\t\tRecommendation \n\t\t\tCategorica
l\ /n\t\t /n\t\t\t /n
t/t/class="normal">2018\ \n/t/t/<!--<td>class="normal">Life&nbsp;--->\n/t/t/
\n\t\t\t<a href="datasets/Drug+Review+Dataset+%28Druglib.co">tr><a href="dataset-%28Druglib.co">tr><a href="dataset-%28Druglib.co">tr><
m%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="normal"
><b><a href="datasets/Drug+Review+Dataset+%28Druglib.com%29">Drug Review Dataset (Druglib.com)</a><
/b>\n\t\t\<!-- <td>The dataset provides patient reviews on specifi
c drugs along with related conditions. Reviews and ratings are grouped into reports on the three aspects benefi
ts, side effects and overall comment.  -->\n\t\tMultivariate, Text 
/p>\n\t\tClassification, Regression, Clustering \n\t\t\t<p class=
"normal">Integer \n\t\t\t4143 \n\t\t\t<p class="norma
l">8\ \n\t\t\t2018\ \n\t\t\t<!-- <td>&nbsp;
; -->\n\t\t\n\t\t<a href="datasets/Drug+Review+Dataset+%28Drugs.com"
%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
<b><a href="datasets/Drug+Review+Dataset+%28Drugs.com%29">Drug Review Dataset (Drugs.com)</a></b>
\n\t\t\<!-- <td>The dataset provides patient reviews on specific dr
ugs along with related conditions and a 10 star patient rating reflecting overall patient satisfaction. 
d> -->\n\t\t\tMultivariate, Text \n\t\t\tClassificatio
n, Regression, Clustering \n\t\tInteger \n\t\t<p cla
```

```
ss="normal">215063\ \n\t\t\t6\ \n\t\t\t6 \n\t\t\t6 
l">2018\ \n\t\t<!-- <td>Life&nbsp; --> \n\t\t bgcolor="DDEE" | l">2018&nbsp;
FF">\n\t\t\t<a href="datasets/Physical+Unclonable+Functions"><img src="assets/MLimages"
/SmallLargedefault.jpg" border=1 /></a>&nbsp;<pclass="normal"><b><a href="datasets/Physical+Unc
lonable+Functions">Physical Unclonable Functions</a></b>\n\t\t\t<!-- <td><p class="
normal">The dataset is generated from Physical Unclonable Functions (PUFs) simulation, specifically XOR Arbi
ter PUFs. PUFs are used for authentication purposes. For more info, refer to our paper below. 
-->\n\t\t\ctd>Multivariate \n\t\t\ctd>Classification <
/p  n \times d > n
/n\t\t129 /n\t\t\t2018 /n\t\t\t2018 /n\t\t\t2018 /n\t\t\t2018 /n\t\t\t2018 /n\t\t\t2018 /n\t\t\t2018 /n\t\t\t2018 /n\t\t\t2018 /n\t\t\t\t2018 /n\t\t\t\t2018 /n\t\t\t\t2018 /n\t\t\t\t2018 /n\t\t\t\t2018 /n\t\t\t\t2018 /n\t\t\t\t2018 /n\t\t\t\t2018 /n\t\t\t\t2018 /n\t\t\t\t\t2018 /n\t\t\t\t\t2018 /n\t\t\t\t\t2018 /n\t\t\t\t\t2018 /n\t\t\t\t\t2018 /n\t\t\t\t\t\t2018 /n\t\t\t\t\t\t\t2018 
<!-- <td>Computer&nbsp; -->\n\t\t\d>\n\t\t\t<a href="data"
sets/Superconductivty+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;
td><b><a href="datasets/Superconductivty+Data">Superconductivty Data</a></b>
\n\t\t<!-- <td>Two file s contain data on 21263 superconductors and their
relevant features.  -->\n\t\t\tMultivariate \n\t\t\t<p cla
ss="normal">Regression \n\t\t\tReal \n\t\t\t<p class=
"normal">21263\ \n\t\t81\ \n\t\t\t
2018 \n\t\t<!-- <td>Physical&nbsp; -->\n\t\t
EFF">\n\t\t\<a href="datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29">
<img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a
href="datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29">WESAD (Wearable Stress and Affect+Detection%29">WESAD (Wearable Stress and Affect+Detection%29">WESAD (Wearable Stress and Affect+Detection%29")
t Detection)</a></b>\n\t\t\t<!-- <td>class="normal">WESAD (Wearable Stress and
Affect Detection) contains data of 15 subjects during a stress-affect lab study, while wearing physiological and
motion sensors.  -->\n\t\t\tMultivariate, Time-Series \n\t\t
\tClassification, Regression \n\t\tReal </p
>ntt\tp class="normal">63000000 nt\t\tp class="normal">12 
ata+Set+2"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="nor
mal"><b><a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2">GNFUV Unmanned
Surface Vehicles Sensor Data Set 2</a></b>\n\t\t<!-- <td>The d
ata-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.  -->\n\t\t\t<t
d>Multivariate, Sequential, Time-Series \n\t\tRegre
ssion \n\t\tReal \n\t\t10190&n
bsp;/nt/tp class="normal">6 /nt/tp class="normal">2018 
>\n\t\t<!-- <td>Computer&nbsp; -->\n\t\t\c\tr>\n\t\t\c\td><t
able><a href="datasets/Student+Academics+Performance"><img src="assets/MLimages/SmallLargede"
fault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Student+Academics+Perfor"
mance">Student Academics Performance</a></b>\n\t\t<!-- <td>
The dataset tried to find the end semester percentage prediction based on different social, economic and acad
emic attributes.   -->\n\t\tMultivariate \n\t\t<p clas
s="normal">Classification \n\t\t\ \n\t\t\
mal">300\ \n\t\t22\ \n\t\t\t2018\ 
bsp;  \ln t < -- < td > --  \ln t  --  \left > --  \left > \le
r><a href="datasets/Online+Shoppers+Purchasing+Intention+Dataset"><img src="assets/MLimages/SmallL
argedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Online+Shoppers+P"><a href="datasets
urchasing+Intention+Dataset">Online Shoppers Purchasing Intention Dataset</a></b>
d>\n\t\t<!-- <td>Of the 12,330 sessions in the dataset,\r\n84.5% (10,422) were negative cla
ss samples that did not\r\nend with shopping, and the rest (1908) were positive class\r\nsamples ending with sh
opping.  -->\n\t\t\ctd>Multivariate \n\t\t\ctd><p class="normal"
al">Classification, Clustering \n\t\t\tInteger, Real \n\t\t\t<t
d>12330 \n\t\t18 \n\t\t18 
="normal">2018\ \n\t\t<!--<td>Business&nbsp; -->\n\t\t\ bg
color="DDEEFF">\n\t\t\t<a href="datasets/PMU-UD"><img src="assets/MLimages/SmallLar"
gedefault.jpg" border=1 /></a>&nbsp;<b><a href="datasets/PMU-UD">PMU-UD</a
></b>\n\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 Univ
ersity - Urdu/Arabic Database (PMU-UD).   -->\n\t\t\tUnivariate </p
>\n\t\t\tClassification \n\t\t\t 
\n\t\t< d>p class="normal">5180&nbsp;\n\t\t< d>p class="normal">9&nbsp;\n\t\t< d>p class="normal">9&nbsp;
class="normal">2018 \n\t\t\t<!-- <td>Computer&nbsp; -->\n\t\t\t
```

```
\n\t\t\t<a href="datasets/Parkinson%27s+Disease+Classification"><img src="assets/MLi
mages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/Parkin"
son%27s+Disease+Classification">Parkinson\'s Disease Classification</a></b></tp>
<!-- <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).  -->\n\t\t\tMulti
variate \n\t\t\tClassification \n\t\t\t<p class="normal"
>Integer, Real \n\t\t756 \n\t\t\t7
54 \n\t\t\t2018 \n\t\t\t<!-- <td>Comput
er  -->\n\t\t\\n\t\t\t<a href="datasets/Electrical+"
Grid+Stability+Simulated+Data+"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;</t
d><b><a href="datasets/Electrical+Grid+Stability+Simulated+Data+">Electrical Grid Sta
bility Simulated Data </a></b>\n\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.   -->\n\t\t\tMultivariate \n\t\t\tMultivariate 
ormal">Classification, Regression \n\t\t\Real \n\t\t\
10000 \n\t\t\t14 \n\t\t\t14 
normal">2018\ \n\t\t<!-- <td>Physical&nbsp; -->\n\t\t
t<a href="datasets/Caesarian+Section+Classification+Dataset"><img src="assets/MLimage"><a href="datasets/Caesarian+Section+Classification+Dataset"><img src="assets/MLimage"><a href="datasets/Caesarian+Section+Classification+Dataset"><a href="datasets/Caesarian+Section+Classification+Dataset"><a href="datasets/Caesarian+Section+Classification+Dataset"><a href="datasets/Caesarian+Section+Classification+Dataset"><a href="datasets/Caesarian+Section+Classification+Dataset"><a href="datasets/Caesarian+Section+Classification+Dataset"><a href="datasets/Caesarian+Section+Classification+Dataset"><a href="datasets/Caesarian+Section+Dataset"><a href="datasets/Caesarian+Section+Dataset"><a href="datasets/Caesarian+Section+Dataset"><a href="datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datasets/Caesarian+Section+Datase
s/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a href="datasets/Caesarian+
Section+Classification+Dataset">Caesarian Section Classification Dataset</a></b>
\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.   -->\n
\t\t\ctd>Univariate \n\t\t\ctd>Classification </t
d>/n\t\tlnteger /n\t\t\t80 /n\t\t\t80 /n\t\t\t80 
d>5 \n\t\t\t2018 \n\t\t\t<!-- <td>2018&nbsp;
ss="normal">Life  -->\n\t\t\t\n\t\t\t<a href="datas"
ets/BAUM-1"><img src="assets/MLimages/SmallLargedefault.ipg" border=1 /></a>&nbsp;<p class="no
rmal"><b><a href="datasets/BAUM-1">BAUM-1</a></b>
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. 
n\t\t\tTime-Series \n\t\tClassification 
/n\t\t /n\t\t\t1184 /n\t\t\t1184 
> \n\t\t2018 \n\t\t<!-- <td><p class
="normal">Computer  -->\n\t\t<\r\>\n\t\t\t<a href="datasets/BAUM-2"><
img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a h
ref="datasets/BAUM-2">BAUM-2</a></b>\n\t\t\t<!-- <td>class="normal">A multilin
gual audio-visual affective face database consisting of 1047 video clips of 286 subjects.   -->\n\t\
t\tTime-Series \n\t\tClassification </t
d>\n\t\t \n\t\t\t1047 \n\t\t\t1047 
p class="normal"> \n\t\t\t2018 \n\t\t\t<!-- <td><p class="n
ormal">Computer  -->\n\t\t\\n\t\t\<a href="data"
sets/Audit+Data"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class
="normal"><b><a href="datasets/Audit+Data">Audit Data</a></b>\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.  -->\n\t\t\t<p class
="normal">Multivariate \n\t\t\tClassification \n\t\t\t<p
class="normal">Real \n\t\t\t777 \n\t\t\t777 
mal">18 \n\t\t2018 \n\t\t\t<!-- <td>Ot
her  -->\n\t\t\t<a href="datasets/BuddyMove+Data+Set"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><a href="
datasets/BuddyMove+Data+Set">BuddyMove Data Set</a></b>\n\t\t\t<!-- <td>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  -->\n\t\t\tMultivariate, Text 
/p>\n\t\tClassification, Clustering \n\t\tRe
al\ \n\t\t\7\ \n\t\t\7\ 
td>n\tp class="normal">2018 \n\t\t\t<!-- <td>p class="normal">Other&nbsp;
-->\n\t\t\ bgcolor="DDEEFF">\n\t\t\<a href="datasets/Real+estate+valuation+data+s">+ valuation+data+s</a>
et"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b
><a href="datasets/Real+estate+valuation+data+set">Real estate valuation data set</a></b>
ble>\n\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.  
>-->\n\t\t\Multivariate&nbsp:\n\t\t\tRegression&nbsp:
```

```
/p>\n\t\tInteger, Real \n\t\t\t414 </p
>\n\t\t7 \n\t\t2018 \n\t\t\t2018 
!-- Business  -->\n\t\t\n\t\t<a href="datas"
ets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connected+speech+Data+Set+"
><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><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></b></
p>/nt/t<!-- <td>.&nbsp;--->\n\t\t\tM
ultivariate \n\t\tClassification \n\t\t\t<p class="normal"
al">Real\ \n\t\t \n\t\t 
/n\t\t2018 /n\t\t\t<!-- <td>Life&nbsp;--- Life&nbsp;--- Life&nbsp;---
->\n\t\t bgcolor="DDEEFF">\n\t\t\t<a href="datasets/Somerville+Happiness+Survey">->\n\t\t\t</a>
"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<b><
a href="datasets/Somerville+Happiness+Survey">Somerville Happiness Survey</a></b><
/td>\n\t\t<!-- <td>A data extract of a non-federal dataset posted here https://catalog.data.go
v/dataset/somerville-happiness-survey-responses-2011-2013-2015  -->\n\t\t\t<p class="nor
mal"> \n\t\t\tI
nteger \n\t\t\t7 \n\t\t\t7 
/p>n/t/tp class="normal">2018 n/t/t<!-- <td>p class="normal">Life&nbsp;
> -->\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t\n\t\t\t\t<tr
mg src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;class="normal"><b><a hr
ef="datasets/2.4+GHZ+Indoor+Channel+Measurements">2.4 GHZ Indoor Channel Measurements</a></b>
>\n\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.&
nbsp;  --> \\ lt  Multivariate & nbsp;  \\ lt  Classi = "normal" > Cla
fication \n\t\t\t7840&n
bsp;nt/tp class="normal">5 nt/t/tp class="normal">2018 
\ \n\t\t\t<!-- <td>Computer&nbsp; -->\n\t\t\t bgcolor="DDEEFF">\n\t\t\t<t
able><a href="datasets/EMG+data+for+gestures"><img src="assets/MLimages/SmallLargedefault.jpg"
border=1 /></a>&nbsp;<b><a href="datasets/EMG+data+for+gestures">EMG data
for gestures</a></b>\n\t\t\t<!-- <td>These are files of raw EMG d
ata recorded by MYO Thalmic bracelet  -->\n\t\t\tTime-Series </p
>\n\t\t\tClassification \n\t\t\tReal 
/n\tp class="normal">30000 /n\t\tp class="normal">6 /n\t\t\t<
td><p\ class="normal">2019\&nbsp;\n\t\t\<!--< td><p\ class="normal">Life\&nbsp;\-->\n\t\t\
\n\t\t\t<a href="datasets/Parking+Birmingham"><img src="assets/MLimages/SmallLarge">
default.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Parking+Birmingham">Pa
rking Birmingham</a></b></tp>\n\t\t<!-- <td>Data collected from car
parks in Birmingham that are operated by NCP from \r\nBirmingham City Council. UK Open Government Licenc
e (OGL).\r\nhttps://data.birmingham.gov.uk/dataset/birmingham-parking  -->\n\t\t\t<p class=
"normal">Multivariate, Univariate, Sequential, Time-Series \n\t\tClassifi
cation, Regression, Clustering \n\t\tReal \n\t\t<p c
lass="normal">35717\ \n\t\t\t4\ \n\t\t\t1000 class="normal">1000 class="no
al">2019 \n\t\t\<!-- <td>Computer&nbsp; -->\n\t\t\
"DDEEFF">\n\t\t\t<a href="datasets/Behavior+of+the+urban+traffic+of+the+city+of+Sao+Pa
ulo+in+Brazil"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a>&nbsp;<p class="n
ormal"><b><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></b>\n\t\t<!-- <td><p class="no
rmal">The database was created with records of behavior of the urban traffic of the city of Sao Paulo in Brazil.&
nbsp; -->\n\t\tMultivariate, Time-Series \n\t\t<p class="n
ormal">Classification, Regression \n\t\tInteger, Real \n\
t/tp class="normal">135 \n/t/tp class="normal">18 \n/t/tp class="normal">18 
tr>\n\t\t\t<a href="datasets/Travel+Reviews"><img src="assets/MLimages/SmallLargedefaul"><a href="datasets/Travel+Reviews"><img src="assets/MLimages/SmallLargedefaul"><a href="datasets/Travel+Reviews"><a href="datasets/Travel+Reviews">
t.jpg" border=1 /></a>&nbsp;<b><a href="datasets/Travel+Reviews">Travel Reviews">Travel Reviews
ws</a></b>/\table>\n\t\t\t<!-- <td>Reviews on destinations in 10 categori
es mentioned across East Asia. Each traveler rating is mapped as Excellent(4), Very Good(3), Average(2), Poo
r(1), and Terrible(0) and average rating is used.  -->\n\t\t\tMultivariate, T
ext \n\t\t\ctd>Classification, Clustering \n\t\t\ctd><p class="n
ormal">Real\ \n\t\t980\ \n\t\t\t11
 \n\t\t\t2018 <math>\n\t\t\t<!-- <td>Other&nb
```

Ratings"> ">Tarvel Review Ratings 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. -->\ n\t\t\tMultivariate, Text \n\t\t\tClassification, Clust ering \n\t\t\tReal \n\t\t\t5456&nbs $p;\n\t\t25 \n\t\t2018 \$ n/t/t<!--< td>Other -->/n/t/tRice Leaf Diseases able>\n\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. -->\n\t\t<p class ="normal">Multivariate \n\t\t\tClassification \n\t\t\t<p class="normal">Integer \n\t\t\t120 \n\t\t\t<p class="n $mputer\ --> \n\t \n \n\n\t \n \n\n\n \n\n\n\r \n\n\n\r$ cellpadding=5 align=center>\n\t\tSupported By:\n mg src="assets/nsfe.gif" height=60 /> \n In Collaboration With: \n\n\n<center>\n\n<a hre f="about.html">About || \nCitation Policy &nbs p;|| \nDonation Policy \nbsp;\n Contact &nca href="http://cml.ics.uci.edu">CML &nc/span> &nc/center> &nc/body> &nca href="http://cml.ics.uci.edu">CML &nc/span> &nc/center> &nc/body> &nca href="http://cml.ics.uci.edu">CML &nc/span> &nc/center> &nc/center> &nca href="http://cml.ics.uci.edu">CML &nc/span> &nc/center> &nca href="http://cml.ics.uci.edu">CML &nc/center> &nca href="http://cml.ics.uci.edu">CML &nc/span> &nc/center> &nca href="http://cml.ics.uci.edu">CML &nc/span> &nc/center> &nca href="http://cml.ics.uci.edu">CML &nc/span> &nc/center> &nca href="http://cml.ics.uci.edu">CML &nc/span> &nc/center> &nca href="http://cml.ics.uci.edu">CML &nc/center> &nca href="http://cml.ics.uci.edu">CML &nca href="http://cml.ics.uci.edu" $n</html>\n'$

```
In [22]:
```

}

```
soup = BeautifulSoup(html)
In [23]:
soup
Out[23]:
<!DOCTYPE HTML>
<a href="https://www.enalth.com/style="color: blue;">httml><body>"-//W3C//DTD HTML 4.01 Transitional//EN\"&gt;</a>
<title>UCI Machine Learning Repository: Data Sets</title>
<!-- Stylesheet link -->
k 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) -->
>
<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>
<span class="whitetext">
<a href="about.html">About</a>
 <a href="citation">citation Policy</a>
 <a href="donation_policy.html">Donate a Data Set</a>
 <a href="contact.html">Contact</a>
</span>
<br/>
<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"</p>
>Repository</font></span>
<input name="sitesearch" type="radio" value=""/> <span class="whitetext"><font size="1">Web</font></span>
```

<img align="middle" alt="Google" border="0" height="27" src="http://

```
www.google.com/logos/Logo_25blk.gif"/></a>
</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/>br/>
<br/>br/>
Browse Through:
 <b>Default Task</b> 
<a href="datasets.php?format=&amp;task=cla&amp;att=&amp;area=&amp;
numAtt=&numIns=&type=&sort=nameUp&view=table">Classification</a> <font color="red">
(350)</font><br/><a href="datasets.php?format=&amp;task=reg&amp;att=&amp;area=&amp;numAtt=&amp;nu
mIns=&type=&sort=nameUp&view=table">Regression</a> <font color="red">(96)</font><br/>>ca
href="datasets.php?format=&task=clu&att=&area=&numAtt=&numIns=&type=&a
mp;sort=nameUp&view=table">Clustering</a> <font color="red">(84)</font><br/><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> 
 <b>Attribute Type</b> 
<a href="datasets.php?format=&amp;task=&amp;att=cat&amp;area=&amp;
numAtt=&numIns=&type=&sort=nameUp&view=table">Categorical</a> <font color="red">(3
8)</font><br/>datasets.php?format=&amp;task=&amp;att=num&amp;area=&amp;numAtt=&amp;numl
ns=&type=&sort=nameUp&view=table">Numerical</a> <font color="red">(307)</font><br/><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> 
 <b>Data Type</b> 
<a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;nu"
mAtt=&numIns=&type=mvar&sort=nameUp&view=table">Multivariate</a> <font color="red"
>(357)</font><br/><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&a
ns=&type=uvar&sort=nameUp&view=table">Univariate</a> <font color="red">(23)</font><br/>>ca
href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=seq&a
mp;sort=nameUp&view=table">Sequential</a> <font color="red">(47)</font><br/><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><br/><a href="datasets.php?format=&amp;task=&a
mp;att=&area=&numAtt=&numIns=&type=text&sort=nameUp&view=table">Text</
a> <font color="red">(53)</font><br/><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;n
umAtt=&numIns=&type=dt&sort=nameUp&view=table">Domain-Theory</a> <font color="re
d">(23)</font><br/><a href="datasets.php?format=&amp;task=&amp;area=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAt
Ins=&type=other&sort=nameUp&view=table">Other</a> <font color="red">(21)</font><br/>
<b>Area</b>
```

```
<a href="datasets.php?format=&amp;task=&amp;att=&amp;area=life&amp;"
numAtt=&numIns=&type=&sort=nameUp&view=table">Life Sciences</a> <font color="red">
(107)</font><br/><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=phys&amp;numAtt=&amp;n
umIns=&type=&sort=nameUp&view=table">Physical Sciences</a> <font color="red">(49)</font>
<br/><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=comp&amp;numAtt=&amp;numIns=&am
p;type=&sort=nameUp&view=table">CS / Engineering</a> <font color="red">(170)</font><br/>><a href
="datasets.php?format=&task=&att=&area=soc&numAtt=&numIns=&type=&
sort=nameUp&view=table">Social Sciences</a> <font color="red">(26)</font><br/><a href="datasets.php?"
format=&task=&att=&area=bus&numAtt=&numIns=&type=&sort=nameUp&a
mp;view=table">Business</a> <font color="red">(29)</font><br/><a href="datasets.php?format=&amp;task=&a
mp;att=&area=game&numAtt=&numIns=&type=&sort=nameUp&view=table">Ga
me</a> <font color="red">(10)</font><br/><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=oth">(10)</font><br/><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=oth">(10)</font><br/><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=oth">(10)</font><br/><a href="datasets.php?format=&amp;task=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&
er&numAtt=&numIns=&type=&sort=nameUp&view=table">Other</a> <font color="red"
>(73)</font>
<b># Attributes</b> 
<a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;nu"
mAtt=less10&numIns=&type=&sort=nameUp&view=table">Less than 10</a> <font color="re
d">(113)</font><br/>ca href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=10to100&
amp;numIns=&type=&sort=nameUp&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=
&type=&sort=nameUp&view=table">Greater than 100</a> <font color="red">(84)</font> 
<b># Instances</b>
<a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=less100&amp;t
ype=&sort=nameUp&view=table">Less than 100</a> <font color="red">(27)</font><br/><a href="data"
sets.php?format=&task=&att=&area=&numAtt=&numIns=100to1000&type=&
sort=nameUp&view=table">100 to 1000</a> <font color="red">(162)</font><br/><a href="datasets.php?for
mat=&task=&att=&area=&numAtt=&numIns=greater1000&type=&sort=name
Up&view=table">Greater than 1000</a> <font color="red">(246)</font> 
<b>Format Type</b> 
<a href="datasets.php?format=mat&amp;task=&amp;att=&amp;area=&amp
;numAtt=&numIns=&type=&sort=nameUp&view=table">Matrix</a> <font color="red">(324)<
/font><br/><a href="datasets.php?format=nonmat&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numI
ns=&type=&sort=nameUp&view=table">Non-Matrix</a> <font color="red">(145)</font> 
<b>469</b> Data Sets
<font color="gray">Table View</font> <a href="datasets.php?format=&am"
p;task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list">Li
st View</a>
```

```
<a href="datasets.php?format=&amp;task=&amp;att"><a href="datasets.php.att"><a href="datasets.php.att">datasets.php.att</a></a></a>
=&area=&numAtt=&numIns=&type=&sort=nameDown&view=table"><b>Name</b
></a>
<!-- <td><b>Abstract</b> -->
<a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;
Att=&numIns=&type=&sort=typeUp&view=table"><b>Data Types</b></a>
<a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;
Att=&numIns=&type=&sort=taskUp&view=table"><b>Default Task</b></a>
<a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;at
Att=&numIns=&type=&sort=attTypeUp&view=table"><b>Attribute Types</b></a>
<a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;
Att=&numIns=&type=&sort=instUp&view=table"><b># Instances</b></a>
<a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;
Att=&numIns=&type=&sort=attUp&view=table"><b># Attributes</b></a>
<a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;num"><a href="datasets.php?format=&amp;task=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp
Att=&numIns=&type=&sort=dateUp&view=table"><b>Year</b></a>
<!-- <td><b>Area</b> -->
<a href="datasets/Abalone"><img border="1" src="assets/MLimages/SmallLarge1.jpg"/></a
> <b><a href="datasets/Abalone">Abalone</a></b>
<!-- <td>Predict the age of abalone from physical measurements&nbsp; -->
Multivariate 
 Classification 
Categorical, Integer, Real 
4177 
8 
1995 
<!-- <td>Life&nbsp; -->
<a href="datasets/Adult"><img border="1" src="assets/MLimages/SmallLarge2.jpg"/></a> <
/td><b><a href="datasets/Adult">Adult</a></b>
<!-- <td>Predict whether income exceeds $50K/yr based on census data. Also known as "C
ensus Income" dataset.  -->
Multivariate 
Classification 
Categorical, Integer 
48842 
14 
1996 
<!-- <td>Social&nbsp; -->
<a href="datasets/Annealing"><img border="1" src="assets/MLimages/SmallLargedefault.jp"
g"/></a> <b><a href="datasets/Annealing">Annealing</a></b>
e>
<!-- <td>Steel annealing data&nbsp; -->
Multivariate 
Classification 
Categorical, Integer, Real 
798 
38 

<!-- <td>Physical&nbsp; -->
<a href="datasets/Anonymous+Microsoft+Web+Data"><img border="1" src="assets/MLima"
ges/SmallLargedefault.jpg"/></a> <b><a href="datasets/Anonymous+Microsoft+W"
eb+Data">Anonymous Microsoft Web Data</a></b>
<!-- <td>Log of anonymous users of www.microsoft.com; predict areas of the web site a use
r visited based on data on other areas the user visited.  -->

 Recommender-Systems 
 <n class="normal">Categorical </n>
```

```
37711 
294 
1998 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Arrhythmia"><img border="1" src="assets/MLimages/SmallLarge5.jpg"/>
</a> <b><a href="datasets/Arrhythmia">Arrhythmia</a></b>
<!-- <td>Distinguish between the presence and absence of cardiac arrhythmia and classify it
in one of the 16 groups.  -->
Multivariate 
Classification 
Categorical, Integer, Real 
452 
279 
1998 
<!-- <td>Life&nbsp; -->
<a href="datasets/Artificial+Characters"><img border="1" src="assets/MLimages/SmallLarg"
e6.jpg"/></a> <b><a href="datasets/Artificial+Characters">Artificial Characters</a>
</b>
<!-- <td>Dataset artificially generated by using first order theory which describes structure of
ten capital letters of English alphabet  -->
Multivariate 
Classification 
Categorical, Integer, Real 
6000 
7 
1992 
<!-- <td>Computer&nbsp; -->
<tm>border="1" src="assets/MLimages/S
mallLarge7.jpg"/></a> <b><a href="datasets/Audiology+%28Original%29">Audiolo
gy (Original)</a></b>
<!-- <td>Nominal audiology dataset from Baylor&nbsp; -->
Multivariate 
Classification 
Categorical 
226 

1987 
<!-- <td>Life&nbsp; -->
<a href="datasets/Audiology+%28Standardized%29"><img border="1" src="assets/MLimag"><a href="datasets/Audiology+%28Standardized%29"><img border="1" src="assets/MLimag"><a href="datasets/Audiology+%28Standardized%29"><a href="datasets/Audiology
es/SmallLarge7.jpg"/></a> class="normal"><b><a href="datasets/Audiology+%28Standardized%29"><b><a href="datasets/Audiology+%28Standardized%29"><b><a href="datasets/Audiology+%28Standardized%29"><b><a href="datasets/Audiology+%28Standardized%29"><b><a href="datasets/Audiology+%28Standardized%29"><b><a href="datasets/Audiology+%28Standardized%29"><b><a href="datasets/Audiology+%28Standardized%29"><b><a href="datasets/Audiology+%28Standardized%29"><b h
">Audiology (Standardized)</a></b>
<!-- <td>Standardized version of the original audiology database&nbsp; -->
Multivariate 
Classification 
Categorical 
226 
69 
1992 
<!-- <td>Life&nbsp; -->
<tm, and the control of the
</a> <b><a href="datasets/Auto+MPG">Auto MPG</a></b><
/td>
<!-- <td>Revised from CMU StatLib library, data concerns city-cycle fuel consumption&nbsp;
 -->
Multivariate 
 tds on alacc-"normal"s Dograccian o/ns o/tds
```

```
<10/>
Categorical, Real 
398 
8 
1993 
<!-- <td>Other&nbsp; -->
<a href="datasets/Automobile"><img border="1" src="assets/MLimages/SmallLarge10.jpg"/
></a> <b><a href="datasets/Automobile">Automobile</a></b>
>
<!-- <td>From 1985 Ward's Automotive Yearbook&nbsp; -->
Multivariate 
Regression 
Categorical, Integer, Real 
205 
26 
1987 
<!-- <td>Other&nbsp; -->
<a href="datasets/Badges"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"
/></a> <b><a href="datasets/Badges">Badges</a></b>
<!-- <td>Badges labeled with a "+" or "-" as a function of a person's name&nbsp; --
>
Univariate, Text 
Classification 

294 
1 
1994 
<!-- <td>Other&nbsp; -->
<a href="datasets/Balance+Scale"><img border="1" src="assets/MLimages/SmallLarge12.ipg">
pg"/></a> <b><a href="datasets/Balance+Scale">Balance Scale</a></b>
<!-- <td>Balance scale weight & distance database&nbsp; -->
Multivariate 
Classification 
Categorical 
625 
4 
1994 
<!-- <td>Social&nbsp; -->
a> <b><a href="datasets/Balloons">Balloons</a></b>
<!-- <td>Data previously used in cognitive psychology experiment; 4 data sets represent diff
erent conditions of an experiment  -->
Multivariate 
Classification 
Categorical 
16 
4 

<!-- <td>Social&nbsp; -->
<a href="datasets/Breast+Cancer"><img border="1" src="assets/MLimages/SmallLarge14.j
pg"/></a> <b><a href="datasets/Breast+Cancer">Breast Cancer</a></b>
<!-- <td>Breast Cancer Data (Restricted Access)&nbsp; -->
Multivariate 
Classification 
Categorical
```

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

```
<ta>1/28 </ta>
6 
1997 
<!-- <td>Other&nbsp; -->
<a href="datasets/Census+Income"><img border="1" src="assets/MLimages/SmallLarge2.j
pg"/></a> <b><a href="datasets/Census+Income">Census Income</a></b></t
d>
<!-- <td>Predict whether income exceeds $50K/yr based on census data. Also known as "A
dult" dataset.  -->
Multivariate 
Classification 
Categorical, Integer 
48842 
14 
1996 
<!-- <td>Social&nbsp; -->
<tmp-Knight%29"><img border="1" src="asse"
ts/MLimages/SmallLarge24.jpg"/></a> <b><a href="datasets/Chess+%28King-Roo
k+vs.+King-Knight%29">Chess (King-Rook vs. King-Knight)</a></b>
<!-- <td>Knight Pin Chess End-Game Database Creator&nbsp; -->
Multivariate, Data-Generator 
Classification 
Categorical, Integer 

22 
1988 
<!-- <td>Game&nbsp; -->
<a href="datasets/Chess+%28King-Rook+vs.+King-Pawn%29"><img border="1" src="asset"
s/MLimages/SmallLarge24.jpg"/></a> class="normal"><b><a href="datasets/Chess+%28King-Rook">
+vs.+King-Pawn%29">Chess (King-Rook vs. King-Pawn)</a></b>
<!-- <td>King+Rook versus King+Pawn on a7 (usually abbreviated KRKPA7).&nbsp;
Multivariate 
Classification 
Categorical 
3196 
36 
1989 
<!-- <td>Game&nbsp; -->
<tmg%29"><img border="1" src="assets/MLi"
mages/SmallLarge24.jpg"/></a> <b><a href="datasets/Chess+%28King-Rook+vs.+"
King%29">Chess (King-Rook vs. King)</a></b>
<!-- <td>Chess Endgame Database for White King and Rook against Black King (KRK).&nbs
p: -->
Multivariate 
Classification 
Categorical, Integer 
28056 
6 
1994 
<!-- <td>Game&nbsp; -->
<a href="datasets/Chess+%28Domain+Theories%29"><img border="1" src="assets/MLima"
ges/SmallLarge24.jpg"/></a> <b><a href="datasets/Chess+%28Domain+Theories">
%29">Chess (Domain Theories)</a></b>
<!-- <td>6 different domain theories for generating legal moves of chess&nbsp; -->
Domain-Theory
```

```
<!-- <td>Game&nbsp; -->
<a href="datasets/Bach+Chorales"><img border="1" src="assets/MLimages/SmallLarge25.i
pg"/></a> class="normal"><b><a href="datasets/Bach+Chorales">Bach Chorales</a></b>
>
<!-- <td>Time-series data based on chorales; challenge is to learn generative grammar; dat
a in Lisp  -->
Univariate, Time-Series 

Categorical, Integer 
100 
6 

<!-- <td>Other&nbsp; -->
<a href="datasets/Connect-4"><img border="1" src="assets/MLimages/SmallLargedefault.j
pg"/></a> <b><a href="datasets/Connect-4">Connect-4</a></b>
ble>
<!-- <td>Contains connect-4 positions&nbsp; -->
Multivariate, Spatial 
Classification 
Categorical 
67557 
42 
1995 
<!-- <td>Game&nbsp; -->
="1" src="assets/MLimages/SmallLargedef">="1" src="assets/MLimages/SmallLargedef">=="1" src="assets/MLimages/SmallArgedef">=="1" src="assets/MLimages/SmallArgedef">=="1" src="assets/MLimages/SmallArgedef">=="1" src="assets/MLimages/SmallArgedef">=="1" src="assets/ML
ault.jpg"/></a> class="normal"><b><a href="datasets/Credit+Approval">Credit Approval</a></b></
p>
<!-- <td>This data concerns credit card applications; good mix of attributes&nbsp; -
Multivariate 
Classification 
Categorical, Integer, Real 
690 
15 

<!-- <td>Financial&nbsp; -->
<a href="datasets/Japanese+Credit+Screening"><img border="1" src="assets/MLimages/S
mallLargedefault.jpg"/></a> <b><a href="datasets/Japanese+Credit+Screening">J
apanese Credit Screening</a></b>
<!-- <td>Includes domain theory (generated by talking to Japanese domain experts); data in
Lisp  -->
Multivariate, Domain-Theory 
Classification 
Categorical, Real, Integer 
125 

1992 
<!-- <td>Financial&nbsp; -->
<a href="datasets/Computer+Hardware"><img border="1" src="assets/MLimages/SmallLar"
ge29.ipg"/></a> class="normal"><b><a href="datasets/Computer+Hardware">Computer Hardware
</a></b>
<!-- <td>Relative CPU Performance Data, described in terms of its cycle time, memory size,
etc.  -->
```

```
Multivariate 
Regression 
Integer 
209 
9 
1987 
<!-- <td>Computer&nbsp; -->
="1" src="assets/MLimages">="1" src="assets/MLimages
/SmallLargedefault.jpg"/></a> <b><a href="datasets/Contraceptive+Method+Choic"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal">
e">Contraceptive Method Choice</a></b>
<!-- <td>Dataset is a subset of the 1987 National Indonesia Contraceptive Prevalence Surve
y.  -->
Multivariate 
Classification 
Categorical, Integer 
1473 
9 
1997 
<!-- <td>Life&nbsp; -->
<tm; datasets/Covertype"><img border="1" src="assets/MLimages/SmallLarge31.jpg"/>
</a> <b><a href="datasets/Covertype">Covertype</a></b></
td>
<!-- <td>Forest CoverType dataset&nbsp; -->
Multivariate 
Classification 
Categorical, Integer 
581012 
54 
1998 
<!-- <td>Life&nbsp; -->
<a href="datasets/Cylinder+Bands"><img border="1" src="assets/MLimages/SmallLargedef"
ault.jpg"/></a> class="normal"><b><a href="datasets/Cylinder+Bands">Cylinder Bands</a></b></p
>
<!-- <td>Used in decision tree induction for mitigating process delays known as "cylinder ban
ds" in rotogravure printing  -->
Multivariate 
Classification 
Categorical, Integer, Real 
512 
39 
1995 
<!-- <td>Physical&nbsp; -->
<a href="datasets/Dermatology"><img border="1" src="assets/MLimages/SmallLargedefaul"><a href="datasets/Dermatology"><timg border="1" src="assets/MLimages/SmallLargedefaul"><a href="datasets/Dermatology"><timg border="1" src="assets/MLimages/SmallLargedefaul"><a href="datasets/Dermatology"><timg border="1" src="assets/MLimages/SmallLargedefaul"><a href="datasets/Dermatology"><timg border="1" src="assets/MLimages/SmallLargedefaul"><t
t.jpg"/></a> <b><a href="datasets/Dermatology">Dermatology</a></b>
r>
<!-- <td>Aim for this dataset is to determine the type of Eryhemato-Squamous Disease.&nbs
p; -->
Multivariate 
Classification 
Categorical, Integer 
366 
33 
1998 
<!-- <td>Life&nbsp; -->
"/></a> <b><a href="datasets/Diabetes">Diabetes</a></b>
td>
```

```
<!-- <td>This diabetes dataset is from AIM '94&nbsp; -->
Multivariate, Time-Series 

Categorical, Integer 

20 

<!-- <td>Life&nbsp; -->
+ Generation + Program > <img border = "1" src
="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/DGP2+-+T">="assets/MLimages/SmallLargedefault.jpg"/></a> 
he+Second+Data+Generation+Program">DGP2 - The Second Data Generation Program</a></b>
>
<!-- <td>Generates application domains based on specific parameters, number of features,
and proportion of positive to negative examples  -->
Data-Generator 

Real 

<!-- <td>Other&nbsp; -->
<a href="datasets/Document+Understanding"><img border="1" src="assets/MLimages/Sm
allLargedefault.jpg"/></a> <b><a href="datasets/Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding">Document+Understanding
ment Understanding</a></b>
<!-- <td>Five concepts, expressed as predicates, to be learned&nbsp; -->

1994 
<!-- <td>Other&nbsp; -->
<a href="datasets/EBL+Domain+Theories"><img border="1" src="assets/MLimages/SmallL"
argedefault.jpg"/></a> <b><a href="datasets/EBL+Domain+Theories">EBL Domain
Theories</a></b>
<!-- <td>Assorted small-scale domain theories&nbsp; -->

<!-- <td>Computer&nbsp; -->
<a href="datasets/Echocardiogram"><img border="1" src="assets/MLimages/SmallLarge38"
.jpg"/></a> <b><a href="datasets/Echocardiogram">Echocardiogram</a></b>
<!-- <td>Data for classifying if patients will survive for at least one year after a heart attack&n
bsp; -->
Multivariate 
Classification 
Categorical, Integer, Real 
132 
12 
1989 
<!-- <td>Life&nbsp; -->
<a href="datasets/Ecoli"><img border="1" src="assets/MLimages/SmallLarge120.jpg"/></a>
class="normal"><b><a href="datasets/Ecoli">Ecoli</a></b>
```

```
<!-- <td>This data contains protein localization sites&nbsp; -->
Multivariate 
Classification 
Real 
336 
8 
1996 
<!-- <td>Life&nbsp; -->
<a href="datasets/Flags"><img border="1" src="assets/MLimages/SmallLarge40.jpg"/></a>
<b><a href="datasets/Flags">Flags</a></b>
<!-- <td>From Collins Gem Guide to Flags, 1986&nbsp; -->
Multivariate 
Classification 
Categorical, Integer 
194 
30 
1990 
<!-- <td>Other&nbsp; -->
<a href="datasets/Function+Finding"><img border="1" src="assets/MLimages/SmallLarge4"
1.jpg"/></a> <b><a href="datasets/Function+Finding">Function Finding</a></b></
p>
<!-- <td>Cases collected mostly from investigations in physical science; intention is to evalua
te function-finding algorithms  -->

Function-Learning 
Real 
352 

1990 
<!-- <td>Physical&nbsp; -->
<t
e42.jpg"/></a> <b><a href="datasets/Glass+Identification">Glass Identification</a>
</b>
<!-- <td>From USA Forensic Science Service; 6 types of glass; defined in terms of their oxid
e content (i.e. Na, Fe, K, etc)  -->
Multivariate 
Classification 
Real 
214 
10 
1987 
<!-- <td>Physical&nbsp; -->
<a href="datasets/Haberman%27s+Survival"><img border="1" src="assets/MLimages/Smal
ILargedefault.jpg"/></a> <b><a href="datasets/Haberman%27s+Survival">Haberm
an's Survival</a></b>
<!-- <td>Dataset contains cases from study conducted on the survival of patients who had u
ndergone surgery for breast cancer  -->
Multivariate 
Classification 
Integer 
306 
3 
1999 
<!-- <td>Life&nbsp; -->
<a href="datasets/Hayes-Roth"><img border="1" src="assets/MLimages/SmallLargedefault."
jpg"/></a> <b><a href="datasets/Hayes-Roth">Hayes-Roth</a></b>
```

```
<!-- <td>Topic: human subjects study&nbsp; -->
Multivariate 
Classification 
Categorical 
160 
5 
1989 
<!-- <td>Social&nbsp; -->
<a href="datasets/Heart+Disease"><img border="1" src="assets/MLimages/SmallLarge45.j
pg"/></a> <b><a href="datasets/Heart+Disease">Heart Disease</a></b>
<!-- <td>4 databases: Cleveland, Hungary, Switzerland, and the VA Long Beach&nbsp;
 -->
Multivariate 
Classification 
Categorical, Integer, Real 
303 
75 
1988 
<!-- <td>Life&nbsp; -->
<a href="datasets/Hepatitis"><img border="1" src="assets/MLimages/SmallLarge46.jpg"/><
/a> <b><a href="datasets/Hepatitis">Hepatitis</a></b>
<!-- <td>From G.Gong: CMU; Mostly Boolean or numeric-valued attribute types; Includes co
st data (donated by Peter Turney)  -->
Multivariate 
Classification 
Categorical, Integer, Real 
155 
19 
1988 
<!-- <td>Life&nbsp; -->
<a href="datasets/Horse+Colic"><img border="1" src="assets/MLimages/SmallLarge47.jpg"
/></a> <b><a href="datasets/Horse+Colic">Horse Colic</a></b>
ble>
<!-- <td>Well documented attributes; 368 instances with 28 attributes (continuous, discrete,
and nominal); 30% missing values  -->
Multivariate 
Classification 
Categorical, Integer, Real 
368 
27 
1989 
<!-- <td>Life&nbsp; -->
<a href="datasets/ICU"><img border="1" src="assets/MLimages/SmallLarge49.jpg"/></a> <
/td><b><a href="datasets/ICU">ICU</a></b>
<!-- <td>Data set prepared for the use of participants for the 1994 AAAI Spring Symposium
on Artificial Intelligence in Medicine.  -->
Multivariate, Time-Series 

Real 

<!-- <td>Life&nbsp; -->
<a href="datasets/Image+Segmentation"><img border="1" src="assets/MLimages/SmallLar
gedefault.jpg"/></a> <b><a href="datasets/Image+Segmentation">Image Segment
ation//a>//h>//n>//td>//tr>//tahla>//td>
```

```
<!-- <td>Image data described by high-level numeric-valued attributes, 7 classes&nbsp;
 -->
Multivariate 
Classification 
Real 
2310 
19 
1990 
<!-- <td>Other&nbsp; -->
<a href="datasets/Internet+Advertisements"><img border="1" src="assets/MLimages/Small"
Large51.jpg"/></a> <b><a href="datasets/Internet+Advertisements">Internet Advertisements
tisements</a></b>
<!-- <td>This dataset represents a set of possible advertisements on Internet pages.&nbsp;
 -->
Multivariate 
Classification 
Categorical, Integer, Real 
3279 
1558 
1998 
<!-- <td>Computer&nbsp; -->
<a href="datasets/lonosphere"><img border="1" src="assets/MLimages/SmallLarge52.jpg"/
></a> <b><a href="datasets/lonosphere">lonosphere</a></b>
<!-- <td>Classification of radar returns from the ionosphere&nbsp; -->
Multivariate 
Classification 
Integer, Real 
351 
34 
1989 
<!-- <td>Physical&nbsp; -->
<tmages/SmallLarge53.jpg"/></a> </
td><b><a href="datasets/lris">lris</a></b>
<!-- <td>Famous database; from Fisher, 1936&nbsp; -->
Multivariate 
Classification 
Real 
150 
4 
1988 
<!-- <td>Life&nbsp; -->
<a href="datasets/ISOLET"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"
/></a> <b><a href="datasets/ISOLET">ISOLET</a></b>
<!-- <td> Goal: Predict which letter-name was spoken--a simple classification task.&nbsp;</p
> -->
Multivariate 
Classification 
Real 
7797 
617 
1994 
<!-- <td>Computer&nbsp; -->
<tm, datasets/Kinship"><img border="1" src="assets/MLimages/SmallLarge55.jpg"/></a
> <b><a href="datasets/Kinship">Kinship</a></b>
  the prolong "normal". Delational detacet about 1/nc 1/th
```

```
<!-- <!u>nelational dataset&nosp,</tu>
Relational 
Relational-Learning 
Categorical 
104 
12 
1990 
<!-- <td>Social&nbsp; -->
<a href="datasets/Labor+Relations"><img border="1" src="assets/MLimages/SmallLargede"
fault.jpg"/></a> <b><a href="datasets/Labor+Relations">Labor Relations</a></b><
/p>
<!-- <td>From Collective Bargaining Review&nbsp; -->
Multivariate 

Categorical, Integer, Real 
57 
16 
1988 
<!-- <td>Social&nbsp; -->
="1" src="assets/MLimages/SmallLar"
gedefault.jpg"/></a> class="normal"><b><a href="datasets/LED+Display+Domain">LED Display Do
main</a></b>
<!-- <td>From Classification and Regression Trees book; We provide here 2 C programs for
generating sample databases  -->
Multivariate, Data-Generator 
Classification 
Categorical 

7 
1988 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Lenses"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/
></a> <b><a href="datasets/Lenses">Lenses</a></b>
<!-- <td>Database for fitting contact lenses&nbsp; -->
Multivariate 
Classification 
Categorical 
24 
4 
1990 
<!-- <td>Other&nbsp; -->
default.jpg"/></a> <b><a href="datasets/Letter+Recognition">Letter Recognition
a></b>
<!-- <td>Database of character image features; try to identify the letter&nbsp; -->
Multivariate 
Classification 
Integer 
20000 
16 
1991 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Liver+Disorders"><img border="1" src="assets/MLimages/SmallLargedef"
ault.jpg"/></a> class="normal"><b><a href="datasets/Liver+Disorders">Liver Disorders</a></b></p
>
<!-- <td>BUPA Medical Research Ltd. database donated by Richard S. Forsyth&nbsp;</
td> -->
```

```
<lu>Mullivariale </lu>

Categorical, Integer, Real 
345 
7 
1990 
<!-- <td>Life&nbsp; -->
<a href="datasets/Logic+Theorist"><img border="1" src="assets/MLimages/SmallLargedefa"
ult.jpg"/></a> <b><a href="datasets/Logic+Theorist">Logic Theorist</a></b></
td>
<!-- <td>All code for Logic Theorist&nbsp; -->
Domain-Theory 

<!-- <td>Computer&nbsp; -->
<a href="datasets/Lung+Cancer"><img border="1" src="assets/MLimages/SmallLarge62.jp"
g"/></a> <b><a href="datasets/Lung+Cancer">Lung Cancer</a></b>
<!-- <td>Lung cancer data; no attribute definitions&nbsp; -->
Multivariate 
Classification 
Integer 
32 
56 
1992 
<!-- <td>Life&nbsp; -->
<a href="datasets/Lymphography"><img border="1" src="assets/MLimages/SmallLargedef"
ault.jpg"/></a> <b><a href="datasets/Lymphography">Lymphography</a></b>
<!-- <td>This lymphography domain was obtained from the University Medical Centre, Institu
te of Oncology, Ljubljana, Yugoslavia. (Restricted access) 
Multivariate 
Classification 
Categorical 
148 
18 
1988 
<!-- <td>Life&nbsp; -->
<a href="datasets/Mechanical+Analysis"><img border="1" src="assets/MLimages/SmallLar
gedefault.jpg"/></a> <b><a href="datasets/Mechanical+Analysis">Mechanical Anal
ysis</a></b>
<!-- <td>Fault diagnosis problem of electromechanical devices; also PUMPS DATA SET is n
ewer version with domain theory and results  -->
Multivariate 
Classification 
Categorical, Integer, Real 
209 
8 
1990 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Meta-data"><img border="1" src="assets/MLimages/SmallLargedefault.j
pg"/></a> <b><a href="datasets/Meta-data">Meta-data</a></b>
ble>
<!-- <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 Statiog project).&hbsp;
Multivariate 
Classification 
Categorical, Integer, Real 
528 
22 
1996 
<!-- <td>Other&nbsp; -->
<a href="datasets/Mobile+Robots"><img border="1" src="assets/MLimages/SmallLarge66.j
pg"/></a> <b><a href="datasets/Mobile+Robots">Mobile Robots</a></b>
<!-- <td>Learning concepts from sensor data of a mobile robot; set of data sets&nbsp;</
td> -->
Domain-Theory 

Categorical, Integer, Real 

1995 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29"><img border=
"1" src="assets/MLimages/SmallLarge67.jpg"/></a> class="normal"><b><a href="datasets/Molecul">
ar+Biology+%28Promoter+Gene+Sequences%29">Molecular Biology (Promoter Gene Sequences)</a>
>
<!-- <td>E. Coli promoter gene sequences (DNA) with partial domain theory&nbsp;
Sequential, Domain-Theory 
Classification 
Categorical 
106 
58 
1990 
<!-- <td>Life&nbsp; -->
<a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%29"><img border
="1" src="assets/MLimages/SmallLarge67.jpg"/></a> <b><a href="datasets/Molecu
lar+Biology+%28Protein+Secondary+Structure%29">Molecular Biology (Protein Secondary Structure)</a></b>
/p>
<!-- <td>From CMU connectionist bench repository; Classifies secondary structure of certain
globular proteins  -->
Sequential 
Classification 
Categorical 
128 

<!-- <td>Life&nbsp; -->
<a href="datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29"><img bo
rder="1" src="assets/MLimages/SmallLarge67.jpg"/></a> <b><a href="datasets/Mo
lecular+Biology+%28Splice-junction+Gene+Sequences%29">Molecular Biology (Splice-junction Gene Sequences
es)</a></b>
<!-- <td>Primate splice-junction gene sequences (DNA) with associated imperfect domain th
eory  -->
Sequential, Domain-Theory 
Classification 
Categorical 
3190 
61 
1992
```

```
<!-- <td>Life&nbsp; -->
="1" src="assets/MLimages/SmallL" src="assets/MLimages/Sma
argedefault.jpg"/></a> <b><a href="datasets/MONK%27s+Problems">MONK's Pro
blems</a></b>
<!-- <td>A set of three artificial domains over the same attribute space; Used to test a wide r
ange of induction algorithms  -->
Multivariate 
Classification 
Categorical 
432 
7 
1992 
<!-- <td>Other&nbsp; -->
<a href="datasets/Moral+Reasoner"><img border="1" src="assets/MLimages/SmallLargede"
fault.jpg"/></a> <b><a href="datasets/Moral+Reasoner">Moral Reasoner</a></b>
<!-- <td>Horn-clause model that qualitatively simulates moral reasoning; Theory includes ne
gated literals  -->
Domain-Theory 

202 

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

<!-- <td>Computer&nbsp; -->
<a href="datasets/Mushroom"><img border="1" src="assets/MLimages/SmallLarge73.jpg"/
></a> class="normal"><b><a href="datasets/Mushroom">Mushroom</a></b>
>
<!-- <td>From Audobon Society Field Guide; mushrooms described in terms of physical char
acteristics; classification: poisonous or edible  -->
Multivariate 
Classification 
Categorical 
8124 
22 
1987 
<!-- <td>Life&nbsp; -->
="1" src="assets/MLimages/Small">= 1" src="assets/MLimages/Small"
|Largedefault.jpg"/></a> <b><a href="datasets/Musk+%28Version+1%29">Musk (V
ersion 1)</a></b>
<!-- <td>The goal is to learn to predict whether new molecules will be musks or non-musks&
nbsp; -->
Multivariate 
Classification 
Integer
```

```
476 
168 
1994 
<!-- <td>Physical&nbsp; -->
="1" src="assets/MLimages/Small">="1" src="assets/MLimages/Small">="1" src="assets/MLimages/Small">="1" src="assets/MLimages/Small">="1" src="assets/MLimages/Small">="1" src="assets/MLimages/Small">="1" src="assets/MLimages/Small">= 1" src="assets/MLimages/Small"
ILargedefault.jpg"/></a> class="normal"><b><a href="datasets/Musk+%28Version+2%29">Musk (V
ersion 2)</a></b>
<!-- <td>The goal is to learn to predict whether new molecules will be musks or non-musks&
nbsp; -->
Multivariate 
Classification 
Integer 
6598 
168 
1994 
<!-- <td>Physical&nbsp; -->
<a href="datasets/Nursery"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"
/></a> <b><a href="datasets/Nursery">Nursery</a></b>
<!-- <td> Nursery Database was derived from a hierarchical decision model originally develo
ped to rank applications for nursery schools.  -->
Multivariate 
Classification 
Categorical 
12960 
8 
1997 
<!-- <td>Social&nbsp; -->
<a href="datasets/Othello+Domain+Theory"><img border="1" src="assets/MLimages/Small"
Largedefault.jpg"/></a> class="normal"><b><a href="datasets/Othello+Domain+Theory">Othello D
omain Theory</a></b>
<!-- <td>Used in research to generate features for an inductive learning system&nbsp;<
/td> -->
Domain-Theory 

1991 
<!-- <td>Game&nbsp; -->
="1" src="assets/MLimages/Sm"
allLargedefault.jpg"/></a> <b><a href="datasets/Page+Blocks+Classification">Pag
e Blocks Classification</a></b>
<!-- <td>The problem consists of classifying all the blocks of the page layout of a document t
hat has been detected by a segmentation process.  -->
Multivariate 
Classification 
Integer, Real 
5473 
10 
1995 
<!-- <td>Computer&nbsp; -->
+ Image: A contract of the con
s/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Optical+Recogniti
on+of+Handwritten+Digits">Optical Recognition of Handwritten Digits</a></b>
<!-- <td>Two versions of this database available; see folder&nbsp; -->
Multivariate
```

```
Classification 
Integer 
5620 
64 
1998 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits"><img border="1" src="a
ssets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Pen-Based+R
ecognition+of+Handwritten+Digits">Pen-Based Recognition of Handwritten Digits</a></b>
<!-- <td>Digit database of 250 samples from 44 writers&nbsp; -->
Multivariate 
Classification 
Integer 
10992 
16 
1998 
<!-- <td>Computer&nbsp; -->
="1" src="assets/MLimages/SmallL
argedefault.jpg"/></a> <b><a href="datasets/Post-Operative+Patient">Post-Operative+Patient">Post-Operative+Patient">Post-Operative+Patient">Post-Operative+Patient
ive Patient</a></b>
<!-- <td>Dataset of patient features&nbsp; -->
Multivariate 
Classification 
Categorical, Integer 
90 
8 
1993 
<!-- <td>Life&nbsp; -->
<a href="datasets/Primary+Tumor"><img border="1" src="assets/MLimages/SmallLargedef"
ault.jpg"/></a> <b><a href="datasets/Primary+Tumor">Primary Tumor</a></b></p
>
<!-- <td>From Ljubljana Oncology Institute&nbsp; -->
Multivariate 
Classification 
Categorical 
339 
17 
1988 
<!-- <td>Life&nbsp; -->
<a href="datasets/Prodigy"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"</a>
/></a> <b><a href="datasets/Prodigy">Prodigy</a></b>
<!-- <td>Assorted domains like blocksworld, eightpuzzle, and schedworld.&nbsp; --
Domain-Theory 

<!-- <td>Other&nbsp; -->
<a href="datasets/Qualitative+Structure+Activity+Relationships"><img border="1" src="asse"
ts/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Qualitative+Struc"><b><a href="datasets/Qualitative+Struc"><b><a href="datasets/Qualitative+Struc"><br/><a href="datasets/Qualitative+Struc"><b><a href="datasets/Qualitative+Struc"><b><a href="datasets/Qualitative+Struc"><b><a href="datasets/Qualitative+Struc"><b><a href="datasets/Qualitative+Struc"><b href="datasets/Qualitativ
ture+Activity+Relationships">Qualitative Structure Activity Relationships</a></b>
<!-- <td>Two sets of datasets are given: pyrimidines and triazines&nbsp; -->
Domain-Theory
```

```
<!-- <td>Physical&nbsp; -->
<a href="datasets/Quadruped+Mammals"><img border="1" src="assets/MLimages/SmallLa"
rge86.jpg"/></a> <b><a href="datasets/Quadruped+Mammals">Quadruped Mam
mals</a></b>
<!-- <td> The file animals.c is a data generator of structured instances representing quadrup
ed animals  -->
Multivariate, Data-Generator 
Classification 
Real 

72 
1992 
<!-- <td>Life&nbsp; -->
<a href="datasets/Servo"><img border="1" src="assets/MLimages/SmallLarge87.jpg"/></a>
<b><a href="datasets/Servo">Servo</a></b>
<!-- <td>Data was from a simulation of a servo system&nbsp; -->
Multivariate 
Regression 
Categorical, Integer 
167 
4 
1993 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Shuttle+Landing+Control"><img border="1" src="assets/MLimages/Small"
Large92.jpg"/></a> class="normal"><b><a href="datasets/Shuttle+Landing+Control">Shuttle Landing+Control">Shuttle Landing+Control<Shuttle Landing+Control<
ng Control</a></b>
<!-- <td>Tiny database; all nominal values&nbsp; -->
Multivariate 
Classification 
Categorical 
15 
6 
1988 
<!-- <td>Physical&nbsp; -->
></a> <b><a href="datasets/Solar+Flare">Solar Flare</a></b>
e>
<!-- <td>Each class attribute counts the number of solar flares of a certain class that occur i
n a 24 hour period  -->
Multivariate 
Regression 
Categorical 
1389 
10 
1989 
<!-- <td>Physical&nbsp; -->
<a href="datasets/Soybean+%28Large%29"><img border="1" src="assets/MLimages/Small"
Large90.jpg"/></a> class="normal"><b><a href="datasets/Soybean+%28Large%29">Soybean (Large90.jpg"/></a>
ge)</a></b>
<!-- <td>Michalski's famous soybean disease database&nbsp; -->
Multivariate 
Classification 
<n class="normal">Categorical </n>
```

```
307 
35 
1988 
<!-- <td>Life&nbsp; -->
<a href="datasets/Soybean+%28Small%29"><img border="1" src="assets/MLimages/Small%29"><img border="1" src="assets/MLimages/Small%29"></img border="1" src="as
Large90.jpg"/></a> class="normal"><b><a href="datasets/Soybean+%28Small%29">Soybean (Sm
all)</a></b>
<!-- <td>Michalski's famous soybean disease database&nbsp; -->
Multivariate 
Classification 
Categorical 
47 
35 
1987 
<!-- <td>Life&nbsp; -->
<a href="datasets/Challenger+USA+Space+Shuttle+O-Ring"><img border="1" src="assets/
MLimages/SmallLarge92.jpg"/></a> class="normal"><b><a href="datasets/Challenger+USA+Space">
+Shuttle+O-Ring">Challenger USA Space Shuttle O-Ring</a></b>
<!-- <td>Task: predict the number of O-rings that experience thermal distress on a flight at 3
1 degrees F given data on the previous 23 shuttle flights  -->
Multivariate 
Regression 
Integer 
23 
4 
1993 
<!-- <td>Physical&nbsp; -->
/SmallLargedefault.jpg"/></a> <b><a href="datasets/Low+Resolution+Spectromete"><br/>
r">Low Resolution Spectrometer</a></b>
<!-- <td>From IRAS data -- NASA Ames Research Center&nbsp; -->
Multivariate 
Classification 
Integer, Real 
531 
102 
1988 
<!-- <td>Physical&nbsp; -->
<a href="datasets/Spambase"><img border="1" src="assets/MLimages/SmallLarge94.jpg"/
></a> <b><a href="datasets/Spambase">Spambase</a></b></table
>
<!-- <td>Classifying Email as Spam or Non-Spam&nbsp; -->
Multivariate 
Classification 
Integer, Real 
4601 
57 
1999 
<!-- <td>Computer&nbsp; -->
<a href="datasets/SPECT+Heart"><img border="1" src="assets/MLimages/SmallLarge45.jp"
g"/></a> <b><a href="datasets/SPECT+Heart">SPECT Heart</a></b>
>
<!-- <td>Data on cardiac Single Proton Emission Computed Tomography (SPECT) images.
Each patient classified into two categories: normal and abnormal.  -->
Multivariate 
Classification 
tds on alace-"normal"s Catagorical o/ns o/tds
```

```
tu>oatogorical </tu>
267 
22 
2001 
<!-- <td>Life&nbsp; -->
<a href="datasets/SPECTF+Heart"><img border="1" src="assets/MLimages/SmallLarge45.j
pg"/></a> <b><a href="datasets/SPECTF+Heart">SPECTF Heart</a></b>
>
<!-- <td>Data on cardiac Single Proton Emission Computed Tomography (SPECT) images.
Each patient classified into two categories: normal and abnormal.  -->
Multivariate 
Classification 
Integer 
267 
44 
2001 
<!-- <td>Life&nbsp; -->
a> <b><a href="datasets/Sponge">Sponge</a></b>
<!-- <td>Data on sponges; Attributes in spanish&nbsp; -->
Multivariate 
Clustering 
Categorical, Integer 
76 
45 

<!-- <td>Life&nbsp; -->
<a href="datasets/Statlog+Project"><img border="1" src="assets/MLimages/SmallLargedef"
ault.jpg"/></a> class="normal"><b><a href="datasets/Statlog+Project">Statlog Project</a></b>
<!-- <td>Various Databases: Vehicle silhouttes, Landsat Sattelite, Shuttle, Australian Credit
Approval, Heart Disease, Image Segmentation, German Credit  -->

1992 
<!-- <td>Other&nbsp; -->
<a href="datasets/Student+Loan+Relational"><img border="1" src="assets/MLimages/Small"><a href="datasets/Student+Loan+Relational"><a href="datasets/Student+Loan+Relational"><a href="datasets/Student+Loan+Relational"><a href="datasets/Student+Loan+Relational"><a href="datasets/Student+Loan+Relational"><a href="datasets/Student+Loan+Relational"><table
ILargedefault.jpg"/></a> <b><a href="datasets/Student+Loan+Relational">Student
Loan Relational</a></b>
<!-- <td>Student Loan Relational Domain&nbsp; -->
Domain-Theory 

1000 

1993 
<!-- <td>Social&nbsp; -->
<tm>border="1" src="assets/MLimages" |
/SmallLargedefault.jpg"/></a> <b><a href="datasets/Teaching+Assistant+Evaluatio"><b><a href="datasets/Teaching+Assistant+Evaluatio"><b><a href="datasets/Teaching+Assistant+Evaluatio"><b><a href="datasets/Teaching+Assistant+Evaluatio"><b><a href="datasets/Teaching+Assistant+Evaluatio"><b><a href="datasets/Teaching+Assistant+Evaluatio"><b><a href="datasets/Teaching+Assistant+Evaluatio"><b href="datasets/Teaching+Assistant+Evaluation"><b href="datasets/Teaching+Assistant+Evaluation"><b href="datasets/Teaching+Assistant+Evaluation"><b href="datasets/Teaching+Assistant+Evaluation"><b href="datasets/Teaching+Assistant+Evaluation"><b href="datasets/Teaching+Assistant+Evaluation"><b href="datasets/Teaching+Basistant+Evaluation"><b href="datasets/Teaching+Basistant+Evaluation"><b href="datasets/Teaching+Basistant+Evaluation"><b href="datasets/Teaching+Basistant+Evaluation"><b href="datasets/Teaching+Basistant+Evaluation"><b href="datasets/Teaching+Basistant+Evaluation"><b href="datasets/Teaching+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basistant+Basis
n">Teaching Assistant Evaluation</a></b>
<!-- <td>The data consist of evaluations of teaching performance; scores are "low", "mediu
m", or "high"  -->
Multivariate 
Classification
```

```
<lu></u>></u>></u></lu>
151 
5 
1997 
<!-- <td>Other&nbsp; -->
="1" src="assets/MLimages/SmallL">= 1" src="assets
arge101.jpg"/></a> <b><a href="datasets/Tic-Tac-Toe+Endgame">Tic-Tac-Toe E
ndgame</a></b>
<!-- <td>Binary classification task on possible configurations of tic-tac-toe game&nbsp;<
/td> -->
Multivariate 
Classification 
Categorical 
958 
9 
1991 
<!-- <td>Game&nbsp; -->
="1" src="assets/MLimages/SmallLarged">="1" src="assets/MLimages/SmallLarged">=="1" src="assets/MLimages/SmallLarged">== 1" src="assets/MLimages/SmallLarged</a>
efault.jpg"/></a> class="normal"><b><a href="datasets/Thyroid+Disease">Thyroid Disease</a></b
>
<!-- <td>10 separate databases from Garavan Institute&nbsp; -->
Multivariate, Domain-Theory 
Classification 
Categorical, Real 
7200 
21 
1987 
<!-- <td>Life&nbsp; -->
a> <b><a href="datasets/Trains">Trains</a></b>
<!-- <td>2 data formats (structured, one-instance-per-line)&nbsp; -->
Multivariate 
Classification 
Categorical 
10 
32 
1994 
<!-- <td>Other&nbsp; -->
<a href="datasets/University"><img border="1" src="assets/MLimages/SmallLarge104.jpg"/
></a> <b><a href="datasets/University">University</a></b>
/td>
<!-- <td>Data in original (LISP-readable) form&nbsp; -->
Multivariate 
Classification 
Categorical, Integer 
285 
17 
1988 
<!-- <td>Other&nbsp; -->
s/SmallLarge105.jpg"/></a> <b><a href="datasets/Congressional+Voting+Records"
">Congressional Voting Records</a></b>
<!-- <td>1984 United Stated Congressional Voting Records; Classify as Republican or Demo
crat  -->
Multivariate 
Classification 
Categorical
```

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

```
1996 
<!-- <td>Life&nbsp; -->
<a href="datasets/Zoo"><img border="1" src="assets/MLimages/SmallLarge111.jpg"/></a>
<b><a href="datasets/Zoo">Zoo</a></b>
<!-- <td>Artificial, 7 classes of animals&nbsp; -->
Multivariate 
Classification 
Categorical, Integer 
101 
17 
1990 
<!-- <td>Life&nbsp; -->
<a href="datasets/Undocumented"><img border="1" src="assets/MLimages/SmallLargedef"
ault.jpg"/></a> class="normal"><b><a href="datasets/Undocumented">Undocumented</a></b>
>
<!-- <td>Various datasets without documentation (feel free to explore!)&nbsp; -->

<!-- <td>Other&nbsp; -->
<a href="datasets/Twenty+Newsgroups"><img border="1" src="assets/MLimages/SmallLar
gedefault.jpg"/></a> class="normal"><b><a href="datasets/Twenty+Newsgroups">Twenty Newsgro
ups</a></b>
<!-- <td>This data set consists of 20000 messages taken from 20 newsgroups.&nbsp;</
td> -->
Text 

20000 

1999 
<!-- <td>Other&nbsp; -->
<a href="datasets/Australian+Sign+Language+signs"><img border="1" src="assets/MLimag"
gns">Australian Sign Language signs</a></b>
<!-- <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.  -->
Multivariate, Time-Series 
Classification 
Categorical, Real 
6650 
15 
1999 
<!-- <td>Other&nbsp; -->
<a href="datasets/Australian+Sign+Language+signs+%28High+Quality%29"><img border="
1" src="assets/MLimages/SmallLarge114.jpg"/></a> class="normal"><b><a href="datasets/Australi"><b>
an+Sign+Language+signs+%28High+Quality%29">Australian Sign Language signs (High Quality)</a></b>
<!-- <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
> -->
Multivariate, Time-Series 
Classification 
Real
```

```
2565 
22 
2002 
<!-- <td>Other&nbsp; -->
<a href="datasets/US+Census+Data+%281990%29"><img border="1" src="assets/MLimag"></a>
es/SmallLarge2.jpg"/></a> class="normal"><b><a href="datasets/US+Census+Data+%281990%29" |
">US Census Data (1990)</a></b>
<!-- <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. 
Multivariate 
Clustering 
Categorical 
2458285 
68 

<!-- <td>Social&nbsp; -->
<a href="datasets/Census-Income+%28KDD%29"><img border="1" src="assets/MLimages/
SmallLarge2.jpg"/></a> <b><a href="datasets/Census-Income+%28KDD%29">Ce
nsus-Income (KDD)</a></b>
<!-- <td>This data set contains weighted census data extracted from the 1994 and 1995 cur
rent population surveys conducted by the U.S. Census Bureau.  -->
Multivariate 
Classification 
Categorical, Integer 
299285 
40 
2000 
<!-- <td>Social&nbsp; -->
SmallLarge118.jpg"/></a> <b><a href="datasets/Coil+1999+Competition+Data">C
oil 1999 Competition Data</a></b>
<!-- <td>This data set is from the 1999 Computational Intelligence and Learning (COIL) com
petition. The data contains measurements of river chemical concentrations and algae densities. 
> -->
Multivariate 

Categorical, Real 
340 
17 
1999 
<!-- <td>Physical&nbsp; -->
="1" src="assets/MLimages/SmallL
arge119.jpg"/></a> class="normal"><b><a href="datasets/Corel+Image+Features">Corel Image F
eatures</a></b>
<!-- <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-occ
urence  -->
Multivariate 

Real 
68040 
89 
1999 
<!-- <td>Other&nbsp; -->
<a href="datasets/E.+Coli+Genes"><img border="1" src="assets/MLimages/SmallLarge120" src="assets/MLimages/SmallAarge120" src="assets/MLimages/SmallAarge120" src="assets/MLimages/SmallAarge120" src="assets/MLimages/SmallAarg
.jpg"/></a> <b><a href="datasets/E.+Coli+Genes">E. Coli Genes</a></b>
```

```
<!-- <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; -->
Relational 

2001 
<!-- <td>Life&nbsp; -->
<a href="datasets/EEG+Database"><img border="1" src="assets/MLimages/SmallLarge121" src="assets/MLimages/SmallCarge121" src="assets/MLimages/SmallCarge
.jpg"/></a> <b><a href="datasets/EEG+Database">EEG Database</a></b>
d>
<!-- <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 
 -->
Multivariate, Time-Series 

Categorical, Integer, Real 
122 
4 
1999 
<!-- <td>Life&nbsp; -->
<a href="datasets/EI+Nino"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"
/></a> <b><a href="datasets/EI+Nino">EI Nino</a></b>
<!-- <td>The data set contains oceanographic and surface meteorological readings taken fr
om a series of buoys positioned throughout the equatorial Pacific.  -->
Spatio-temporal 

Integer, Real 
178080 
12 
1999 
<!-- <td>Physical&nbsp; -->
<a href="datasets/Entree+Chicago+Recommendation+Data"><img border="1" src="assets/
MLimages/SmallLarge123.jpg"/></a> <b><a href="datasets/Entree+Chicago+Reco
mmendation+Data">Entree Chicago Recommendation Data</a></b>
<!-- <td>This data contains a record of user interactions with the Entree Chicago restaurant
recommendation system.  -->
Transactional, Sequential 
Recommender-Systems 
Categorical 
50672 

2000 
<!-- <td>Other&nbsp; -->
="1" src="assets/MLimages/SmallLarg">="1" src="assets/MLimages/SmallLarg">="1" src="assets/MLimages/SmallLarg">="1" src="assets/MLimages/SmallLarg">="1" src="assets/MLimages/SmallLarg">="1" src="assets/MLimages/SmallLarg">="1" src="assets/MLimages/SmallLarg">= 1" src="assets/MLimages/SmallLarg= 1" src="assets/MLimages/SmallLarg</tr
e124.jpg"/></a> class="normal"><b><a href="datasets/CMU+Face+Images">CMU Face Images</a
></b>
<!-- <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  -->
Image 
Classification 
Integer 
640 

1999
```

```
<!-- <td>Other&nbsp; -->
<a href="datasets/Insurance+Company+Benchmark+%28COIL+2000%29"><img border="1"><a href="datasets/Insurance+Company+Benchmark+%28COIL+2000%29"><img border="1"><a href="datasets/Insurance+Company+Benchmark+%28COIL+2000%29"><a href="datasets/Insurance+Co
" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Insura
nce+Company+Benchmark+%28COIL+2000%29">Insurance Company Benchmark (COIL 2000)</a></b>
td>
<!-- <td>This data set used in the ColL 2000 Challenge contains information on customers o
f an insurance company. The data consists of 86 variables and includes product usage data and socio-demogr
aphic data  -->
Multivariate 
Regression, Description 
Categorical, Integer 
9000 
86 
2000 
<!-- <td>Social&nbsp; -->
<a href="datasets/Internet+Usage+Data"><img border="1" src="assets/MLimages/SmallLar"></a>
gedefault.jpg"/></a> class="normal"><b><a href="datasets/Internet+Usage+Data">Internet Usage
Data</a></b>
<!-- <td>This data contains general demographic information on internet users in 1997.&nbs
p; -->
Multivariate 

Categorical, Integer 
10104 
72 
1999 
<!-- <td>Computer&nbsp; -->
<a href="datasets/IPUMS+Census+Database"><img border="1" src="assets/MLimages/Sm
allLarge2.jpg"/></a> <b><a href="datasets/IPUMS+Census+Database">IPUMS Ce
nsus Database</a></b>
<!-- <td>This data set contains unweighted PUMS census data from the Los Angeles and Lo
ng Beach areas for the years 1970, 1980, and 1990. 
Multivariate 

Categorical, Integer 
256932 
61 
1999 
<!-- <td>Social&nbsp; -->
="1" src="assets/MLimages/SmallLarge">="1" src="assets/MLimages/SmallLarge">= 1" src="assets/MLimages/SmallLarge
default.jpg"/></a> <b><a href="datasets/Japanese+Vowels">Japanese Vowels</a>
</b>
<!-- <td>This dataset records 640 time series of 12 LPC cepstrum coefficients taken from ni
ne male speakers.  -->
Multivariate, Time-Series 
Classification 
Real 
640 
12 

<!-- <td>Other&nbsp; -->
<a href="datasets/KDD+Cup+1998+Data"><img border="1" src="assets/MLimages/SmallLa"
rgedefault.jpg"/></a> class="normal"><b><a href="datasets/KDD+Cup+1998+Data">KDD Cup 199
8 Data</a></b>
<!-- <td>This is the data set used for The Second International Knowledge Discovery and D
ata Mining Tools Competition, which was held in conjunction with KDD-98
```

< class="normal">Multivariate

```
Regression 
Categorical, Integer 
191779 
481 
1998 
<!-- <td>Other&nbsp; -->
<a href="datasets/KDD+Cup+1999+Data"><img border="1" src="assets/MLimages/SmallLa"
rgedefault.jpg"/></a> class="normal"><b><a href="datasets/KDD+Cup+1999+Data">KDD Cup 199
9 Data</a></b>
<!-- <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 
Multivariate 
Classification 
Categorical, Integer 
4000000 
42 
1999 
<!-- <td>Computer&nbsp; -->
="1" src="assets/MLimages/Small">= "assets/MLimages/Small">= "assets/MLimage
Large131.jpg"/></a> class="normal"><b><a href="datasets/M.+Tuberculosis+Genes">M. Tubercul
osis Genes</a></b>
<!-- <td> Data giving characteristics of each ORF (potential gene) in the M. tuberculosis bact
erium. Sequence, homology (similarity to other genes) and structural information, and function (if known) are pr
ovided  -->
Relational 

2001 
<!-- <td>Life&nbsp; -->
<a href="datasets/Movie"><img border="1" src="assets/MLimages/SmallLarge132.jpg"/></a
> <b><a href="datasets/Movie">Movie</a></b>
<!-- <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. 
Multivariate, Relational 

10000 

1999 
<!-- <td>Other&nbsp; -->
<a href="datasets/MSNBC.com+Anonymous+Web+Data"><img border="1" src="assets/MLi
mages/SmallLargedefault.jpg"/></a> <b><a href="datasets/MSNBC.com+Anonymo"><br/>
us+Web+Data">MSNBC.com Anonymous Web Data</a></b>
<!-- <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.&nb
sp; -->
Sequential 

Categorical 
989818 

<!-- <td>Computer&nbsp; -->
<a href="datasets/NSF+Research+Award+Abstracts+1990-2003"><img border="1" src="as
sats/MI images/Smalll arge134 ing"/>/a> //td>td>>n class="normal">>h>>a href="datasets/NSF+Research+A
```

```
storial integral of the integral of the star of the st
ward+Abstracts+1990-2003">NSF Research Award Abstracts 1990-2003</a></b>
<!-- <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  -->
Text 

129000 

2003 
<!-- <td>Other&nbsp; -->
<a href="datasets/Pioneer-1+Mobile+Robot+Data"><img border="1" src="assets/MLimages"
/SmallLargedefault.jpg"/></a> <b><a href="datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Datasets/Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+Mobile+Robot+Pioneer-1+
a">Pioneer-1 Mobile Robot Data</a></b>
<!-- <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  -->
Multivariate, Time-Series 

Categorical, Real 

1999 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Pseudo+Periodic+Synthetic+Time+Series"><img border="1" src="assets/
MLimages/SmallLarge136.jpg"/></a> class="normal"><b><a href="datasets/Pseudo+Periodic+Synt">
hetic+Time+Series">Pseudo Periodic Synthetic Time Series</a></b>
<!-- <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.  -->
Univariate, Time-Series 

100000 

1999 
<!-- <td>Other&nbsp; -->
<a href="datasets/Reuters-21578+Text+Categorization+Collection"><img border="1" src="a"
ssets/MLimages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Reuters-2157"><b><a href="datasets/Reuters-2157"><b href="datasets/R
8+Text+Categorization+Collection">Reuters-21578 Text Categorization Collection</a></b>
>
<!-- <td>This is a collection of documents that appeared on Reuters newswire in 1987. The
documents were assembled and indexed with categories.  -->
Text 
Classification 
Categorical 
21578 
5 
1997 
<!-- <td>Other&nbsp; -->
="1" src="assets/MLimages/Sm
allLarge138.jpg"/></a> <b><a href="datasets/Robot+Execution+Failures">Robot E
xecution Failures</a></b>
<!-- <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 
Multivariate, Time-Series 
Classification 
         an along "normal", Intoger a/n. a/td.
```

```
463 
90 
1999 
<!-- <td>Physical&nbsp; -->
<a href="datasets/Synthetic+Control+Chart+Time+Series"><img border="1" src="assets/ML
images/SmallLargedefault.jpg"/></a> <b><a href="datasets/Synthetic+Control+Cha"><br/>td><a href="datasets/Synthetic+Control+Cha"><b href="datasets/Synthetic+Cha"><b href="
rt+Time+Series">Synthetic Control Chart Time Series</a></b>
<!-- <td>This data consists of synthetically generated control charts.&nbsp; -->
Time-Series 
Classification, Clustering 
Real 
600 

1999 
<!-- <td>Other&nbsp; -->
<a href="datasets/Syskill+and+Webert+Web+Page+Ratings"><img border="1" src="assets/
MLimages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Syskill+and+Webert">
+Web+Page+Ratings">Syskill and Webert Web Page Ratings</a></b>
<!-- <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)  -->
Multivariate, Text 
Classification 
Categorical 
332 
5 
1998 
<!-- <td>Computer&nbsp; -->
<a href="datasets/UNIX+User+Data"><img border="1" src="assets/MLimages/SmallLarge1" src="assets/MLimages/SmallArge1" src="assets/
41.jpg"/></a> <b><a href="datasets/UNIX+User+Data">UNIX User Data</a></b></
p>
<!-- <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.  -->
Text, Sequential 

<!-- <td>Computer&nbsp; -->
<a href="datasets/Volcanoes+on+Venus+-+JARtool+experiment"><img border="1" src="ass
ets/MLimages/SmallLarge142.jpg"/></a> class="normal"><b><a href="datasets/Volcanoes+on+Ve"><b><a href="datasets/Volcanoes+on+Ve"><b><a href="datasets/Volcanoes+on+Ve"><b><a href="datasets/Volcanoes+on+Ve"><b><a href="datasets/Volcanoes+on+Ve"><b ><a href="datasets/Volcanoes+on+Ve"><a href="datasets/V
nus+-+JARtool+experiment">Volcanoes on Venus - JARtool experiment</a></b>
<!-- <td>The JARtool project was a pioneering effort to develop an automatic system for cat
aloging small volcanoes in the large set of Venus images returned by the Magellan spacecraft. 
-->
Image 
Classification 

<!-- <td>Physical&nbsp; -->
<a href="datasets/Statlog+%28Australian+Credit+Approval%29"><img border="1" src="ass
ets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Statlog+%28Au"><br/>/28Au
stralian+Credit+Approval%29">Statlog (Australian Credit Approval)</a></b>
```

```
<!-- <lu>1 nls lile concerns credit card applications. This database exists elsewhere in the re
pository (Credit Screening Database) in a slightly different form 
Multivariate 
Classification 
Categorical, Integer, Real 
690 
14 

<!-- <td>Financial&nbsp; -->
<a href="datasets/Statlog+%28German+Credit+Data%29"><img border="1" src="assets/ML
images/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Statlog+%28German+"
Credit+Data%29">Statlog (German Credit Data)</a></b>
<!-- <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 
Multivariate 
Classification 
Categorical, Integer 
1000 
20 
1994 
<!-- <td>Financial&nbsp; -->
<a href="datasets/Statlog+%28Heart%29"><img border="1" src="assets/MLimages/SmallLa"
rge45.jpg"/></a> class="normal"><b><a href="datasets/Statlog+%28Heart%29">Statlog (Heart)</a>
></b>
<!-- <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  -->
Multivariate 
Classification 
Categorical, Real 
270 
13 

<!-- <td>Life&nbsp; -->
<a href="datasets/Statlog+%28Landsat+Satellite%29"><img border="1" src="assets/MLima"
ges/SmallLarge146.jpg"/></a> <b><a href="datasets/Statlog+%28Landsat+Satellit
e%29">Statlog (Landsat Satellite)</a></b>
<!-- <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  -->
Multivariate 
Classification 
Integer 
6435 
36 
1993 
<!-- <td>Physical&nbsp; -->
<a href="datasets/Statlog+%28Image+Segmentation%29"><img border="1" src="assets/ML
images/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Statlog+%28Image+Se"
gmentation%29">Statlog (Image Segmentation)</a></b>
<!-- <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.  -->
Multivariate 
Classification 
Real 
2310 
19 
1990 
<!-- <td>Other&nbsp; -->
```

```
</td
Large92.jpg"/></a> class="normal"><b><a href="datasets/Statlog+%28Shuttle%29">Statlog (Shuttle%29">Statlog (Shuttle%29")
e)</a></b>
<!-- <td>The shuttle dataset contains 9 attributes all of which are numerical. Approximately 8
0% of the data belongs to class 1  -->
Multivariate 
Classification 
Integer 
58000 
9 

<!-- <td>Physical&nbsp; -->
<a href="datasets/Statlog+%28Vehicle+Silhouettes%29"><img border="1" src="assets/MLi
mages/SmallLarge149.jpg"/></a> class="normal"><b><a href="datasets/Statlog+%28Vehicle+Silho">
uettes%29">Statlog (Vehicle Silhouettes)</a></b>
<!-- <td>3D objects within a 2D image by application of an ensemble of shape feature extrac
tors to the 2D silhouettes of the objects.  -->
Multivariate 
Classification 
Integer 
946 
18 

<!-- <td>Other&nbsp; -->
<a href="datasets/Connectionist+Bench+%28Nettalk+Corpus%29"><img border="1" src="a
ssets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Connectionist"><b><a href="datasets/Connectionist"><b><a href="datasets/Connectionist"><b /><a href="datasets/Connectionist"><a href="datasets/Connection
+Bench+%28Nettalk+Corpus%29">Connectionist Bench (Nettalk Corpus)</a></b></d>
<!-- <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 
Multivariate 

Categorical 
20008 
4 

<!-- <td>Other&nbsp; -->
+W28Sonar%2C+Mines+vs.+Rocks%29"><img bor
der="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets"
/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29">Connectionist Bench (Sonar, Mines vs. Rocks)<
/a></b>
<!-- <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.  -->
Multivariate 
Classification 
Real 
208 
60 

<!-- <td>Physical&nbsp; -->
<a href="datasets/Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29"><
img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="d
atasets/Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29">Connectionist Bench (Vowel Re
cognition - Deterding Data)</a></b>
<!-- <td>Speaker independent recognition of the eleven steady state vowels of British Englis
h using a specified training set of lpc derived log area ratios.  -->

Classification 
Real
```

```
528 
10 

<!-- <td>Other&nbsp; -->
="1" src="assets/MLimages/SmallLar"
ge153.jpg"/></a> <b><a href="datasets/Economic+Sanctions">Economic Sanction
s</a></b>
<!-- <td>Domain Theory on Economic Sanctions; Undocumented&nbsp; -->
Domain-Theory 

<!-- <td>Financial&nbsp; -->
g"/></a> <b><a href="datasets/Protein+Data">Protein Data</a></b>
<!-- <td>Undocumented&nbsp; -->

<!-- <td>Life&nbsp; -->
<a href="datasets/Cloud"><img border="1" src="assets/MLimages/SmallLarge155.jpg"/></a
> <b><a href="datasets/Cloud">Cloud</a></b>
<!-- <td>Little Documentation&nbsp; -->
Multivariate 

Real 
1024 
10 
1989 
<!-- <td>Physical&nbsp; -->
s/SmallLarge156.jpg"/></a> <b><a href="datasets/Callt2+Building+People+Counts"
">Callt2 Building People Counts</a></b>
<!-- <td>This data comes from the main door of the Callt2 building at UCI.&nbsp; -
->
Multivariate, Time-Series 

Categorical, Integer 
10080 
4 
2006 
<!-- <td>Other&nbsp; -->
="1" src="assets/MLimages/SmallL
arge157.jpg"/></a> class="normal"><b><a href="datasets/Dodgers+Loop+Sensor">Dodgers Loop
Sensor</a></b>
<!-- <td>Loop sensor data was collected for the Glendale on ramp for the 101 North freeway
in Los Angeles  -->
Multivariate, Time-Series 

Categorical, Integer 
50400
```

```
3 
2006 
<!-- <td>Other&nbsp; -->
<a href="datasets/Poker+Hand"><img border="1" src="assets/MLimages/SmallLarge158.jp"
g"/></a> <b><a href="datasets/Poker+Hand">Poker Hand</a></b>
table>
<!-- <td>Purpose is to predict poker hands&nbsp; -->
Multivariate 
Classification 
Categorical, Integer 
1025010 
11 
2007 
<!-- <td>Game&nbsp; -->
<tm>border="1" src="assets/MLimages/S
mallLarge159.jpg"/></a> class="normal"><b><a href="datasets/MAGIC+Gamma+Telescope">MAG
IC Gamma Telescope</a></b>
<!-- <td>Data are MC generated to simulate registration of high energy gamma particles in a
n atmospheric Cherenkov telescope  -->
Multivariate 
Classification 
Real 
19020 
11 
2007 
<!-- <td>Physical&nbsp; -->
<a href="datasets/UJI+Pen+Characters"><img border="1" src="assets/MLimages/SmallLar"
ge160.jpg"/></a> <b><a href="datasets/UJI+Pen+Characters">UJI Pen Characters
</a></b>
<!-- <td>Data consists of written characters in a UNIPEN-like format&nbsp; -->
Multivariate, Sequential 
Classification 
Integer 
1364 

2007 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Mammographic+Mass"><img border="1" src="assets/MLimages/SmallLa"
rgedefault.jpg"/></a> <b><a href="datasets/Mammographic+Mass">Mammographi
c Mass</a></b>
<!-- <td>Discrimination of benign and malignant mammographic masses based on BI-RADS
attributes and the patient's age.  -->
Multivariate 
Classification 
Integer 
961 
6 
2007 
<!-- <td>Life&nbsp; -->
<a href="datasets/Forest+Fires"><img border="1" src="assets/MLimages/SmallLarge162.jp"
g"/></a> <b><a href="datasets/Forest+Fires">Forest Fires</a></b>
/table>
<!-- <td>This is a difficult regression task, where the aim is to predict the burned area of fore
st fires, in the northeast region of Portugal, by using meteorological and other data (see details at: http://www.d
si.uminho.pt/~pcortez/forestfires).  -->
Multivariate 
Regression
```

```
Real 
517 
13 
2008 
<!-- <td>Physical&nbsp; -->
SmallLargedefault.jpg"/></a> <b><a href="datasets/Reuters+Transcribed+Subset"
>Reuters Transcribed Subset</a></b>
<!-- <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.  -->
Text 
Classification 

200 

2008 
<!-- <td>Business&nbsp; -->
<a href="datasets/Bag+of+Words"><img border="1" src="assets/MLimages/SmallLargedefa"
ult.jpg"/></a> <b><a href="datasets/Bag+of+Words">Bag of Words</a></b></t
d>
<!-- <td>This data set contains five text collections in the form of bags-of-words.&nbsp;
/td> -->
Text 
Clustering 
Integer 
8000000 
100000 
2008 
<!-- <td>Other&nbsp; -->
<t
es/SmallLarge165.jpg"/></a> class="normal"><b><a href="datasets/Concrete+Compressive+Stren"><br/>class="normal"><b><a href="datasets/Concrete+Compressive+Stren"><br/>class="normal"><b><a href="datasets/Concrete+Compressive+Stren"><br/>class="normal"><b><a href="datasets/Concrete+Compressive+Stren"><br/>class="normal"><b><a href="datasets/Concrete+Compressive+Stren"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>clas
gth">Concrete Compressive Strength</a></b>
<!-- <td>Concrete is the most important material in civil engineering. The concrete compress
ive strength is a highly nonlinear function of age and ingredients.   -->
Multivariate 
Regression 
Real 
1030 
9 
2007 
<!-- <td>Physical&nbsp; -->
<a href="datasets/Hill-Valley"><img border="1" src="assets/MLimages/SmallLarge166.jpg"/
></a> <b><a href="datasets/Hill-Valley">Hill-Valley</a></b><
/td>
<!-- <td>Each record represents 100 points on a two-dimensional graph. When plotted in or
der (from 1 through 100) as the Y co-ordinate, the points will create either a Hill (a $\phi\bunder \text{bump}$ in the terrain) or a
Valley (a �dip� in the terrain).  -->
Sequential 
Classification 
Real 
606 
101 
2008 
<!-- <td>Other&nbsp; -->
<a href="datasets/Arcene"><img border="1" src="assets/MLimages/SmallLarge167.jpg"/></
a> <b><a href="datasets/Arcene">Arcene</a></b>
```

```
<!-- <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.  -->
Multivariate 
Classification 
Real 
900 
10000 
2008 
<!-- <td>Life&nbsp; -->
<a href="datasets/Dexter"><img border="1" src="assets/MLimages/SmallLarge168.jpg"/></
a> <b><a href="datasets/Dexter">Dexter</a></b>
<!-- <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.
  -->
Multivariate 
Classification 
Integer 
2600 
20000 
2008 
<!-- <td>Other&nbsp; -->
></a> class="normal"><b><a href="datasets/Dorothea">Dorothea</a></b></t
<!-- <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.  -->
Multivariate 
Classification 
Integer 
1950 
100000 
2008 
<!-- <td>Life&nbsp; -->
<a href="datasets/Gisette"><img border="1" src="assets/MLimages/SmallLarge170.jpg"/></
a> <b><a href="datasets/Gisette">Gisette</a></b>
<!-- <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
  -->
Multivariate 
Classification 
Integer 
13500 
5000 
2008 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Madelon"><img border="1" src="assets/MLimages/SmallLarge171.jpg"/>
</a> <b><a href="datasets/Madelon">Madelon</a></b>
<!-- <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.   -->
Multivariate 
Classification 
Real 
<n class="normal">4400 </n>
```

```
500 
2008 
<!-- <td>Other&nbsp; -->
="1" src="assets/MLimages/Small
Large172.jpg"/></a> class="normal"><b><a href="datasets/Ozone+Level+Detection">Ozone Level
Detection</a></b>
<!-- <td>Two ground ozone level data sets are included in this collection. One is the eight ho
ur peak set (eighthr.data), the other is the one hour peak set (onehr.data). Those data were collected from 199
8 to 2004 at the Houston, Galveston and Brazoria area.  -->
Multivariate, Sequential, Time-Series 
Classification 
Real 
2536 
73 
2008 
<!-- <td>Physical&nbsp; -->
<a href="datasets/Abscisic+Acid+Signaling+Network"><img border="1" src="assets/MLimag"
es/SmallLargedefault.jpg"/></a> <b><a href="datasets/Abscisic+Acid+Signaling+N"><b><a href="datasets/Abscisic+Acid+Signaling+N"><b><a href="datasets/Abscisic+Acid+Signaling+N"><b><a href="datasets/Abscisic+Acid+Signaling+N"><b><a href="datasets/Abscisic+Acid+Signaling+N"><b><a href="datasets/Abscisic+Acid+Signaling+N"><b><a href="datasets/Abscisic+Acid+Signaling+N"><b><a href="datasets/Abscisic+Acid+Signaling+N"><b><a href="datasets/Abscisic+Acid+Signaling+N"><b href="datasets/Abscisic+Acid+N"><b href="datasets/Abscisic+Acid+Signaling+N"><b href="datasets/Abscisic+Acid+N"><b href="datasets/Abscisic+Acid+N"><b href="d
etwork">Abscisic Acid Signaling Network</a></b>
<!-- <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 si
mulations using an asynchronous update scheme.   -->
Multivariate 
Causal-Discovery 
Integer 
300 
43 
2008 
<!-- <td>Life&nbsp; -->
<a href="datasets/Parkinsons"><img border="1" src="assets/MLimages/SmallLarge174.jpg"
/></a> <b><a href="datasets/Parkinsons">Parkinsons</a></b>
e>
<!-- <td>Oxford Parkinson's Disease Detection Dataset&nbsp; -->
Multivariate 
Classification 
Real 
197 
23 
2008 
<!-- <td>Life&nbsp; -->
="1" src="assets/MLimages/SmallL
arge175.jpg"/></a> <b><a href="datasets/Character+Trajectories">Character Traj
ectories</a></b>
<!-- <td>Multiple, labelled samples of pen tip trajectories recorded whilst writing individual ch
aracters. All samples are from the same writer, for the purposes of primitive extraction. Only characters with a s
ingle pen-down segment were considered.  -->
Time-Series 
Classification, Clustering 
Real 
2858 
3 
2008 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Blood+Transfusion+Service+Center"><img border="1" src="assets/MLim"></a>
ages/SmallLarge176.jpg"/></a> <b><a href="datasets/Blood+Transfusion+Service"><br/>tasets/Blood+Transfusion+Service
+Center">Blood Transfusion Service Center</a></b>
   stds on place-"normal". Data taken from the Placed Transfusion Service Center in Hein Chu City in Taiwan
```

```
- tu>-p diaso- normal >bata taken nom the blood Translasion oct vice ochter in risin ond oity
this is a classification problem.   -->
Multivariate 
Classification 
Real 
748 
5 
2008 
<!-- <td>Business&nbsp; -->
<ty><tm, description of the state of the s
MLimages/SmallLarge160.jpg"/></a> class="normal"><b><a href="datasets/UJI+Pen+Characters+"
%28Version+2%29">UJI Pen Characters (Version 2)</a></b>
<!-- <td>A pen-based database with more than 11k isolated handwritten characters&nbsp;</
p> -->
Multivariate, Sequential 
Classification 
Integer 
11640 

2009 
<!-- <td>Computer&nbsp; -->
<tm, big it = 1 min set = 
mallLarge178.jpg"/></a> class="normal"><b><a href="datasets/Semeion+Handwritten+Digit">Sem
eion Handwritten Digit</a></b>
<!-- <td>1593 handwritten digits from around 80 persons were scanned, stretched in a recta
ngular box 16x16 in a gray scale of 256 values.  -->
Multivariate 
Classification 
Integer 
1593 
256 
2008 
<!-- <td>Computer&nbsp; -->
</a> <b><a href="datasets/SECOM">SECOM</a></b>
<!-- <td>Data from a semi-conductor manufacturing process&nbsp; -->
Multivariate 
Classification, Causal-Discovery 
Real 
1567 
591 
2008 
<!-- <td>Computer&nbsp; -->
a> <b><a href="datasets/Plants">Plants</a></b>
<!-- <td>Data has been extracted from the USDA plants database. It contains all plants (spe
cies and genera) in the database and the states of USA and Canada where they occur.  -->
Multivariate 
Clustering 
Categorical 
22632 
70 
2008 
<!-- <td>Life&nbsp; -->
<a href="datasets/Libras+Movement"><img border="1" src="assets/MLimages/SmallLarge1" src="assets/MLimages/SmallArge1" src="assets/MLim
81.jpg"/></a> <b><a href="datasets/Libras+Movement">Libras Movement</a></b>
andalma 45 alabasa of 04 inota
```

```
<!-- <lu>Class= normal > rne data set contains ro classes of 24 instances each. Each class references to a
hand movement type in LIBRAS (Portuguese
name 'Lngua BRAsileira de Sinais', oficial brazilian signal language). 
Multivariate, Sequential 
Classification, Clustering 
Real 
360 
91 
2009 
<!-- <td>Other&nbsp; -->
<a href="datasets/Concrete+Slump+Test"><img border="1" src="assets/MLimages/SmallLa"
rge165.jpg"/></a> class="normal"><b><a href="datasets/Concrete+Slump+Test">Concrete Slump
Test</a></b>
<!-- <td>Concrete is a highly complex material. The slump flow of concrete is not only deter
mined by the water content, but that is also influenced by other concrete ingredients. 
Multivariate 
Regression 
Real 
103 
10 
2009 
<!-- <td>Computer&nbsp; -->
="1" src="assets/MLimages/Smaller">="1" src="assets/MLimages/Smaller">="1" src="assets/MLimages/Smaller">="1" src="assets/MLimages/Smaller">="1" src="assets/MLimages/Smaller">="1" src="assets/MLimages/Smaller">="1" src="assets/MLimages/Smaller">="1" src="assets/MLimages/Smaller">="1" src="assets/MLimages/Smaller">="1" src="assets/MLimages/Smaller">= 1" src="assets/MLimages/Smaller
ILarge 183.jpg"/></a><b><a href="datasets/Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime">Communities+and+Crime</a>
es and Crime</a></b>
<!-- <td>Communities within the United States. The data combines socio-economic data fro
m the 1990 US Census, law enforcement data from the 1990 US LEMAS survey, and crime data from the 1995
FBI UCR.  -->
Multivariate 
Regression 
Real 
1994 
128 
2009 
<!-- <td>Social&nbsp; -->
<a href="datasets/Acute+Inflammations"><img border="1" src="assets/MLimages/SmallLar"
ge184.jpg"/></a> <b><a href="datasets/Acute+Inflammations">Acute Inflammation
s</a></b>
<!-- <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.
  -->
Multivariate 
Classification 
Categorical, Integer 
120 
6 
2009 
<!-- <td>Life&nbsp; -->
<a href="datasets/Wine+Quality"><img border="1" src="assets/MLimages/SmallLarge186.j
pg"/></a> <b><a href="datasets/Wine+Quality">Wine Quality</a></b>
<!-- <td>Two datasets are included, related to red and white vinho verde wine samples, fro
m the north of Portugal. The goal is to model wine quality based on physicochemical tests (see [Cortez et al., 2
009], http://www3.dsi.uminho.pt/pcortez/wine/).  -->
Multivariate 
Classification, Regression 
Real 
4898
```

```
<ta>12 </ta>
2009 
<!-- <td>Business&nbsp; -->
<a href="datasets/URL+Reputation"><img border="1" src="assets/MLimages/SmallLarge18" |
7.jpg"/></a> <b><a href="datasets/URL+Reputation">URL Reputation</a></b>
<!-- <td>Anonymized 120-day subset of the ICML-09 URL data containing 2.4 million examp
les and 3.2 million features.  -->
Multivariate, Time-Series 
Classification 
Integer, Real 
2396130 
3231961 
2009 
<!-- <td>Computer&nbsp; -->
<a href="datasets/p53+Mutants"><img border="1" src="assets/MLimages/SmallLarge188.jp"
<!-- <td>The goal is to model mutant p53 transcriptional activity (active vs inactive) based on
data extracted from biophysical simulations.
  -->
Multivariate 
Classification 
Real 
16772 
5409 
2010 
<!-- <td>Life&nbsp; -->
<a href="datasets/Parkinsons+Telemonitoring"><img border="1" src="assets/MLimages/Sm
allLarge174.jpg"/></a> <b><a href="datasets/Parkinsons+Telemonitoring">Parkins
ons Telemonitoring</a></b>
<!-- <td>Oxford Parkinson's Disease Telemonitoring Dataset&nbsp; -->
Multivariate 
Regression 
Integer, Real 
5875 
26 
2009 
<!-- <td>Life&nbsp; -->
<a href="datasets/Demospongiae"><img border="1" src="assets/MLimages/SmallLarge190"
.ipg"/></a> <b><a href="datasets/Demospongiae">Demospongiae</a></b></t
d>
<!-- <td>Marine sponges of the Demospongiae class classification domain.&nbsp;
Multivariate 
Classification 
Integer 
503 

2010 
<!-- <td>Life&nbsp; -->
<a href="datasets/Opinosis+Opinion+%26frasl%3B+Review"><img border="1" src="assets/
MLimages/SmallLarge191.jpg"/></a> class="normal"><b><a href="datasets/Opinosis+Opinion+%2">
6frasl%3B+Review">Opinosis Opinion / Review</a></b>
<!-- <td>This dataset contains sentences extracted from user reviews on a given topic. Exa
mple topics are "performance of Toyota Camry" and "sound quality of ipod nano".  
Text
```

```
51 

2010 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Breast+Tissue"><img border="1" src="assets/MLimages/SmallLargedefa"
ult.jpg"/></a> <b><a href="datasets/Breast+Tissue">Breast Tissue</a></b></t
d>
<!-- <td>Dataset with electrical impedance measurements of freshly excised tissue samples
from the breast.  -->
Multivariate 
Classification 
Real 
106 
10 
2010 
<!-- <td>Life&nbsp; -->
efault.jpg"/></a> class="normal"><b><a href="datasets/Cardiotocography">Cardiotocography</a><
/b>
<!-- <td>The dataset consists of measurements of fetal heart rate (FHR) and uterine contrac
tion (UC) features on cardiotocograms classified by expert obstetricians. 
Multivariate 
Classification 
Real 
2126 
23 
2010 
<!-- <td>Life&nbsp; -->
<a href="datasets/Wall-Following+Robot+Navigation+Data"><img border="1" src="assets/M
Limages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Wall-Following+Robot">+Robot</a>
+Navigation+Data">Wall-Following Robot Navigation Data</a></b>
<!-- <td>The data were collected as the SCITOS G5 robot navigates through the room follo
wing the wall in a clockwise direction, for 4 rounds, using 24 ultrasound sensors arranged circularly around its '
waist'.  -->
Multivariate, Sequential 
Classification 
Real 
5456 
24 
2010 
<!-- <td>Computer&nbsp; -->
="1" src="assets/MLimages/SmallLar"
ge195.jpg"/></a> <b><a href="datasets/Spoken+Arabic+Digit">Spoken Arabic Digit
</a></b>
<!-- <td>This dataset contains timeseries of mel-frequency cepstrum coefficients (MFCCs) c
orresponding to spoken Arabic digits. Includes data from 44 male and 44 female native Arabic speakers. & nbsp
; -->
Multivariate, Time-Series 
Classification 
Real 
8800 
13 
2010 
<!-- <td>Other&nbsp; -->
<a href="datasets/Localization+Data+for+Person+Activity"><img border="1" src="assets/ML"
```

```
images/SmallLargedefault.jpg"/></a> <b><a href="datasets/Localization+Data+for+"
Person+Activity">Localization Data for Person Activity</a></b>
<!-- <td>Data contains recordings of five people performing different activities. Each person
wore four sensors (tags) while performing the same scenario five times.  
Univariate, Sequential, Time-Series 
Classification 
Real 
164860 
8 
2010 
<!-- <td>Life&nbsp; -->
<a href="datasets/AutoUniv"><img border="1" src="assets/MLimages/SmallLargedefault.jp"
g"/></a> <b><a href="datasets/AutoUniv">AutoUniv</a></b>
<!-- <td>AutoUniv is an advanced data generator for classifications tasks. The aim is to refle
ct the nuances and heterogeneity of real data. Data can be generated in .csv, ARFF or C4.5 formats. </
p> -->
Multivariate 
Classification 
Categorical, Integer, Real 

2010 
<!-- <td>Other&nbsp; -->
edefault.jpg"/></a> <b><a href="datasets/Steel+Plates+Faults">Steel Plates Faults
</a></b>
<!-- <td>A dataset of steel plates' faults, classified into 7 different types.
The goal was to train machine learning for automatic pattern recognition.
  -->
Multivariate 
Classification 
Integer, Real 
1941 
27 
2010 
<!-- <td>Physical&nbsp; -->
es/SmallLargedefault.jpg"/></a> <b><a href="datasets/MiniBooNE+particle+identifi
cation">MiniBooNE particle identification</a></b>
<!-- <td>This dataset is taken from the MiniBooNE experiment and is used to distinguish ele
ctron neutrinos (signal) from muon neutrinos (background).  -->
Multivariate 
Classification 
Real 
130065 
50 
2010 
<!-- <td>Physical&nbsp; -->
="1" src="assets/MLimages/SmallLarg">="1" src="assets/MLimages/SmallLarg">= 1" src="assets/MLimages/SmallLarg
e203.jpg"/></a> class="normal"><b><a href="datasets/YearPredictionMSD">YearPredictionMSD</
a></b>
<!-- <td>Prediction of the release year of a song from audio features. Songs are mostly west
ern, commercial tracks ranging from 1922 to 2011, with a peak in the year 2000s.  -->
Multivariate 
Regression 
Real 
515345
```

```
90 
2011 
<!-- <td>Other&nbsp; -->
<a href="datasets/PEMS-SF"><img border="1" src="assets/MLimages/SmallLargedefault.jp"
g"/></a> <b><a href="datasets/PEMS-SF">PEMS-SF</a></b>
<!-- <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. 
Multivariate, Time-Series 
Classification 
Real 
440 
138672 
2011 
<!-- <td>Computer&nbsp; -->
<a href="datasets/OpinRank+Review+Dataset"><img border="1" src="assets/MLimages/S
mallLargedefault.jpg"/></a> <b><a href="datasets/OpinRank+Review+Dataset">O
pinRank Review Dataset</a></b>
<!-- <td>This data set contains user reviews of cars and and hotels collected from Tripadvis
or (~259,000
reviews) and Edmunds (~42,230 reviews).   -->
Text 

2011 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Relative+location+of+CT+slices+on+axial+axis"><img border="1" src="as
sets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Relative+locati
on+of+CT+slices+on+axial+axis">Relative location of CT slices on axial axis</a></b>
<!-- <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. 
Domain-Theory 
Regression 
Real 
53500 
386 
2011 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Online+Handwritten+Assamese+Characters+Dataset"><img border="1" s
rc="assets/MLimages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Online+"
Handwritten+Assamese+Characters+Dataset">Online Handwritten Assamese Characters Dataset</a></b>
<!-- <td>This is a dataset of 8235 online handwritten assamese characters. The "online" pro
cess involves capturing of data as text is written on a digitizing tablet with an electronic pen. 
Multivariate, Sequential 
Classification 
Integer 
8235 

2011 
<!-- <td>Computer&nbsp; -->
<a href="datasets/PubChem+Bioassay+Data"><img border="1" src="assets/MLimages/Sm
allLargedefault.ipg"/></a> class="normal"><b><a href="datasets/PubChem+Bioassay+Data">PubC
```

```
hem Bioassay Data</a></b>
<!-- <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. 
->
Multivariate 
Classification 
Integer, Real 

2011 
<!-- <td>Life&nbsp; -->
Limages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Record+Linkage+Co"><a href="datasets/
mparison+Patterns">Record Linkage Comparison Patterns</a></b>
<!-- <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. &nbsp.
; -->
Multivariate 
Classification 
Real 
5749132 
12 
2011 
<!-- <td>Other&nbsp; -->
<tme+Unnormalized"><img border="1" src="assets/
MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Communities+and+"><b><a href="datasets/Communities+and+"><b><a href="datasets/Communities+and+"><b href="datasets/Communitie
Crime+Unnormalized">Communities and Crime Unnormalized</a></b>
<!-- <td>Communities in the US. Data combines socio-economic data from the '90 Census, I
aw enforcement data from the 1990 Law Enforcement Management and Admin Stats survey, and crime data fr
om the 1995 FBI UCR  -->
Multivariate 
Regression 
Real 
2215 
147 
2011 
<!-- <td>Social&nbsp; -->
="1" src="assets/MLimages/SmallLarge">="1" src="assets/MLimages/SmallLarge">="2" src="assets/MLimages/SmallLarge">="3" src="assets/MLimages/SmallLarge="3" src="assets/MLimages/Small-arge="3" src="assets/MLimages/Small-arge="3" src="assets/MLimages/Small-arge=
default.jpg"/></a> <b><a href="datasets/Vertebral+Column">Vertebral Column</a>
</b>
<!-- <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> -->
Multivariate 
Classification 
Real 
310 
6 
2011 
<!-- <td>&nbsp; -->
<tm>border="1" src="assets/MLimag" src="assets
es/SmallLargedefault.jpg"/></a> <b><a href="datasets/EMG+Physical+Action+Data"
+Set">EMG Physical Action Data Set</a></b>
<!-- <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.  -->
Time-Series 
ctd>class="normal">Classification
```

```
Real 
10000 
8 
2011 
<!-- <td>Physical&nbsp; -->
<a href="datasets/Vicon+Physical+Action+Data+Set"><img border="1" src="assets/MLimag"
es/SmallLargedefault.jpg"/></a> <b><a href="datasets/Vicon+Physical+Action+Datasets/">
a+Set">Vicon Physical Action Data Set</a></b>
<!-- <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 10 subjects using the Vicon 3D tracker.&nb
sp; -->
Time-Series 
Classification 
Real 
3000 
27 
2011 
<!-- <td>Physical&nbsp; -->
<a href="datasets/Amazon+Commerce+reviews+set"><img border="1" src="assets/MLimag
es/SmallLargedefault.jpg"/></a> <b><a href="datasets/Amazon+Commerce+revie"><b><a href="datasets/Amazon+Commerce+revie"><a href="datasets/Amazon+revie"><a href=
ws+set">Amazon Commerce reviews set</a></b>
<!-- <td>The dataset is used for authorship identification in online Writeprint which is a new r
esearch field of pattern recognition.   -->
Multivariate, Text, Domain-Theory 
Classification 
Real 
1500 
10000 
2011 
<!-- <td>Physical&nbsp; -->
="1" src="assets/MLimages/Sm
allLargedefault.jpg"/></a> <b><a href="datasets/Amazon+Access+Samples">Amaz
on Access Samples</a></b>
<!-- <td>Amazon's InfoSec is getting smarter about the way Access data is leveraged. This i
s an anonymized sample of access provisioned within the company.   -->
Time-Series, Domain-Theory 
Regression, Clustering, Causal-Discovery 

30000 
20000 
2011 
<!-- <td>Business&nbsp; -->
<a href="datasets/Reuter 50 50"><img border="1" src="assets/MLimages/SmallLargedefa"
ult.jpg"/></a> <b><a href="datasets/Reuter_50_50">Reuter_50_50</a></b></t
d>
<!-- <td>The dataset is used for authorship identification in online Writeprint which is a new r
esearch field of pattern recognition.   -->
Multivariate, Text, Domain-Theory 
Classification, Clustering 
Real 
2500 
10000 
2011 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Farm+Ads"><img border="1" src="assets/MLimages/SmallLargedefault.j
pg"/></a> <b><a href="datasets/Farm+Ads">Farm Ads</a></b>
|__/td_
```

```
<!-- <td>This data was collected from text ads found on twelve websites that deal with variou
s farm animal related topics. The binary labels are based on whether or not the content owner approves of the
ad.  -->
Text 
Classification 

4143 
54877 
2011 
<!-- <td>Business&nbsp; -->
efault.jpg"/></a> class="normal"><b><a href="datasets/DBWorld+e-mails">DBWorld e-mails</a></
b>
<!-- <td>It contains 64 e-mails which I have manually collected from DBWorld mailing list. Th
ey are classified in: 'announces of conferences' and 'everything else'.  -->
Text 
Classification 

64 
4702 
2011 
<!-- <td>Computer&nbsp; -->
<a href="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29"><img border="1"><a href="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29"><img border="1"><a href="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29"><img border="1"><a href="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29"><img border="1"><a href="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29"><img border="1"><a href="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29"><<a href="datasets/KEGG+Metabolic+Rela
src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/KEGG"
+Metabolic+Relation+Network+%28Directed%29">KEGG Metabolic Relation Network (Directed)</a></b>
d>
<!-- <td>KEGG Metabolic pathways modeled as directed relation network. Variety of graphic
al features presented.  -->
Multivariate, Univariate, Text 
Classification, Regression, Clustering 
Integer, Real 
53414 
24 
2011 
<!-- <td>Life&nbsp; -->
<a href="datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29"><img border
="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/KE"</pre>
GG+Metabolic+Reaction+Network+%28Undirected%29">KEGG Metabolic Reaction Network (Undirected)</a>
b>
<!-- <td>KEGG Metabolic pathways modeled as un-directed reaction network. Variety of gra
phical features presented.  -->
Multivariate, Univariate, Text 
Classification, Regression, Clustering 
Integer, Real 
65554 
29 
2011 
<!-- <td>Life&nbsp; -->
<a href="datasets/Bank+Marketing"><img border="1" src="assets/MLimages/SmallLargede"
fault.jpg"/></a> <b><a href="datasets/Bank+Marketing">Bank Marketing</a></b></
p>
<!-- <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).&nbs
p; -->
Multivariate 
Classification 
Real 
 td. in along "normal", 15011 ./n. ./ta
```

```
<u><u><u><u><u><u><<u><<u><<u><<u><<u><<u><<u><<u><<u><<u><<u><<u><<u><<u><<u><<u><<u><<u><<u><<u><<u><</p></u>
17 
2012 
<!-- <td>Business&nbsp; -->
<a href="datasets/YouTube+Comedy+Slam+Preference+Data"><img border="1" src="asset"
s/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/YouTube+Comed"><br/>to=+Comed
y+Slam+Preference+Data">YouTube Comedy Slam Preference Data</a></b>
<!-- <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 metad
ata.  -->
Text 
Classification 

1138562 
3 
2012 
<!-- <td>Computer&nbsp; -->
es/SmallLargedefault.jpg"/></a> <b><a href="datasets/Gas+Sensor+Array+Drift+D"><br/>the contract of the contract of
ataset">Gas Sensor Array Drift Dataset</a></b>
<!-- <td>This archive contains 13910 measurements from 16 chemical sensors utilized in si
mulations for drift compensation in a discrimination task of 6 gases at various levels of concentrations. </
p> -->
Multivariate 
Classification 
Real 
13910 
128 
2012 
<!-- <td>Computer&nbsp; -->
<a href="datasets/ILPD+%28Indian+Liver+Patient+Dataset%29"><img border="1" src="ass
ets/MLimages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/ILPD+%28India"><b>
n+Liver+Patient+Dataset%29">ILPD (Indian Liver Patient Dataset)</a></b>
<!-- <td>This data set contains 10 variables that are age, gender, total Bilirubin, direct Bilirub
in, total proteins, albumin, A/G ratio, SGPT, SGOT and Alkphos. 
Multivariate 
Classification 
Integer, Real 
583 
10 
2012 
<!-- <td>Life&nbsp; -->
<a href="datasets/OPPORTUNITY+Activity+Recognition"><img border="1" src="assets/MLi
mages/SmallLarge226.jpg"/></a> class="normal"><b><a href="datasets/OPPORTUNITY+Activity+">+Activity+</a>
Recognition">OPPORTUNITY Activity Recognition</a></b>
<!-- <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). 
Multivariate, Time-Series 
Classification 
Real 
2551 
242 
2012 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Nomao"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/
></a> <b><a href="datasets/Nomao">Nomao</a></b>
```

```
<!-- <ld>class= normal >normal 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.  -->
Univariate 
Classification 
Real 
34465 
120 
2012 
<!-- <td>Computer&nbsp; -->
<a href="datasets/SMS+Spam+Collection"><img border="1" src="assets/MLimages/SmallL
argedefault.jpg"/></a> <b><a href="datasets/SMS+Spam+Collection">SMS Spam
Collection</a></b>
<!-- <td>The SMS Spam Collection is a public set of SMS labeled messages that have been
collected for mobile phone spam research.  -->
Multivariate, Text, Domain-Theory 
Classification, Clustering 
Real 
5574 

2012 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Skin+Segmentation"><img border="1" src="assets/MLimages/SmallLarg"
edefault.jpg"/></a> <b><a href="datasets/Skin+Segmentation">Skin Segmentation">
</a></b>
<!-- <td>The Skin Segmentation dataset is constructed over B, G, R color space. Skin and N
onskin dataset is generated using skin textures from face images of diversity of age, gender, and race people.&
nbsp; -->
Univariate 
Classification 
Real 
245057 
4 
2012 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Planning+Relax"><img border="1" src="assets/MLimages/SmallLargedef"
ault.jpg"/></a> class="normal"><b><a href="datasets/Planning+Relax">Planning Relax</a></b>
>
<!-- <td>The dataset concerns with the classification of two mental stages from recorded EE
G signals: Planning (during imagination of motor act) and Relax state.   -->
Univariate 
Classification 
Real 
182 
13 
2012 
<!-- <td>Computer&nbsp; -->
<tm; = "datasets/PAMAP2+Physical+Activity+Monitoring"><img border="1" src="assets/ML"
images/SmallLargedefault.jpg"/></a> <b><a href="datasets/PAMAP2+Physical+Act"><a href="datasets/PAMAP2+Physical+Act">datasets/PAMAP2+Physical+Act</a></a>
ivity+Monitoring">PAMAP2 Physical Activity Monitoring</a></b></to>
cal activities, performed by 9 subjects wearing 3 inertial measurement units and a heart rate monitor. </p
Multivariate, Time-Series 
Classification 
Real 
3850505
```

```
<pcd><pcd>
2012 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Restaurant+%26+consumer+data"><img border="1" src="assets/MLimag"></a>
es/SmallLargedefault.jpg"/></a> <b><a href="datasets/Restaurant+%26+consumer"><b><a href="datasets/Restaurant+%26+consumer"><b><a href="datasets/Restaurant+%26+consumer"><b><a href="datasets/Restaurant+%26+consumer"><b><a href="datasets/Restaurant+%26+consumer"><b><a href="datasets/Restaurant+%26+consumer"><b><a href="datasets/Restaurant+%26+consumer"><b><a href="datasets/Restaurant+%26+consumer"><b href="datasets/Restaurant
+data">Restaurant & consumer data</a></b>
<!-- <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.  
Multivariate 

138 
47 
2012 
<!-- <td>Computer&nbsp; -->
<a href="datasets/CNAE-9"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"</a>
/></a> <d>><b><a href="datasets/CNAE-9">CNAE-9</a></b>
<!-- <td>This is a data set containing 1080 documents of free text business descriptions of B
razilian companies categorized into a
subset of 9 categories  -->
Multivariate, Text 
Classification 
Integer 
1080 
857 
2012 
<!-- <td>Business&nbsp; -->
+ lose to the consumption of the consum
c="assets/MLimages/SmallLargedefault.jpg"/></a> c class="normal"><b><a href="datasets/Individual"><b><a href="datasets/Individual"><b class="normal"><b><a href="datasets/Individual"><b class="normal"><b class="normal"><
+household+electric+power+consumption">Individual household electric power consumption</a></b>
<!-- <td>Measurements of electric power consumption in one household with a one-minute s
ampling rate over a period of almost 4 years. Different electrical quantities and some sub-metering values are a
vailable.  -->
Multivariate, Time-Series 
Regression, Clustering 
Real 
2075259 
9 
2012 
<!-- <td>Physical&nbsp; -->
<tmages/SmallLargedefault.jpg"/>
</a> <b><a href="datasets/seeds">seeds</a></b>
<!-- <td>Measurements of geometrical properties of kernels belonging to three different vari
eties of wheat. A soft X-ray technique and GRAINS package were used to construct all seven, real-valued attrib
utes.  -->
Multivariate 
Classification, Clustering 
Real 
210 
7 
2012 
<!-- <td>Life&nbsp; -->
<a href="datasets/Northix"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/
></a> <b><a href="datasets/Northix">Northix"
<!-- <td>Northix is designed to be a schema matching benchmark problem for data integrati
```

```
on of two entity relationship databases.   -->
Multivariate, Univariate, Text 
Classification 
Integer, Real 
115 
200 
2012 
<!-- <td>Computer&nbsp; -->
efault.jpg"/></a> class="normal"><b><a href="datasets/QtyT40I10D100K">QtyT40I10D100K</a></
b>
<!-- <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 
Sequential 

Integer 
3960456 
4 
2012 
<!-- <td>&nbsp; -->
<a href="datasets/Legal+Case+Reports"><img border="1" src="assets/MLimages/SmallLar"
gedefault.jpg"/></a> <b><a href="datasets/Legal+Case+Reports">Legal Case Rep
orts</a></b>
<!-- <td>A textual corpus of 4000 legal cases for automatic summarization and citation analy
sis. For each document we collect catchphrases, citations sentences, citation catchphrases and citation classes
.  -->
Text 
Classification 

2012 
<!-- <td>Other&nbsp; -->
<a href="datasets/Human+Activity+Recognition+Using+Smartphones"><img border="1" src
="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Human+A"
ctivity+Recognition+Using+Smartphones">Human Activity Recognition Using Smartphones</a></b>
r>
<!-- <td>Human Activity Recognition database built from the recordings of 30 subjects perfor
ming activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors.
  -->
Multivariate, Time-Series 
Classification, Clustering 

10299 
561 
2012 
<!-- <td>Computer&nbsp; -->
<a href="datasets/One-hundred+plant+species+leaves+data+set"><img border="1" src="as
sets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/One-hundred+"
plant+species+leaves+data+set">One-hundred plant species leaves data set</a></b>
<!-- <td>Sixteen samples of leaf each of one-hundred plant species. For each sample, a sha
pe descriptor, fine scale margin and texture histogram are given.  -->

Classification 
Real 
1600 
64
```

```
2012 
<!-- <td>Life&nbsp; -->
<a href="datasets/Energy+efficiency"><img border="1" src="assets/MLimages/SmallLarged"
efault.jpg"/></a> class="normal"><b><a href="datasets/Energy+efficiency">Energy efficiency</a></
b>
<!-- <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.   -->
Multivariate 
Classification, Regression 
Integer, Real 
768 
8 
2012 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Yacht+Hydrodynamics"><img border="1" src="assets/MLimages/SmallLa
rgedefault.jpg"/></a> <b><a href="datasets/Yacht+Hydrodynamics">Yacht Hydrod
ynamics</a></b>
<!-- <td>Delft data set, used to predict the hydodynamic performance of sailing yachts from
dimensions and velocity.  -->
Multivariate 
Regression 
Real 
308 
7 
2013 
<!-- <td>Physical&nbsp; -->
<a href="datasets/Fertility"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/
></a> <b><a href="datasets/Fertility">Fertility</a></b>
<!-- <td>100 volunteers provide a semen sample analyzed according to the WHO 2010 crite
ria. Sperm concentration are related to socio-demographic data, environmental factors, health status, and life h
abits  -->
Multivariate 
Classification, Regression 
Real 
100 
10 
2013 
<!-- <td>Life&nbsp; -->
<a href="datasets/Daphnet+Freezing+of+Gait"><img border="1" src="assets/MLimages/Sm
allLargedefault.jpg"/></a> <b><a href="datasets/Daphnet+Freezing+of+Gait">Dap
hnet Freezing of Gait</a></b>
<!-- <td>This dataset contains the annotated readings of 3 acceleration sensors at the hip a
nd leg of Parkinson's disease patients that experience freezing of gait (FoG) during walking tasks.
  -->
Multivariate, Time-Series 
Classification 
Real 
237 
9 
2013 
<!-- <td>Life&nbsp; -->
<a href="datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29"><img border
="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/3D"</pre>
+Road+Network+%28North+Jutland%2C+Denmark%29">3D Road Network (North Jutland, Denmark)</a></b>
/p>
<!-- <td>3D road network with highly accurate elevation information (+-20cm) from Denmark
used in eco-routing and fuel/Co2-estimation routing algorithms.  -->
```

```
Sequential, Text 
Regression, Clustering 
Real 
434874 
4 
2013 
<!-- <td>Computer&nbsp; -->
es/SmallLargedefault.jpg"/></a> <b><a href="datasets/ISTANBUL+STOCK+EXCH"><a href="datasets/ISTANBUL+STOCK+E
ANGE">ISTANBUL STOCK EXCHANGE</a></b>
<!-- <td>Data sets includes returns of Istanbul Stock Exchange with seven other international
I index; SP, DAX, FTSE, NIKKEI, BOVESPA, MSCE EU, MSCI EM from Jun 5, 2009 to Feb 22, 2011. 
> -->
Multivariate, Univariate, Time-Series 
Classification, Regression 
Real 
536 
8 
2013 
<!-- <td>Business&nbsp; -->
<a href="datasets/Buzz+in+social+media+"><img border="1" src="assets/MLimages/SmallL"
argedefault.jpg"/></a> <b><a href="datasets/Buzz+in+social+media+">Buzz in soci
al media </a></b>
<!-- <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.&
nbsp; -->
Time-Series, Multivariate 
Regression, Classification 
Integer, Real 
140000 
77 
2013 
<!-- <td>Computer&nbsp; -->
<tmp border="1" src="assets/MLimages/S" |
mallLargedefault.jpg"/></a> <b><a href="datasets/First-order+theorem+proving">F
irst-order theorem proving</a></b>
<!-- <td>Given a theorem, predict which of five heuristics will give the fastest proof when use
d by a first-order prover. A sixth prediction declines to attempt a proof, should the theorem be too difficult.&nbs
p; -->
Multivariate 
Classification 
Real 
6118 
51 
2013 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Move"><a href="datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Move*]</a>
ments+%28PUC-Rio%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <p cl
ass="normal"><b><a href="datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Movem">
ents+%28PUC-Rio%29">Wearable Computing: Classification of Body Postures and Movements (PUC-Rio)</a>
</b>
<!-- <td>A dataset with 5 classes (sitting-down, standing-up, standing, walking, and sitting) c
ollected on 8 hours of activities of 4 healthy subjects. We also established a baseline performance index. 
 -->
Sequential 
Classification 
Integer, Real 
165632
```

```
18 
2013 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Gas+sensor+arrays+in+open+sampling+settings"><img border="1" src="
assets/MLimages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Gas+sensor">
+arrays+in+open+sampling+settings">Gas sensor arrays in open sampling settings</a></b>
e>
<!-- <td>The dataset contains 18000 time-series recordings from a chemical detection platfo
rm at six different locations in a wind tunnel facility in response to ten high-priority chemical gaseous substance
s  -->
Multivariate, Time-Series 
Classification 
Real 
18000 
1950000 
2013 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Climate+Model+Simulation+Crashes"><img border="1" src="assets/MLi
mages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Climate+Model+Simula"><b><a href="datasets/Climate+Model+Simula"><b><a href="datasets/Climate+Model+Simula"><b><a href="datasets/Climate+Model+Simula"><b><a href="datasets/Climate+Model+Simula"><b><a href="datasets/Climate+Model+Simula"><b><a href="datasets/Climate+Model+Simula"><b><a href="datasets/Climate+Model+Simula"><b href="datasets/Climate+Model+Simula"><b href="datasets/Climate+Model+Simula"><a href="datasets/Climate+Model+Simula">
tion+Crashes">Climate Model Simulation Crashes</a></b>
<!-- <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.&nb
sp: -->
Multivariate 
Classification 
Real 
540 
18 
2013 
<!-- <td>Physical&nbsp; -->
<a href="datasets/MicroMass"><img border="1" src="assets/MLimages/SmallLargedefault.j
pg"/></a> <b><a href="datasets/MicroMass">MicroMass</a></b>
<!-- <td>A dataset to explore machine learning approaches for the identification of microorg
anisms from mass-spectrometry data.  -->
Multivariate 
Classification 
Real 
931 
1300 
2013 
<!-- <td>Life&nbsp; -->
<a href="datasets/QSAR+biodegradation"><img border="1" src="assets/MLimages/SmallLa"
rgedefault.jpg"/></a> <b><a href="datasets/QSAR+biodegradation">QSAR biodegr
adation</a></b>
<!-- <td>Data set containing values for 41 attributes (molecular descriptors) used to classify
1055 chemicals into 2 classes (ready and not ready biodegradable).  -->
Multivariate 
Classification 
Integer, Real 
1055 
41 
2013 
<!-- <td>Other&nbsp; -->
<a href="datasets/BLOGGER"><img border="1" src="assets/MLimages/SmallLargedefault.j
pg"/></a> <b><a href="datasets/BLOGGER">BLOGGER</a></b>
```

```
<!-- <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  -->
Multivariate 
Classification 

100 
6 
2013 
<!-- <td>Computer&nbsp; -->
<tm> border="1" src="assets/MLimages/S
mallLargedefault.jpg"/></a> <b><a href="datasets/Daily+and+Sports+Activities">D
aily and Sports Activities</a></b>
<!-- <td>The dataset comprises motion sensor data of 19 daily and sports activities each per
formed by 8 subjects in their own style for 5 minutes. Five Xsens MTx units are used on the torso, arms, and le
gs.
  -->
Multivariate, Time-Series 
Classification, Clustering 
Real 
9120 
5625 
2013 
<!-- <td>Computer&nbsp; -->
<tmg border="1" src="assets/MLimages/S
mallLargedefault.jpg"/></a> <b><a href="datasets/User+Knowledge+Modeling">Us
er Knowledge Modeling</a></b>
<!-- <td>It is the real dataset about the students' knowledge status about the subject of Elect
rical DC Machines. The dataset had been obtained from Ph.D. Thesis. 
Multivariate 
Classification, Clustering 
Integer 
403 
5 
2013 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+"><a href="datasets/Reuters+RCV1+RCV2+Multiview+Text+Categorization+"><a href="datasets/Reuters+RCV1+RCV2+Multiview+Text+Categorization+"><a href="datasets/Reuters+RCV1+RCV2+Multiview+Text+Categorization+"><a href="datasets/Reuters+RCV1+RCV2+Multiview+Text+Categorization+"><a href="datasets/Reuters+RCV1+RCV2+Multiview+Text+Categorization+"><a href="datasets/Reuters+RCV1+RCV2+Multiview+Text+Categorization+"><a href="datasets/Reuters+RCV1+RCV2+Multiview+Text+Categorization+"><a href="datasets/Reuters+RCV1+RCV2+Multiview+Text+Categorization+"><a href="datasets/Reuters+RCV1+RCV2+Multiview+Text+Categorization+"><a href="datasets/Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reute
Test+collection"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> 
mal"><b><a href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+colle
ction">Reuters RCV1 RCV2 Multilingual, Multiview Text Categorization Test collection</a></b>
able>
<!-- <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.  
Multivariate 
Classification 
Real 
111740 

2013 
<!-- <td>Life&nbsp; -->
<a href="datasets/NYSK"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
</a> <b><a href="datasets/NYSK">NYSK</a></b>
<!-- <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 20
11).  -->
Multivariate, Sequential, Text 
Clustering 
tds on alacc-"normal"s o/ns o/tds
```

```
10421 
7 
2013 
<!-- <td>Social&nbsp; -->
<a href="datasets/Turkiye+Student+Evaluation"><img border="1" src="assets/MLimages/S
mallLargedefault.jpg"/></a> <b><a href="datasets/Turkiye+Student+Evaluation">T
urkiye Student Evaluation</a></b>
<!-- <td>This data set contains a total 5820 evaluation scores provided by students from Ga
zi University in Ankara (Turkey). There is a total of 28 course specific questions and additional 5 attributes. &nbs
p; -->
Multivariate 
Classification, Clustering 

5820 
33 
2013 
<!-- <td>Other&nbsp; -->
<a href="datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+"
d><b><a href="datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Le
vels+on+DC+Electrical+Machines%29">ser Knowledge Modeling Data (Students' Knowledge Levels on DC Elec
trical Machines)</a></b>
<!-- <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.  -->
Multivariate 
Classification 
Real 
403 
5 
2013 
<!-- <td>Computer&nbsp; -->
<a href="datasets/EEG+Eye+State"><img border="1" src="assets/MLimages/SmallLargedef"
ault.jpg"/></a> <b><a href="datasets/EEG+Eye+State">EEG Eye State</a></b></p
>
<!-- <td>The data set consists of 14 EEG values and a value indicating the eye state.&nbsp;
 -->
Multivariate, Sequential, Time-Series 
Classification 
Integer, Real 
14980 
15 
2013 
<!-- <td>Life&nbsp; -->
+Properties+of+Protein+Tertiary+Structure"><img borde
r="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/P"
hysicochemical+Properties+of+Protein+Tertiary+Structure">Physicochemical Properties of Protein Tertiary Stru
cture</a></b>
<!-- <td>This is a data set of Physicochemical Properties of Protein Tertiary Structure. The d
ata set is taken from CASP 5-9. There are 45730 decoys and size varying from 0 to 21 armstrong. 
td> -->
Multivariate 
Regression 
Real 
45730 
9 
2013
```

```
<!-- <\u)<p \( \text{id>} \) = \( \text{id>} \) \( \text{id>} \) \( \text{-->} \)
<a href="datasets/seismic-bumps"><img border="1" src="assets/MLimages/SmallLargedef"
ault.jpg"/></a> <b><a href="datasets/seismic-bumps">seismic-bumps</a></b>
<!-- <td>The data describe the problem of high energy (higher than 10^4 J) seismic bumps f
orecasting in a coal
mine. Data come from two of longwalls located in a Polish coal mine.  
Multivariate 
Classification 
Real 
2584 
19 
2013 
<!-- <td>Other&nbsp; -->
<a href="datasets/banknote+authentication"><img border="1" src="assets/MLimages/Small"
Largedefault.jpg"/></a> <b><a href="datasets/banknote+authentication">banknote
authentication</a></b>
<!-- <td>Data were extracted from images that were taken for the evaluation of an authentic
ation procedure for banknotes.  -->
Multivariate 
Classification 
Real 
1372 
5 
2013 
<!-- <td>Computer&nbsp; -->
<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> <b><a href="datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harv"><b><a href="datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harv"><b><a href="datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harv"><b><a href="datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harv"><b><a href="datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harv"><b><a href="datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harv"><b><a href="datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harv"><b href="datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harv"></b href="datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harv"><b href="datasets/USPTO+Algorithm+Challeng
ard+Tournament+Lab+and+TopCoder++++Problem%3A+Pat">USPTO Algorithm Challenge, run by NASA-Harv
ard Tournament Lab and TopCoder Problem: Pat</a></b>
<!-- <td>Data used for USPTO Algorithm Competition. Contains drawing pages from US pat
ents with manually labeled figure and part labels.   -->
Domain-Theory 
Classification 
Integer 
306 
5 
2013 
<!-- <td>Other&nbsp; -->
<a href="datasets/YouTube+Multiview+Video+Games+Dataset"><img border="1" src="asse
ts/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/YouTube+Multivi"><b><a href="datasets/YouTube+Multivi"><b><a href="datasets/YouTube+Multivi"><b><a href="datasets/YouTube+Multivi"><b><a href="datasets/YouTube+Multivi"><b><a href="datasets/YouTube+Multivi"><b><a href="datasets/YouTube+Multivi"><b><a href="datasets/YouTube+Multivi"><b><a href="datasets/YouTube+Multivi"><b href="datasets/YouTube+
ew+Video+Games+Dataset">YouTube Multiview Video Games Dataset</a></b>
<!-- <td>This dataset contains about 120k instances, each described by 13 feature types, wi
th class information, specially useful for exploring multiview topics (cotraining, ensembles, clustering,..). <
Multivariate, Text 
Classification, Clustering 
Integer, Real 
120000 
1000000 
2013 
<!-- <td>Computer&nbsp; -->
+at+Different+Concentrations"><img b
order="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datase"
ts/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations">Gas Sensor Array Drift Dataset at Different
Concentrations</a></b>
```

```
<!-- <ta>1 his archive contains 13910 measurements from 16 chemical sensors exposed to
6 different gases at various concentration levels.  -->
Multivariate, Time-Series 
Classification, Regression, Clustering, Causa 
Real 
13910 
129 
2013 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Se"><a href="datasets/Activities+of+Daily+Living+Maily+Binary+Se"><a href="datasets/Activities+of+Daily+Living+Maily+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+Binary+B
nsors"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><
a href="datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors">Activities of D
aily Living (ADLs) Recognition Using Binary Sensors</a></b>
<!-- <td>This dataset comprises information regarding the ADLs performed by two users on
a daily basis in their
own homes.   -->
Multivariate, Sequential, Time-Series 
Classification, Clustering 

2747 

2013 
<!-- <td>Computer&nbsp; -->
<a href="datasets/SkillCraft1+Master+Table+Dataset"><img border="1" src="assets/MLima"
ges/SmallLargedefault.jpg"/></a> <b><a href="datasets/SkillCraft1+Master+Table+"><b><a href="datasets/SkillCraft1+Master+Table+"><b><a href="datasets/SkillCraft1+Master+Table+"><b><a href="datasets/SkillCraft1+Master+Table+"><b><a href="datasets/SkillCraft1+Master+Table+"><b><a href="datasets/SkillCraft1+Master+Table+"><b><a href="datasets/SkillCraft1+Master+Table+"><b><a href="datasets/SkillCraft1+Master+Table+"><b><a href="datasets/SkillCraft1+Master+Table+"><b href="datasets/SkillC
Dataset">SkillCraft1 Master Table Dataset</a></b>
<!-- <td>This data was used in Thompson et al. (2013). A list of possible game actions is dis
cussed in Thompson, Blair, Chen, & Henrey (2013).  -->
Multivariate 
Regression 
Integer, Real 
3395 
20 
2013 
<!-- <td>Game&nbsp; -->
td>td><a href="datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units"</a>
><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href
="datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units">Weight Lifting Exercises m
onitored with Inertial Measurement Units</a></b>
<!-- <td>Six young health subjects were asked to perform 5 variations of the biceps curl wei
ght lifting exercise. One of the variations is the one predicted by the health professional. 
Multivariate 
Classification 
Real 
39242 
152 
2013 
<!-- <td>Physical&nbsp; -->
<a href="datasets/SML2010"><img border="1" src="assets/MLimages/SmallLargedefault.jp"
>
<!-- <td>This dataset is collected from a monitor system mounted in a domotic house. It corr
esponds to approximately 40 days of monitoring data.  -->
Multivariate, Sequential, Time-Series, Text 
Regression 
Real 
4137 
24
```

```
2014 
<!-- <td>Computer&nbsp; -->
<tm, description of the state of the stat
rgedefault.jpg"/></a> class="normal"><b><a href="datasets/Bike+Sharing+Dataset">Bike Sharing
Dataset</a></b>
<!-- <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. 
Univariate 
Regression 
Integer, Real 
17389 
16 
2013 
<!-- <td>Social&nbsp; -->
" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Predic
t+keywords+activities+in+a+online+social+media">Predict keywords activities in a online social media</a></b></
p>
<!-- <td>The data from Twitter was collected during 360 consecutive days. It was done by q
uerying 1497 English keywords sampled from Wikipedia. This dataset is proposed in a Learning to rank setting.
  -->
Multivariate, Sequential, Time-Series 

Integer, Real 
51 
35 
2013 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Thoracic+Surgery+Data"><img border="1" src="assets/MLimages/SmallL"
argedefault.jpg"/></a> <b><a href="datasets/Thoracic+Surgery+Data">Thoracic S
urgery Data</a></b>
<!-- <td>The data is dedicated to classification problem related to the post-operative life exp
ectancy in the lung cancer patients: class 1 - death within one year after surgery, class 2 - survival. 
Multivariate 
Classification 
Integer, Real 
470 
17 
2013 
<!-- <td>Life&nbsp; -->
<a href="datasets/EMG+dataset+in+Lower+Limb"><img border="1" src="assets/MLimages/
SmallLargedefault.jpg"/></a> <b><a href="datasets/EMG+dataset+in+Lower+Limb"><b><a href="datasets/EMG+dataset+in+Lower+Limb"><b href="datasets/EMG+dataset+in+Lower+Limb"><b href="datasets/EMG+dataset+in+Lower+Limb"><b href="datasets/EMG+dataset+in+Lower+Limb"><b href="datasets/EMG+dataset+in+Lower+Limb"><b href="datasets/EMG+dataset+in+Lower+Limb"><b href="dataset-in+Lower+Limb"><b href="dataset-in+L
">EMG dataset in Lower Limb</a></b>
<!-- <td>3 different exercises: sitting, standing and walking in the muscles: biceps femoris, v
astus medialis, rectus femoris and semitendinosus addition to goniometry in the exercises.  -->
Multivariate, Time-Series 

Real 
132 
5 
2014 
<!-- <td>Computer&nbsp; -->
<tmages/SmallLargedefault.jpg"/>
</a> <b><a href="datasets/SUSY">SUSY</a></b>
<!-- <td>This is a classification problem to distinguish between a signal process which produ
```

```
ces supersymmetric particles and a background process which does not.  -->

Classification 
Real 
5000000 
18 
2014 
<!-- <td>Physical&nbsp; -->
<a href="datasets/HIGGS"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/
></a> <b><a href="datasets/HIGGS">HIGGS</a></b>
<!-- <td>This is a classification problem to distinguish between a signal process which produ
ces Higgs bosons and a background process which does not.   -->

Classification 
Real 
11000000 
28 
2014 
<!-- <td>Physical&nbsp; -->
="1" src="assets/MLimages/SmallL"
argedefault.jpg"/></a> <b><a href="datasets/Qualitative_Bankruptcy">Qualitative_
Bankruptcy</a></b>
<!-- <td>Predict the Bankruptcy from Qualitative parameters from experts.&nbsp; -
->
Multivariate 
Classification 

250 
7 
2014 
<!-- <td>Computer&nbsp; -->
<tmp border="1" src="assets/MLimages/Sm.
allLargedefault.jpg"/></a> <b><a href="datasets/LSVT+Voice+Rehabilitation">LSV
T Voice Rehabilitation</a></b>
<!-- <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).  -->
Multivariate 
Classification 
Real 
126 
309 
2014 
<!-- <td>Life&nbsp; -->
+with+Wrist-worn+Accelerometer"><img b
order="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datase"><a href="datase"><b href="datase"><a href="datase"><
ts/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer">Dataset for ADL Recognition with Wrist-worn
n Accelerometer</a></b>
<!-- <td>Recordings of 16 volunteers performing 14 Activities of Daily Living (ADL) while carr
ying a single wrist-worn tri-axial accelerometer.  
Multivariate, Time-Series 
Classification, Clustering 

3 
2014 
<!-- <td>Computer&nbsp; -->
```

```
<tmages/SmallLargedefault.jpg"/></
a> <b><a href="datasets/Wilt">Wilt</a></b>
<!-- <td>High-resolution Remote Sensing data set (Quickbird). Small number of training sam
ples of diseased trees, large number for other land cover. Testing data set from stratified random sample of im
age.&nbsp: -->
Multivariate 
Classification 

4889 
6 
2014 
<!-- <td>Life&nbsp; -->
+ Cable - 
s/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/User+Identificatio"><b><a href="datasets/User+Identificatio"><b href=
n+From+Walking+Activity">User Identification From Walking Activity</a></b>
<!-- <td>The dataset collects data from an Android smartphone positioned in the chest pock
et from 22 participants walking in the wild over a predefined path.
  -->
Univariate, Sequential, Time-Series 
Classification, Clustering 
Real 

2014 
<!-- <td>Other&nbsp; -->
+Chest-Mounted+Accelerometer
border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datas"
ets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer">Activity Recognition from Single Chest-
Mounted Accelerometer</a></b>
<!-- <td>The dataset collects data from a wearable accelerometer mounted on the chest. Th
e dataset is intended for Activity Recognition research purposes.  -->
Univariate, Sequential, Time-Series 
Classification, Clustering 
Real 

2014 
<!-- <td>Other&nbsp; -->
<a href="datasets/Leaf"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/><
/a> <b><a href="datasets/Leaf">Leaf</a></b>
<!-- <td>This dataset consists in a collection of shape and texture features extracted from di
gital images of leaf specimens originating from a total of 40 different plant species.  
Multivariate 
Classification 
Real 
340 
16 
2014 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Dresses_Attribute_Sales"><img border="1" src="assets/MLimages/Small"
Largedefault.jpg"/></a> <b><a href="datasets/Dresses">Dresses
Attribute Sales</a></b>
<!-- <td>This dataset contain Attributes of dresses and their recommendations according to
their sales. Sales are monitor on the basis of alternate days.   -->
Text 
Classification, Clustering 

501
```

```
13 
2014 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Tamilnadu+Electricity+Board+Hourly+Readings"><img border="1" src="a
ssets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Tamilnadu+El
ectricity+Board+Hourly+Readings">Tamilnadu Electricity Board Hourly Readings</a></b>
<!-- <td>This data can be effectively produced the result to fewer parameter of the Load pro
file can be reduced in the Database   -->
Multivariate 
Classification, Regression, Clustering 
Real 
45781 
5 
2013 
<!-- <td>Life&nbsp; -->
<a href="datasets/Airfoil+Self-Noise"><img border="1" src="assets/MLimages/SmallLarged"><a href="datasets/Airfoil+Self-Noise"><img border="1" src="assets/MLimages/SmallLarged"></a>
efault.jpg"/></a> <b><a href="datasets/Airfoil+Self-Noise">Airfoil Self-Noise</a></b
>
<!-- <td>NASA data set, obtained from a series of aerodynamic and acoustic tests of two an
d three-dimensional airfoil blade sections conducted in an anechoic wind tunnel. 
Multivariate 
Regression 
Real 
1503 
6 
2014 
<!-- <td>Physical&nbsp; -->
<a href="datasets/Wholesale+customers"><img border="1" src="assets/MLimages/SmallLa"
rgedefault.jpg"/></a> <b><a href="datasets/Wholesale+customers">Wholesale cus
tomers</a></b>
<!-- <td>The data set refers to clients of a wholesale distributor. It includes the annual spend
ing in monetary units (m.u.) on diverse product categories  -->
Multivariate 
Classification, Clustering 
Integer 
440 
8 
2014 
<!-- <td>Business&nbsp; -->
<a href="datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis"><img border="1" src=
"assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Twitter+Dat"
a+set+for+Arabic+Sentiment+Analysis">Twitter Data set for Arabic Sentiment Analysis</a></b>
able>
<!-- <td>This problem of Sentiment Analysis (SA) has been studied well on the English langu
age but not Arabic one. Two main approaches have been devised: corpus-based and lexicon-based.  
Text 
Classification 

2000 
2 
2014 
<!-- <td>Social&nbsp; -->
<a href="datasets/Combined+Cycle+Power+Plant"><img border="1" src="assets/MLimages"><a href="datasets/Combined+Cycle+Power+Plant"><a href="datasets/Combined+Cycle+Power+Power+Plant"><a href="datasets/Combined+Cycle+Power+Plant"><
/SmallLargedefault.jpg"/></a> <b><a href="datasets/Combined+Cycle+Power+Plan"><br/>
t">Combined Cycle Power Plant</a></b></b>
```

```
<!-- <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.   -->
Multivariate 
Regression 
Real 
9568 
4 
2014 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Urban+Land+Cover"><img border="1" src="assets/MLimages/SmallLarg"
edefault.jpg"/></a> <b><a href="datasets/Urban+Land+Cover">Urban Land Cover">Urban Land Cover</a>
</a></b>
<!-- <td>Classification of urban land cover using high resolution aerial imagery. Intended to
assist sustainable urban planning efforts.  -->
Multivariate 
Classification 

168 
148 
2014 
<!-- <td>Physical&nbsp; -->
<a href="datasets/Diabetes+130-US+hospitals+for+years+1999-2008"><img border="1" src
="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Diabetes+"
130-US+hospitals+for+years+1999-2008">Diabetes 130-US hospitals for years 1999-2008</a></b>
r>
<!-- <td>This data has been prepared to analyze factors related to readmission as well as ot
outcomes pertaining to patients with diabetes.  -->
Multivariate 
Classification, Clustering 
Integer 
100000 
55 
2014 
<!-- <td>Life&nbsp; -->
<a href="datasets/Bach+Choral+Harmony"><img border="1" src="assets/MLimages/SmallL
argedefault.jpg"/></a> <b><a href="datasets/Bach+Choral+Harmony">Bach Choral
Harmony</a></b>
<!-- <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.  -->
Sequential 
Classification 

5665 
17 
2014 
<!-- <td>Other&nbsp; -->
<a href="datasets/StoneFlakes"><img border="1" src="assets/MLimages/SmallLargedefault"><a href="datasets/StoneFlakes"><img border="1" src="assets/MLimages/SmallLargedefault"><a href="datasets/StoneFlakes"><img border="1" src="assets/MLimages/SmallLargedefault"></a>
.jpg"/></a> <b><a href="datasets/StoneFlakes">StoneFlakes</a></b>
>
<!-- <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.  
Multivariate 
Classification, Clustering, Causal-Discovery 
n class="normal">Real </n>/td>
```

```
79 
8 
2014 
<!-- <td>Other&nbsp; -->
<a href="datasets/Tennis+Major+Tournament+Match+Statistics"><img border="1" src="ass
ets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Tennis+Major+"
Tournament+Match+Statistics">Tennis Major Tournament Match Statistics</a></b>
<!-- <td>This is a collection of 8 files containing the match statistics for both women and me
n at the four major tennis tournaments of the year 2013. Each file has 42 columns and a minimum of 76 rows.&
nbsp; -->
Multivariate 
Classification, Regression, Clustering 
Integer, Real 
127 
42 
2014 
<!-- <td>Other&nbsp; -->
+Multiple+Types+of+Sound+Recordin
gs"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a h
ref="datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings">Parkinson Speech D
ataset with Multiple Types of Sound Recordings</a></b>
<!-- <td>The training data belongs to 20 Parkinson's Disease (PD) patients and 20 healthy s
ubjects. From all subjects, multiple types of sound recordings (26) are taken. 
Multivariate 
Classification, Regression 
Integer, Real 
1040 
26 
2014 
<!-- <td>Life&nbsp; -->
<tm>border="1" src="assets/MLimages" |
/SmallLargedefault.jpg"/></a> <b><a href="datasets/Gesture+Phase+Segmentatio">
n">Gesture Phase Segmentation</a></b>
<!-- <td>The dataset is composed by features extracted from 7 videos with people gesticulat
ing, aiming at studying Gesture Phase Segmentation. It contains 50 attributes divided into two files for each vid
eo.  -->
Multivariate, Sequential, Time-Series 
Classification, Clustering 
Real 
9900 
50 
2014 
<!-- <td>Other&nbsp; -->
<a href="datasets/Perfume+Data"><img border="1" src="assets/MLimages/SmallLargedefa"
ult.jpg"/></a> <b><a href="datasets/Perfume+Data">Perfume Data</a></b></t
d>
<!-- <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. 
Univariate, Domain-Theory 
Classification, Clustering 
Integer 
560 
2 
2014 
<!-- <td>Computer&nbsp; -->
<a href="datasets/BlogFeedback"><img border="1" src="assets/MLimages/SmallLargedefa"
ultipa"/, /ax //td. td. in along "normal", ib. ia brof "datagets/PlagEagdhealt", PlagEagdhealt /ax /bx /b
```

```
uit.jpg /></a> </tu><u
d>
<!-- <td>Instances in this dataset contain features extracted from blog posts. The task assoc
iated with the data is to predict how many comments the post will receive.   -->
Multivariate 
Regression 
Integer, Real 
60021 
281 
2014 
<!-- <td>Social&nbsp; -->
<a href="datasets/REALDISP+Activity+Recognition+Dataset"><img border="1" src="assets/
MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/REALDISP+Activity"><b><a href="datasets/REALDISP+Activity"><b href="datasets/
+Recognition+Dataset">REALDISP Activity Recognition Dataset</a></b>
<!-- <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 algorit
hms   -->
Multivariate, Time-Series 
Classification 
Real 
1419 
120 
2014 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Newspaper+and+magazine+images+segmentation+dataset"><img bord
er="1" src="assets/MLimages/SmallLargedefault.ipg"/></a> <b><a href="datasets/
Newspaper+and+magazine+images+segmentation+dataset">Newspaper and magazine images segmentation
dataset</a></b>
<!-- <td>Dataset is well suited for segmentation tasks. It contains 101 scanned pages from d
ifferent newspapers and magazines in Russian with ground truth pixel-based masks.  -->

Classification 

101 

2014 
<!-- <td>Computer&nbsp; -->
<a href="datasets/AAAI+2014+Accepted+Papers"><img border="1" src="assets/MLimages/
SmallLargedefault.jpg"/></a> <b><a href="datasets/AAAI+2014+Accepted+Papers""
>AAAI 2014 Accepted Papers</a></b>
<!-- <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. 
Multivariate 
Clustering 

399 
6 
2014 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Gas+sensor+array+under+flow+modulation"><img border="1" src="asset"
s/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Gas+sensor+arra"><b><a href="datasets/Gas+sensor+arra"><b><a href="datasets/Gas+sensor+arra"><b><a href="datasets/Gas+sensor+arra"><b><a href="datasets/Gas+sensor+arra"><b><a href="datasets/Gas+sensor+arra"><b><a href="datasets/Gas+sensor+arra"><b><a href="datasets/Gas+sensor+arra"><b><a href="datasets/Gas+sensor+arra"><b href="datasets/Gas+senso
y+under+flow+modulation">Gas sensor array under flow modulation</a></b>
<!-- <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 eth
anol.  -->
Multivariate, Time-Series 
Classification, Regression 
Real
```

```
120432 
2014 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures"><img border="
1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Gas"
+sensor+array+exposed+to+turbulent+gas+mixtures">Gas sensor array exposed to turbulent gas mixtures</a>
</b>
<!-- <td>A chemical detection platform composed of 8 chemoresistive gas sensors was expo
sed to turbulent gas mixtures generated naturally in a wind tunnel. The acquired time series of the sensors are
provided.  -->
Multivariate, Time-Series 
Classification, Regression 
Real 
180 
150000 
2014 
<!-- <td>Computer&nbsp; -->
<a href="datasets/UJIIndoorLoc"><img border="1" src="assets/MLimages/SmallLargedefau"
lt.jpg"/></a> <b><a href="datasets/UJIIndoorLoc">UJIIndoorLoc</a></b>
<!-- <td>The UJIIndoorLoc is a Multi-Building Multi-Floor indoor localization database to test
Indoor Positioning System that rely on WLAN/WiFi fingerprint.  -->
Multivariate 
Classification, Regression 
Integer, Real 
21048 
529 
2014 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Sentence+Classification"><img border="1" src="assets/MLimages/SmallL"
argedefault.jpg"/></a> <b><a href="datasets/Sentence+Classification">Sentence C
lassification</a></b>
<!-- <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.  -->
Text 
Classification 
Integer 

2014 
<!-- <td>Other&nbsp; -->
default.jpg"/></a> <b><a href="datasets/Dow+Jones+Index">Dow Jones Index</a>
</b>
<!-- <td>This dataset contains weekly data for the Dow Jones Industrial Index. It has been u
sed in computational investing research.  -->
Time-Series 
Classification, Clustering 
Integer, Real 
750 
16 
2014 
<!-- <td>Business&nbsp; -->
<tmp border="1" src="assets/MLi" src="ass
mages/SmallLargedefault.jpg"/></a> <b><a href="datasets/sEMG+for+Basic+Hand"><b><a href="datasets/sEMG+for+Basic+Hand"><b><a href="datasets/sEMG+for+Basic+Hand"><b href="datasets/sEMG+fo
```

<lu>58 </lu>

```
+movements">sEMG for Basic Hand movements</a></b>
<!-- <td>The "sEMG for Basic Hand movements" includes 2 databases of surface electromy
ographic signals of 6 hand movements using Delsys' EMG System. Healthy subjects conducted six daily life gra
sps.  -->
Time-Series 
Classification 
Real 
3000 
2500 
2014 
<!-- <td>Life&nbsp; -->
<a href="datasets/AAAI+2013+Accepted+Papers"><img border="1" src="assets/MLimages/
SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/AAAI+2013+Accepted+Papers"
>AAAI 2013 Accepted Papers</a></b>
<!-- <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.  
Multivariate 
Clustering 

150 
5 
2014 
<!-- <td>Computer&nbsp; -->
<tmp border="1" src="assets/MLimag" src="assets/ML
es/SmallLargedefault.jpg"/></a> <b><a href="datasets/Geographical+Original+of+"
Music">Geographical Original of Music</a></b>
<!-- <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.
  -->
Multivariate 
Classification, Regression 
Real 
1059 
68 
2014 
<!-- <td>Other&nbsp; -->
+Naval+Propulsion+Plants"><img bord
er="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/
Condition+Based+Maintenance+of+Naval+Propulsion+Plants">Condition Based Maintenance of Naval Propulsi
on Plants</a></b>
<!-- <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.
  -->
Multivariate 
Regression 
Real 
11934 
16 
2014 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Grammatical+Facial+Expressions"><img border="1" src="assets/MLimag">
es/SmallLargedefault.jpg"/></a> <b><a href="datasets/Grammatical+Facial+Expre"><b><a href="datasets/Grammatical+Facial+Expre"><b><a href="datasets/Grammatical+Facial+Expre"><b><a href="datasets/Grammatical+Facial+Expre"><b><a href="datasets/Grammatical+Facial+Expre"><b><a href="datasets/Grammatical+Facial+Expre"><b><a href="datasets/Grammatical+Facial+Expre"><b><a href="datasets/Grammatical+Facial+Expre"><b><a href="datasets/Grammatical+Facial+Expre"><a href="datasets/Grammatical+Expre"><a href="datasets/Gra
ssions">Grammatical Facial Expressions</a></b>
<!-- <td>This dataset supports the development of models that make possible to interpret Gr
ammatical Facial Expressions from Brazilian Sign Language (Libras).  -->
Multivariate, Sequential 
Classification, Clustering
```

```
Real 
27965 
100 
2014 
<!-- <td>Computer&nbsp; -->
MLimages/SmallLargedefault.i
pg"/></a> class="normal"><b><a href="datasets/NoisyOffice">NoisyOffice</a></b>
able>
<!-- <td>Corpus intended to do cleaning (or binarization) and enhancement of noisy graysca
le printed text images using supervised learning methods. Noisy images and their corresponding ground truth p
rovided.  -->
Multivariate 
Classification, Regression 
Real 
216 
216 
2015 
<!-- <td>Computer&nbsp; -->
<a href="datasets/MHEALTH+Dataset"><img border="1" src="assets/MLimages/SmallLarg"><a href="datasets/MHEALTH+Dataset"><img border="1" src="assets/MLimages/SmallLarg"><a href="datasets/MHEALTH+Dataset"><a href="datasets/MHEALTH+Dataset"><a href="datasets/MLimages/SmallLarg"><a href="datasets/MLimages/SmallLarg"><a href="datasets/MHEALTH+Dataset"><a href="datasets/MLimages/SmallLarg"><a href="datasets/MHEALTH+Dataset"><a href="datasets/MLimages/SmallLarg"><a 
edefault.jpg"/></a> <b><a href="datasets/MHEALTH+Dataset">MHEALTH Dataset
</a></b>
<!-- <td>The MHEALTH (Mobile Health) dataset is devised to benchmark techniques dealing
with human behavior analysis based on multimodal body sensing.  -->
Multivariate, Time-Series 
Classification 
Real 
120 
23 
2014 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Student+Performance"><img border="1" src="assets/MLimages/SmallLa"
rgedefault.jpg"/></a> <b><a href="datasets/Student+Performance">Student Perfor
mance</a></b>
<!-- <td>Predict student performance in secondary education (high school). &nbsp;
Multivariate 
Classification, Regression 
Integer 
649 
33 
2014 
<!-- <td>Social&nbsp; -->
<a href="datasets/ElectricityLoadDiagrams20112014"><img border="1" src="assets/MLima"><a href="datasets/ElectricityLoadDiagrams20112014"><tm><a href="datasets/ElectricityLoadDiagrams20112014"><a href="datasets/ElectricityLoadDiagrams20112014"><a href="datasets/ElectricityLoadDiagrams20112014"><a href="datasets/ElectricityLoadDiagrams20112014"><a href="datasets/ElectricityLoadDiagrams20112014"><a href="datasets/ElectricityLoadDiagrams20112014"><a href="datasets/ElectricityLoadDiagrams20112014"><a href="datasets/ElectricityLoadDiagrams20112014"><a href="datasets/Electricity
ges/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/ElectricityLoadDiagrams20"><b><a href="datasets/ElectricityLoadDiagrams20"><b><a href="datasets/ElectricityLoadDiagrams20"><b><a href="datasets/ElectricityLoadDiagrams20"><b><a href="datasets/ElectricityLoadDiagrams20"><b><a href="datasets/ElectricityLoadDiagrams20"><b><a href="datasets/ElectricityLoadDiagrams20"><b><a href="datasets/ElectricityLoadDiagrams20"><b><a href="datasets/ElectricityLoadDiagrams20"><b href="datasets/Electrici
112014">ElectricityLoadDiagrams20112014</a></b>
<!-- <td>This data set contains electricity consumption of 370 points/clients.
  -->
Time-Series 
Regression, Clustering 
Real 
370 
140256 
2015 
<!-- <td>Computer&nbsp; -->
+ dynamic+gas+mixtures"><img border="1" src
="assets/MLimages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Gas+sens"
or+array+under+dynamic+gas+mixtures">Gas sensor array under dynamic gas mixtures</a></b>
```

```
<!-- <td>The data set contains the recordings of 16 chemical sensors exposed to two dynam
ic gas mixtures at varying concentrations. For each mixture, signals were acquired continuously during 12 hour
s.  -->
Multivariate, Time-Series 
Classification, Regression 
Real 
4178504 
19 
2015 
<!-- <td>Computer&nbsp; -->
<a href="datasets/microblogPCU"><img border="1" src="assets/MLimages/SmallLargedefa"
ult.jpg"/></a> <b><a href="datasets/microblogPCU">microblogPCU</a></b></t
d>
<!-- <td>MicroblogPCU data is crawled from sina weibo microblog[http://weibo.com/]. This d
ata can be used to study machine learning methods as well as do some social network research.  
td> -->
Multivariate, Univariate, Sequential, Text 
Classification, Causal-Discovery 
Integer, Real 
221579 
20 
2015 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Firm-Teacher_Clave-Direction_Classification"><img border="1" src="ass
ets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Firm-Teacher_
Clave-Direction_Classification">Firm-Teacher_Clave-Direction_Classification</a></b>
<!-- <td>The data are binary attack-point vectors and their clave-direction class(es) according
g to the partido-alto-based paradigm.  -->
Multivariate 
Classification 

10800 
20 
2015 
<!-- <td>Other&nbsp; -->
<tm> datasets/Dataset+for+Sensorless+Drive+Diagnosis"><img border="1" src="assets/
MLimages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Dataset+for+Senso"
rless+Drive+Diagnosis">Dataset for Sensorless Drive Diagnosis</a></b>
<!-- <td>Features are extracted from motor current. The motor has intact and defective com
ponents. This results in 11 different classes with different conditions.  
Multivariate 
Classification 
Real 
58509 
49 
2015 
<!-- <td>Computer&nbsp; -->
<a href="datasets/TV+News+Channel+Commercial+Detection+Dataset"><img border="1" s
rc="assets/MLimages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/TV+New"><a href="datasets/TV+New"><b href="datasets/TV+New"><b href="datasets/TV+New"><a href="datasets/TV+New"><a href="datasets/TV+New"><b href="datasets/TV+New"><b href="datasets/TV+New"><a href="datasets/TV+New">><a href="datasets/TV+New"><a href="datasets/TV+New">><a href="datasets/TV+N
s+Channel+Commercial+Detection+Dataset">TV News Channel Commercial Detection Dataset</a></b></t
d>
<!-- <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)
.   -->
Multivariate 
Classification, Clustering
```

```
Real 
129685 
12 
2015 
<!-- <td>Computer&nbsp; -->
default.jpg"/></a> <b><a href="datasets/Phishing+Websites">Phishing Websites
a></b>
<!-- <td>This dataset collected mainly from: PhishTank archive, MillerSmiles archive, Google
's searching operators.  -->

Classification 
Integer 
2456 
30 
2015 
<!-- <td>Computer Security&nbsp; -->
>><a href="datasets/Greenhouse+Gas+Observing+Network"><img border="1" src="assets/M
Limages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Greenhouse+Gas+O"><br/>class="normal"><b><a href="datasets/Greenhouse+Gas+O"><b href="datas
bserving+Network">Greenhouse Gas Observing Network</a></b>
<!-- <td>Design an observing network to monitor emissions of a greenhouse gas (GHG) in C
alifornia given time series of synthetic observations and tracers from weather model simulations.
  -->
Multivariate, Time-Series 
Regression 
Real 
2921 
5232 
2015 
<!-- <td>Physical&nbsp; -->
<te+Set"><img border="1" src="asset"
s/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Diabetic+Retinop"><b><a href="datasets/Diabetic+Retinop"><b href="datasets/Diabetic+Retinop"
athy+Debrecen+Data+Set">Diabetic Retinopathy Debrecen Data Set</a></b>
<!-- <td>This dataset contains features extracted from the Messidor image set to predict wh
ether an image contains signs of diabetic retinopathy or not.   -->
Multivariate 
Classification 
Integer, Real 
1151 
20 
2014 
<!-- <td>Life&nbsp; -->
="1" src="assets/MLimages/Sma">="1" src="assets/MLimages/Sma">
IILargedefault.jpg"/></a> <b><a href="datasets/HIV-1+protease+cleavage">HIV-1
protease cleavage</a></b>
<!-- <td>The data contains lists of octamers (8 amino acids) and a flag (-1 or 1) depending o
n whether HIV-1 protease will cleave in the central position (between amino acids 4 and 5). 
Multivariate 
Classification 
Categorical 
6590 
1 
2015 
<!-- <td>Life&nbsp; -->
<a href="datasets/Sentiment+Labelled+Sentences"><img border="1" src="assets/MLimage"></a>
s/SmallLargedefault.jpg"/></a> <b><a href="datasets/Sentiment+Labelled+Senten"><b><a href="datasets/Sentiment+Labelled+Senten"><b href="datasets/Sentiment+Labelled+Senten"><b href="datasets/Sentiment+Labelled+Senten"><b href="datasets/Sentiment+Labelled+Senten"><b href="datasets/Sentiment+Labelled+Senten"><b href="datasets/Sentiment+Labelled+Senten"><b href="datasets/Sentiment+Labelled+Senten"><b href="datasets/Sentiment+Labelled+Senten"><b href="datasets/Senten"><b href="datasets/Senten"><
ces">Sentiment Labelled Sentences</a></b>
```

```
<!-- <td>The dataset contains sentences labelled with positive or negative sentiment.&nbsp;
 -->
Text 
Classification 

3000 

2015 
<!-- <td>Other&nbsp; -->
<a href="datasets/Online+News+Popularity"><img border="1" src="assets/MLimages/Small"
Largedefault.jpg"/></a> <b><a href="datasets/Online+News+Popularity">Online Ne
ws Popularity</a></b>
<!-- <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
).  -->
Multivariate 
Classification, Regression 
Integer, Real 
39797 
61 
2015 
<!-- <td>Business&nbsp; -->
<tm, with the control of the contro
rgedefault.jpg"/></a> <b><a href="datasets/Forest+type+mapping">Forest type m
apping</a></b>
<!-- <td>Multi-temporal remote sensing data of a forested area in Japan. The goal is to map
different forest types using spectral data.  -->
Multivariate 
Classification 

326 
27 
2015 
<!-- <td>Life&nbsp; -->
<tm, as expected by the contraction of the contraction
/></a> <b><a href="datasets/wiki4HE">wiki4HE</a></b>
<!-- <td>Survey of faculty members from two Spanish universities on teaching uses of Wikip
edia  -->
Multivariate 
Regression, Clustering, Causal-Discovery 

913 
53 
2015 
<!-- <td>Social&nbsp; -->
<a href="datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset"><img b
order="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datase"><a href="datase"><b href="datase"><a href="datase"><
ts/Online+Video+Characteristics+and+Transcoding+Time+Dataset">Online Video Characteristics and Transcodi
ng Time Dataset</a></b>
<!-- <td>The dataset contains a million randomly sampled video instances listing 10 fundam
ental video characteristics along with the YouTube video ID.   -->
Multivariate 
Regression 
Integer, Real 
168286 
11 
<n class="normal">2015 </n>
```

```
<!-- <td>Computer&nbsp; -->
<a href="datasets/Chronic_Kidney_Disease"><img border="1" src="assets/MLimages/Small"
Largedefault.jpg"/></a> <b><a href="datasets/Chronic Kidney Disease">Chronic Kidney Disease</br>
Kidney_Disease</a></b>
<!-- <td>This dataset can be used to predict the chronic kidney disease and it can be collect
ed from the hospital nearly 2 months of period.  -->
Multivariate 
Classification 
Real 
400 
25 
2015 
<!-- <td>Other&nbsp; -->
<a href="datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+201"><a href="datasets/Machine+Learning+based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Based+Base
2-2014"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b>
<a href="datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014">Machine L
earning based ZZAlpha Ltd. Stock Recommendations 2012-2014</a></b>
<!-- <td>The data here are the ZZAlpha® machine learning recommendations made for vari
ous US traded stock portfolios the morning of each day during the 3 year period Jan 1, 2012 - Dec 31, 2014. &
nbsp; -->
Sequential, Time-Series 
Classification 
Real 
314080 
0 
2015 
<!-- <td>Business&nbsp; -->
<tmages/SmallLargedefault.jpg"/><
/a> <b><a href="datasets/Folio">Folio</a></b>
<!-- <td>20 photos of leaves for each of 32 different species.&nbsp; -->
Multivariate 
Classification, Clustering 

637 
20 
2015 
<!-- <td>Other&nbsp; -->
15"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a h
ref="datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015">Taxi Service Traject
ory - Prediction Challenge, ECML PKDD 2015</a></b>
<!-- <td>An accurate dataset describing trajectories performed by all the 442 taxis running in
the city of Porto, in Portugal.
  -->
Multivariate, Sequential, Time-Series, Domain-Theory 
Clustering, Causal-Discovery 
Real 
1710671 
9 
2015 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Cuff-Less+Blood+Pressure+Estimation"><img border="1" src="assets/ML"
images/SmallLargedefault.jpg"/></a> <b><a href="datasets/Cuff-Less+Blood+Pres"
sure+Estimation">Cuff-Less Blood Pressure Estimation</a></b>
<!-- <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. & nbsp;
tds on alace "normal"s Multivariate o/ns o/tds
```

```
Classification, Regression 
Real 
12000 
3 
2015 
<!-- <td>Life&nbsp; -->
+ Human + Activities + and + Postural + Tr
ansitions"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <
b><a href="datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions">Smartp
hone-Based Recognition of Human Activities and Postural Transitions</a></b>
<!-- <td>Activity recognition data set built from the recordings of 30 subjects performing basi
c activities and postural transitions while carrying a waist-mounted smartphone with embedded inertial sensors.
  -->
Multivariate, Time-Series 
Classification 
Real 
10929 
561 
2015 
<!-- <td>Life&nbsp; -->
<a href="datasets/Mice+Protein+Expression"><img border="1" src="assets/MLimages/Small">
ILargedefault.jpg"/></a> <b><a href="datasets/Mice+Protein+Expression">Mice Pr
otein Expression</a></b>
<!-- <td>Expression levels of 77 proteins measured in the cerebral cortex of 8 classes of con
trol and Down syndrome mice exposed to context fear conditioning, a task used to assess associative learning.
  -->
Multivariate 
Classification, Clustering 
Real 
1080 
82 
2015 
<!-- <td>Life&nbsp; -->
="1" src="assets/MLimages/SmallLarge">="1" src="assets/MLimages/SmallLarge">="2" src="assets/MLimages/SmallLarge">="3" src="assets/MLimages/SmallLarge">="4" src="assets/MLimages/SmallLarge">="4" src="assets/MLimages/SmallLarge">="4" src="assets/MLimages/SmallLarge">="4" src="assets/MLimages/SmallLarge">="4" src="assets/MLimages/SmallLarge">="4" src="assets/MLimages/SmallLarge">="4" src="assets/MLimages/SmallLarge">="4" src="assets/MLimages/SmallLarge">= 4" src="assets/MLimages/SmallLarge
default.jpg"/></a> <b><a href="datasets/UJIIndoorLoc-Mag">UJIIndoorLoc-Mag">UJIIndoorLoc-Mag</
a></b>
<!-- <td>The UJIIndoorLoc-Mag is an indoor localization database to test Indoor Positioning
System that rely on Earth's magnetic field variations.  -->
Multivariate, Sequential, Time-Series 
Classification, Regression, Clustering 
Integer, Real 
40000 
13 
2015 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Heterogeneity+Activity+Recognition"><img border="1" src="assets/MLim"
ages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Heterogeneity+Activity+R"
ecognition">Heterogeneity Activity Recognition</a></b>
<!-- <td>The Heterogeneity Human Activity Recognition (HHAR) dataset from Smartphones
and Smartwatches is a dataset devised to benchmark human activity recognition algorithms (classification, auto
matic data segmentation, sensor fusion, feature extraction, etc.) in real-world contexts; specifically, the dataset i
s gathered with a variety of different device models and use-scenarios, in order to reflect sensing heterogeneiti
es to be expected in real deployments.  -->
Multivariate, Time-Series 
Classification, Clustering 
Real 
43930257
```

```
<lu><lu></lu>
2015 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+"><a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+"><a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+"><a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+"><a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+"><a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+"><a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+"><a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+"><a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+"><a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Mining+%28EPM%29%3A+A+Learning+Mining+%28EPM%29%3A+A+Learning+Mining+%28EPM%29%3A+A+Learning+Mining+%28EPM%29%3A+A+Learning+Mining+%28EPM%29%3A+A+Learning+Mining+%28EPM%29%3A+A+Learning+Mining+%28EPM%29%3A+A+Learning+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining+Mining
Data+Set"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <</p>
b><a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set">Educa
tional Process Mining (EPM): A Learning Analytics Data Set</a></b>
<!-- <td>Educational Process Mining data set is built from the recordings of 115 subjects' act
ivities through a logging application while learning with an educational simulator.  -->
Multivariate, Sequential, Time-Series 
Classification, Regression, Clustering 
Integer 
230318 
13 
2015 
<!-- <td>Computer&nbsp; -->
<a href="datasets/HEPMASS"><img border="1" src="assets/MLimages/SmallLargedefault.j
pg"/></a> <b><a href="datasets/HEPMASS">HEPMASS</a></b>
ble>
<!-- <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. </p
> -->
Multivariate 
Classification 
Real 
10500000 
28 
2016 
<!-- <td>Physical&nbsp; -->
+ ref="datasets/Indoor+User+Movement+Prediction+from+RSS+data"><img border="1" s
rc="assets/MLimages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Indoor+"
User+Movement+Prediction+from+RSS+data">Indoor User Movement Prediction from RSS data</a></b>
td>
<!-- <td>This dataset contains temporal data from a Wireless Sensor Network deployed in re
al-world office environments. The task is intended as real-life benchmark in the area of Ambient Assisted Living.
  -->
Multivariate, Sequential, Time-Series 
Classification 
Real 
13197 
4 
2016 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Open+University+Learning+Analytics+dataset"><img border="1" src="as
sets/MLimages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Open+Universi"><b><a href="datasets/Open+Universi"><b><a href="datasets/Open+Universi"><b><a href="datasets/Open+Universi"><b><a href="datasets/Open+Universi"><b><a href="datasets/Open+Universi"><b><a href="datasets/Open+Universi"><b><a href="datasets/Open+Universi"><b><a href="datasets/Open+Universi"><b><a href="datasets/Open+Universi"><b href="datasets/Open+Universi"><
ty+Learning+Analytics+dataset">Open University Learning Analytics dataset</a></b>
<!-- <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 stude
nts.  -->
Multivariate, Sequential, Time-Series 
Classification, Regression, Clustering 
Integer 

2015 
<!-- <td>Computer&nbsp; -->
```

```
<ta><to><ta>table><to><ta><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta>fo><ta
SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/default+of+credit+card+clients"
>default of credit card clients</a></b>
<!-- <td>This research aimed at the case of customers' default payments in Taiwan and co
mpares the predictive accuracy of probability of default among six data mining methods.  -->
Multivariate 
Classification 
Integer, Real 
30000 
24 
2016 
<!-- <td>Business&nbsp; -->
<a href="datasets/Mesothelioma%E2%80%99s+disease+data+set+"><img border="1" src="
assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Mesotheliom"><b><a href="datasets/Mesotheliom"><b><a href="datasets/Mesotheliom"><b href="dat
a%E2%80%99s+disease+data+set+">Mesothelioma's disease data set </a></b>
<!-- <td>Mesothelioma's disease data set were prepared at Dicle University Faculty of Medic
ine in Turkey.
Three hundred and twenty-four Mesothelioma patient data. In the dataset, all samples have 34 features.  
 -->
Multivariate 
Classification 
Real 
324 
34 
2016 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Online+Retail"><img border="1" src="assets/MLimages/SmallLargedefau"
lt.jpg"/></a> class="normal"><b><a href="datasets/Online+Retail">Online Retail</a></b><
/tr>
<!-- <td>This is a transnational data set which contains all the transactions occurring betwee
n 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail.  
Multivariate, Sequential, Time-Series 
Classification, Clustering 
Integer, Real 
541909 
8 
2015 
<!-- <td>Business&nbsp; -->
<a href="datasets/SIFT10M"><img border="1" src="assets/MLimages/SmallLargedefault.jp"
g''/></a> <b><a href="datasets/SIFT10M">SIFT10M</a></b>
>
<!-- <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. </t
d> -->
Multivariate 
Causal-Discovery 
Integer 
11164866 
128 
2016 
<!-- <td>Computer&nbsp; -->
efault.jpg"/></a> class="normal"><b><a href="datasets/GPS+Trajectories">GPS Trajectories</a></
b>
<!-- <td>The dataset has been feed by Android app called Go!Track. It is available at Goolg
e Play Store(https://play.google.com/store/apps/details?id=com.go.router).   -->
Multivariate 
Classification, Regression
```

```
Real 
163 
15 
2016 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Detect+Malacious+Executable%28AntiVirus%29"><img border="1" src="
assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Detect+Mala"><b><a href="datasets/Detect+Mala"><b><a href="datasets/Detect+Mala"><b href="dat
cious+Executable%28AntiVirus%29">Detect Malacious Executable(AntiVirus)</a></b>
<!-- <td>| extract features from malacious and non-malacious and create and training datas
et to teach svm classifier. Dataset made of unknown executable to detect if it is virus or normal safe executable.
  -->
Multivariate 
Classification 
Real 
373 
513 
2016 
<!-- <td>Computer&nbsp; -->
="1" src="assets/MLimages/SmallL
argedefault.jpg"/></a> <b><a href="datasets/Occupancy+Detection+">Occupancy
Detection </a></b>
<!-- <td>Experimental data used for binary classification (room occupancy) from Temperatur
e, Humidity, Light and CO2. Ground-truth occupancy was obtained from time stamped pictures that were taken e
very minute.  -->
Multivariate, Time-Series 
Classification 
Real 
20560 
7 
2016 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring"><a href="datasets/Improved+Graphics+Tablet+for+Monitoring"><a href="datasets/Improved+Graphics+Tablet+for+Monitorin
+Parkinson%E2%80%99s+Disease"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> 
><b><a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+M"><b><a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+M"><b><a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+M"><b><a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+M"><b><a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+M"><b href="datasets/Improved+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+
onitoring+Parkinson%E2%80%99s+Disease">Improved Spiral Test Using Digitized Graphics Tablet for Monitori
ng Parkinson's Disease</a></b>
<!-- <td>Handwriting database consists of 25 PWP(People with Parkinson) and 15 healthy in
dividuals. Three types of recordings (Static Spiral Test, Dynamic Spiral Test and Stability Test) are taken. & nbsp;
 -->
Multivariate 
Classification, Regression, Clustering 
Real 
40 
7 
2016 
<!-- <td>Computer&nbsp; -->
<a href="datasets/News+Aggregator"><img border="1" src="assets/MLimages/SmallLarged"><a href="datasets/News+Aggregator"><img border="1" src="assets/MLimages/SmallLarged"><a href="datasets/News+Aggregator"><a href="dat
efault.jpg"/></a> class="normal"><b><a href="datasets/News+Aggregator">News Aggregator</a><
/b>
<!-- <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 s
ame story.  -->
Multivariate 
Classification, Clustering 

422937 
5
```

```
2016 
<!-- <td>Other&nbsp; -->
<a href="datasets/Air+Quality"><img border="1" src="assets/MLimages/SmallLargedefault.j
pg"/></a> <b><a href="datasets/Air+Quality">Air Quality</a></b>
ble>
<!-- <td>Contains the responses of a gas multisensor device deployed on the field in an Itali
an city. Hourly responses averages are recorded along with gas concentrations references from a certified anal
yzer.   -->
Multivariate, Time-Series 
Regression 
Real 
9358 
15 
2016 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Twin+gas+sensor+arrays"><img border="1" src="assets/MLimages/Small
ILargedefault.jpg"/></a> <b><a href="datasets/Twin+gas+sensor+arrays">Twin ga
s sensor arrays</a></b>
<!-- <td>5 replicates of an 8-MOX gas sensor array were exposed to different gas conditions
(4 volatiles at 10 concentration levels each).   -->
Multivariate, Time-Series, Domain-Theory 
Classification, Regression 
Real 
640 
480000 
2016 
<!-- <td>Computer&nbsp; -->
<tip><tip><timg</td>border="1" src="asse
ts/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Gas+sensors+for"><b><a href="datasets/Gas+sensors+for"><b href="datas
+home+activity+monitoring">Gas sensors for home activity monitoring</a></b>
<!-- <td>100 recordings of a sensor array under different conditions in a home setting: back
ground, wine and banana presentations. The array includes 8 MOX gas sensors, and humidity and temperatur
e sensors.
  -->
Multivariate, Time-Series 
Classification 
Real 
919438 
11 
2016 
<!-- <td>Computer&nbsp; -->
<tmp border="1" src="assets/M"
Limages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Facebook+Comment"><b><a href="datasets/Facebook+Comment"><br/>class="normal"><b><a href="datasets/Facebook+Comment"><br/>class="normal"><b><a href="datasets/Facebook+Comment"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>cla
+Volume+Dataset">Facebook Comment Volume Dataset</a></b>
<!-- <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.   -->
Multivariate 
Regression 
Integer, Real 
40949 
54 
2016 
<!-- <td>Other&nbsp; -->
+ Recognition + %28HAR %29 + in Recognition + %28HAR %
+Ambient+Assisted+Living+%28AAL%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a
> <b><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></b>
<!-- <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.   -->
Time-Series 
Classification 
Real 
5744 
561 
2016 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Polish+companies+bankruptcy+data"><img border="1" src="assets/MLim
ages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Polish+companies+bankr"><b><a href="datasets/Polish+companies+bankr"><b><a href="datasets/Polish+companies+bankr"><b><a href="datasets/Polish+companies+bankr"><b><a href="datasets/Polish+companies+bankr"><b><a href="datasets/Polish+companies+bankr"><b><a href="datasets/Polish+companies+bankr"><b><a href="datasets/Polish+companies+bankr"><b><a href="datasets/Polish+companies+bankr"><a href="datasets/Polish+c
uptcy+data">Polish companies bankruptcy data</a></b>
<!-- <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.&nbsp: -->
Multivariate 
Classification 
Real 
10503 
64 
2016 
<!-- <td>Business&nbsp; -->
<a href="datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28A"
ReM%29"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <
b><a href="datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29">Activit
y Recognition system based on Multisensor data fusion (AReM)</a></b>
<!-- <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. 
Multivariate, Sequential, Time-Series 
Classification 
Real 
42240 
6 
2016 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Dota2+Games+Results"><img border="1" src="assets/MLimages/SmallL
argedefault.jpg"/></a> <b><a href="datasets/Dota2+Games+Results">Dota2 Gam
es Results</a></b>
<!-- <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.  -->
Multivariate 
Classification 

102944 
116 
2016 
<!-- <td>Game&nbsp; -->
="1" src="assets/MLimages/SmallLarge">="1" src="assets/MLimages/SmallLarge">= 1" src="assets/MLimages/SmallLarge=="1" src="assets/MLimages/SmallArge=="1" src="assets/MLimages/SmallArge=="1" src="assets/MLimages/SmallArge=="1" src="assets/MLim
default.jpg"/></a> <b><a href="datasets/Facebook+metrics">Facebook metrics</a
></b>
<!-- <td>Facebook performance metrics of a renowned cosmetic's brand Facebook page.&n
bsp; -->
Multivariate 
Regression 
Integer 
500
```

```
19 
2016 
<!-- <td>Business&nbsp; -->
/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/UbiqLog+%28smar"><b><a href="datasets/UbiqLog+%28smar"><b><a href="datasets/UbiqLog+%28smar"><b><a href="datasets/UbiqLog+%28smar"><b><a href="datasets/UbiqLog+%28smar"><b><a href="datasets/UbiqLog+%28smar"><b><a href="datasets/UbiqLog+%28smar"><b><a href="datasets/UbiqLog+%28smar"><b href="datasets/UbiqLog+%28sm
tphone+lifelogging%29">UbiqLog (smartphone lifelogging)</a></b>
<!-- <td>UbiqLog is the smartphone lifelogging tool that runs on the smartphone of 35 users
for about 2 months.
  -->
Multivariate 
Causal-Discovery 

9782222 

2016 
<!-- <td>Computer&nbsp; -->
<a href="datasets/NIPS+Conference+Papers+1987-2015"><img border="1" src="assets/ML">
images/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/NIPS+Conference+Pa">this images/SmallLargedefault.jpg"/></a> 
pers+1987-2015">NIPS Conference Papers 1987-2015</a></b>
<!-- <td>This data set contains the distribution of words in the full text of the NIPS conference
e papers published from 1987 to 2015.  -->
Text 
Clustering 
Integer 
11463 
5812 
2016 
<!-- <td>Computer&nbsp; -->
></a> class="normal"><b><a href="datasets/HTRU2">HTRU2</a></b>
<!-- <td>Pulsar candidates collected during the HTRU survey. Pulsars are a type of star, of c
onsiderable scientific interest. Candidates must be classified in to pulsar and non-pulsar classes to aid discover
y.  -->
Multivariate 
Classification, Clustering 
Real 
17898 
9 
2017 
<!-- <td>Physical&nbsp; -->
<a href="datasets/Drug+consumption+%28quantified%29"><img border="1" src="assets/M
Limages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Drug+consumption+">consumption+</a>
%28quantified%29">Drug consumption (quantified)</a></b>
<!-- <td>Classify type of drug consumer by personality data&nbsp; -->
Multivariate 
Classification 
Real 
1885 
32 
2016 
<!-- <td>Social&nbsp; -->
<tm>border="1" src="assets/MLimages/
SmallLargedefault.jpg"/></a> <b><a href="datasets/Appliances+energy+prediction""><b><a href="datasets/Appliances+energy+prediction""><b><a href="datasets/Appliances+energy+prediction""><b><a href="datasets/Appliances+energy+prediction"</a>
>Appliances energy prediction</a></b>
<!-- <td>Experimental data used to create regression models of appliances energy use in a l
ow energy building.  -->
```

class="normal">MultivariateTime-Series

```
Regression 
Real 
19735 
29 
2017 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Miskolc+IIS+Hybrid+IPS"><img border="1" src="assets/MLimages/SmallL"
argedefault.jpg"/></a> <b><a href="datasets/Miskolc+IIS+Hybrid+IPS">Miskolc IIS
Hybrid IPS</a></b>
<!-- <td>The dataset was created for the comparison and evaluation of hybrid indoor positio
ning methods. The dataset presented contains data from W-LAN and Bluetooth interfaces, and Magnetometer.
  -->
Text 
Classification, Clustering, Causal-Discovery 
Integer 
1540 
67 
2016 
<!-- <td>Computer&nbsp; -->
<a href="datasets/KDC-4007+dataset+Collection"><img border="1" src="assets/MLimages/
SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/KDC-4007+dataset+Collection"
>KDC-4007 dataset Collection</a></b>
<!-- <td>KDC-4007 dataset Collection is the Kurdish Documents Classification text used in c
ategories regarding Kurdish Sorani news and articles.  -->
Multivariate, Text 
Classification, Regression 
Integer 
4007 

2017 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Geo-Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+"><a href="datasets/Geo-Magnetic+field+and+WLAN+dataset+for+indoor+field+and+"><a href="datasets/Geo-Magnetic+field+and+dataset+for+indoor+field+and+dataset+for+indoor+field+and+dataset+for+indoor+field+and+dataset+for+indoor+field+and+dataset+for+indoor+field+and+dataset+for+indoor+field+and+dataset+for+indoor+field+and+dataset+for+indoor+field+and+dataset+for+indoor+field+and+dataset+for+indoor+field+and+dataset+for+indoor+field+and+dataset+for+indoor+field+and+dataset+for+indoor+field+and+dataset+for+indoor+field+and+dataset+for+indoor+field+and+dataset+for+indoor+field+and+dataset+for+indoor+field+and+dataset+for+indoor+field+and+dataset+for+indoor+field+and+dataset+for+indoor+field+and+dataset+for+indoor+field+and+dataset+for+indoor+field+and+dataset+for+indoor+field+and+field+and+field+and+field+and+field+and+field+and+field+and+field+and+fie
wristband+and+smartphone"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <
p class="normal"><b><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 an
d smartphone</a></b>
<!-- <td>A multisource and multivariate dataset for indoor localisation methods based on WL
AN and Geo-Magnetic field fingerprinting  -->
Multivariate, Sequential, Time-Series 
Classification, Regression, Clustering 
Integer, Real 
153540 
25 
2017 
<!-- <td>Computer&nbsp; -->
g"/></a> <b><a href="datasets/DrivFace">DrivFace</a></b>
<!-- <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 feat
ures.  -->
Multivariate 
Classification, Regression, Clustering 
Real 
606 
6400 
2016 
~!-- <n class="normal"<Computer&nhen:</pre>//n> -->
```

```
efault.jpg"/></a> class="normal"><b><a href="datasets/Website+Phishing">Website Phishing</a></
b>
<!-- <td>
  -->
Multivariate 
Classification 
Integer 
1353 
10 
2016 
<!-- <td>Computer&nbsp; -->
<tmp border="1" src="assets/MLimages/Sm" src="assets/MLimag
allLargedefault.jpg"/></a> <b><a href="datasets/YouTube+Spam+Collection">You
Tube Spam Collection</a></b>
<!-- <td>It is a public set of comments collected for spam research. It has five datasets com
posed by 1,956 real messages extracted from five videos that were among the 10 most viewed on the collectio
n period.  -->
Text 
Classification 

1956 
5 
2017 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Beijing+PM2.5+Data"><img border="1" src="assets/MLimages/SmallLarg"><a href="datasets/Beijing+PM2.5+Data"><img border="1" src="assets/MLimages/SmallLarg"><a href="datasets/Beijing+PM2.5+Data"><a href="datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Beijing+PM2.5+Datasets/Be
edefault.jpg"/></a> class="normal"><b><a href="datasets/Beijing+PM2.5+Data">Beijing PM2.5 Dat
a</a></b>
<!-- <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.   -->
Multivariate, Time-Series 
Regression 
Integer, Real 
43824 
13 
2017 
<!-- <td>Physical&nbsp; -->
<a href="datasets/Cargo+2000+Freight+Tracking+and+Tracing"><img border="1" src="ass
ets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Cargo+2000+Fr"><br/>/as ets/MLimages/SmallLargedefault.jpg"/></a> 
eight+Tracking+and+Tracing">Cargo 2000 Freight Tracking and Tracing</a></b>
<!-- <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). &
nbsp; -->
Multivariate, Sequential 
Classification, Regression 
Integer 
3942 
98 
2016 
<!-- <td>Business&nbsp; -->
<a href="datasets/Cervical+cancer+%28Risk+Factors%29"><img border="1" src="assets/M
Limages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Cervical+cancer+%28">-cancer+%28
Risk+Factors%29">Cervical cancer (Risk Factors)</a></b>
<!-- <td>This dataset focuses on the prediction of indicators/diagnosis of cervical cancer. Th
e features cover demographic information, habits, and historic medical records.  -->
        an along "normal", Multivariate a/n. a/td-
```

```
<u><u>Clu>Mullivariate </tu>
Classification 
Integer, Real 
858 
36 
2017 
<!-- <td>Life&nbsp; -->
<to><a href="datasets/Quality+Assessment+of+Digital+Colposcopies"><img border="1" src="as
sets/MLimages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Quality+Asses"><br/>datasets/Quality+Asses
sment+of+Digital+Colposcopies">Quality Assessment of Digital Colposcopies</a></b>
<!-- <td>This dataset explores the subjective quality assessment of digital colposcopies.&nb
sp; -->
Multivariate 
Classification 
Real 
287 
69 
2017 
<!-- <td>Life&nbsp; -->
<a href="datasets/KASANDR"><img border="1" src="assets/MLimages/SmallLargedefault.j
pg"/></a> <b><a href="datasets/KASANDR">KASANDR</a></b>
ble>
<!-- <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.  
-->
Multivariate 
Causal-Discovery 
Integer 
17764280 
2158859 
2017 
<!-- <td>Life&nbsp; -->
<a href="datasets/FMA%3A+A+Dataset+For+Music+Analysis"><img border="1" src="assets"
/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/FMA%3A+A+Datas"><b><a href="datasets/FMA%3A+A+Datas"><b><a href="datasets/FMA%3A+A+Datas"><b><a href="datasets/FMA%3A+A+Datas"><b><a href="datasets/FMA%3A+A+Datas"><b><a href="datasets/FMA%3A+A+Datas"><b href="datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMA%3A+A+Datasets/FMAA+A+Datasets/FMAA+A+Datasets/FMAA+A+Datasets/FMAA+A+Datasets/FMAA+A+Datasets/
et+For+Music+Analysis">FMA: A Dataset For Music Analysis</a></b>
<!-- <td>FMA features 106,574 tracks and includes song title, album, artist, genres; play cou
nts, favorites, comments; description, biography, tags; together with audio (343 days, 917 GiB) and features.&n
bsp; -->
Multivariate, Time-Series 
Classification, Clustering 
Real 
106574 
518 
2017 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Air+quality"><img border="1" src="assets/MLimages/SmallLargedefault.j
pg"/></a> <b><a href="datasets/Air+quality">Air quality</a></b>
le>
<!-- <td> Contains the responses of a gas multisensor device deployed on the field in an Itali
an city.   -->
Multivariate, Time-Series 
Regression 
Real 
9358 
15 
2016 
<!-- <td>Other&nbsp; -->
```

```
</11></11>
<a href="datasets/Epileptic+Seizure+Recognition"><img border="1" src="assets/MLimages/
SmallLargedefault.jpg"/></a> <b><a href="datasets/Epileptic+Seizure+Recognition"><b
">Epileptic Seizure Recognition</a></b>
<!-- <td>This dataset is a pre-processed and re-structured/reshaped version of a very com-
monly used dataset featuring epileptic seizure detection.   -->
Multivariate, Time-Series 
Classification, Clustering 
Integer, Real 
11500 
179 
2017 
<!-- <td>Life&nbsp; -->
<a href="datasets/Devanagari+Handwritten+Character+Dataset"><img border="1" src="ass
ets/MLimages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Devanagari+Ha"
ndwritten+Character+Dataset">Devanagari Handwritten Character Dataset</a></b>
<!-- <td>This is an image database of Handwritten Devanagari characters. There are 46 cla
sses of characters with 2000 examples each. The dataset is split into training set(85%) and testing set(15%). &
nbsp; -->

Classification 
Integer 
92000 

2016 
<!-- <td>Computer&nbsp; -->
<tmg border="1" src="assets/MLimages/S
mallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Stock+portfolio+performance">S
tock portfolio performance</a></b>
<!-- <td>The data set of performances of weighted scoring stock portfolios are obtained with
mixture design from the US stock market historical database.   -->
Multivariate 
Regression 
Real 
315 
12 
2016 
<!-- <td>Business&nbsp; -->
="1" src="assets/MLimages/Small
Largedefault.jpg"/></a> <b><a href="datasets/MoCap+Hand+Postures">MoCap H
and Postures</a></b>
<!-- <td>5 types of hand postures from 12 users were recorded using unlabeled markers att
ached to fingers of a glove in a motion capture environment. Due to resolution and occlusion, missing values ar
e common.  -->
Multivariate 
Classification, Clustering 
Integer, Real 
78095 
38 
2016 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+conn"><a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+biomarkers+of+Parkinson%92s+disease+biomarkers+of+Parkinson%92s+disease+biomarkers+of+Parkinson%92s+disease+biomarkers+of+Parkinson%92s+disease+biomarkers+of+Parkinson%92s+disease+biomarkers+of+Parkinson%92s+disease+biomarkers+of+Parkinson%92s+disease+biomarkers+of+Parkinson%92s+disease+biomarkers+of+Parkinso
ected+speech"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <p class="norm"
al"><b><a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speec
h">Early biomarkers of Parkinson so disease based on natural connected speech</a></b>
>
<!-- <td>Predict a pattern of neurodegeneration in the dataset of speech features obtained f
rom patients with early untreated Parkinson's disease and patients at high risk developing Parkinson's disease.
```

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

```
2017 
<!-- <td>&nbsp; -->
efault.jpg"/></a> class="normal"><b><a href="datasets/Las+Vegas+Strip">Las Vegas Strip</a></b
>
<!-- <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). 
->

Classification, Regression 
Integer 
504 
20 
2017 
<!-- <td>Business&nbsp; -->
<a href="datasets/Eco-hotel"><img border="1" src="assets/MLimages/SmallLargedefault.jp"
g"/></a> <b><a href="datasets/Eco-hotel">Eco-hotel</a> </b>  
<!-- <td>This dataset includes Online Textual Reviews from both online (e.g., TripAdvisor) a
nd offline (e.g., Guests' book) sources from the Areias do Seixo Eco-Resort. 
Text 

401 
1 
2017 
<!-- <td>Business&nbsp; -->
<a href="datasets/MEU-Mobile+KSD"><img border="1" src="assets/MLimages/SmallLarged"><a href="datasets/MEU-Mobile+KSD"><img border="1" src="assets/MLimages/SmallLarged"><a href="datasets/MEU-Mobile+KSD"><timg border="1" src="assets/MLimages/SmallLarged"><a href="datasets/MEU-Mobile+KSD"><timg border="1" src="assets/MLimages/SmallLarged"><a href="datasets/MEU-Mobile+KSD"><timg border="1" src="assets/MLimages/SmallLarged"><a href="datasets/MEU-Mobile+KSD">>time states >>time states ><table
efault.jpg"/></a> class="normal"><b><a href="datasets/MEU-Mobile+KSD">MEU-Mobile KSD</a><
/b>
<!-- <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.  
Multivariate 
Classification 
Integer, Real 
2856 
71 
2016 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Crowdsourced+Mapping"><img border="1" src="assets/MLimages/Small"
Largedefault.jpg"/></a> class="normal"><b><a href="datasets/Crowdsourced+Mapping">Crowdso
urced Mapping</a></b>
<!-- <td>Crowdsourced data from OpenStreetMap is used to automate the classification of s
atellite images into different land cover classes (impervious, farm, forest, grass, orchard, water).  
d> -->
Multivariate 
Classification 

10546 
29 
2016 
<!-- <td>Physical&nbsp; -->
<a href="datasets/gene+expression+cancer+RNA-Seq"><img border="1" src="assets/MLim"
ages/SmallLargedefault.jpg"/></a> <b><a href="datasets/gene+expression+cancer"><b><a href="datasets/gene+expression+cancer"><b><a href="datasets/gene+expression+cancer"><b><a href="datasets/gene+expression+cancer"><b><a href="datasets/gene+expression+cancer"><b><a href="datasets/gene+expression+cancer"><b><a href="datasets/gene+expression+cancer"><b><a href="datasets/gene+expression+cancer"><a href="datasets/gene+expres
+RNA-Seg">gene expression cancer RNA-Seg</a></b>
<!-- <td>This collection of data is part of the RNA-Seq (HiSeq) PANCAN data set, it is a rand
om extraction of gene expressions of patients having different types of tumor: BRCA, KIRC, COAD, LUAD and
```

```
PRAD.  -->
Multivariate 
Classification, Clustering 
Real 
801 
20531 
2016 
<!-- <td>Life&nbsp; -->
<a href="datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+an"><a href="datasets/Hybrid+Indoor+Positioning+Dataset+from+Indoor+Positioning+Dataset+from+Indoor+Positioning+Dataset+from+Indoor+Positioning+Dataset+from+Indoor+Positioning+Dataset+from+Indoor+Positioning+Dataset+from+Indoor+Positioning+Dataset+from+Indoor+Positioning+Dataset+from+Indoor+Positioning+Dataset+from+Indoor+Positioning+Dataset+from+Indoor+Positioning+Dataset+from+Indoor+Positioning+Dataset+from+Indoor+Positioning+Dataset+from+Indoor+Positioning+Dataset+from+Positioning+Dataset+from+Positioning+Dataset+from+Positioning+Dataset+from+Positioning+from+Positioning+from+Positioning+from+Positioning+from+Positioning+from+Positioning+from+Positioning+from+Positioning+
d+magnetometer"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <p class="no
rmal"><b><a href="datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnet">
ometer">Hybrid Indoor Positioning Dataset from WiFi RSSI, Bluetooth and magnetometer</a></b>
>
<!-- <td>The dataset was created for the comparison and evaluation of hybrid indoor positio
ning methods. The dataset presented contains data from W-LAN and Bluetooth interfaces, and Magnetometer.
  -->
Multivariate, Sequential, Time-Series 
Classification 
Real 
1540 
65 
2016 
<!-- <td>Computer&nbsp; -->
<a href="datasets/chestnut+%E2%80%93+LARVIC"><img border="1" src="assets/MLimage"
s/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/chestnut+%E2%80%93+LAR">
VIC">chestnut - LARVIC</a></b>
<!-- <td>The research project presents this database, shows the images of chestnuts that wi
Il be processed to determine the presence or absence of defects  -->

Classification, Clustering 

1451 
3 
2017 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+"><a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+"><a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+"><a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+"><a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+"><a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+"><a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+"><a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+"><a href="datasets/Burst+"><a 
Switching+%28OBS%29+Network"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> 
<b><a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical"><b><a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical"><b style="datasets/burst-Header-packet-keeps"><a href="datasets/Burst+Header-packet-keeps"><a href="datasets/Burst-Header-packet-keeps"><a href="datasets/Bu
+Burst+Switching+%28OBS%29+Network">Burst Header Packet (BHP) flooding attack on Optical Burst Switchin
g (OBS) Network</a></b>
<!-- <td>One of the primary challenges in identifying the risks of the Burst Header Packet (B
HP) flood attacks in Optical Burst Switching networks (OBS) is the scarcity of reliable historical data.  </p
> -->
Text 
Classification 
Integer 
1075 
22 
2017 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Motion+Capture+Hand+Postures"><img border="1" src="assets/MLimag"
es/SmallLargedefault.jpg"/></a> <b><a href="datasets/Motion+Capture+Hand+Pos"
tures">Motion Capture Hand Postures</a></b>
<!-- <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 commo
n.  -->
Multivariate 
Classification, Clustering
```

```
Real 
78095 
38 
2017 
<!-- <td>Computer&nbsp; -->
<tmg border="1" src="assets/MLimages"
/SmallLargedefault.jpg"/></a> <b><a href="datasets/Anuran+Calls+%28MFCCs%2"><br/>/SmallLargedefault.jpg"/></a> 
9">Anuran Calls (MFCCs)</a></b>
<!-- <td>Acoustic features extracted from syllables of anuran (frogs) calls, including the famil
y, the genus, and the species labels (multilabel).   -->
Multivariate 
Classification, Clustering 
Real 
7195 
22 
2017 
<!-- <td>Life&nbsp; -->
+dataset+for+Turkish+text+categorization"><i
mg border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="d
atasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization">TTC-3600: Benchmark dataset f
or Turkish text categorization</a></b>
<!-- <td>The TTC-3600 data set is a collection of Turkish news and articles including catego
rized 3,600 documents from 6 well-known portals in Turkey. It has 4 different forms in ARFF Weka format. &nbs
p: -->
Text 
Classification, Clustering 
Integer 
3600 
4814 
2017 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy"><img border="1" sr
c="assets/MLimages/SmallLargedefault.jpg"/></a> c class="normal"><b><a href="datasets/Gastroint"><b><a href="datasets/Gastroint"><b style="color: blue;"><b style="co
estinal+Lesions+in+Regular+Colonoscopy">Gastrointestinal Lesions in Regular Colonoscopy</a></b>
<!-- <td>This dataset contains features extracted from colonoscopy videos used to detect ga
strointestinal lesions. It contains 76 lesions: 15 serrated adenomas, 21 hyperplastic lesions and 40 adenoma. &
nbsp; -->
Multivariate 
Classification 
Real 
76 
698 
2016 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Daily+Demand+Forecasting+Orders"><img border="1" src="assets/MLim"
ages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Daily+Demand+Forecasti
ng+Orders">Daily Demand Forecasting Orders</a></b>
<!-- <td>The dataset was collected during 60 days, this is a real database of a brazilian logis
tics company.  -->
Time-Series 
Regression 
Integer 
60 
13 
2017 
<!-- <td>Business&nbsp; -->
```

```
<a href="datasets/Paper+Reviews"><img border="1" src="assets/MLimages/SmallLargedef"
ault.jpg"/></a> <b><a href="datasets/Paper+Reviews">Paper Reviews</a></b></p
>
<!-- <td>This sentiment analysis data set contains scientific paper reviews from an internatio
nal conference on computing and informatics. The task is to predict the orientation or the evaluation of a review
.  -->
Text 
Classification, Regression 
Integer 
405 
10 
2017 
<!-- <td>Computer&nbsp; -->
<a href="datasets/extention+of+Z-Alizadeh+sani+dataset"><img border="1" src="assets/ML"
images/SmallLargedefault.jpg"/></a> <b><a href="datasets/extention+of+Z-Alizade"><br/>
h+sani+dataset">extention of Z-Alizadeh sani dataset</a></b>
<!-- <td>It was collected for CAD diagnosis.&nbsp; -->

Classification 
Integer, Real 
303 
59 
2017 
<!-- <td>Life&nbsp; -->
<a href="datasets/Z-Alizadeh+Sani"><img border="1" src="assets/MLimages/SmallLargede"
fault.jpg"/></a> <b><a href="datasets/Z-Alizadeh+Sani">Z-Alizadeh Sani</a></b><
/p>
<!-- <td>It was collected for CAD diagnosis.&nbsp; -->

Classification 
Integer, Real 
303 
56 
2017 
<!-- <td>Life&nbsp; -->
<a href="datasets/Dynamic+Features+of+VirusShare+Executables"><img border="1" src="
assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Dynamic+Fe"><b><a href="datasets/Dynamic+Fe"><b href="datase
atures+of+VirusShare+Executables">Dynamic Features of VirusShare Executables</a></b>
e>
<!-- <td>This dataset contains the dynamic features of 107,888 executables, collected by Vir
usShare from Nov/2010 to Jul/2014.  -->
Multivariate, Time-Series 
Classification, Regression 
Integer 
107888 
482 
2017 
<!-- <td>Computer&nbsp; -->
<a href="datasets/IDA2016Challenge"><img border="1" src="assets/MLimages/SmallLarge">
default.jpg"/></a> <b><a href="datasets/IDA2016Challenge">IDA2016Challenge
a></b>
<!-- <td>The dataset consists of data collected from heavy Scania trucks in everyday usage.
  -->
Multivariate 
Classification 
Integer 
76000 
<n class="normal">171 </n>
```

```
2017 
<!-- <td>Computer&nbsp; -->
es/SmallLargedefault.jpg"/></a> <b><a href="datasets/DSRC+Vehicle+Communica"><b><a href="datasets/DSRC+Vehicle+Communica"><b><a href="datasets/DSRC+Vehicle+Communica"><b><a href="datasets/DSRC+Vehicle+Communica"><b><a href="datasets/DSRC+Vehicle+Communica"><b><a href="datasets/DSRC+Vehicle+Communica"><b><a href="datasets/DSRC+Vehicle+Communica"><b><a href="datasets/DSRC+Vehicle+Communica"><b href="datasets/DSRC+Vehicle+
tions">DSRC Vehicle Communications</a></b>
<!-- <td>This set Provides data regarding wireless communications between vehicles and ro
ad side units. two separate data sets are provided (normal scenario) and in the presence of attacker (jammer).
  -->
Sequential, Text 
Clustering 
Real 
10000 
5 
2017 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Mturk+User-Perceived+Clusters+over+Images"><img border="1" src="as
sets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Mturk+User-Pe"><b><a href="datasets/Mturk+User-Pe"><b><a href="datasets/Mturk+User-Pe"><b><a href="datasets/Mturk+User-Pe"><b><a href="datasets/Mturk+User-Pe"><b><a href="datasets/Mturk+User-Pe"><b><a href="datasets/Mturk+User-Pe"><b><a href="datasets/Mturk+User-Pe"><b><a href="datasets/Mturk+User-Pe"><b><a href="datasets/Mturk+User-Pe"><b href="datasets/Mturk+User-Pe"
rceived+Clusters+over+Images">Mturk User-Perceived Clusters over Images</a></b>
<!-- <td>This dataset was collected by Shan-Hung Wu and DataLab members at NTHU, Tai
wan. There're 325 user-perceived clusters from 100 users and their corresponding descriptions. 
Multivariate, Text 
Clustering 
Integer 
180 
500 
2016 
<!-- <td>Computer&nbsp; -->
="1" src="assets/MLimages/Small
Largedefault.jpg"/></a> <b><a href="datasets/Character+Font+Images">Character
Font Images</a></b>
<!-- <td>Character images from scanned and computer generated fonts.&nbsp; --
>
Multivariate 
Classification 
Integer, Real 
745000 
411 
2016 
<!-- <td>Computer&nbsp; -->
<a href="datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning">the Annual Companies of the Annual Compan
s="normal"><b><a href="datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+wit
h+Instance+Labels">DeliciousMIL: A Data Set for Multi-Label Multi-Instance Learning with Instance Labels</a>
</b>
<!-- <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 docume
nts.  -->
Text 
Classification 
Integer 
12234 
8519 
2016 
<!-- <td>Computer&nbsp; -->
stds stables stroughts as brof-"datacate/Autictic, Spectrum, Disorder, Spragning, Data, for, Children, , "s sima bor
```

```
<l
der="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets"
/Autistic+Spectrum+Disorder+Screening+Data+for+Children++">Autistic Spectrum Disorder Screening Data for
Children </a></b>
<!-- <td>Children screening data for autism suitable for classification and predictive tasks &n
bsp; -->
Multivariate 
Classification 
Integer 
292 
21 
2017 
<!-- <td>Life&nbsp; -->
<a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++"><img
border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datas"
ets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++">Autistic Spectrum Disorder Screening D
ata for Adolescent </a></b>
<!-- <td>Autistic Spectrum Disorder Screening Data for Adolescent. This dataset is related to
classification and predictive tasks.  -->
Multivariate 
Classification 
Integer 
104 
21 
2017 
<!-- <td>Life&nbsp; -->
<a href="datasets/APS+Failure+at+Scania+Trucks"><img border="1" src="assets/MLimage"><a href="datasets/APS+Failure+at+Scania+Trucks"><img border="1" src="assets/MLimage"><a href="datasets/APS+Failure+at+Scania+Trucks"><img border="1" src="assets/MLimage"><a href="datasets/APS+Failure+at+Scania+Trucks"><a href="datasets/APS+Failure+at
s/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/APS+Failure+at+Scania+Tru"><b>
cks">APS Failure at Scania Trucks</a></b>
<!-- <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.&n
bsp; -->
Multivariate 
Classification 
Integer, Real 
60000 
171 
2017 
<!-- <td>Computer&nbsp; -->
<tmp border="1" src="assets/MLimages/"
SmallLargedefault.jpg"/></a> <b><a href="datasets/Wireless+Indoor+Localization""><b><a href="datasets/Wireless+Indoor+Localization""><b><a href="datasets/Wireless+Indoor+Localization"><b><a href="datasets/Wireless+Indoor+Localization"><b><a href="datasets/Wireless+Indoor+Localization"><b><a href="datasets/Wireless+Indoor+Localization"><b><a href="datasets/Wireless+Indoor+Localization"><b><a href="datasets/Wireless+Indoor+Localization"><b><a href="datasets/Wireless+Indoor+Localization"><b href="datasets/Wireless+Indoor+Localiz
>Wireless Indoor Localization</a></b>
<!-- <td>Collected in indoor space by observing signal strengths of seven WiFi signals visible
e on a smartphone. The decision variable is one of the four rooms.   -->
Multivariate 
Classification 
Real 
2000 
7 
2017 
<!-- <td>Computer&nbsp; -->
="1" src="assets/MLimages/SmallLargedefa"
ult.jpg"/></a> <b><a href="datasets/HCC+Survival">HCC Survival</a></b>
>
<!-- <td>Hepatocellular Carcinoma dataset (HCC dataset) was collected at a University Hos
pital in Portugal. It contains real clinical data of 165 patients diagnosed with HCC. 
Multivariate 
Classification
```

```
<\u>\u><p \u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u2016\u201
165 
49 
2017 
<!-- <td>Life&nbsp; -->
<a href="datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+"><a href="datasets/CSM+%28Conventional+and+Social+And+Social+Media+Movies%29+"><a href="datasets/CSM+%28Conventional+and+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+And+Social+An
and+2015"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> 
<b><a href="datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015">CS
M (Conventional and Social Media Movies) Dataset 2014 and 2015</a></b>
<!-- <td>12 features categorized as conventional and social media features. Both conventional and social media features.
nal features, collected from movies databases on Web as well as social media features(YouTube,Twitter).&nbs
p; -->
Multivariate 
Classification, Regression 
Integer 
217 
12 
2017 
<!-- <td>Computer&nbsp; -->
<a href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29"><
img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="d
atasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29">University of Tehran Question D
ataset 2016 (UTQD.2016)</a></b>
<!-- <td>Persian questions gathered from a jeopardy game broadcasted on Iranian national
television.   -->
Text 
Classification 

1175 
3 
2017 
<!-- <td>Other&nbsp; -->
="1" src="assets/MLimages/Small">= "1" src="assets
Largedefault.jpg"/></a> <b><a href="datasets/Autism+Screening+Adult">Adult">Autism Sc
reening Adult</a></b>
<!-- <td>Autistic Spectrum Disorder Screening Data for Adult. This dataset is related to class
ification and predictive tasks.  -->

Classification 
Integer 
704 
21 
2017 
<!-- <td>Social&nbsp; -->
+ using + a + batteryless + we
arable+sensor"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <p class="norm"
al"><b><a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sens"
or">Activity recognition with healthy older people using a batteryless wearable sensor</a></b>
able>
<!-- <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.&n
bsp; -->
Sequential 
Classification 
Real 
75128 
9 
2016
```

```
<!-- <ta>Life&nosp;</ta> -->
="1" src="assets/MLimages/Small
Largedefault.jpg"/></a> <b><a href="datasets/Immunotherapy+Dataset">Immunotherapy+Dataset">Immunotherapy+Dataset
herapy Dataset</a></b>
<!-- <td>This dataset contains information about wart treatment results of 90 patients using i
mmunotherapy.  -->
Univariate 
Classification 
Integer, Real 
90 
8 
2018 
<!-- <td>Life&nbsp; -->
="1" src="assets/MLimages/SmallLa" |
rgedefault.jpg"/></a> <b><a href="datasets/Cryotherapy+Dataset+">Cryotherapy
Dataset </a></b>
<!-- <td>This dataset contains information about wart treatment results of 90 patients using
cryotherapy.  -->
Univariate 
Classification 
Integer, Real 
90 
7 
2018 
<!-- <td>Life&nbsp; -->
<a href="datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes"><im
g border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="dat
asets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes">OCT data & amp; Color Fundus Ima
ges of Left & amp; Right Eyes</a></b>
<!-- <td>This dataset contains OCT data (in mat format) and color fundus data (in jpg format
) of left & right eyes of 50 healthy persons.  -->
Multivariate 
Classification 
Real 
50 
2 
2016 
<!-- <td>Computer&nbsp; -->
/SmallLargedefault.jpg"/></a> <b><a href="datasets/Discrete+Tone+Image+Datas"><b><a href="datasets/Discrete+Tone+Image+Datas"><b><a href="datasets/Discrete+Tone+Image+Datas"><b><a href="datasets/Discrete+Tone+Image+Datas"><b><a href="datasets/Discrete+Tone+Image+Datas"><b><a href="datasets/Discrete+Tone+Image+Datas"><b><a href="datasets/Discrete+Tone+Image+Datas"><b><a href="datasets/Discrete+Tone+Image+Datas"><b><a href="datasets/Discrete+Tone+Image+Datas"><b href="datasets/Discrete+Tone+Image+Datas"><b href="datasets/Discrete+Tone+Image+Datas"><b href="datasets/Discrete+Tone+Image+Datas"><b href="datasets/Discrete+Tone+Image+Datas"><b href="datasets/Discrete+Tone+Image+Datas"><b href="datasets/Discrete+Tone+Image+Datas"><b href="datasets/Discrete+Tone+Image+Datas"><b href="datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discrete+Tone+Image+Datasets/Discre
et">Discrete Tone Image Dataset</a></b>
<!-- <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.
  -->
Multivariate 
Classification 

71 
11 
2018 
<!-- <td>Computer&nbsp; -->
<a href="datasets/News+Popularity+in+Multiple+Social+Media+Platforms"><img border="1"
src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/News+"
Popularity+in+Multiple+Social+Media+Platforms">News Popularity in Multiple Social Media Platforms</a>
p>
<!-- <td>Large data set of news items and their respective social feedback on multiple platfo
rms: Facebook, Google+ and LinkedIn.  -->
```

```
Multivariate, Time-Series, Text 
Regression 
Integer, Real 
93239 
11 
2018 
<!-- <td>Computer&nbsp; -->
es/SmallLargedefault.jpg"/></a> <b><a href="datasets/Ultrasonic+flowmeter+diagn"><b><a href="datasets/Ultrasonic+flowmeter+diagn"><b><a href="datasets/Ultrasonic+flowmeter+diagn"><b><a href="datasets/Ultrasonic+flowmeter+diagn"><b><a href="datasets/Ultrasonic+flowmeter+diagn"><b><a href="datasets/Ultrasonic+flowmeter+diagn"><b><a href="datasets/Ultrasonic+flowmeter+diagn"><b><a href="datasets/Ultrasonic+flowmeter+diagn"><b><a href="datasets/Ultrasonic+flowmeter+diagn"><b href="datasets/Ultrasonic
ostics">Ultrasonic flowmeter diagnostics</a></b>
<!-- <td>Fault diagnosis of four liquid ultrasonic flowmeters&nbsp; -->
Multivariate 
Classification 
Real 
540 
173 
2018 
<!-- <td>Computer&nbsp; -->
<a href="datasets/ICMLA+2014+Accepted+Papers+Data+Set"><img border="1" src="asset"></a>
s/MLimages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/ICMLA+2014+Acc
epted+Papers+Data+Set">ICMLA 2014 Accepted Papers Data Set</a></b>
<!-- <td>This data set compromises the metadata for the 2014 ICMLA conference's accepte
d papers, including ID, paper titles, author's keywords, abstracts and sessions in which they were exposed.&nb
sp; -->
Multivariate 
Classification, Clustering 

105 
5 
2018 
<!-- <td>Other&nbsp; -->
<a href="datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation"><img bord
er="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/
BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation">BLE RSSI Dataset for Indoor localization and Na
vigation</a></b>
<!-- <td>This dataset contains RSSI readings gathered from an array of Bluetooth Low Ener
gy (BLE) iBeacons in a real-world and operational indoor environment for localization and navigation purposes.
  -->
Multivariate, Sequential, Time-Series 
Classification, Clustering 
Integer 
6611 
15 
2018 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Container+Crane+Controller+Data+Set"><img border="1" src="assets/M
Limages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Container+Crane+Co</pre>
ntroller+Data+Set">Container Crane Controller Data Set</a></b>
<!-- <td>A container crane has the function of transporting containers from one point to anot
her point.  -->
Univariate, Domain-Theory 
Classification, Regression 
Real 
15 
3 
2018 
<!-- <td>Computer&nbsp; -->
```

```
<a href="datasets/Residential+Building+Data+Set"><img border="1" src="assets/MLimages"
/SmallLargedefault.jpg"/></a> <b><a href="datasets/Residential+Building+Data+Se"
t">Residential Building Data Set</a></b>
<!-- <td>Data set includes construction cost, sale prices, project variables, and economic var
iables corresponding to real estate single-family residential apartments in Tehran, Iran.   -->
Multivariate 
Regression 
Real 
372 
105 
2018 
<!-- <td>Computer&nbsp; -->
="1" src="assets/MLimages/Small">="1" src="assets/MLimages/Small">="1" src="assets/MLimages/Small">="1" src="assets/MLimages/Small">="1" src="assets/MLimages/Small">="1" src="assets/MLimages/Small">="1" src="assets/MLimages/Small">= 1" src="assets/MLimages/Small"
Largedefault.jpg"/></a> class="normal"><b><a href="datasets/Health+News+in+Twitter">Health Ne
ws in Twitter</a></b>
<!-- <td>The data was collected in 2015 using Twitter API. This dataset contains health new
s from more than 15 major health news agencies such as BBC, CNN, and NYT.   -->
Text 
Clustering 
Real 
58000 
25000 
2018 
<!-- <td>Computer&nbsp; -->
<a href="datasets/chipseq"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"
/></a> <b><a href="datasets/chipseq">chipseq</a></b>
<!-- <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.   -->
Sequential 
Classification 
Integer 
4960 

2018 
<!-- <td>Life&nbsp; -->
<a href="datasets/SGEMM+GPU+kernel+performance"><img border="1" src="assets/MLim"
ages/SmallLargedefault.jpg"/></a> <b><a href="datasets/SGEMM+GPU+kernel+p"><b><a href="datasets/SGEMM+GPU+kernel+p"><b style="text-align: center;"><b style="te
erformance">SGEMM GPU kernel performance</a></b>
<!-- <td>Running times for multiplying two 2048 x 2048 matrices using a GPU OpenCL SGE
MM kernel with varying parameters (using the library 'CLTune'). 
Multivariate 
Regression 
Integer 
241600 
18 
2018 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Repeat+Consumption+Matrices"><img border="1" src="assets/MLimage"></a>
s/SmallLargedefault.jpg"/></a> <b><a href="datasets/Repeat+Consumption+Matric"><b><a href="datasets/Repeat+Consumption+Matric"><b><a href="datasets/Repeat+Consumption+Matric"><b><a href="datasets/Repeat+Consumption+Matric"><b><a href="datasets/Repeat+Consumption+Matric"><b><a href="datasets/Repeat+Consumption+Matric"><b><a href="datasets/Repeat+Consumption+Matric"><b><a href="datasets/Repeat+Consumption+Matric"><b><a href="datasets/Repeat+Consumption+Matric"><b href="datasets/Repeat+C
es">Repeat Consumption Matrices</a></b>
<!-- <td>The dataset contains 7 datasets of User - Item matrices, where each entry represe
nts how many times a user consumed an item. Item is used as an umbrella term for various categories. <
/p> -->
Multivariate 
Clustering 
Real 
130000
```

```
21000 
2018 
<!-- <td>Computer&nbsp; -->
<a href="datasets/detection of IoT botnet attacks N BaIoT"><img border="1" src="asset"
s/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/detection_of_loT"><b><a href="datasets/detection_of_loT"><b><a href="datasets/detection_of_loT"><b><a href="datasets/detection_of_loT"><b><a href="datasets/detection_of_loT"><b><a href="datasets/detection_of_loT"><b><a href="datasets/detection_of_loT"><b><a href="datasets/detection_of_loT"><b><a href="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_loT"><base="datasets/detection_of_l
botnet_attacks_N_BaloT">detection_of_loT_botnet_attacks_N_BaloT</a></b>
<!-- <td>This dataset addresses the lack of public botnet datasets, especially for the IoT. It s
uggests *real* traffic data, gathered from 9 commercial IoT devices authentically infected by Mirai and BASHLIT
E.  -->
Multivariate, Sequential 
Classification, Clustering 
Real 
7062606 
115 
2018 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Absenteeism+at+work"><img border="1" src="assets/MLimages/SmallLa
rgedefault.jpg"/></a> <b><a href="datasets/Absenteeism+at+work">Absenteeism a
t work</a></b>
<!-- <td>The database was created with records of absenteeism at work from July 2007 to J
uly 2010 at a courier company in Brazil.  -->
Multivariate, Time-Series 
Classification, Clustering 
Integer, Real 
740 
21 
2018 
<!-- <td>Business&nbsp; -->
<a href="datasets/SCADI"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/
></a> <b><a href="datasets/SCADI">SCADI</a></b>
<!-- <td>First self-care activities dataset based on ICF-CY.&nbsp; -->
Multivariate 
Classification, Clustering 

70 
206 
2018 
<!-- <td>Life&nbsp; -->
ets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Condition+monit"><b><a href="datasets/Condition+monit"><b href="datasets/Condition+monit</b></b href="datasets/Condition+monit</b href="datasets
oring+of+hydraulic+systems">Condition monitoring of hydraulic systems</a></b>
<!-- <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 quantificat
ion.  -->
Multivariate, Time-Series 
Classification, Regression 
Real 
2205 
43680 
2018 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Carbon+Nanotubes"><img border="1" src="assets/MLimages/SmallLarg"
edefault.jpg"/></a> <b><a href="datasets/Carbon+Nanotubes">Carbon Nanotubes
</a></b>
<!-- <td>This dataset contains 10721 initial and calculated atomic coordinates of carbon nan
otubes.  -->
Univariate
```

```
Regression 
Real 
10721 
8 
2018 
<!-- <td>Computer&nbsp; -->
<tm> border="1" src="assets/MLima" 
ges/SmallLargedefault.jpg"/></a> <b><a href="datasets/Optical+Interconnection+N"><b><a href="datasets/Optical+Interconnection+N"><b href="datase
etwork+">Optical Interconnection Network </a></b>
<!-- <td>This dataset contains 640 performance measurements from a simulation of 2-Dime
nsional Multiprocessor Optical Interconnection Network.  
Multivariate 
Classification, Regression 
Integer, Real 
640 
10 
2018 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Sports+articles+for+objectivity+analysis"><img border="1" src="assets/M
Limages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Sports+articles+for+o"><b</a>
bjectivity+analysis">Sports articles for objectivity analysis</a></b>
<!-- <td>1000 sports articles were labeled using Amazon Mechanical Turk as objective or su
bjective. The raw texts, extracted features, and the URLs from which the articles were retrieved are provided.&
nbsp; -->
Multivariate, Text 
Classification 
Integer 
1000 
59 
2018 
<!-- <td>Social&nbsp; -->
<a href="datasets/Breast+Cancer+Coimbra"><img border="1" src="assets/MLimages/Small"
Largedefault.jpg"/></a> <b><a href="datasets/Breast+Cancer+Coimbra">Breast C
ancer Coimbra</a></b>
<!-- <td>Clinical features were observed or measured for 64 patients with breast cancer and
52 healthy controls.   -->
Multivariate 
Classification 
Integer 
116 
10 
2018 
<!-- <td>Life&nbsp; -->
src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/GNFUV"><a href="datasets/GNFUV"><b href="dataset
+Unmanned+Surface+Vehicles+Sensor+Data">GNFUV Unmanned Surface Vehicles Sensor Data</a></b>
<!-- <td>The data-set contains four (4) sets of mobile sensor readings data (humidity, tempe
rature) corresponding to a swarm of four (4) Unmanned Surface Vehicles (USVs) in a test-bed in Athens (Gree
ce).   -->
Multivariate, Time-Series 
Regression 
Real 
1672 
5 
2018 
<!-- <td>Computer&nbsp; -->
```

```
<a href="datasets/Dishonest+Internet+users+Dataset"><img border="1" src="assets/MLima"
ges/SmallLargedefault.jpg"/></a> <b><a href="datasets/Dishonest+Internet+users"
+Dataset">Dishonest Internet users Dataset</a></b>
<!-- <td>The dataset was used to test an architecture based on a trust model capable to cop
e with the evaluation of the trustworthiness of users interacting in pervasive environments. 
Multivariate 
Classification, Clustering 

322 
5 
2018 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Victorian+Era+Authorship+Attribution"><img border="1" src="assets/MLi
mages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Victorian+Era+Authors"
hip+Attribution">Victorian Era Authorship Attribution</a></b>
<!-- <td>To create the largest authorship attribution dataset, we extracted works of 50 well-k
nown authors. To have a non-exhaustive learning, in training there are 45 authors whereas, in the testing, it's 5
0  -->
Text 
Classification 

93600 
1000 
2018 
<!-- <td>Computer&nbsp; -->
" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Simul"
ated+Falls+and+Daily+Living+Activities+Data+Set">Simulated Falls and Daily Living Activities Data Set</a></b>
<!-- <td>20 falls and 16 daily living activities were performed by 17 volunteers with 5 repetiti
ons while wearing 6 sensors (3.060 instances) that attached to their head, chest, waist, wrist, thigh and ankle.&
nbsp; -->
Time-Series 
Classification 
Integer 
3060 
138 
2018 
<!-- <td>Life&nbsp; -->
<a href="datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing"><img
border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datas"
ets/Multimodal+Damage+Identification+for+Humanitarian+Computing">Multimodal Damage Identification for Hu
manitarian Computing</a></b>
<!-- <td>5879 captioned images (image and text) from social media related to damage durin
g natural disasters/wars, and belong to 6 classes: Fires, Floods, Natural landscape, Infrastructural, Human, No
n-damage.  -->
Multivariate, Text 
Classification 
Integer 
5879 

2018 
<!-- <td>Social&nbsp; -->
+ Visual + Evoked + Potential + Signals" > <img border = "1"
src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/EEG+S"
teady-State+Visual+Evoked+Potential+Signals">EEG Steady-State Visual Evoked Potential Signals</a></b>
```

```
ady State Visual Evoked Potentials (BCI-SSVEP).   -->
Multivariate, Time-Series 
Classification, Regression 
Integer 
9200 
16 
2018 
<!-- <td>Life&nbsp; -->
="1" src="assets/MLimages/Small">= 1" src="assets/MLimages/Small"
Largedefault.jpg"/></a> <b><a href="datasets/Roman+Urdu+Data+Set">Roman Ur
du Data Set</a></b>
<!-- <td>Roman Urdu (the scripting style for Urdu language) is one of the limited resource la
nguages. A data corpus comprising of more than 20000 records was collected.  
Text 
Classification 

20000 
2 
2018 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Avila"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/><
/a> <b><a href="datasets/Avila">Avila</a></b>
<!-- <td>The Avila data set has been extracted from 800 images of the 'Avila Bible', an XII c
entury giant Latin copy of the Bible. The prediction task consists in associating each pattern to a copyist. 
 -->
Multivariate 
Classification 
Real 
20867 
10 
2018 
<!-- <td>Computer&nbsp; -->
<a href="datasets/PANDOR"><img border="1" src="assets/MLimages/SmallLargedefault.jp"
g"/></a> <b><a href="datasets/PANDOR">PANDOR</a></b>
>
<!-- <td>PANDOR is a novel and publicly available dataset for online recommendation provi
ded by Purch (http://www.purch.com/).   -->
Multivariate 
Recommendation 
Categorical 

2018 
<!-- <td>Life&nbsp; -->
<a href="datasets/Drug+Review+Dataset+%28Druglib.com%29"><img border="1" src="ass
ets/MLimages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Drug+Review+D"><b><a href="datasets/Drug+Review+D"><br/>
ataset+%28Druglib.com%29">Drug Review Dataset (Druglib.com)</a></b>
<!-- <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.
  -->
Multivariate, Text 
Classification, Regression, Clustering 
Integer 
4143 
8 
2018 
<!-- <td>&nbsp; -->
```

```
<a href="datasets/Drug+Review+Dataset+%28Drugs.com%29"><img border="1" src="asset"
s/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Drug+Review+Dat"><b><a href="datasets/Drug+Review+Dat"><b><a href="datasets/Drug+Review+Dat"><b><a href="datasets/Drug+Review+Dat"><b><a href="datasets/Drug+Review+Dat"><b><a href="datasets/Drug+Review+Dat"><b><a href="datasets/Drug+Review+Dat"><b><a href="datasets/Drug+Review+Dat"><b><a href="datasets/Drug+Review+Dat"><b href="datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Review+Datasets/Drug+Revi
aset+%28Drugs.com%29">Drug Review Dataset (Drugs.com)</a></b>
<!-- <td>The dataset provides patient reviews on specific drugs along with related conditions
and a 10 star patient rating reflecting overall patient satisfaction.   -->
Multivariate, Text 
Classification, Regression, Clustering 
Integer 
215063 
6 
2018 
<!-- <td>Life&nbsp; -->
<a href="datasets/Physical+Unclonable+Functions"><img border="1" src="assets/MLimage"></a>
s/SmallLargedefault.jpg"/></a> <b><a href="datasets/Physical+Unclonable+Functi
ons">Physical Unclonable Functions</a></b>
<!-- <td>The dataset is generated from Physical Unclonable Functions (PUFs) simulation, sp
ecifically XOR Arbiter PUFs. PUFs are used for authentication purposes. For more info, refer to our paper belo
w.  -->
Multivariate 
Classification 
Integer 
6000000 
129 
2018 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Superconductivty+Data"><img border="1" src="assets/MLimages/SmallL
argedefault.jpg"/></a> <b><a href="datasets/Superconductivty+Data">Superconductivty+Data">Superconductivty+Data">Superconductivty+Data">Superconductivty+Data">Superconductivty+Data">Superconductivty+Data">Superconductivty+Data">Superconductivty+Data">Superconductivty+Data">Superconductivty+Data">Superconductivty+Data">Superconductivty+Data">Superconductivty+Data">Superconductivty+Data">Superconductivty+Data">Superconductivty+Data">Superconductivty+Data">Superconductivty+Data">Superconductivty+Data">Superconductivty+Data</a>
ctivty Data</a></b>
<!-- <td>Two file s contain data on 21263 superconductors and their relevant features.&nbs
p; -->
Multivariate 
Regression 
Real 
21263 
81 
2018 
<!-- <td>Physical&nbsp; -->
to><a href="datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29"><img border</a>
="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/W"
ESAD+%28Wearable+Stress+and+Affect+Detection%29">WESAD (Wearable Stress and Affect Detection)</a>
/b>
<!-- <td>WESAD (Wearable Stress and Affect Detection) contains data of 15 subjects during
a stress-affect lab study, while wearing physiological and motion sensors.  -->
Multivariate, Time-Series 
Classification, Regression 
Real 
63000000 
12 
2018 
<!-- <td>Computer&nbsp; -->
<a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2"><img bord
er="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/
GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2">GNFUV Unmanned Surface Vehicles Sensor Dat
a Set 2</a></b>
<!-- <td>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.  -->
```

```
<ta>iviullivariale, 5equential, 1 ime-5eries </ta>
Regression 
Real 
10190 
6 
2018 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Student+Academics+Performance"><img border="1" src="assets/MLima"
ges/SmallLargedefault.jpg"/></a> <b><a href="datasets/Student+Academics+Perfo"><br/>
rmance">Student Academics Performance</a></b>
<!-- <td>The dataset tried to find the end semester percentage prediction based on different
social, economic and academic attributes.   -->
Multivariate 
Classification 

300 
22 
2018 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Online+Shoppers+Purchasing+Intention+Dataset"><img border="1" src="
assets/MLimages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Online+Sho">
ppers+Purchasing+Intention+Dataset">Online Shoppers Purchasing Intention Dataset</a></b>
<!-- <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.  -->
Multivariate 
Classification, Clustering 
Integer, Real 
12330 
18 
2018 
<!-- <td>Business&nbsp; -->
"/></a> <b><a href="datasets/PMU-UD">PMU-UD</a></b>
<!-- <td>The handwritten dataset was collected from 170 participants with a total of 5,180 nu
meral patterns. The dataset is named Prince Mohammad Bin Fahd University - Urdu/Arabic Database (PMU-U
D).   -->
Univariate 
Classification 

5180 
9 
2018 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Parkinson%27s+Disease+Classification"><img border="1" src="assets/M
Limages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Parkinson%27s+Dise"><a hr
ase+Classification">Parkinson's Disease Classification</a></b>
<!-- <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). 
Multivariate 
Classification 
Integer, Real 
756 
754 
2018
```

```
<!-- <td>Computer&nbsp; -->
<a href="datasets/Electrical+Grid+Stability+Simulated+Data+"><imq border="1" src="assets"
/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Electrical+Grid+Sta"
bility+Simulated+Data+">Electrical Grid Stability Simulated Data </a></b>
<!-- <td>The local stability analysis of the 4-node star system (electricity producer is in the c
enter) implementing Decentral Smart Grid Control concept.   -->
Multivariate 
Classification, Regression 
Real 
10000 
14 
2018 
<!-- <td>Physical&nbsp; -->
><a href="datasets/Caesarian+Section+Classification+Dataset"><img border="1" src="assets">
/MLimages/SmallLargedefault.jpg"/></a> <b><a href="datasets/Caesarian+Section"><a href="dataset
+Classification+Dataset">Caesarian Section Classification Dataset</a></b>
<!-- <td>This dataset contains information about caesarian section results of 80 pregnant w
omen with the most important characteristics of delivery problems in the medical field.  
Univariate 
Classification 
Integer 
80 
5 
2018 
<!-- <td>Life&nbsp; -->
"/></a> <b><a href="datasets/BAUM-1">BAUM-1</a></b></t
d>
<!-- <td>BAUM-1 dataset contains 1184 multimodal facial video clips collected from 31 subje
cts. The 1184 video clips contain spontaneous facial expressions and speech of 13 emotional and mental state
s.  -->
Time-Series 
Classification 

1184 

2018 
<!-- <td>Computer&nbsp; -->
"/></a> <b><a href="datasets/BAUM-2">BAUM-2</a></b></t
<!-- <td>A multilingual audio-visual affective face database consisting of 1047 video clips of
286 subjects.   -->
Time-Series 
Classification 

1047 

2018 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Audit+Data"><img border="1" src="assets/MLimages/SmallLargedefault.j
pg"/></a> <b><a href="datasets/Audit+Data">Audit Data</a></b>
<!-- <td>Exhaustive one year non-confidential data in the year 2015 to 2016 of firms is colle
cted from the Auditor Office of India to build a predictor for classifying suspicious firms.  -->
Multivariate 
Classification
```

```
Real 
777 
18 
2018 
<!-- <td>Other&nbsp; -->
<tm, description of the state of the stat
rgedefault.jpg"/></a> <b><a href="datasets/BuddyMove+Data+Set">BuddyMove D
ata Set</a></b>
<!-- <td>User interest information extracted from user reviews published in holidayiq.com ab
out various types of point of interests in South India  -->
Multivariate, Text 
Classification, Clustering 
Real 
249 
7 
2018 
<!-- <td>Other&nbsp; -->
<a href="datasets/Real+estate+valuation+data+set"><img border="1" src="assets/MLimage"
s/SmallLargedefault.jpg"/></a> <b><a href="datasets/Real+estate+valuation+data+"
set">Real estate valuation data set</a></b>
<!-- <td>The "real estate valuation" is a regression problem. The market historical data set o
f real estate valuation are collected from Sindian Dist., New Taipei City, Taiwan.   -->
Multivariate 
Regression 
Integer, Real 
414 
7 
2018 
<!-- <td>Business&nbsp; -->
<a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natu"><a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natu*</a>
ral+connected+speech+Data+Set+"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> 
<b><a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+o
n+natural+connected+speech+Data+Set+">Early biomarkers of Parkinson's disease based on natural connecte
d speech Data Set </a></b>
<!-- <td>.&nbsp; -->
Multivariate 
Classification 
Real 

2018 
<!-- <td>Life&nbsp; -->
<a href="datasets/Somerville+Happiness+Survey"><img border="1" src="assets/MLimages/
SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/Somerville+Happiness+Survey"><br/>
">Somerville Happiness Survey</a></b>
<!-- <td>A data extract of a non-federal dataset posted here https://catalog.data.gov/dataset
/somerville-happiness-survey-responses-2011-2013-2015 

Classification 
Integer 
143 
7 
2018 
<!-- <td>Life&nbsp; -->
<a href="datasets/2.4+GHZ+Indoor+Channel+Measurements"><img border="1" src="asset"
s/MLimages/SmallLargedefault.jpg"/></a> class="normal"><b><a href="datasets/2.4+GHZ+Indoor+"
Channel+Measurements">2.4 GHZ Indoor Channel Measurements</a></b>
```

```
<!-- <td>Measurement of the S21, consists of 10 sweeps, each sweep contains 601 frequence
y points with spacing of 0.167MHz to cover a 100MHz band centered at 2.4GHz. 
Multivariate 
Classification 
Real 
7840 
5 
2018 
<!-- <td>Computer&nbsp; -->
<a href="datasets/EMG+data+for+gestures"><img border="1" src="assets/MLimages/Small
Largedefault.jpg"/></a> <b><a href="datasets/EMG+data+for+gestures">EMG dat
a for gestures</a></b>
<!-- <td>These are files of raw EMG data recorded by MYO Thalmic bracelet&nbsp;
Time-Series 
Classification 
Real 
30000 
6 
2019 
<!-- <td>Life&nbsp; -->
<a href="datasets/Parking+Birmingham"><img border="1" src="assets/MLimages/SmallLar
gedefault.jpg"/></a> <b><a href="datasets/Parking+Birmingham">Parking Birming
ham</a></b>
<!-- <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  -->
Multivariate, Univariate, Sequential, Time-Series 
Classification, Regression, Clustering 
Real 
35717 
4 
2019 
<!-- <td>Computer&nbsp; -->
+ the + city + of + Sao + Paulo + in + Brazil" > <i
mg border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a> <b><a href="d
atasets/Behavior+of+the+urban+traffic+of+the+city+of+Sao+Paulo+in+Brazil">Behavior of the urban traffic of the
e city of Sao Paulo in Brazil</a></b>
<!-- <td>The database was created with records of behavior of the urban traffic of the city of
Sao Paulo in Brazil.  -->
Multivariate, Time-Series 
Classification, Regression 
Integer, Real 
135 
18 
2018 
<!-- <td>Computer&nbsp; -->
<a href="datasets/Travel+Reviews"><img border="1" src="assets/MLimages/SmallLargedef"
ault.jpg"/></a> <b><a href="datasets/Travel+Reviews">Travel Reviews</a></b></p
>
<!-- <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.  -->
Multivariate, Text 
Classification, Clustering 
Real 
980 
11
```

```
2018 
<!-- <td>Other&nbsp; -->
<a href="datasets/Tarvel+Review+Ratings"><img border="1" src="assets/MLimages/SmallL
argedefault.jpg"/></a> <b><a href="datasets/Tarvel+Review+Ratings">Tarvel Revi
ew Ratings</a></b>
<!-- <td>Google reviews on attractions from 24 categories across Europe are considered. G
oogle user rating ranges from 1 to 5 and average user rating per category is calculated. 
Multivariate, Text 
Classification, Clustering 
Real 
5456 
25 
2018 
<!-- <td>Other&nbsp; -->
="1" src="assets/MLimages/SmallLarg">="1" src="assets/MLimages/SmallLarg
edefault.jpg"/></a> <b><a href="datasets/Rice+Leaf+Diseases">Rice Leaf Disease
s</a></b>
<!-- <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.   -->
Multivariate 
Classification 
Integer 
120 

2019 
<!-- <td>Computer&nbsp; -->
Supported By:
<img height="60" src="assets/nsfe.gif"/> 
 In Collaboration With:
<img src="assets/rexaSmall.jpg"/>
<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 [24]:
print(soup.prettify())
<!DOCTYPE HTML>
<html>
<body>
>
 "-//W3C//DTD HTML 4.01 Transitional//EN\">
```

<title>

</title>

UCI Machine Learning Repository: Data Sets

```
<!-- Stylesheet link -->
 k 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) -->
 <span class="normal">
   <a alt="Home" href="index.html">
    <img border="0" src="assets/logo.gif"/>
```

```
</a>
   <br/>br/>
   <a href="http://cml.ics.uci.edu">
    <font color="FFDD33">
    Center for Machine Learning and Intelligent Systems
   </a>
   </span>
  <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 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>
```

```
<br/>
      <br/>
   Browse Through:
            <b>
               Default Task
             </b>
            <a href="datasets.php?format=&amp;task=cla&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;t
ype=&sort=nameUp&view=table">
              Classification
             </a>
             <font color="red">
              (350)
             </font>
             <br/>
             <a href="datasets.php?format=&amp;task=reg&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;
type=&sort=nameUp&view=table">
              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;t
ype=&sort=nameUp&view=table">
               Clustering
             </a>
             <font color="red">
              (84)
             </font>
             <br/>
             <a href="datasets.php?format=&amp;task=other&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;area=&amp;numAtt=&amp;numIns=&amp;area=&amp;numAtt=&amp;numIns=&amp;area=&amp;numAtt=&amp;numIns=&amp;area=&amp;numAtt=&amp;numIns=&amp;area=&amp;numAtt=&amp;numIns=&amp;area=&amp;numAtt=&amp;numIns=&amp;area=&amp;numAtt=&amp;area=&amp;numAtt=&amp;numIns=&amp;area=&amp;numAtt=&amp;numIns=&amp;area=&amp;numAtt=&amp;numIns=&amp;area=&amp;numIns=&amp;area=&amp;numIns=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;area=&amp;ar
p;type=&sort=nameUp&view=table">
              Other
             </a>
             <font color="red">
               (55)
             </font>
            </n>
```

```
<b>
              Attribute Type
             </b>
            <a href="datasets.php?format=&amp;task=&amp;att=cat&amp;area=&amp;numAtt=&amp;numIns=&amp;t
ype=&sort=nameUp&view=table">
              Categorical
             </a>
             <font color="red">
              (38)
             </font>
             <br/>
             <a href="datasets.php?format=&amp;task=&amp;att=num&amp;area=&amp;numAtt=&amp;numIns=&amp;att=num&amp;area=&amp;numAtt=&amp;numIns=&amp;att=num&amp;area=&amp;numAtt=&amp;numIns=&amp;att=num&amp;area=&amp;numAtt=&amp;numIns=&amp;att=num&amp;area=&amp;numAtt=&amp;numIns=&amp;att=num&amp;att=num&amp;att=numAtt=&amp;numIns=&amp;att=num&amp;att=numAtt=&amp;numIns=&amp;att=num&amp;att=numAtt=&amp;numIns=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numAtt=&amp;att=numA
p;type=&sort=nameUp&view=table">
              Numerical
             </a>
             <font color="red">
              (307)
             </font>
             <br/>
             <a href="datasets.php?format=&amp;task=&amp;att=mix&amp;area=&amp;numAtt=&amp;numIns=&amp;
type=&sort=nameUp&view=table">
              Mixed
             </a>
             <font color="red">
              (55)
             </font>
            <b>
              Data Type
             </b>
            <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;typ">amp;typ</a>
e=mvar&sort=nameUp&view=table">
              Multivariate
             </a>
             <font color="red">
               (357)
             </font>
             <br/>br/>
             <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;typ">amp;typ</a>
```

ar & ampreart_named In & amprejow_table">

```
Univariate
                      </a>
                      <font color="red">
                        (23)
                      </font>
                      <br/>
                      <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;typ">a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;typ">a href="datasets.php?format=&amp;task=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;typ">a href="datasets.php?format=&amp;task=&amp;task=&amp;att=&amp;att=&amp;numIns=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp
e=seg&sort=nameUp&view=table">
                        Sequential
                      </a>
                      <font color="red">
                        (47)
                      </font>
                      <br/>
                      <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;typ">a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;typ">a href="datasets.php?format=&amp;task=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;typ">a href="datasets.php?format=&amp;task=&amp;task=&amp;att=&amp;att=&amp;numIns=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp
e=ts&sort=nameUp&view=table">
                        Time-Series
                      </a>
                      <font color="red">
                        (91)
                      </font>
                      <br/>
                      <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;typ">amp;typ</a>
e=text&sort=nameUp&view=table">
                        Text
                      </a>
                      <font color="red">
                        (53)
                      </font>
                      <br/>
                      <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;typ">amp;typ</a>
e=dt&sort=nameUp&view=table">
                        Domain-Theory
                      </a>
                      <font color="red">
                        (23)
                      </font>
                      <br/>
                      <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;typ">amp;typ</a>
e=other&sort=nameUp&view=table">
                        Other
                      </a>
                      <font color="red">
                        (21)
                      </font>
                      <br/>br/>
                   <b>
                        Area
                      </b>
                   م 0 المام مرمور و المراب مرمور و المرمور و والمراب
```

```
<a firei= datasets.prip:format=aamp,task=aamp,att=aamp,area=iiieaamp,numAtt=aamp,numiiis=aamp,t
ype=&sort=nameUp&view=table">
                             Life Sciences
                           </a>
                           <font color="red">
                             (107)
                           </font>
                           <br/>
                           <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=phys&amp;numAtt=&amp;numIns=&amp;att=&amp;area=phys&amp;numAtt=&amp;numIns=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;
p;type=&sort=nameUp&view=table">
                             Physical Sciences
                           </a>
                           <font color="red">
                             (49)
                           </font>
                           <br/>
                           <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=comp&amp;numAtt=&amp;numIns=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;at
p;type=&sort=nameUp&view=table">
                              CS / Engineering
                           </a>
                           <font color="red">
                             (170)
                           </font>
                           <br/>
                           <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=soc&amp;numAtt=&amp;numIns=&amp;
type=&sort=nameUp&view=table">
                              Social Sciences
                           </a>
                           <font color="red">
                             (26)
                           </font>
                           <br/>
                           <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=bus&amp;numAtt=&amp;numIns=&amp;
type=&sort=nameUp&view=table">
                             Business
                           </a>
                           <font color="red">
                             (29)
                           </font>
                           <br/>br/>
                           <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=game&amp;numAtt=&amp;numIns=&a
mp;type=&sort=nameUp&view=table">
                             Game
                           </a>
                           <font color="red">
                             (10)
                           </font>
                            <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=other&amp;numAtt=&amp;numIns=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;a
p;type=&sort=nameUp&view=table">
                             Other
                           </a>
                           <font color="red">
                             (73)
                           </font>
                       <b>
```

```
# Attributes
             </b>
           <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=less10&amp;numIns=&a
mp;type=&sort=nameUp&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=&sort=nameUp&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">a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=greater100&amp;numIns">a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=greater100&amp;numIns">a href="datasets.php?format=&amp;task=&amp;att=&amp;att=&amp;att=&amp;numAtt=greater100&amp;numIns">a href="datasets.php?format=&amp;task=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&amp;att=&a
=&type=&sort=nameUp&view=table">
              Greater than 100
             </a>
             <font color="red">
              (84)
             </font>
           <b>
              # Instances
             </b>
           <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=less100&
amp;type=&sort=nameUp&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&type=&sort=nameUp&view=table">
              100 to 1000
             </a>
             <font color="red">
              (162)
             </font>
             <br/>
```

```
<a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=greater10
00&type=&sort=nameUp&view=table">
              Greater than 1000
             </a>
             <font color="red">
              (246)
             </font>
            <b>
              Format Type
             </b>
            <a href="datasets.php?format=mat&amp;task=&amp;att=&amp;numAtt=&amp;numIns=&amp;
;type=&sort=nameUp&view=table">
               Matrix
             </a>
             <font color="red">
              (324)
             </font>
             <br/>
             <a href="datasets.php?format=nonmat&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&
amp;type=&sort=nameUp&view=table">
              Non-Matrix
             </a>
             <font color="red">
              (145)
             </font>
            <b>
               469
             </b>
             Data Sets
            <font color="gray">
              Table View
             </font>
             <a href="datasets.php?format=&amp;task=&amp;att=&amp;numAtt=&amp;numIns=&amp;typ">a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;typ">a href="datasets.php?format=&amp;task=&amp;att=&amp;att=&amp;numAtt=&amp;numIns=&amp;typ">a href="datasets.php?format=&amp;task=&amp;att=&amp;att=&amp;numAtt=&amp;numIns=&amp;typ">a href="datasets.php?format=&amp;task=&amp;att=&amp;att=&amp;att=&amp;numIns=&amp;typ">a href="datasets.php?format=&amp;typ">a href="da
e=&sort=nameUp&view=list">
              List View
             </a>
```

```
<a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;typ">amp;typ</a>
e=&sort=nameDown&view=table">
                     <b>
                      Name
                     </b>
                   </a>
                <!-- <td><b>Abstract</b> -->
               <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;typ
e=&sort=typeUp&view=table">
                     <b>
                       Data Types
                     </b>
                   </a>
                <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;typ">a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;typ">a href="datasets.php?format=&amp;task=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;typ">a href="datasets.php?format=&amp;task=&amp;task=&amp;att=&amp;att=&amp;numIns=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;typ">a href="datasets.php?format=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp
e=&sort=taskUp&view=table">
                     <b>
                      Default Task
                     </b>
                   </a>
                 <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;typ">amp;typ</a>
e=&sort=attTypeUp&view=table">
                     <b>
                       Attribute Types
                     </b>
                   </a>
                <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;typ">amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;task=&amp;
e=&sort=instUp&view=table">
                     <b>>
                       # Instances
                     </b>
                   </a>
                <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;typ
e=&sort=attUp&view=table">
                     <b>
```

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

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

```
<!-- <td>Social&nbsp; -->
<a href="datasets/Annealing">
 <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
 </a>
 <b>
  <a href="datasets/Annealing">
  Annealing
  </a>
 </b>
 <!-- <td>Steel annealing data&nbsp; -->
Multivariate
Classification
Categorical, Integer, Real
798
38
<!-- <td>Physical&nbsp; -->
<a href="datasets/Anonymous+Microsoft+Web+Data">
 -img horder="1" erc="assets/MI images/Small argedefault ing"/>
```

```
</a>
   <b>
    <a href="datasets/Anonymous+Microsoft+Web+Data">
    Anonymous Microsoft Web Data
    </b>
   <!-- <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.  -->
  Recommender-Systems
  Categorical
  37711
  294
  1998
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Arrhythmia">
    <img border="1" src="assets/MLimages/SmallLarge5.jpg"/>
   </a>
   <a href="datasets/Arrhythmia">
    Arrhythmia
```

```
</a>
   </b>
   <!-- <td>Distinguish between the presence and absence of cardiac arrhythmia and clas
sify it in one of the 16 groups.  -->
  Multivariate
  Classification
  Categorical, Integer, Real
  452
  279
  1998
  <!-- <td>Life&nbsp; -->
  <a href="datasets/Artificial+Characters">
   <img border="1" src="assets/MLimages/SmallLarge6.jpg"/>
   </a>
   <a href="datasets/Artificial+Characters">
    Artificial Characters
    </a>
   </b>
```

```
<!-- <ld>Dataset artificially generated by using first order theory which describes structu
re of ten capital letters of English alphabet  -->
  Multivariate
  Classification
  Categorical, Integer, Real
  6000
  1992
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Audiology+%28Original%29">
    <img border="1" src="assets/MLimages/SmallLarge7.jpg"/>
   </a>
   <a href="datasets/Audiology+%28Original%29">
    Audiology (Original)
    </a>
    </b>
   <!-- <td>Nominal audiology dataset from Baylor&nbsp; -->
  Multivariate
```

```
Classification
Categorical
226
1987
<!-- <td>Life&nbsp; -->
<a href="datasets/Audiology+%28Standardized%29">
 <img border="1" src="assets/MLimages/SmallLarge7.jpg"/>
 </a>
 <a href="datasets/Audiology+%28Standardized%29">
  Audiology (Standardized)
  </a>
 </b>
 <!-- <td>Standardized version of the original audiology database&nbsp; -->
Multivariate
Classification
Categorical
```

```
226
  69
  1992
  <!-- <td>Life&nbsp; -->
 <a href="datasets/Auto+MPG">
   <img border="1" src="assets/MLimages/SmallLarge9.jpg"/>
   </a>
  <b>
   <a href="datasets/Auto+MPG">
   Auto MPG
   </a>
   </b>
   <!-- <td>Revised from CMU StatLib library, data concerns city-cycle fuel consumption&n
bsp; -->
 >
  Multivariate
  Regression
  Categorical, Real
  398
```

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

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

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

```
</a>
   </b>
   <!-- <td>Data previously used in cognitive psychology experiment; 4 data sets represent
different conditions of an experiment  -->
  Multivariate
  Classification
  Categorical
  16
  4
  <!-- <td>Social&nbsp; -->
 <a href="datasets/Breast+Cancer">
   <img border="1" src="assets/MLimages/SmallLarge14.jpg"/>
   </a>
   <a href="datasets/Breast+Cancer">
    Breast Cancer
   </a>
   </b>
```

on place-"normal"> Projet Cancor Data (Postrioted Access) & phon:

```
Multivariate
Classification
Categorical
286
9
1988
<!-- <td>Life&nbsp; -->
<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>
 </b>
 <!-- <td>Original Wisconsin Breast Cancer Database&nbsp; -->
Multivariate
```

```
Integer
699
10
1992
<!-- <td>Life&nbsp; -->
<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>
 </b>
 <!-- <td>Prognostic Wisconsin Breast Cancer Database&nbsp; -->
Multivariate
Classification, Regression
Real
```

Ciassilication

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

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

```
<!-- <td>Other&nbsp; -->
 <a href="datasets/Car+Evaluation">
   <img border="1" src="assets/MLimages/SmallLarge19.jpg"/>
   </a>
   <a href="datasets/Car+Evaluation">
    Car Evaluation
   </a>
   </b>
   <!-- <td>Derived from simple hierarchical decision model, this database may be useful f
or testing constructive induction and structure discovery methods. 
  Multivariate
  Classification
  Categorical
  1728
  6
  1997
  <!-- <td>Other&nbsp; -->
```

```
<a href="datasets/Census+Income">
   <img border="1" src="assets/MLimages/SmallLarge2.jpg"/>
   <b>
    <a href="datasets/Census+Income">
    Census Income
    </b>
   <!-- <td>Predict whether income exceeds $50K/yr based on census data. Also known a
s "Adult" dataset.  -->
  Multivariate
  Classification
  Categorical, Integer
  48842
  14
  1996
  <!-- <td>Social&nbsp; -->
  <a href="datasets/Chess+%28King-Rook+vs.+King-Knight%29">
    <img border="1" src="assets/MLimages/SmallLarge24.jpg"/>
   </a>
   <b>
```

```
<a href="datasets/Chess+%28King-Rook+vs.+King-Knight%29">
  Chess (King-Rook vs. King-Knight)
  </a>
 </b>
 <!-- <td>Knight Pin Chess End-Game Database Creator&nbsp; -->
Multivariate, Data-Generator
Classification
Categorical, Integer
22
1988
<!-- <td>Game&nbsp; -->
<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>
 </b>
```

```
<!-- <td>King+Rook versus King+Pawn on a7 (usually abbreviated KRKPA7).&nbsp;
 -->
  Multivariate
  Classification
  Categorical
  3196
  36
  1989
  <!-- <td>Game&nbsp; -->
  <a href="datasets/Chess+%28King-Rook+vs.+King%29">
   <img border="1" src="assets/MLimages/SmallLarge24.ipg"/>
   </a>
   <a href="datasets/Chess+%28King-Rook+vs.+King%29">
    Chess (King-Rook vs. King)
    </a>
   </b>
   <!-- <td>Chess Endgame Database for White King and Rook against Black King (KRK).
  -->
  Multivariate
  /td>
```

```
Classification
  Categorical, Integer
  28056
  6
  1994
  <!-- <td>Game&nbsp; -->
 <a href="datasets/Chess+%28Domain+Theories%29">
   <img border="1" src="assets/MLimages/SmallLarge24.jpg"/>
   </a>
  <b>
   <a href="datasets/Chess+%28Domain+Theories%29">
    Chess (Domain Theories)
   </a>
   </b>
   <!-- <td>6 different domain theories for generating legal moves of chess&nbsp;
> -->
  Domain-Theory
```

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

```
</lu>
<!-- <td>Other&nbsp; -->
<a href="datasets/Connect-4">
 <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
 <b>
 <a href="datasets/Connect-4">
  Connect-4
 </a>
 </b>
 <!-- <td>Contains connect-4 positions&nbsp; -->
Multivariate, Spatial
Classification
Categorical
67557
42
1995
<!-- <td>Game&nbsp; -->
```

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

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

```
<!-- <td>Relative CPU Performance Data, described in terms of its cycle time, memory s
ize, etc.  -->
  Multivariate
  Regression
  Integer
  209
  9
  1987
  <!-- <td>Computer&nbsp; -->
  <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>
   </b>
   <!-- <td>Dataset is a subset of the 1987 National Indonesia Contraceptive Prevalence S
urvey.  -->
```

```
Multivariate
Classification
Categorical, Integer
1473
9
1997
<!-- <td>Life&nbsp; -->
<a href="datasets/Covertype">
 <img border="1" src="assets/MLimages/SmallLarge31.jpg"/>
 </a>
 <a href="datasets/Covertype">
  Covertype
 </a>
 </b>
 <!-- <td>Forest CoverType dataset&nbsp; -->
Multivariate
Classification
<a></a>
```

```
Categorical, Integer
  581012
  54
  1998
  <!-- <td>Life&nbsp; -->
 <a href="datasets/Cylinder+Bands">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Cylinder+Bands">
    Cylinder Bands
    </a>
   </b>
   <!-- <td>Used in decision tree induction for mitigating process delays known as "cylinder
bands" in rotogravure printing  -->
  Multivariate
  Classification
  Categorical, Integer, Real
  <ht>td>
```

```
512
  39
  1995
  <!-- <td>Physical&nbsp; -->
 <a href="datasets/Dermatology">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
  <a href="datasets/Dermatology">
   Dermatology
   </a>
   </b>
   <!-- <td>Aim for this dataset is to determine the type of Eryhemato-Squamous Disease.
  -->
  Multivariate
  Classification
  Categorical, Integer
  366
```

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

```
<a href="datasets/DGP2+-+The+Second+Data+Generation+Program">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/DGP2+-+The+Second+Data+Generation+Program">
    DGP2 - The Second Data Generation Program
    </b>
   <!-- <td>Generates application domains based on specific parameters, number of featu
res, and proportion of positive to negative examples  -->
  Data-Generator
  Real
  <!-- <td>Other&nbsp; -->
  <a href="datasets/Document+Understanding">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
```

<:u>

```
<a nref="datasets/Document+Understanding">
  Document Understanding
  </a>
 </b>
 <!-- <td>Five concepts, expressed as predicates, to be learned&nbsp; -->
1994
<!-- <td>Other&nbsp; -->
<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>
 </b>
 <!-- <td>Assorted small-scale domain theories&nbsp; -->
```

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

```
132
12
1989
<!-- <td>Life&nbsp; -->
<a href="datasets/Ecoli">
 <img border="1" src="assets/MLimages/SmallLarge120.jpg"/>
 </a>
 <b>
 <a href="datasets/Ecoli">
 Ecoli
 </a>
 </b>
 <!-- <td>This data contains protein localization sites&nbsp; -->
Multivariate
Classification
Real
336
8
```

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

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

```
<a href="datasets/Glass+Identification">
    Glass Identification
    </a>
    </b>
   <!-- <td>From USA Forensic Science Service; 6 types of glass; defined in terms of their
oxide content (i.e. Na, Fe, K, etc)  -->
  Multivariate
  Classification
  Real
  214
  10
  1987
  <!-- <td>Physical&nbsp; -->
  <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>
```

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

```
Multivariate
  Classification
  Categorical
  160
  5
  1989
  <!-- <td>Social&nbsp; -->
 <a href="datasets/Heart+Disease">
   <img border="1" src="assets/MLimages/SmallLarge45.jpg"/>
   </a>
  <a href="datasets/Heart+Disease">
   Heart Disease
   </a>
   </b>
   <!-- <td>4 databases: Cleveland, Hungary, Switzerland, and the VA Long Beach&nbsp;
 -->
 Multivariate
```

```
Classification
  Categorical, Integer, Real
  303
  75
  1988
  <!-- <td>Life&nbsp; -->
 <a href="datasets/Hepatitis">
   <img border="1" src="assets/MLimages/SmallLarge46.jpg"/>
   </a>
   <a href="datasets/Hepatitis">
    Hepatitis
    </a>
   </b>
   <!-- <td>From G.Gong: CMU; Mostly Boolean or numeric-valued attribute types; Include
s cost data (donated by Peter Turney)  -->
  Multivariate
  Classification
  Categorical, Integer, Real
```

```
155
  19
  1988
  <!-- <td>Life&nbsp; -->
 <a href="datasets/Horse+Colic">
   <img border="1" src="assets/MLimages/SmallLarge47.jpg"/>
   </a>
   <b>
   <a href="datasets/Horse+Colic">
    Horse Colic
   </a>
   </b>
   <!-- <td>Well documented attributes; 368 instances with 28 attributes (continuous, discr
ete, and nominal); 30% missing values  -->
  Multivariate
  Classification
  Categorical, Integer, Real
  368
```

```
27
  1989
  <!-- <td>Life&nbsp; -->
 <a href="datasets/ICU">
   <img border="1" src="assets/MLimages/SmallLarge49.jpg"/>
  </a>
  <a href="datasets/ICU">
   ICU
   </a>
   </b>
  <!-- <td>Data set prepared for the use of participants for the 1994 AAAI Spring Symposi
um on Artificial Intelligence in Medicine.  -->
 Multivariate, Time-Series
  Real
  <!-- <td>Life&nbsp; -->
```

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

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

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

```
Multivariate
  Classification
  Real
  150
  4
  1988
  <!-- <td>Life&nbsp; -->
 <a href="datasets/ISOLET">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
   <a href="datasets/ISOLET">
    ISOLET
   </a>
   </b>
   <!-- <td> Goal: Predict which letter-name was spoken--a simple classification task.&nbs
p; -->
  Multivariate
  n clace_"normal"
```

```
~p 01433= 110111141 >
Classification
Real
7797
617
1994
<!-- <td>Computer&nbsp; -->
<a href="datasets/Kinship">
 <img border="1" src="assets/MLimages/SmallLarge55.jpg"/>
 </a>
 <a href="datasets/Kinship">
  Kinship
 </a>
 </b>
 <!-- <td>Relational dataset&nbsp; -->
Relational
Relational-Learning
Categorical
```

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

</lu>

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

```
<a href="datasets/Lenses">
 <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
 </a>
 <b>
  <a href="datasets/Lenses">
  Lenses
  </a>
 </b>
 <!-- <td>Database for fitting contact lenses&nbsp; -->
Multivariate
Classification
Categorical
24
4
1990
<!-- <td>Other&nbsp; -->
<a href="datasets/Letter+Recognition">
 <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
```

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

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

```
Domain-Theory
<!-- <td>Computer&nbsp; -->
<a href="datasets/Lung+Cancer">
 <img border="1" src="assets/MLimages/SmallLarge62.jpg"/>
 </a>
 <a href="datasets/Lung+Cancer">
 Lung Cancer
 </a>
 </b>
 <!-- <td>Lung cancer data; no attribute definitions&nbsp; -->
Multivariate
Classification
Integer
```

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

/td>

```
18
  1988
  <!-- <td>Life&nbsp; -->
  <a href="datasets/Mechanical+Analysis">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Mechanical+Analysis">
    Mechanical Analysis
    </a>
    </b>
   <!-- <td>Fault diagnosis problem of electromechanical devices; also PUMPS DATA SET
is newer version with domain theory and results  -->
  Multivariate
  Classification
  Categorical, Integer, Real
  209
  8
  alaga "narmal".
```

```
1990
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Meta-data">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <a href="datasets/Meta-data">
    Meta-data
    </a>
   </b>
   <!-- <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). 
  Multivariate
  Classification
  Categorical, Integer, Real
  528
  22
  1996
  <!-- <td>Other&nbsp; -->
```

```
<lu>
  <a href="datasets/Mobile+Robots">
    <img border="1" src="assets/MLimages/SmallLarge66.jpg"/>
   </a>
   <a href="datasets/Mobile+Robots">
    Mobile Robots
    </a>
    </b>
   <!-- <td>Learning concepts from sensor data of a mobile robot; set of data sets&nbsp;</
p> -->
  Domain-Theory
  Categorical, Integer, Real
  1995
  <!-- <td>Computer&nbsp; -->
  <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>
    </b>
    <!-- <td>E. Coli promoter gene sequences (DNA) with partial domain theory&nbsp;
 -->
  Sequential, Domain-Theory
  Classification
  Categorical
  106
  58
  1990
  <!-- <td>Life&nbsp; -->
  <a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%29">
    <img border="1" src="assets/MLimages/SmallLarge67.jpg"/>
    </a>
   <b>
    <a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%29">
     Molecular Biology (Protein Secondary Structure)
    </a>
    </b>
```

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

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

```
Categorical
  432
  1992
  <!-- <td>Other&nbsp; -->
 >
  <a href="datasets/Moral+Reasoner">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
  <a href="datasets/Moral+Reasoner">
   Moral Reasoner
   </a>
   </b>
   <!-- <td>Horn-clause model that qualitatively simulates moral reasoning; Theory include
s negated literals  -->
 Domain-Theory
  202
```

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

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

```
<a href="datasets/Musk+%28Version+1%29">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <b>
    <a href="datasets/Musk+%28Version+1%29">
    Musk (Version 1)
    </a>
    </b>
   <!-- <td>The goal is to learn to predict whether new molecules will be musks or non-mu
sks  -->
  Multivariate
  Classification
  Integer
  476
  168
  1994
  <!-- <td>Physical&nbsp; -->
  <a href="datasets/Musk+%28Version+2%29">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
```

```
<b>
    <a href="datasets/Musk+%28Version+2%29">
    Musk (Version 2)
    </a>
   </b>
   <!-- <td>The goal is to learn to predict whether new molecules will be musks or non-mu
sks  -->
  Multivariate
  Classification
  Integer
  6598
  168
  1994
  <!-- <td>Physical&nbsp; -->
 <a href="datasets/Nursery">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Nursery">
    Nursery
    </a>
   </b>
```

```
</ta>
   <!-- <td> Nursery Database was derived from a hierarchical decision model originally de
veloped to rank applications for nursery schools.  -->
  Multivariate
  Classification
  Categorical
  12960
  8
  1997
  <!-- <td>Social&nbsp; -->
  <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>
    </b>
   <!-- <td>Used in research to generate features for an inductive learning system&nbsp;<
/p> -->
```

```
Domain-Theory
  1991
  <!-- <td>Game&nbsp; -->
 <a href="datasets/Page+Blocks+Classification">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
   <a href="datasets/Page+Blocks+Classification">
    Page Blocks Classification
   </a>
   </b>
   <!-- <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.  -->
  Multivariate
  Classification
```

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

```
64
1998
<!-- <td>Computer&nbsp; -->
<a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits">
 <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
 </a>
 <b>
  <a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits">
  Pen-Based Recognition of Handwritten Digits
  </a>
 </b>
 <!-- <td>Digit database of 250 samples from 44 writers&nbsp; -->
Multivariate
Classification
Integer
10992
16
1998
```

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

```
<a href="datasets/Primary+Tumor">
 <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
 </a>
 <b>
  <a href="datasets/Primary+Tumor">
  Primary Tumor
  </a>
 </b>
 <!-- <td>From Ljubljana Oncology Institute&nbsp; -->
Multivariate
Classification
Categorical
339
17
1988
<!-- <td>Life&nbsp; -->
<a href="datasets/Prodigy">
 <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
 </a>
```

```
<a href="datasets/Prodigy">
    Prodigy
    </a>
    </b>
    <!-- <td>Assorted domains like blocksworld, eightpuzzle, and schedworld.&nbsp;</t
d> -->
  Domain-Theory
  <!-- <td>Other&nbsp; -->
  <a href="datasets/Qualitative+Structure+Activity+Relationships">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Qualitative+Structure+Activity+Relationships">
    Qualitative Structure Activity Relationships
    </a>
    </b>
    <!-- <td>Two sets of datasets are given: pyrimidines and triazines&nbsp; -->
```

```
Domain-Theory
  <!-- <td>Physical&nbsp; -->
 <a href="datasets/Quadruped+Mammals">
   <img border="1" src="assets/MLimages/SmallLarge86.jpg"/>
   </a>
  <a href="datasets/Quadruped+Mammals">
   Quadruped Mammals
   </a>
   </b>
   <!-- <td> The file animals.c is a data generator of structured instances representing qua
druped animals  -->
  Multivariate, Data-Generator
  Classification
  Real
```

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

```
1993
<!-- <td>Computer&nbsp; -->
<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>
 </b>
 <!-- <td>Tiny database; all nominal values&nbsp; -->
Multivariate
Classification
Categorical
15
6
1988
```

```
<!-- <td>Physical&nbsp; -->
 <a href="datasets/Solar+Flare">
   <img border="1" src="assets/MLimages/SmallLarge89.jpg"/>
   </a>
   <b>
   <a href="datasets/Solar+Flare">
    Solar Flare
   </a>
   </b>
   <!-- <td>Each class attribute counts the number of solar flares of a certain class that oc
cur in a 24 hour period  -->
  Multivariate
  Regression
  Categorical
  1389
  10
  1989
  <!-- <td>Physical&nbsp; -->
```

```
<a href="datasets/Soybean+%28Large%29">
 <img border="1" src="assets/MLimages/SmallLarge90.jpg"/>
 </a>
 <a href="datasets/Soybean+%28Large%29">
  Soybean (Large)
  </a>
 </b>
 <!-- <td>Michalski's famous soybean disease database&nbsp; -->
Multivariate
Classification
Categorical
307
35
1988
<!-- <td>Life&nbsp; -->
<a href="datasets/Soybean+%28Small%29">
 <img border="1" src="assets/MLimages/SmallLarge90.jpg"/>
 </a>
 <a href="datasets/Soybean+%28Small%29">
```

```
Soybean (Small)
  </a>
 </b>
 <!-- <td>Michalski's famous soybean disease database&nbsp; -->
Multivariate
Classification
Categorical
47
35
1987
<!-- <td>Life&nbsp; -->
<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>
 </b>
```

```
<!-- <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  -->
  Multivariate
  Regression
  Integer
  23
  4
  1993
  <!-- <td>Physical&nbsp; -->
  <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>
    </b>
   <!-- <td>From IRAS data -- NASA Ames Research Center&nbsp; -->
  Multivariate
  <ht>td>
```

```
Classification
Integer, Real
531
102
1988
<!-- <td>Physical&nbsp; -->
<a href="datasets/Spambase">
 <img border="1" src="assets/MLimages/SmallLarge94.jpg"/>
 <a href="datasets/Spambase">
  Spambase
 </a>
 </b>
 <!-- <td>Classifying Email as Spam or Non-Spam&nbsp; -->
Multivariate
Classification
Integer, Real
```

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

```
22
  2001
  <!-- <td>Life&nbsp; -->
 <a href="datasets/SPECTF+Heart">
   <img border="1" src="assets/MLimages/SmallLarge45.jpg"/>
   <b>
   <a href="datasets/SPECTF+Heart">
    SPECTF Heart
   </a>
   </b>
   <!-- <td>Data on cardiac Single Proton Emission Computed Tomography (SPECT) imag
es. Each patient classified into two categories: normal and abnormal. 
  Multivariate
  Classification
  Integer
  267
  44
```

```
200 I
<!-- <td>Life&nbsp; -->
<a href="datasets/Sponge">
 <img border="1" src="assets/MLimages/SmallLarge97.jpg"/>
 <a href="datasets/Sponge">
 Sponge
 </a>
 </b>
 <!-- <td>Data on sponges; Attributes in spanish&nbsp; -->
Multivariate
Clustering
Categorical, Integer
76
45
<!-- <td>Life&nbsp; -->
```

```
<a href="datasets/Statlog+Project">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <a href="datasets/Statlog+Project">
     Statlog Project
    </a>
    </b>
    <!-- <td>Various Databases: Vehicle silhouttes, Landsat Sattelite, Shuttle, Australian Cr
edit Approval, Heart Disease, Image Segmentation, German Credit  -->
  1992
  <!-- <td>Other&nbsp; -->
  <a href="datasets/Student+Loan+Relational">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
    <a href="datasets/Student+Loan+Relational">
     Student Loan Relational
    </a>
    </b>
```

```
<!-- <td>Student Loan Relational Domain&nbsp; -->
  Domain-Theory
  1000
  1993
  <!-- <td>Social&nbsp; -->
 <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>
   </b>
   <!-- <td>The data consist of evaluations of teaching performance; scores are "low", "me
dium", or "high"  -->
  Multivariate
```

```
Classification
  Categorical, Integer
  151
  5
  1997
  <!-- <td>Other&nbsp; -->
 <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>
   </b>
   <!-- <td>Binary classification task on possible configurations of tic-tac-toe game&nbsp;<
/p> -->
  Multivariate
  Classification
```

```
Categorical
958
9
1991
<!-- <td>Game&nbsp; -->
<a href="datasets/Thyroid+Disease">
 <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
 </a>
 <a href="datasets/Thyroid+Disease">
  Thyroid Disease
 </a>
 </b>
 <!-- <td>10 separate databases from Garavan Institute&nbsp; -->
Multivariate, Domain-Theory
Classification
Categorical, Real
7200
</n>
```

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

```
1994
<!-- <td>Other&nbsp; -->
<a href="datasets/University">
 <img border="1" src="assets/MLimages/SmallLarge104.jpg"/>
 </a>
 <b>
 <a href="datasets/University">
  University
 </a>
 </b>
 <!-- <td>Data in original (LISP-readable) form&nbsp; -->
Multivariate
Classification
Categorical, Integer
285
17
1988
<!-- <td>Other&nbsp; -->
```

```
<a href="datasets/Congressional+Voting+Records">
   <img border="1" src="assets/MLimages/SmallLarge105.jpg"/>
   </a>
   <b>
    <a href="datasets/Congressional+Voting+Records">
    Congressional Voting Records
    </a>
    </b>
   <!-- <td>1984 United Stated Congressional Voting Records; Classify as Republican or D
emocrat  -->
  Multivariate
  Classification
  Categorical
  435
  16
  1987
  <!-- <td>Social&nbsp; -->
  <a href="datasets/Water+Treatment+Plant">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
```

<1abic>

```
<a href="datasets/Water+Treatment+Plant">
  Water Treatment Plant
  </a>
  </b>
 <!-- <td>Multiple classes predict plant state&nbsp; -->
Multivariate
Clustering
Integer, Real
527
38
1993
<!-- <td>Physical&nbsp; -->
<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>
  </b>
```

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

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

```
Integer, Real
178
13
1991
<!-- <td>Physical&nbsp; -->
>
 <a href="datasets/Yeast">
 <img border="1" src="assets/MLimages/SmallLarge110.jpg"/>
 </a>
 <b>
 <a href="datasets/Yeast">
 Yeast
 </a>
 </b>
 <!-- <td>Predicting the Cellular Localization Sites of Proteins&nbsp; -->
Multivariate
Classification
Real
1484
```

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

```
1990
<!-- <td>Life&nbsp; -->
<a href="datasets/Undocumented">
 <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
 </a>
 <b>
  <a href="datasets/Undocumented">
  Undocumented
 </a>
 </b>
 <!-- <td>Various datasets without documentation (feel free to explore!)&nbsp;
<!-- <td>Other&nbsp; -->
<a href="datasets/Twenty+Newsgroups">
 <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
```

```
<b>
   <a href="datasets/Twenty+Newsgroups">
    Twenty Newsgroups
   </a>
   </b>
   <!-- <td>This data set consists of 20000 messages taken from 20 newsgroups.&nbsp;</
p> -->
  Text
  20000
  1999
  <!-- <td>Other&nbsp; -->
 <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>
   </b>
```

·/+d>

```
<!-- <td>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. 
  Multivariate, Time-Series
  Classification
   Categorical, Real
   6650
   15
   1999
   <!-- <td>Other&nbsp; -->
  <a href="datasets/Australian+Sign+Language+signs+%28High+Quality%29">
    <img border="1" src="assets/MLimages/SmallLarge114.jpg"/>
    </a>
    <a href="datasets/Australian+Sign+Language+signs+%28High+Quality%29">
     Australian Sign Language signs (High Quality)
     </a>
    </b>
    <!-- <td>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&nbsp
; -->
```

```
Multivariate, Time-Series
  Classification
  Real
  2565
  22
  2002
  <!-- <td>Other&nbsp; -->
 <a href="datasets/US+Census+Data+%281990%29">
   <img border="1" src="assets/MLimages/SmallLarge2.jpg"/>
   </a>
   <b>
    <a href="datasets/US+Census+Data+%281990%29">
    US Census Data (1990)
    </a>
   </b>
   <!-- <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. 
  Multivariate
```

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

```
299285
  40
  2000
  <!-- <td>Social&nbsp; -->
  <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>
   </b>
   <!-- <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. 
 -->
  Multivariate
  Categorical, Real
  340
```

```
17
  1999
  <!-- <td>Physical&nbsp; -->
  <a href="datasets/Corel+Image+Features">
    <img border="1" src="assets/MLimages/SmallLarge119.jpg"/>
   <a href="datasets/Corel+Image+Features">
    Corel Image Features
    </a>
    </b>
   <!-- <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 c
o-occurence  -->
  Multivariate
  Real
  68040
  89
  1999
```

```
<!-- <td>Other&nbsp; -->
  <a href="datasets/E.+Coli+Genes">
    <img border="1" src="assets/MLimages/SmallLarge120.jpg"/>
   <a href="datasets/E.+Coli+Genes">
    E. Coli Genes
    </a>
    </b>
   <!-- <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.
  -->
  Relational
  2001
  <!-- <td>Life&nbsp; -->
  <a href="datasets/EEG+Database">
    <img border="1" src="assets/MLimages/SmallLarge121.ipg"/>
```

```
</a>
   <b>
    <a href="datasets/EEG+Database">
    EEG Database
    </a>
    </b>
   <!-- <td>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; -->
  Multivariate, Time-Series
  Categorical, Integer, Real
  122
  4
  1999
  <!-- <td>Life&nbsp; -->
  <a href="datasets/El+Nino">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/EI+Nino">
    FI Nino
```

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

-- -- class="normal". This data contains a record of user interactions with the Entree Chicago restau

```
rant recommendation system.  -->
  Transactional, Sequential
  Recommender-Systems
  Categorical
  50672
  2000
  <!-- <td>Other&nbsp; -->
  <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>
    </b>
   <!-- <td>This data consists of 640 black and white face images of people taken with var
ying pose (straight, left, right, up), expression (neutral, happy, sad, angry), eyes (wearing sunglasses or not), a
nd size  -->
  Image
```

```
Classification
  Integer
  640
  1999
  <!-- <td>Other&nbsp; -->
  <a href="datasets/Insurance+Company+Benchmark+%28COIL+2000%29">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Insurance+Company+Benchmark+%28COIL+2000%29">
    Insurance Company Benchmark (COIL 2000)
    </a>
    </b>
   <!-- <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  -->
  Multivariate
  Regression, Description
```

```
Categorical, integer
  9000
  86
  2000
  <!-- <td>Social&nbsp; -->
 <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>
   </b>
   <!-- <td>This data contains general demographic information on internet users in 1997.
  -->
  Multivariate
  Categorical, Integer
  10104
```

```
72
  1999
  <!-- <td>Computer&nbsp; -->
 <a href="datasets/IPUMS+Census+Database">
   <img border="1" src="assets/MLimages/SmallLarge2.jpg"/>
   </a>
   <b>
   <a href="datasets/IPUMS+Census+Database">
    IPUMS Census Database
   </a>
   </b>
   <!-- <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. 
  Multivariate
  Categorical, Integer
  256932
  61
  1999
```

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

```
<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>
    </b>
   <!-- <td>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 
  Multivariate
  Regression
  Categorical, Integer
  191779
  481
  1998
  <!-- <td>Other&nbsp; -->
  <a href="datasets/KDD+Cup+1999+Data">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
```

```
<b>
    <a href="datasets/KDD+Cup+1999+Data">
    KDD Cup 1999 Data
    </a>
    </b>
   <!-- <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 
  Multivariate
  Classification
  Categorical, Integer
  4000000
  42
  1999
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/M.+Tuberculosis+Genes">
    <img border="1" src="assets/MLimages/SmallLarge131.jpg"/>
   </a>
   <a href="datasets/M.+Tuberculosis+Genes">
    M. Tuberculosis Genes
    </b>
```

```
<!-- <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) ar
e provided  -->
  Relational
  2001
  <!-- <td>Life&nbsp; -->
  <a href="datasets/Movie">
    <img border="1" src="assets/MLimages/SmallLarge132.jpg"/>
    </a>
   <b>
    <a href="datasets/Movie">
     Movie
    </a>
    </b>
    <!-- <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. 
  Multivariate, Relational
```

```
10000
  1999
  <!-- <td>Other&nbsp; -->
  <a href="datasets/MSNBC.com+Anonymous+Web+Data">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/MSNBC.com+Anonymous+Web+Data">
    MSNBC.com Anonymous Web Data
    </a>
   </b>
   <!-- <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.
  -->
  Sequential
  Categorical
```

-/+ds

```
989818
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/NSF+Research+Award+Abstracts+1990-2003">
    <img border="1" src="assets/MLimages/SmallLarge134.jpg"/>
   </a>
   <b>
    <a href="datasets/NSF+Research+Award+Abstracts+1990-2003">
    NSF Research Award Abstracts 1990-2003
    </a>
    </b>
   <!-- <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  -->
  Text
  129000
```

```
2003
  <!-- <td>Other&nbsp; -->
  <a href="datasets/Pioneer-1+Mobile+Robot+Data">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <b>
    <a href="datasets/Pioneer-1+Mobile+Robot+Data">
    Pioneer-1 Mobile Robot Data
    </a>
    </b>
   <!-- <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  -->
  Multivariate, Time-Series
  Categorical, Real
  1999
  <!-- <td>Computer&nbsp; -->
```

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

```
<a href="datasets/Reuters-21578+Text+Categorization+Collection">
     Reuters-21578 Text Categorization Collection
    </b>
    <!-- <td>This is a collection of documents that appeared on Reuters newswire in 1987.
The documents were assembled and indexed with categories.  -->
  Text
  Classification
  Categorical
  21578
  5
  1997
  <!-- <td>Other&nbsp; -->
  >
    <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>
    </b>
```

```
<!-- <td>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 
 -->
  Multivariate, Time-Series
  Classification
  Integer
  463
  90
  1999
  <!-- <td>Physical&nbsp; -->
  <a href="datasets/Synthetic+Control+Chart+Time+Series">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
    <a href="datasets/Synthetic+Control+Chart+Time+Series">
     Synthetic Control Chart Time Series
    </a>
    </b>
    <!-- <td>This data consists of synthetically generated control charts.&nbsp; -->
  Time-Series
```

```
Classification, Clustering
  Real
  600
  1999
  <!-- <td>Other&nbsp; -->
  <a href="datasets/Syskill+and+Webert+Web+Page+Ratings">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Syskill+and+Webert+Web+Page+Ratings">
    Syskill and Webert Web Page Ratings
    </a>
    </b>
   <!-- <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; a
nd BioMedical)  -->
  Multivariate, Text
  Classification
```

```
Categorical
  332
  5
  1998
  <!-- <td>Computer&nbsp; -->
 <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>
   </b>
   <!-- <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.  -->
  Text, Sequential
```

```
<!-- <td>Computer&nbsp; -->
  >
   <a href="datasets/Volcanoes+on+Venus+-+JARtool+experiment">
   <img border="1" src="assets/MLimages/SmallLarge142.jpg"/>
   <b>
    <a href="datasets/Volcanoes+on+Venus+-+JARtool+experiment">
    Volcanoes on Venus - JARtool experiment
    </a>
   </b>
   <!-- <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. </t
d> -->
  Image
  Classification
  <!-- <td>Physical&nbsp; -->
```

```
<a href="datasets/Statlog+%28Australian+Credit+Approval%29">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <b>
    <a href="datasets/Statlog+%28Australian+Credit+Approval%29">
     Statlog (Australian Credit Approval)
    </b>
    <!-- <td>This file concerns credit card applications. This database exists elsewhere in the
e repository (Credit Screening Database) in a slightly different form  -->
  Multivariate
  Classification
  Categorical, Integer, Real
  690
  14
  <!-- <td>Financial&nbsp; -->
  <a href="datasets/Statlog+%28German+Credit+Data%29">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
```

```
</a>
   <b>
    <a href="datasets/Statlog+%28German+Credit+Data%29">
     Statlog (German Credit Data)
    </a>
    </b>
    <!-- <td>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 
  Multivariate
  Classification
  Categorical, Integer
  1000
  20
  1994
  <!-- <td>Financial&nbsp; -->
  <a href="datasets/Statlog+%28Heart%29">
    <img border="1" src="assets/MLimages/SmallLarge45.jpg"/>
    </a>
   <a href="datasets/Statlog+%28Heart%29">
     Statlog (Heart)
```

```
</a>
    </b>
    <!-- <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  -->
  Multivariate
  Classification
  Categorical, Real
  270
  13
  <!-- <td>Life&nbsp; -->
  <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>
    </b>
    <!-- <td>Multi-spectral values of pixels in 3x3 neighbourhoods in a satellite image, and t
```

```
he classification associated with the central pixel in each neighbourhood  -->
  Multivariate
  Classification
  Integer
  6435
  36
  1993
  <!-- <td>Physical&nbsp; -->
  <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>
    </b>
    <!-- <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. 
  Multivariate
```

```
Classification
  Real
  2310
  19
  1990
  <!-- <td>Other&nbsp; -->
 <a href="datasets/Statlog+%28Shuttle%29">
   <img border="1" src="assets/MLimages/SmallLarge92.jpg"/>
   </a>
   <a href="datasets/Statlog+%28Shuttle%29">
    Statlog (Shuttle)
    </a>
   </b>
   <!-- <td>The shuttle dataset contains 9 attributes all of which are numerical. Approximat
ely 80% of the data belongs to class 1  -->
  Multivariate
  Classification
  Integer
```

```
58000
  9
  <!-- <td>Physical&nbsp; -->
  <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>
   </b>
   <!-- <td>3D objects within a 2D image by application of an ensemble of shape feature e
xtractors to the 2D silhouettes of the objects.  -->
  Multivariate
  Classification
  Integer
  946
```

```
18
  <!-- <td>Other&nbsp; -->
 <a href="datasets/Connectionist+Bench+%28Nettalk+Corpus%29">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <b>>
    <a href="datasets/Connectionist+Bench+%28Nettalk+Corpus%29">
    Connectionist Bench (Nettalk Corpus)
   </b>
   <!-- <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 
  Multivariate
  Categorical
  20008
  4
```

```
<!-- <td>Other&nbsp; -->
  <a href="datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <b>
    <a href="datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29">
     Connectionist Bench (Sonar, Mines vs. Rocks)
    </b>
    <!-- <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.  -->
  Multivariate
  Classification
  Real
  208
  60
  <!-- <td>Physical&nbsp; -->
  <a href="datasets/Connectionist+Bench+%28Vowel+Becognition+-+Deterding+Data%29">
```

```
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
    <a href="datasets/Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29">
     Connectionist Bench (Vowel Recognition - Deterding Data)
    </a>
    </b>
    <!-- <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.  -->
  Classification
  Real
  528
  10
  <!-- <td>Other&nbsp; -->
  <a href="datasets/Economic+Sanctions">
    <img border="1" src="assets/MLimages/SmallLarge153.jpg"/>
    </a>
   <a href="datasets/Economic+Sanctions">
     Economic Sanctions
```

```
</b>
 <!-- <td>Domain Theory on Economic Sanctions; Undocumented&nbsp; -->
Domain-Theory
<!-- <td>Financial&nbsp; -->
<a href="datasets/Protein+Data">
 <img border="1" src="assets/MLimages/SmallLarge154.jpg"/>
 </a>
 <a href="datasets/Protein+Data">
 Protein Data
 </a>
 </b>
 <!-- <td>Undocumented&nbsp; -->
```

```
<!-- <td>Life&nbsp; -->
<a href="datasets/Cloud">
 <img border="1" src="assets/MLimages/SmallLarge155.jpg"/>
</a>
<a href="datasets/Cloud">
 Cloud
 </a>
 </b>
<!-- <td>Little Documentation&nbsp; -->
Multivariate
Real
1024
```

```
10
  1989
  <!-- <td>Physical&nbsp; -->
 <a href="datasets/Callt2+Building+People+Counts">
   <img border="1" src="assets/MLimages/SmallLarge156.jpg"/>
   <a href="datasets/Callt2+Building+People+Counts">
    Callt2 Building People Counts
   </a>
   </b>
   <!-- <td>This data comes from the main door of the Callt2 building at UCI.&nbsp;</t
d> -->
  Multivariate, Time-Series
  Categorical, Integer
  10080
  4
  2006
```

```
<!-- <td>Other&nbsp; -->
 <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>
   </b>
   <!-- <td>Loop sensor data was collected for the Glendale on ramp for the 101 North fre
eway in Los Angeles  -->
  Multivariate, Time-Series
  Categorical, Integer
  50400
  3
  2006
  <!-- <td>Other&nbsp; -->
```

```
<a href="datasets/Poker+Hand">
 <img border="1" src="assets/MLimages/SmallLarge158.jpg"/>
 </a>
 <b>
  <a href="datasets/Poker+Hand">
  Poker Hand
  </a>
 </b>
 <!-- <td>Purpose is to predict poker hands&nbsp; -->
Multivariate
Classification
Categorical, Integer
1025010
11
2007
<!-- <td>Game&nbsp; -->
<a href="datasets/MAGIC+Gamma+Telescope">
 <img border="1" src="assets/MLimages/SmallLarge159.jpg"/>
 </a>
 <b>
```

```
<a href="datasets/MAGIC+Gamma+Telescope">
    MAGIC Gamma Telescope
    </b>
   <!-- <td>Data are MC generated to simulate registration of high energy gamma particle
s in an atmospheric Cherenkov telescope  -->
  Multivariate
  Classification
  Real
  19020
  11
  2007
  <!-- <td>Physical&nbsp; -->
  <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>
    </b>
```

```
<!-- <td>Data consists of written characters in a UNIPEN-like format&nbsp; -->
  Multivariate, Sequential
  Classification
  Integer
  1364
  2007
  <!-- <td>Computer&nbsp; -->
 <a href="datasets/Mammographic+Mass">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Mammographic+Mass">
    Mammographic Mass
    </a>
   </b>
   <!-- <td>Discrimination of benign and malignant mammographic masses based on BI-R
ADS attributes and the patient's age.  -->
  Multivariate
```

```
Classification
  Integer
  961
  6
  2007
  <!-- <td>Life&nbsp; -->
  <a href="datasets/Forest+Fires">
    <img border="1" src="assets/MLimages/SmallLarge162.jpg"/>
   </a>
   <b>
    <a href="datasets/Forest+Fires">
    Forest Fires
    </a>
    </b>
   <!-- <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://ww
w.dsi.uminho.pt/~pcortez/forestfires).  -->
  Multivariate
  Regression
  ~td>
```

```
Real
  517
  13
  2008
  <!-- <td>Physical&nbsp; -->
 <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>
   </b>
   <!-- <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.  -->
  Text
  Classification
```

```
200
  2008
  <!-- <td>Business&nbsp; -->
 <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>
   </b>
   <!-- <td>This data set contains five text collections in the form of bags-of-words.&nbsp;<
/p> -->
  Text
  Clustering
  Integer
  8000000
  100000
```

```
2008
  <!-- <td>Other&nbsp; -->
  <a href="datasets/Concrete+Compressive+Strength">
    <img border="1" src="assets/MLimages/SmallLarge165.jpg"/>
   <b>
    <a href="datasets/Concrete+Compressive+Strength">
    Concrete Compressive Strength
    </a>
    </b>
   <!-- <td>Concrete is the most important material in civil engineering. The concrete com
pressive strength is a highly nonlinear function of age and ingredients.   -->
  Multivariate
  Regression
  Real
  1030
  9
  2007
  <!-- <td>Physical&nbsp; -->
```

```
<a href="datasets/Hill-Valley">
    <img border="1" src="assets/MLimages/SmallLarge166.jpg"/>
   <a href="datasets/Hill-Valley">
    Hill-Valley
    </a>
    </b>
   <!-- <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 �bump� in the terrain)
or a Valley (a �dip� in the terrain).  -->
  Sequential
  Classification
  Real
  606
  101
  2008
  <!-- <td>Other&nbsp; -->
  <a href="datasets/Arcene">
```

```
<img border="1" src="assets/MLimages/SmallLarge167.jpg"/>
   <b>
    <a href="datasets/Arcene">
    Arcene
    </a>
    </b>
    <!-- <td>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. 
  Multivariate
  Classification
  Real
  900
  10000
  2008
  <!-- <td>Life&nbsp; -->
  <a href="datasets/Dexter">
    <img border="1" src="assets/MLimages/SmallLarge168.jpg"/>
    </a>
   <b>
```

```
<a href="datasets/Dexter">
     Dexter
    </a>
    </b>
    <!-- <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.
  -->
  Multivariate
   Classification
   Integer
   2600
   20000
   2008
   <!-- <td>Other&nbsp; -->
  <a href="datasets/Dorothea">
    <img border="1" src="assets/MLimages/SmallLarge169.jpg"/>
    </a>
   <b>
    <a href="datasets/Dorothea">
     Dorothea
    </a>
```



```
<!-- <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.    -->
  Multivariate
   Classification
   Integer
   1950
   100000
   2008
   <!-- <td>Life&nbsp; -->
  <a href="datasets/Gisette">
    <img border="1" src="assets/MLimages/SmallLarge170.jpg"/>
    </a>
    <a href="datasets/Gisette">
     Gisette
     </a>
    </b>
    <!-- <td>GISETTE is a handwritten digit recognition problem. The problem is to separat
e the highly confusible digits '4' and '9'. This dataset is one of five datasets of the NIPS 2003 feature selection c
```

```
hallenge.
  -->
  Multivariate
  Classification
  Integer
  13500
  5000
  2008
  <!-- <td>Computer&nbsp; -->
  >
    <a href="datasets/Madelon">
    <img border="1" src="assets/MLimages/SmallLarge171.jpg"/>
    </a>
   <b>
    <a href="datasets/Madelon">
     Madelon
    </a>
    </b>
    <!-- <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.   -->
  Multivariate
   </n>
```

```
Classification
  Real
  4400
  500
  2008
  <!-- <td>Other&nbsp; -->
  <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>
    </b>
    <!-- <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.  -->
  Multivariate, Sequential, Time-Series
  Classification
   /+d·
```

```
Real
  2536
  73
  2008
  <!-- <td>Physical&nbsp; -->
  <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>
    </b>
    <!-- <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.   -->
  Multivariate
  Causal-Discovery
  Integer
```

```
300
43
2008
<!-- <td>Life&nbsp; -->
<a href="datasets/Parkinsons">
 <img border="1" src="assets/MLimages/SmallLarge174.jpg"/>
 <b>
 <a href="datasets/Parkinsons">
 Parkinsons
 </a>
 </b>
 <!-- <td>Oxford Parkinson's Disease Detection Dataset&nbsp; -->
Multivariate
Classification
Real
197
23
```

```
2008
  <!-- <td>Life&nbsp; -->
  <a href="datasets/Character+Trajectories">
    <img border="1" src="assets/MLimages/SmallLarge175.jpg"/>
   </a>
   <b>
    <a href="datasets/Character+Trajectories">
    Character Trajectories
    </a>
    </b>
    <!-- <td>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.  -->
  Time-Series
  Classification, Clustering
  Real
  2858
  3
  2008
```

```
<!-- <td>Computer&nbsp; -->
 <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>
   </b>
   <!-- <td>Data taken from the Blood Transfusion Service Center in Hsin-Chu City in Taiw
an -- this is a classification problem.   -->
  Multivariate
  Classification
  Real
  748
  5
  2008
  <!-- <td>Business&nbsp; -->
```

```
<a href="datasets/UJI+Pen+Characters+%28Version+2%29">
    <img border="1" src="assets/MLimages/SmallLarge160.jpg"/>
   </a>
   <b>
    <a href="datasets/UJI+Pen+Characters+%28Version+2%29">
    UJI Pen Characters (Version 2)
    </a>
    </b>
   <!-- <td>A pen-based database with more than 11k isolated handwritten characters&nb
sp; -->
  Multivariate, Sequential
  Classification
  Integer
  11640
  2009
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Semeion+Handwritten+Digit">
    <img border="1" src="assets/MLimages/SmallLarge178.jpg"/>
   </a>
   <b>
```

```
<a href="datasets/Semeion+Handwritten+Digit">
    Semeion Handwritten Digit
    </b>
   <!-- <td>1593 handwritten digits from around 80 persons were scanned, stretched in a r
ectangular box 16x16 in a gray scale of 256 values.  -->
  Multivariate
  Classification
  Integer
  1593
  256
  2008
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/SECOM">
    <img border="1" src="assets/MLimages/SmallLarge179.jpg"/>
   </a>
   <a href="datasets/SECOM">
    SECOM
    </a>
    </b>
```

```
<!-- <td>Data from a semi-conductor manufacturing process&nbsp; -->
  Multivariate
  Classification, Causal-Discovery
  Real
  1567
  591
  2008
  <!-- <td>Computer&nbsp; -->
 <a href="datasets/Plants">
   <img border="1" src="assets/MLimages/SmallLarge180.jpg"/>
   </a>
   <a href="datasets/Plants">
    Plants
    </a>
   </b>
   <!-- <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. 
  Multivariate
  </n>
```

```
Clustering
  Categorical
  22632
  70
  2008
  <!-- <td>Life&nbsp; -->
  <a href="datasets/Libras+Movement">
    <img border="1" src="assets/MLimages/SmallLarge181.jpg"/>
   </a>
   <a href="datasets/Libras+Movement">
    Libras Movement
    </a>
    </b>
   <!-- <td>The data set contains 15 classes of 24 instances each. Each class references t
o a hand movement type in LIBRAS (Portuguese
name 'Longua BRAsileira de Sinais', oficial brazilian signal language). 
  Multivariate, Sequential
  Classification, Clustering
  _/td>
```

```
Real
  360
  91
  2009
  <!-- <td>Other&nbsp; -->
 <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>
   </b>
   <!-- <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. 
  Multivariate
  Regression
  Real
  alaga "narmal".
```

```
103
  10
  2009
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Communities+and+Crime">
   <img border="1" src="assets/MLimages/SmallLarge183.jpg"/>
   </a>
   <b>
    <a href="datasets/Communities+and+Crime">
    Communities and Crime
    </a>
   </b>
   <!-- <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.  -->
  Multivariate
  Regression
  Real
  1994
```

```
120
  2009
  <!-- <td>Social&nbsp; -->
  <a href="datasets/Acute+Inflammations">
    <img border="1" src="assets/MLimages/SmallLarge184.jpg"/>
   </a>
   <a href="datasets/Acute+Inflammations">
    Acute Inflammations
    </a>
    </b>
   <!-- <td>The data was created by a medical expert as a data set to test the expert syste
which will perform the presumptive diagnosis of two diseases of the urinary system.
  -->
  Multivariate
  Classification
  Categorical, Integer
  120
  6
```

m,

```
2009
  <!-- <td>Life&nbsp; -->
  <a href="datasets/Wine+Quality">
    <img border="1" src="assets/MLimages/SmallLarge186.jpg"/>
   </a>
   <a href="datasets/Wine+Quality">
    Wine Quality
    </a>
    </b>
   <!-- <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/).  -->
  Multivariate
  Classification, Regression
  Real
  4898
  12
  2009
  <!-- <td>Business&nbsp; -->
```

```
<a href="datasets/URL+Reputation">
    <img border="1" src="assets/MLimages/SmallLarge187.jpg"/>
   </a>
   <a href="datasets/URL+Reputation">
    URL Reputation
    </a>
    </b>
   <!-- <td>Anonymized 120-day subset of the ICML-09 URL data containing 2.4 million ex
amples and 3.2 million features.   -->
  Multivariate, Time-Series
  Classification
  Integer, Real
  2396130
  3231961
  2009
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/p53+Mutants">
    <img border="1" src="assets/MLimages/SmallLarge188.jpg"/>
   </a>
```

```
<b>
    <a href="datasets/p53+Mutants">
    p53 Mutants
    </a>
    </b>
   <!-- <td>The goal is to model mutant p53 transcriptional activity (active vs inactive) base
d on data extracted from biophysical simulations.
  -->
  Multivariate
  Classification
  Real
  16772
  5409
  2010
  <!-- <td>Life&nbsp; -->
  <a href="datasets/Parkinsons+Telemonitoring">
    <img border="1" src="assets/MLimages/SmallLarge174.jpg"/>
   </a>
   <a href="datasets/Parkinsons+Telemonitoring">
    Parkinsons Telemonitoring
```

```
</a>
 </b>
 <!-- <td>Oxford Parkinson's Disease Telemonitoring Dataset&nbsp; -->
Multivariate
Regression
Integer, Real
5875
26
2009
<!-- <td>Life&nbsp; -->
<a href="datasets/Demospongiae">
 <img border="1" src="assets/MLimages/SmallLarge190.jpg"/>
 </a>
 <a href="datasets/Demospongiae">
  Demospongiae
  </a>
 </b>
 <!-- <td>Marine sponges of the Demospongiae class classification domain.&nbsp:
```

```
td> -->
  Multivariate
  Classification
  Integer
  503
  2010
  <!-- <td>Life&nbsp; -->
  <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>
    </b>
   <!-- <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".   -->
  Text
  class="normal">
```

```
51
  2010
  <!-- <td>Computer&nbsp; -->
 <a href="datasets/Breast+Tissue">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
  <a href="datasets/Breast+Tissue">
    Breast Tissue
   </a>
   </b>
   <!-- <td>Dataset with electrical impedance measurements of freshly excised tissue sam
ples from the breast.  -->
  Multivariate
  Classification
  Real
  n clace_"normal"s
```

```
106
  10
  2010
  <!-- <td>Life&nbsp; -->
 <a href="datasets/Cardiotocography">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Cardiotocography">
    Cardiotocography
    </a>
   </b>
   <!-- <td>The dataset consists of measurements of fetal heart rate (FHR) and uterine co
ntraction (UC) features on cardiotocograms classified by expert obstetricians. 
  Multivariate
  Classification
  Real
  2126
  23
```

```
2010
  <!-- <td>Life&nbsp; -->
  <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>
    </b>
    <!-- <td>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'.  -->
  Multivariate, Sequential
  Classification
  Real
  5456
  24
  2010
```

```
<!-- <td>Computer&nbsp; -->
  <a href="datasets/Spoken+Arabic+Digit">
    <img border="1" src="assets/MLimages/SmallLarge195.jpg"/>
   </a>
   <b>
    <a href="datasets/Spoken+Arabic+Digit">
    Spoken Arabic Digit
    </a>
    </b>
   <!-- <td>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; -->
  Multivariate, Time-Series
  Classification
  Real
  8800
  13
  2010
  <!-- <td>Other&nbsp; -->
```

```
<a href="datasets/Localization+Data+for+Person+Activity">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <a href="datasets/Localization+Data+for+Person+Activity">
     Localization Data for Person Activity
    </b>
    <!-- <td>Data contains recordings of five people performing different activities. Each per
son wore four sensors (tags) while performing the same scenario five times.   -->
  Univariate, Sequential, Time-Series
  Classification
  Real
  164860
  8
  2010
  <!-- <td>Life&nbsp; -->
  <a href="datasets/AutoUniv">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
```

```
<a href="datasets/AutoUniv">
     AutoUniv
    </a>
    </b>
    <!-- <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.&nbs
p; -->
  Multivariate
  Classification
  Categorical, Integer, Real
  2010
  <!-- <td>Other&nbsp; -->
  <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>
    </b>
```

```
<!-- <td>A dataset of steel plates' faults, classified into 7 different types.
The goal was to train machine learning for automatic pattern recognition.
  -->
  Multivariate
  Classification
  Integer, Real
  1941
  27
  2010
  <!-- <td>Physical&nbsp; -->
  <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>
    </b>
    <!-- <td>This dataset is taken from the MiniBooNE experiment and is used to distinguish
electron neutrinos (signal) from muon neutrinos (background).
```

```
Multivariate
  Classification
  Real
  130065
  50
  2010
  <!-- <td>Physical&nbsp; -->
 <a href="datasets/YearPredictionMSD">
   <img border="1" src="assets/MLimages/SmallLarge203.jpg"/>
   </a>
   <a href="datasets/YearPredictionMSD">
    YearPredictionMSD
    </a>
   </b>
   <!-- <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. 
  Multivariate
```

Regression

```
Real
  515345
  90
  2011
  <!-- <td>Other&nbsp; -->
  >
   <a href="datasets/PEMS-SF">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/PEMS-SF">
    PEMS-SF
    </a>
   </b>
   <!-- <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. </t
d> -->
  Multivariate, Time-Series
  Classification
  Real
```

```
440
  138672
  2011
  <!-- <td>Computer&nbsp; -->
 <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>
   </b>
   <!-- <td>This data set contains user reviews of cars and and hotels collected from Tripa
dvisor (~259,000
reviews) and Edmunds (~42,230 reviews).   -->
  Text
```

```
2011
  <!-- <td>Computer&nbsp; -->
  <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>
    </b>
   <!-- <td>The dataset consists of 384 features extracted from CT images. The class vari
able is numeric and denotes the relative location of the CT slice on the axial axis of the human body. </p
> -->
  Domain-Theory
  Regression
  Real
  53500
  386
  2011
```

```
<!-- <td>Computer&nbsp; -->
  <a href="datasets/Online+Handwritten+Assamese+Characters+Dataset">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Online+Handwritten+Assamese+Characters+Dataset">
    Online Handwritten Assamese Characters Dataset
    </a>
    </b>
   <!-- <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. 
-->
  Multivariate, Sequential
  Classification
  Integer
  8235
  2011
  <!-- <td>Computer&nbsp; -->
```

```
>
    <a href="datasets/PubChem+Bioassay+Data">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
    <a href="datasets/PubChem+Bioassay+Data">
     PubChem Bioassay Data
    </a>
    </b>
    <!-- <td>These highly imbalanced bioassay datasets are from the differing types of scre
ening that can be performed using HTS technology. 21 datasets were created from 12 bioassays. 
d> -->
  Multivariate
  Classification
  Integer, Real
  2011
  <!-- <td>Life&nbsp; -->
  <a href="datasets/Record+Linkage+Comparison+Patterns">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   < b>
```

```
<a href="datasets/Record+Linkage+Comparison+Patterns">
     Record Linkage Comparison Patterns
    </b>
    <!-- <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; -->
  Multivariate
  Classification
  Real
  5749132
  12
  2011
  <!-- <td>Other&nbsp; -->
  <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>
    </b>
```

```
<!-- <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  -->
   Multivariate
   Regression
   Real
   2215
   147
   2011
   <!-- <td>Social&nbsp; -->
  <a href="datasets/Vertebral+Column">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
     <a href="datasets/Vertebral+Column">
     Vertebral Column
     </a>
    </b>
    <!-- <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; -->
```

```
Multivariate
  Classification
  Real
  310
  6
  2011
  <!-- <td>&nbsp; -->
  <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>
    </b>
   <!-- <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.  -->
  Time-Series
```

```
Classification
  Real
  10000
  8
  2011
  <!-- <td>Physical&nbsp; -->
  <a href="datasets/Vicon+Physical+Action+Data+Set">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Vicon+Physical+Action+Data+Set">
     Vicon Physical Action Data Set
    </a>
    </b>
   <!-- <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.  -->
  Time-Series
  Classification
  class="normal">
```

```
Real
  3000
  27
  2011
  <!-- <td>Physical&nbsp; -->
  <a href="datasets/Amazon+Commerce+reviews+set">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Amazon+Commerce+reviews+set">
    Amazon Commerce reviews set
    </a>
   </b>
   <!-- <td>The dataset is used for authorship identification in online Writeprint which is a n
ew research field of pattern recognition.   -->
  Multivariate, Text, Domain-Theory
  Classification
  Real
  1500
```

```
10000
  2011
  <!-- <td>Physical&nbsp; -->
  <a href="datasets/Amazon+Access+Samples">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Amazon+Access+Samples">
    Amazon Access Samples
    </a>
    </b>
   <!-- <td>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. 
  Time-Series, Domain-Theory
  Regression, Clustering, Causal-Discovery
  30000
  20000
```

```
2011
  <!-- <td>Business&nbsp; -->
 <a href="datasets/Reuter_50_50">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Reuter_50_50">
    Reuter_50_50
    </a>
   </b>
   <!-- <td>The dataset is used for authorship identification in online Writeprint which is a n
ew research field of pattern recognition.   -->
  Multivariate, Text, Domain-Theory
  Classification, Clustering
  Real
  2500
  10000
  2011
  <!-- <td>Computer&nbsp; -->
```

```
<ta>
   <a href="datasets/Farm+Ads">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Farm+Ads">
    Farm Ads
    </a>
    </b>
   <!-- <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.  -->
  Text
  Classification
  4143
  54877
  2011
  <!-- <td>Business&nbsp; -->
  <a href="datasets/DBWorld+e-mails">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
```

```
<b>
    <a href="datasets/DBWorld+e-mails">
     DBWorld e-mails
    </a>
    </b>
    <!-- <td>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'.  -->
  Text
  Classification
  64
  4702
  2011
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
    <a href="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29">
     KEGG Metabolic Relation Network (Directed)
    </a>
    </b>
```

```
<!-- <td>KEGG Metabolic pathways modeled as directed relation network. Variety of gra
phical features presented.  -->
  Multivariate, Univariate, Text
  Classification, Regression, Clustering
  Integer, Real
  53414
  24
  2011
  <!-- <td>Life&nbsp; -->
  <a href="datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <a href="datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29">
     KEGG Metabolic Reaction Network (Undirected)
    </b>
    <!-- <td>KEGG Metabolic pathways modeled as un-directed reaction network. Variety of
graphical features presented.  -->
```

```
Multivariate, Univariate, Text
  Classification, Regression, Clustering
  Integer, Real
  65554
  29
  2011
  <!-- <td>Life&nbsp; -->
  <a href="datasets/Bank+Marketing">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Bank+Marketing">
     Bank Marketing
    </a>
    </b>
   <!-- <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).
  -->
  Multivariate
```

```
Classification
  Real
  45211
  17
  2012
  <!-- <td>Business&nbsp; -->
  <a href="datasets/YouTube+Comedy+Slam+Preference+Data">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/YouTube+Comedy+Slam+Preference+Data">
    YouTube Comedy Slam Preference Data
    </b>
   <!-- <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.  -->
  Text
  Classification
```

```
1138562
  3
  2012
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Gas+Sensor+Array+Drift+Dataset">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Gas+Sensor+Array+Drift+Dataset">
    Gas Sensor Array Drift Dataset
    </a>
    </b>
   <!-- <td>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; -->
  Multivariate
  Classification
  Real
  13910
```

//n>

```
128
  2012
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/ILPD+%28Indian+Liver+Patient+Dataset%29">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/ILPD+%28Indian+Liver+Patient+Dataset%29">
    ILPD (Indian Liver Patient Dataset)
    </a>
    </b>
   <!-- <td>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. 
  Multivariate
  Classification
  Integer, Real
  583
  10
```

```
2012
  <!-- <td>Life&nbsp; -->
  <a href="datasets/OPPORTUNITY+Activity+Recognition">
    <img border="1" src="assets/MLimages/SmallLarge226.jpg"/>
    </a>
   <a href="datasets/OPPORTUNITY+Activity+Recognition">
     OPPORTUNITY Activity Recognition
    </b>
    <!-- <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).  -->
  Multivariate, Time-Series
  Classification
  Real
  2551
  242
  2012
  <!-- <td>Computer&nbsp; -->
```

```
<a href="datasets/Nomao">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Nomao">
    Nomao
    </a>
   </b>
   <!-- <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.  -->
  Univariate
  Classification
  Real
  34465
  120
  2012
  <!-- <td>Computer&nbsp; -->
```

```
>
   <a href="datasets/SMS+Spam+Collection">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/SMS+Spam+Collection">
    SMS Spam Collection
    </a>
    </b>
   <!-- <td>The SMS Spam Collection is a public set of SMS labeled messages that have b
een collected for mobile phone spam research.  -->
  Multivariate, Text, Domain-Theory
  Classification, Clustering
  Real
  5574
  2012
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Skin+Segmentation">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
```

```
<a href="datasets/Skin+Segmentation">
     Skin Segmentation
    </a>
    </b>
   <!-- <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.  -->
  Univariate
  Classification
  Real
  245057
  4
  2012
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Planning+Relax">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Planning+Relax">
     Planning Relax
    </a>
    </b>
```

```
<!-- <td>The dataset concerns with the classification of two mental stages from recorde
d EEG signals: Planning (during imagination of motor act) and Relax state.   -->
  Univariate
   Classification
   Real
   182
   13
   2012
   <!-- <td>Computer&nbsp; -->
  <a href="datasets/PAMAP2+Physical+Activity+Monitoring">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
    <a href="datasets/PAMAP2+Physical+Activity+Monitoring">
     PAMAP2 Physical Activity Monitoring
     </a>
    </b>
    <!-- <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. &nbs
p; -->
```

```
Multivariate, Time-Series
  Classification
  Real
  3850505
  52
  2012
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Restaurant+%26+consumer+data">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Restaurant+%26+consumer+data">
    Restaurant & amp; consumer data
    </a>
   </b>
   <!-- <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.  
  Multivariate
```

<a>

```
138
  47
  2012
  <!-- <td>Computer&nbsp; -->
 <a href="datasets/CNAE-9">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/CNAE-9">
    CNAE-9
   </a>
   </b>
   <!-- <td>This is a data set containing 1080 documents of free text business descriptions
of Brazilian companies categorized into a
subset of 9 categories  -->
  Multivariate, Text
  Classification
  Integer
  <ht>td>
```

```
1080
  857
  2012
  <!-- <td>Business&nbsp; -->
  <a href="datasets/Individual+household+electric+power+consumption">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <a href="datasets/Individual+household+electric+power+consumption">
     Individual household electric power consumption
    </a>
    </b>
    <!-- <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.  -->
  Multivariate, Time-Series
  Regression, Clustering
  Real
  2075259
  n alace_"normal"s
```

```
~p 01435= 110111141 /
   9
  2012
  <!-- <td>Physical&nbsp; -->
  <a href="datasets/seeds">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/seeds">
    seeds
    </a>
    </b>
   <!-- <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.  -->
  Multivariate
  Classification, Clustering
  Real
  210
```

```
2012
  <!-- <td>Life&nbsp; -->
 <a href="datasets/Northix">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Northix">
    Northix
    </a>
   </b>
   <!-- <td>Northix is designed to be a schema matching benchmark problem for data inte
gration of two entity relationship databases.   -->
  Multivariate, Univariate, Text
  Classification
  Integer, Real
  115
  200
  2012
  <!-- <td>Computer&nbsp; -->
```

```
<a href="datasets/QtyT40I10D100K">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/QtyT40I10D100K">
    QtyT40I10D100K
    </a>
    </b>
   <!-- <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 
  Sequential
  Integer
  3960456
  4
  2012
  <!-- <td>&nbsp; -->
  <a href="datasets/Legal+Case+Reports">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
```

```
<b>
    <a href="datasets/Legal+Case+Reports">
     Legal Case Reports
    </a>
    </b>
   <!-- <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 cl
asses.  -->
  Text
  Classification
  2012
  <!-- <td>Other&nbsp; -->
  <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>
    </b>
```

```
<!-- <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.  -->
  Multivariate, Time-Series
   Classification, Clustering
   10299
   561
   2012
   <!-- <td>Computer&nbsp; -->
  <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>
    </b>
    <!-- <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.
```

```
Classification
  Real
  1600
  64
  2012
  <!-- <td>Life&nbsp; -->
 >
  <a href="datasets/Energy+efficiency">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Energy+efficiency">
    Energy efficiency
    </a>
   </b>
   <!-- <td>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.  
  Multivariate
  Classification, Regression
```

```
Integer, Real
  768
  8
  2012
  <!-- <td>Computer&nbsp; -->
 <a href="datasets/Yacht+Hydrodynamics">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Yacht+Hydrodynamics">
    Yacht Hydrodynamics
    </a>
   </b>
   <!-- <td>Delft data set, used to predict the hydodynamic performance of sailing yachts fr
om dimensions and velocity.  -->
  Multivariate
  Regression
  Real
  308
```

```
2013
  <!-- <td>Physical&nbsp; -->
  <a href="datasets/Fertility">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Fertility">
    Fertility
    </a>
    </b>
   <!-- <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 li
fe habits  -->
  Multivariate
  Classification, Regression
  Real
  100
  10
```

/n.

```
2013
  <!-- <td>Life&nbsp; -->
  <a href="datasets/Daphnet+Freezing+of+Gait">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <a href="datasets/Daphnet+Freezing+of+Gait">
    Daphnet Freezing of Gait
    </a>
    </b>
   <!-- <td>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.
  -->
  Multivariate, Time-Series
  Classification
  Real
  237
  9
  2013
```

```
<!-- <td>Life&nbsp; -->
  <a href="datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29">
    3D Road Network (North Jutland, Denmark)
    </a>
    </b>
   <!-- <td>3D road network with highly accurate elevation information (+-20cm) from Den
mark used in eco-routing and fuel/Co2-estimation routing algorithms. 
  Sequential, Text
  Regression, Clustering
  Real
  434874
  4
  2013
  <!-- <td>Computer&nbsp; -->
```

```
<a href="datasets/ISTANBUL+STOCK+EXCHANGE">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <b>
    <a href="datasets/ISTANBUL+STOCK+EXCHANGE">
     ISTANBUL STOCK EXCHANGE
    </a>
    </b>
    <!-- <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; -->
  Multivariate, Univariate, Time-Series
  Classification, Regression
  Real
  536
  8
  2013
  <!-- <td>Business&nbsp; -->
  <a href="datasets/Buzz+in+social+media+">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
```

```
<b>
     <a href="datasets/Buzz+in+social+media+">
     Buzz in social media
     </a>
    </b>
    <!-- <td>This data-set contains examples of buzz events from two different social netwo
rks: Twitter, and Tom's Hardware, a forum network focusing on new technology with more conservative dynami
cs.  -->
  Time-Series, Multivariate
  Regression, Classification
  Integer, Real
   140000
   77
   2013
   <!-- <td>Computer&nbsp; -->
  <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>
    </b>
```

```
<!-- <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.&
nbsp; -->
                    Multivariate
                      Classification
                      Real
                      6118
                      51
                      2013
                      <!-- <td>Computer&nbsp; -->
                 <a href="datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28P"><a href="datasets/wear
UC-Rio%29">
                                 <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
                              </a>
                            <b>
                                   <a href="datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28"><a href="datasets/Wearable+Computing%3A+Clas
PUC-Rio%29">
                                      Wearable Computing: Classification of Body Postures and Movements (PUC-Rio)
                                   </a>
                                 </b>
```

```
<!-- <td>A dataset with 5 classes (sitting-down, standing-up, standing, walking, and sitti
ng) collected on 8 hours of activities of 4 healthy subjects. We also established a baseline performance index.&
nbsp; -->
  Sequential
   Classification
   Integer, Real
   165632
   18
   2013
   <!-- <td>Computer&nbsp; -->
  <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>
    </b>
    <!-- <td>The dataset contains 18000 time-series recordings from a chemical detection p
latform at six different locations in a wind tunnel facility in response to ten high-priority chemical gaseous substa
nces  -->
```

```
Multivariate, Time-Series
  Classification
  Real
  18000
  1950000
  2013
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Climate+Model+Simulation+Crashes">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Climate+Model+Simulation+Crashes">
    Climate Model Simulation Crashes
    </a>
    </b>
   <!-- <td>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
.  -->
  Multivariate
  Classification
```

```
Real
  540
  18
  2013
  <!-- <td>Physical&nbsp; -->
 >
   <a href="datasets/MicroMass">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/MicroMass">
    MicroMass
   </a>
   </b>
   <!-- <td>A dataset to explore machine learning approaches for the identification of micr
oorganisms from mass-spectrometry data.  -->
  Multivariate
  Classification
  Real
```

```
931
  1300
  2013
  <!-- <td>Life&nbsp; -->
 <a href="datasets/QSAR+biodegradation">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/QSAR+biodegradation">
    QSAR biodegradation
    </a>
   </b>
   <!-- <td>Data set containing values for 41 attributes (molecular descriptors) used to clas
sify 1055 chemicals into 2 classes (ready and not ready biodegradable).  -->
  Multivariate
  Classification
  Integer, Real
  1055
```

n alace_"normal"s

```
41
  2013
  <!-- <td>Other&nbsp; -->
 <a href="datasets/BLOGGER">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/BLOGGER">
    BLOGGER
   </a>
   </b>
   <!-- <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  -->
  Multivariate
  Classification
  100
  6
  2013
```

```
<!-- <td>Computer&nbsp; -->
  <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>
    </b>
   <!-- <td>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.
  -->
  Multivariate, Time-Series
  Classification, Clustering
  Real
  9120
  5625
  2013
  <!-- <td>Computer&nbsp; -->
```

```
<ta>
   <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>
    </b>
    <!-- <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. 
  Multivariate
  Classification, Clustering
  Integer
  403
  5
  2013
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collecti
on">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
```

```
</a>
                 <b>
                     <a href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collect"><a href="datasets/Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reuters+Reut
ion">
                       Reuters RCV1 RCV2 Multilingual, Multiview Text Categorization Test collection
                     </a>
                    </b>
                  <!-- <td>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.   -->
            Multivariate
             Classification
             Real
             111740
             2013
             <!-- <td>Life&nbsp; -->
          <a href="datasets/NYSK">
                    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
                  </a>
                 <b>
                     <a href="datasets/NYSK">
                       NYSK
```

```
</a>
    </b>
    <!-- <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).  -->
  Multivariate, Sequential, Text
  Clustering
  10421
  2013
  <!-- <td>Social&nbsp; -->
  <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>
    </b>
```

```
<!-- <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
.  -->
   Multivariate
   Classification, Clustering
   5820
   33
   2013
   <!-- <td>Other&nbsp; -->
  <a href="datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electr
ical+Machines%29">
     <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
    <b>
     <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>
     </b>
    <!-- <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.  -->
```

```
Multivariate
  Classification
  Real
  403
  2013
  <!-- <td>Computer&nbsp; -->
 <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>
   </b>
   <!-- <td>The data set consists of 14 EEG values and a value indicating the eye state.&n
bsp; -->
  Multivariate, Sequential, Time-Series
  Classification
  </n>
```

```
Integer, Real
  14980
  15
  2013
  <!-- <td>Life&nbsp; -->
  <a href="datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
    <a href="datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure">
     Physicochemical Properties of Protein Tertiary Structure
    </a>
    </b>
    <!-- <td>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> -->
  Multivariate
  Regression
  Real
```

/td>

```
45730
  9
  2013
  <!-- <td>Life&nbsp; -->
 <a href="datasets/seismic-bumps">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/seismic-bumps">
    seismic-bumps
    </a>
   </b>
   <!-- <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.  
  Multivariate
  Classification
  Real
  2584
```

```
19
  2013
  <!-- <td>Other&nbsp; -->
 <a href="datasets/banknote+authentication">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <a href="datasets/banknote+authentication">
    banknote authentication
   </a>
   </b>
   <!-- <td>Data were extracted from images that were taken for the evaluation of an auth
entication procedure for banknotes.  -->
  Multivariate
  Classification
  Real
  1372
  5
```

```
2013
   <!-- <td>Computer&nbsp; -->
  <a href="datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+T">+ Lab+and+T</a>
opCoder++++Problem%3A+Pat">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <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>
    <!-- <td>Data used for USPTO Algorithm Competition. Contains drawing pages from US
patents with manually labeled figure and part labels.  -->
  Domain-Theory
  Classification
  Integer
  306
  5
  2013
  <!-- <td>Other&nbsp; -->
```

```
<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
    </b>
    <!-- <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; -->
  Multivariate, Text
  Classification, Clustering
  Integer, Real
  120000
  1000000
  2013
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations">
```

```
<img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    <b>
     <a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations">
     Gas Sensor Array Drift Dataset at Different Concentrations
     </a>
    </b>
    <!-- <td>This archive contains 13910 measurements from 16 chemical sensors expose
d to 6 different gases at various concentration levels.  -->
   Multivariate, Time-Series
   Classification, Regression, Clustering, Causa
   Real
   13910
   129
   2013
   <!-- <td>Computer&nbsp; -->
  <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
    </b>
   <!-- <td>This dataset comprises information regarding the ADLs performed by two user
s on a daily basis in their
own homes.   -->
  Multivariate, Sequential, Time-Series
  Classification, Clustering
  2747
  2013
  <!-- <td>Computer&nbsp; -->
  <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>
    </b>
```

```
<!-- <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). 
  Multivariate
   Regression
   Integer, Real
   3395
   20
   2013
   <!-- <td>Game&nbsp; -->
  <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>
    </b>
    <!-- <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. 
  Multivariate
```

```
Classification
  Real
  39242
  152
  2013
  <!-- <td>Physical&nbsp; -->
 <a href="datasets/SML2010">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/SML2010">
    SML2010
    </a>
   </b>
   <!-- <td>This dataset is collected from a monitor system mounted in a domotic house. It
corresponds to approximately 40 days of monitoring data.  -->
  Multivariate, Sequential, Time-Series, Text
  Regression
  class="normal">
```

```
Real
  4137
  24
  2014
  <!-- <td>Computer&nbsp; -->
  <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>
    </b>
   <!-- <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. </
p> -->
  Univariate
  Regression
  Integer, Real
```

17200

```
16
  2013
  <!-- <td>Social&nbsp; -->
  <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>
    </b>
    <!-- <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 set
ting.  -->
  Multivariate, Sequential, Time-Series
  Integer, Real
  51
  35
```

```
</tu>
  2013
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Thoracic+Surgery+Data">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Thoracic+Surgery+Data">
    Thoracic Surgery Data
    </a>
    </b>
   <!-- <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. </
p> -->
  Multivariate
  Classification
  Integer, Real
  470
  17
  2013
```

```
<!-- <ta>Life&nosp;</ta> -->
  <a href="datasets/EMG+dataset+in+Lower+Limb">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <b>
    <a href="datasets/EMG+dataset+in+Lower+Limb">
    EMG dataset in Lower Limb
    </a>
    </b>
   <!-- <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. 
-->
  Multivariate, Time-Series
  Real
  132
  5
  2014
  <!-- <td>Computer&nbsp; -->
```

```
<a href="datasets/SUSY">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <b>
    <a href="datasets/SUSY">
    SUSY
    </a>
    </b>
   <!-- <td>This is a classification problem to distinguish between a signal process which p
roduces supersymmetric particles and a background process which does not.  -->
  Classification
  Real
  5000000
  18
  2014
  <!-- <td>Physical&nbsp; -->
  <a href="datasets/HIGGS">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/HIGGS">
```

```
HIGGS
    </a>
    </b>
   <!-- <td>This is a classification problem to distinguish between a signal process which p
roduces Higgs bosons and a background process which does not.   -->
  Classification
  Real
  11000000
  28
  2014
  <!-- <td>Physical&nbsp; -->
  <a href="datasets/Qualitative_Bankruptcy">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Qualitative_Bankruptcy">
    Qualitative_Bankruptcy
    </a>
    </b>
```

```
<!-- <td>Predict the Bankruptcy from Qualitative parameters from experts.&nbsp;</t
d> -->
  Multivariate
  Classification
  250
  2014
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/LSVT+Voice+Rehabilitation">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
    <a href="datasets/LSVT+Voice+Rehabilitation">
     LSVT Voice Rehabilitation
    </a>
    </b>
    <!-- <td>126 samples from 14 participants, 309 features. Aim: assess whether voice reh
abilitation treatment lead to phonations considered 'acceptable' or 'unacceptable' (binary class classification pro
blem).  -->
  Multivariate
```

```
Classification
  Real
  126
  309
  2014
  <!-- <td>Life&nbsp; -->
  <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>
    </b>
   <!-- <td>Recordings of 16 volunteers performing 14 Activities of Daily Living (ADL) while
carrying a single wrist-worn tri-axial accelerometer.  -->
  Multivariate, Time-Series
  Classification, Clustering
```

```
3
  2014
  <!-- <td>Computer&nbsp; -->
 <a href="datasets/Wilt">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <a href="datasets/Wilt">
    Wilt
    </a>
   </b>
   <!-- <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 o
f image.  -->
  Multivariate
  Classification
  4889
  ~td>
```

```
2014
  <!-- <td>Life&nbsp; -->
  <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>
    </b>
   <!-- <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.
  -->
  Univariate, Sequential, Time-Series
  Classification, Clustering
  Real
  2014
```

```
<!-- <td>Other&nbsp; -->
  <a href="datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer">
    Activity Recognition from Single Chest-Mounted Accelerometer
    </b>
   <!-- <td>The dataset collects data from a wearable accelerometer mounted on the ches
t. The dataset is intended for Activity Recognition research purposes. 
  Univariate, Sequential, Time-Series
  Classification, Clustering
  Real
  2014
  <!-- <td>Other&nbsp; -->
```

f H.L. L. . . L. /L .

```
<a nrei= datasets/Lear >
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Leaf">
    Leaf
    </a>
    </b>
   <!-- <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. 
  Multivariate
  Classification
  Real
  340
  16
  2014
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Dresses_Attribute_Sales">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
```

```
<a href="datasets/Dresses_Attribute_Sales">
     Dresses_Attribute_Sales
     </a>
    </b>
    <!-- <td>This dataset contain Attributes of dresses and their recommendations according
g to their sales. Sales are monitor on the basis of alternate days.   -->
   Text
   Classification, Clustering
   501
   13
   2014
   <!-- <td>Computer&nbsp; -->
  <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>
    </b>
```

```
<!-- <td>This data can be effectively produced the result to fewer parameter of the Load
profile can be reduced in the Database   -->
  Multivariate
  Classification, Regression, Clustering
  Real
  45781
  5
  2013
  <!-- <td>Life&nbsp; -->
  <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>
    </b>
    <!-- <td>NASA data set, obtained from a series of aerodynamic and acoustic tests of tw
o and three-dimensional airfoil blade sections conducted in an anechoic wind tunnel. 
  Multivariate
```

```
Regression
  Real
  1503
  6
  2014
  <!-- <td>Physical&nbsp; -->
 <a href="datasets/Wholesale+customers">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Wholesale+customers">
    Wholesale customers
    </a>
   </b>
   <!-- <td>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 
  Multivariate
  Classification, Clustering
```

```
Integer
  440
  8
  2014
  <!-- <td>Business&nbsp; -->
  <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>
    </b>
   <!-- <td>This problem of Sentiment Analysis (SA) has been studied well on the English I
anguage but not Arabic one. Two main approaches have been devised: corpus-based and lexicon-based. &nbs
p; -->
  Text
  Classification
  2000
```

```
2
  2014
  <!-- <td>Social&nbsp; -->
 <a href="datasets/Combined+Cycle+Power+Plant">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Combined+Cycle+Power+Plant">
    Combined Cycle Power Plant
    </a>
   </b>
   <!-- <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.  
  Multivariate
  Regression
  Real
  9568
  4
```

```
2014
  <!-- <td>Computer&nbsp; -->
 <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>
   </b>
   <!-- <td>Classification of urban land cover using high resolution aerial imagery. Intende
d to assist sustainable urban planning efforts.  -->
  Multivariate
  Classification
  168
  148
  2014
  <!-- <td>Physical&nbsp; -->
```

```
<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>
    </b>
    <!-- <td>This data has been prepared to analyze factors related to readmission as well
as other
outcomes pertaining to patients with diabetes.  -->
  Multivariate
  Classification, Clustering
  Integer
  100000
  55
  2014
  <!-- <td>Life&nbsp; -->
  f llalada a ada /Da ala . Ola aval . I la visa a istill
```

```
<a nite = ualasels/bach+Gnoral+hannony >
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <a href="datasets/Bach+Choral+Harmony">
     Bach Choral Harmony
    </a>
    </b>
    <!-- <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.  -->
  Sequential
  Classification
  5665
  17
  2014
  <!-- <td>Other&nbsp; -->
  <a href="datasets/StoneFlakes">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
```

```
<a href="datasets/StoneFlakes">
     StoneFlakes
    </a>
    </b>
    <!-- <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. 
  Multivariate
   Classification, Clustering, Causal-Discovery
   Real
   79
   8
   2014
   <!-- <td>Other&nbsp; -->
  <a href="datasets/Tennis+Major+Tournament+Match+Statistics">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
    <a href="datasets/Tennis+Major+Tournament+Match+Statistics">
     Tennis Major Tournament Match Statistics
    </a>
    </b>
```

```
<!-- <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 row
s.  -->
  Multivariate
   Classification, Regression, Clustering
   Integer, Real
   127
   42
   2014
   <!-- <td>Other&nbsp; -->
  <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 Speech Dataset with Multiple Types of Sound Recordings
     </a>
    </b>
    <!-- <td>The training data belongs to 20 Parkinson's Disease (PD) patients and 20 healt
hy subjects. From all subjects, multiple types of sound recordings (26) are taken.
```

```
Multivariate
  Classification, Regression
  Integer, Real
  1040
  26
  2014
  <!-- <td>Life&nbsp; -->
  <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>
    </b>
   <!-- <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.  -->
  Multivariate, Sequential, Time-Series
```

```
Classification, Clustering
  Real
  9900
  50
  2014
  <!-- <td>Other&nbsp; -->
  <a href="datasets/Perfume+Data">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Perfume+Data">
    Perfume Data
    </a>
    </b>
   <!-- <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.  -->
  Univariate, Domain-Theory
  Classification, Clustering
  Integer
```

```
560
  2
  2014
  <!-- <td>Computer&nbsp; -->
 <a href="datasets/BlogFeedback">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/BlogFeedback">
    BlogFeedback
    </a>
   </b>
   <!-- <td>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.  -->
  Multivariate
  Regression
  Integer, Real
  60021
```

```
281
  2014
  <!-- <td>Social&nbsp; -->
  <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>
    </b>
    <!-- <td>The REALDISP dataset is devised to evaluate techniques dealing with the effect
ts of sensor displacement in wearable activity recognition as well as to benchmark general activity recognition al
gorithms   -->
  Multivariate, Time-Series
  Classification
  Real
  1419
  120
   ~td~
```

```
2014
  <!-- <td>Computer&nbsp; -->
 <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
   </b>
   <!-- <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. 
  Classification
  101
  2014
  <!-- <td>Computer&nbsp; -->
```

```
ヘロン
   <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
    </b>
   <!-- <td>This data set compromises the metadata for the 2014 AAAI conference's acce
pted papers, including paper titles, authors, abstracts, and keywords of varying granularity. 
  Multivariate
  Clustering
  399
  6
  2014
  <!-- <td>Computer&nbsp; -->
  <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
    </b>
    <!-- <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.  -->
  Multivariate, Time-Series
   Classification, Regression
   Real
   58
   120432
   2014
   <!-- <td>Computer&nbsp; -->
  <a href="datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
    <b>
     <a href="datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures">
     Gas sensor array exposed to turbulent gas mixtures
     </a>
    </b>
```

```
<!-- <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.  -->
  Multivariate, Time-Series
   Classification, Regression
   Real
   180
   150000
   2014
   <!-- <td>Computer&nbsp; -->
  <a href="datasets/UJIIndoorLoc">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
    <a href="datasets/UJIIndoorLoc">
     UJIIndoorLoc
    </a>
    </b>
    <!-- <td>The UJIIndoorLoc is a Multi-Building Multi-Floor indoor localization database to
test Indoor Positioning System that rely on WLAN/WiFi fingerprint.
```

```
Multivariate
  Classification, Regression
  Integer, Real
  21048
  529
  2014
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Sentence+Classification">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Sentence+Classification">
    Sentence Classification
    </a>
    </b>
   <!-- <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.  -->
  Text
```

```
Classification
  Integer
  2014
  <!-- <td>Other&nbsp; -->
 <a href="datasets/Dow+Jones+Index">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
   <a href="datasets/Dow+Jones+Index">
    Dow Jones Index
   </a>
   </b>
   <!-- <td>This dataset contains weekly data for the Dow Jones Industrial Index. It has be
en used in computational investing research.  -->
  Time-Series
  Classification, Clustering
  Integer, Real
```

```
750
  16
  2014
  <!-- <td>Business&nbsp; -->
  >
   <a href="datasets/sEMG+for+Basic+Hand+movements">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/sEMG+for+Basic+Hand+movements">
    sEMG for Basic Hand movements
    </a>
    </b>
   <!-- <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.  -->
  Time-Series
  Classification
  Real
  3000
```

```
2500
  2014
  <!-- <td>Life&nbsp; -->
 <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>
   </b>
   <!-- <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. 
 -->
  Multivariate
  Clustering
  150
  5
  2014
```

```
<!-- <td>Computer&nbsp; -->
  <a href="datasets/Geographical+Original+of+Music">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Geographical+Original+of+Music">
    Geographical Original of Music
    </a>
    </b>
   <!-- <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.
  -->
  Multivariate
  Classification, Regression
  Real
  1059
  68
  2014
  <!-- <td>Other&nbsp; -->
```

```
<a href="datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    >
    <a href="datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants">
     Condition Based Maintenance of Naval Propulsion Plants
    </b>
    <!-- <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.  -->
  Multivariate
   Regression
   Real
   11934
   16
   2014
   <!-- <td>Computer&nbsp; -->
  <a href="datasets/Grammatical+Facial+Expressions">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
```

```
</a>
   <b>
    <a href="datasets/Grammatical+Facial+Expressions">
    Grammatical Facial Expressions
    </b>
   <!-- <td>This dataset supports the development of models that make possible to interpr
et Grammatical Facial Expressions from Brazilian Sign Language (Libras). 
  Multivariate, Sequential
  Classification, Clustering
  Real
  27965
  100
  2014
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/NoisyOffice">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/NoisyOffice">
    NoisyOffice
```

```
</b>
   <!-- <td>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.  -->
  Multivariate
  Classification, Regression
  Real
  216
  216
  2015
  <!-- <td>Computer&nbsp; -->
  >
  <a href="datasets/MHEALTH+Dataset">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/MHEALTH+Dataset">
    MHEALTH Dataset
    </a>
    </b>
```



```
<!-- <td>The MHEALTH (Mobile Health) dataset is devised to benchmark techniques de
aling with human behavior analysis based on multimodal body sensing.  -->
  Multivariate, Time-Series
  Classification
  Real
  120
  23
  2014
  <!-- <td>Computer&nbsp; -->
 <a href="datasets/Student+Performance">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Student+Performance">
    Student Performance
    </a>
    </b>
   <!-- <td>Predict student performance in secondary education (high school). &nbsp;
 -->
  Multivariate
```

```
Classification, Regression
  Integer
  649
  33
  2014
  <!-- <td>Social&nbsp; -->
 <a href="datasets/ElectricityLoadDiagrams20112014">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/ElectricityLoadDiagrams20112014">
    ElectricityLoadDiagrams20112014
    </a>
   </b>
   <!-- <td>This data set contains electricity consumption of 370 points/clients.
  -->
  Time-Series
  Regression, Clustering
```

```
Real
  370
  140256
  2015
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Gas+sensor+array+under+dynamic+gas+mixtures">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Gas+sensor+array+under+dynamic+gas+mixtures">
    Gas sensor array under dynamic gas mixtures
    </a>
    </b>
   <!-- <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.  -->
  Multivariate, Time-Series
  Classification, Regression
  Real
```

```
4178504
  19
  2015
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/microblogPCU">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/microblogPCU">
    microblogPCU
    </a>
    </b>
   <!-- <td>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.  </p
> -->
  Multivariate, Univariate, Sequential, Text
  Classification, Causal-Discovery
  Integer, Real
  221579
  20
```

```
2015
  <!-- <td>Computer&nbsp; -->
 <a href="datasets/Firm-Teacher_Clave-Direction_Classification">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Firm-Teacher_Clave-Direction_Classification">
    Firm-Teacher_Clave-Direction_Classification
   </b>
   <!-- <td>The data are binary attack-point vectors and their clave-direction class(es) acc
ording to the partido-alto-based paradigm.   -->
  Multivariate
  Classification
  10800
  20
  2015
```

```
<a href="datasets/Dataset+for+Sensorless+Drive+Diagnosis">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <a href="datasets/Dataset+for+Sensorless+Drive+Diagnosis">
    Dataset for Sensorless Drive Diagnosis
    </b>
   <!-- <td>Features are extracted from motor current. The motor has intact and defective
components. This results in 11 different classes with different conditions.  
  Multivariate
  Classification
  Real
  58509
  49
  2015
  <!-- <td>Computer&nbsp; -->
  hraf "datacata/TV: Nava : Channel : Commorcial : Dataction : Datacat"
```

```
<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>
    </b>
   <!-- <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).   -->
  Multivariate
  Classification, Clustering
  Real
  129685
  12
  2015
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Phishing+Websites">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
```

```
<a href="datasets/Phishing+Websites">
    Phishing Websites
    </a>
    </b>
   <!-- <td>This dataset collected mainly from: PhishTank archive, MillerSmiles archive, G
oogle's searching operators.  -->
  Classification
  Integer
  2456
  30
  2015
  <!-- <td>Computer Security&nbsp; -->
  <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>
    </b>
```

```
<!-- <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.
  -->
  Multivariate, Time-Series
   Regression
   Real
   2921
   5232
   2015
   <!-- <td>Physical&nbsp; -->
  <a href="datasets/Diabetic+Retinopathy+Debrecen+Data+Set">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
    <b>
     <a href="datasets/Diabetic+Retinopathy+Debrecen+Data+Set">
     Diabetic Retinopathy Debrecen Data Set
     </a>
    </b>
    <!-- <td>This dataset contains features extracted from the Messidor image set to predic
t whether an image contains signs of diabetic retinopathy or not.   -->
```

```
Multivariate
  Classification
  Integer, Real
  1151
  20
  2014
  <!-- <td>Life&nbsp; -->
  <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>
    </b>
   <!-- <td>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). 
> -->
  Multivariate
  Classification
```

```
Categorical
  6590
  2015
  <!-- <td>Life&nbsp; -->
 >
  <a href="datasets/Sentiment+Labelled+Sentences">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
   <a href="datasets/Sentiment+Labelled+Sentences">
    Sentiment Labelled Sentences
   </a>
   </b>
   <!-- <td>The dataset contains sentences labelled with positive or negative sentiment.&n
bsp; -->
  Text
  Classification
```

```
3000
  2015
  <!-- <td>Other&nbsp; -->
  <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>
    </b>
   <!-- <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).  -->
  Multivariate
  Classification, Regression
  Integer, Real
  39797
  61
```

```
2015
  <!-- <td>Business&nbsp; -->
  <a href="datasets/Forest+type+mapping">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Forest+type+mapping">
    Forest type mapping
    </a>
    </b>
   <!-- <td>Multi-temporal remote sensing data of a forested area in Japan. The goal is to
map different forest types using spectral data.  -->
  Multivariate
  Classification
  326
  27
  2015
  <!-- <td><n class="normal">| ife&nhsn:</n> -->
```

```
<a href="datasets/wiki4HE">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/wiki4HE">
    wiki4HE
    </a>
    </b>
   <!-- <td>Survey of faculty members from two Spanish universities on teaching uses of
Wikipedia  -->
  Multivariate
  Regression, Clustering, Causal-Discovery
  913
  53
  2015
  <!-- <td>Social&nbsp; -->
  <a href="datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset">
    sima bordor_"1" cro_"accots/MI imagos/Small argodofault ing"/s
```

```
</a>
   <b>
    <a href="datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset">
    Online Video Characteristics and Transcoding Time Dataset
    </b>
    <!-- <td>The dataset contains a million randomly sampled video instances listing 10 fun
damental video characteristics along with the YouTube video ID.  
  Multivariate
  Regression
  Integer, Real
  168286
  11
  2015
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Chronic_Kidney_Disease">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Chronic_Kidney_Disease">
```

```
Childrid_Numey_bisease
    </a>
    </b>
    <!-- <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.  -->
  Multivariate
  Classification
  Real
  400
  25
  2015
  <!-- <td>Other&nbsp; -->
  <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>
    </b>
```

```
<!-- <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
  -->
  Sequential, Time-Series
  Classification
  Real
  314080
  0
  2015
  <!-- <td>Business&nbsp; -->
  <a href="datasets/Folio">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Folio">
    Folio
    </a>
    </b>
   <!-- <td>20 photos of leaves for each of 32 different species.&nbsp; -->
  Multivariate
```

```
Classification, Clustering
  637
  20
  2015
  <!-- <td>Other&nbsp; -->
  <a href="datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
    <a href="datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015">
     Taxi Service Trajectory - Prediction Challenge, ECML PKDD 2015
    </a>
    </b>
    <!-- <td>An accurate dataset describing trajectories performed by all the 442 taxis runni
ng in the city of Porto, in Portugal.
  -->
  Multivariate, Sequential, Time-Series, Domain-Theory
  Clustering, Causal-Discovery
```

```
Real
  1710671
  2015
  <!-- <td>Computer&nbsp; -->
  <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>
    </b>
   <!-- <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. 
  Multivariate
  Classification, Regression
  Real
  12000
```

```
3
   2015
   <!-- <td>Life&nbsp; -->
  <a href="datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
     <a href="datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions">
     Smartphone-Based Recognition of Human Activities and Postural Transitions
     </a>
    </b>
    <!-- <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 sens
  -->
  Multivariate, Time-Series
   Classification
   Real
   10929
   561
```

```
2015
  <!-- <td>Life&nbsp; -->
  <a href="datasets/Mice+Protein+Expression">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Mice+Protein+Expression">
    Mice Protein Expression
    </a>
    </b>
   <!-- <td>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.  -->
  Multivariate
  Classification, Clustering
  Real
  1080
  82
  2015
  </n>
```

```
<!-- <td>Life&nbsp; -->
 <a href="datasets/UJIIndoorLoc-Mag">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <b>
    <a href="datasets/UJIIndoorLoc-Mag">
    UJIIndoorLoc-Mag
    </a>
   </b>
   <!-- <td>The UJIIndoorLoc-Mag is an indoor localization database to test Indoor Positio
ning System that rely on Earth's magnetic field variations.  -->
  Multivariate, Sequential, Time-Series
  Classification, Regression, Clustering
  Integer, Real
  40000
  13
  2015
  <!-- <td>Computer&nbsp; -->
```

```
<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>
     </b>
    <!-- <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.  -->
   Multivariate, Time-Series
   Classification, Clustering
   Real
   43930257
   16
   2015
   <!-- <td>Computer&nbsp; -->
  <a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set">
     <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
```

```
</a>
   <b>
    <a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set">
     Educational Process Mining (EPM): A Learning Analytics Data Set
    </a>
    </b>
    <!-- <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. & nbsp; 
  Multivariate, Sequential, Time-Series
  Classification, Regression, Clustering
  Integer
  230318
  13
  2015
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/HEPMASS">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <a href="datasets/HEPMASS">
     HEPMASS
```

```
</a>
    </b>
    <!-- <td>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: -->
  Multivariate
  Classification
  Real
  10500000
  28
  2016
  <!-- <td>Physical&nbsp; -->
  <a href="datasets/Indoor+User+Movement+Prediction+from+RSS+data">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
    <a href="datasets/Indoor+User+Movement+Prediction+from+RSS+data">
     Indoor User Movement Prediction from RSS data
    </a>
    </b>
```

```
<!-- <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 Li
ving.  -->
  Multivariate, Sequential, Time-Series
  Classification
   Real
   13197
   4
   2016
   <!-- <td>Computer&nbsp; -->
  <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>
    </b>
    <!-- <td>Open University Learning Analytics Dataset contains data about courses, stude
nts and their interactions with Virtual Learning Environment for seven selected courses and more than 30000 st
udents.  -->
```

```
Multivariate, Sequential, Time-Series
  Classification, Regression, Clustering
  Integer
  2015
  <!-- <td>Computer&nbsp; -->
  <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>
    </b>
   <!-- <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. 
  Multivariate
  Classification
```

```
Integer, Real
  30000
  24
  2016
  <!-- <td>Business&nbsp; -->
  <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>
    </b>
   <!-- <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. & nbsp;
 -->
  Multivariate
  Classification
  Real
```

```
324
  34
  2016
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Online+Retail">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Online+Retail">
    Online Retail
    </a>
    </b>
   <!-- <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. 
  Multivariate, Sequential, Time-Series
  Classification, Clustering
  Integer, Real
  541909
```

```
2015
  <!-- <td>Business&nbsp; -->
  <a href="datasets/SIFT10M">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/SIFT10M">
    SIFT10M
    </a>
   </b>
   <!-- <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. </
p> -->
  Multivariate
  Causal-Discovery
  Integer
  11164866
  128
  2016
  e/n
```

```
<!-- <td>Computer&nbsp; -->
 <a href="datasets/GPS+Trajectories">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <b>
    <a href="datasets/GPS+Trajectories">
    GPS Trajectories
    </a>
   </b>
   <!-- <td>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).  
  Multivariate
  Classification, Regression
  Real
  163
  15
  2016
  <!-- <td>Computer&nbsp; -->
```

```
<a href="datasets/Detect+Malacious+Executable%28AntiVirus%29">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    <a href="datasets/Detect+Malacious+Executable%28AntiVirus%29">
     Detect Malacious Executable(AntiVirus)
    </b>
    <!-- <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.  -->
  Multivariate
   Classification
   Real
   373
   513
   2016
   <!-- <td>Computer&nbsp; -->
  <a href="datasets/Occupancy+Detection+">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
```

```
<a href="datasets/Occupancy+Detection+">
    Occupancy Detection
    </a>
   </b>
   <!-- <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.  -->
  Multivariate, Time-Series
  Classification
  Real
  20560
  2016
  <!-- <td>Computer&nbsp; -->
  <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"/>
   %E2%80%99s+Disease">
```

```
improved Spiral Lest Using Digitized Graphics Lablet for Monitoring Parkinson's Disease
     </a>
    </b>
    <!-- <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; -->
  >
   Multivariate
   Classification, Regression, Clustering
  Real
   40
   2016
   <!-- <td>Computer&nbsp; -->
  <a href="datasets/News+Aggregator">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <a href="datasets/News+Aggregator">
     News Aggregator
     </a>
    </b>
```

```
<!-- <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.  -->
  Multivariate
   Classification, Clustering
   422937
   5
   2016
   <!-- <td>Other&nbsp; -->
  <a href="datasets/Air+Quality">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <a href="datasets/Air+Quality">
     Air Quality
    </a>
    </b>
    <!-- <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.   -->
```

```
Multivariate, Time-Series
  Regression
  Real
  9358
  15
  2016
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Twin+gas+sensor+arrays">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Twin+gas+sensor+arrays">
    Twin gas sensor arrays
    </a>
    </b>
   <!-- <td>5 replicates of an 8-MOX gas sensor array were exposed to different gas condi
tions (4 volatiles at 10 concentration levels each).  -->
  Multivariate, Time-Series, Domain-Theory
  Classification, Regression
```

```
Real
  640
  480000
  2016
  <!-- <td>Computer&nbsp; -->
  <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>
    </b>
   <!-- <td>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.
  -->
  Multivariate, Time-Series
  Classification
  Real
```

```
919438
  11
  2016
  <!-- <td>Computer&nbsp; -->
 <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
   </b>
   <!-- <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.  
  Multivariate
  Regression
  Integer, Real
  40949
  >
```

```
54
  2016
  <!-- <td>Other&nbsp; -->
  <a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+As">–28HAR%29+in+Ambient+As</a>
sisted+Living+%28AAL%29">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   Assisted+Living+%28AAL%29">
     Smartphone Dataset for Human Activity Recognition (HAR) in Ambient Assisted Living (AAL)
    </a>
    </b>
   <!-- <td>This data is an addition to an existing dataset on UCI. We collected more data t
o improve the accuracy of our human activity recognition algorithms applied in the domain of Ambient Assisted
Living.   -->
  Time-Series
  Classification
  Real
  5744
  561
  /td>
```

```
2016
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Polish+companies+bankruptcy+data">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
    <a href="datasets/Polish+companies+bankruptcy+data">
     Polish companies bankruptcy data
    </a>
    </b>
    <!-- <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.  -->
  Multivariate
  Classification
  Real
  10503
  64
  2016
  n alaga "narmal", Duainaga anhany //n. //td.
```

```
<a href="datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29">
    <img border="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)
    </b>
   <!-- <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. 
  Multivariate, Sequential, Time-Series
  Classification
  Real
  42240
  6
  2016
  <!-- <td>Computer&nbsp; -->
```

```
<a nrei= datasets/Dota2+Games+Results >
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
    <a href="datasets/Dota2+Games+Results">
     Dota2 Games Results
    </b>
    <!-- <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. 
  Multivariate
  Classification
  102944
  116
  2016
  <!-- <td>Game&nbsp; -->
  <a href="datasets/Facebook+metrics">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <a href="datasets/Facebook+metrics">
```

```
Facebook metrics
    </a>
   </b>
   <!-- <td>Facebook performance metrics of a renowned cosmetic's brand Facebook pag
e.  -->
  Multivariate
  Regression
  Integer
  500
  19
  2016
  <!-- <td>Business&nbsp; -->
  <a href="datasets/UbiqLog+%28smartphone+lifelogging%29">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/UbiqLog+%28smartphone+lifelogging%29">
    UbiqLog (smartphone lifelogging)
    </a>
   </b>
```

```
<!-- <td>UbiqLog is the smartphone lifelogging tool that runs on the smartphone of 35 u
sers for about 2 months.
  -->
  Multivariate
  Causal-Discovery
  9782222
  2016
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/NIPS+Conference+Papers+1987-2015">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/NIPS+Conference+Papers+1987-2015">
    NIPS Conference Papers 1987-2015
    </a>
    </b>
   <!-- <td>This data set contains the distribution of words in the full text of the NIPS confe
rence papers published from 1987 to 2015.  -->
  Text
```

```
Clustering
  Integer
  11463
  5812
  2016
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/HTRU2">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>>
    <a href="datasets/HTRU2">
    HTRU2
    </a>
    </b>
   <!-- <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.  -->
  >
  Multivariate
  Classification, Clustering
```

```
Real
17898
2017
<!-- <td>Physical&nbsp; -->
>
<a href="datasets/Drug+consumption+%28quantified%29">
 <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
 </a>
 <a href="datasets/Drug+consumption+%28quantified%29">
  Drug consumption (quantified)
  </a>
 </b>
 <!-- <td>Classify type of drug consumer by personality data&nbsp; -->
Multivariate
Classification
Real
1885
```

```
32
  2016
  <!-- <td>Social&nbsp; -->
  <a href="datasets/Appliances+energy+prediction">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Appliances+energy+prediction">
    Appliances energy prediction
    </a>
   </b>
   <!-- <td>Experimental data used to create regression models of appliances energy use i
n a low energy building.  -->
  Multivariate, Time-Series
  Regression
  Real
  19735
  29
  <ht>td>
```

```
2017
  <!-- <td>Computer&nbsp; -->
  <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>
    </b>
   <!-- <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.   -->
  Text
  Classification, Clustering, Causal-Discovery
  Integer
  1540
  67
  2016
  <!-- <td>Computer&nbsp; -->
```

```
<a href="datasets/KDC-4007+dataset+Collection">
                   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
                  </a>
                <b>
                     <a href="datasets/KDC-4007+dataset+Collection">
                      KDC-4007 dataset Collection
                     </a>
                   </b>
                  <!-- <td>KDC-4007 dataset Collection is the Kurdish Documents Classification text used
in categories regarding Kurdish Sorani news and articles.  -->
            Multivariate, Text
             Classification, Regression
             Integer
             4007
             2017
             <!-- <td>Computer&nbsp; -->
          <a href="datasets/Geo-Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland+wland
d+smartphone">
```

"It I are "a cata / Miliana area / Cara III area a la facilità in a "I'/

```
<iiiig border= 1 SiC= assets/iviLimages/SinaiiLargederauit.jpg />
                <a href="datasets/Geo-Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+a">- a href="datasets/Geo-Magnetic+field+and+WLAN+dataset+for+indoor+for+field+and+a">- a href="datasets/Geo-Magnetic+field+and+a">- a 
nd+smartphone">
                      Geo-Magnetic field and WLAN dataset for indoor localisation from wristband and smartphone
                   </b>
                 <!-- <td>A multisource and multivariate dataset for indoor localisation methods based o
n WLAN and Geo-Magnetic field fingerprinting  -->
           Multivariate, Sequential, Time-Series
            Classification, Regression, Clustering
            Integer, Real
            153540
            25
            2017
            <!-- <td>Computer&nbsp; -->
          <a href="datasets/DrivFace">
                   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
                 </a>
                <b>
```

```
<a nref="datasets/DrivFace">
     DrivFace
    </a>
    </b>
    <!-- <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.  -->
  Multivariate
  Classification, Regression, Clustering
  Real
  606
  6400
  2016
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Website+Phishing">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <a href="datasets/Website+Phishing">
     Website Phishing
    </a>
    </b>
```

```
<!-- <td>
  -->
  Multivariate
  Classification
  Integer
  1353
  10
  2016
  <!-- <td>Computer&nbsp; -->
  <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>
    </b>
   <!-- <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 colle
```

ction period. -->

```
Text
  Classification
  1956
  2017
  <!-- <td>Computer&nbsp; -->
 >
   <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>
   </b>
   <!-- <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.  
  Multivariate, Time-Series
  Regression
```

```
Integer, Real
  43824
  13
  2017
  <!-- <td>Physical&nbsp; -->
  <a href="datasets/Cargo+2000+Freight+Tracking+and+Tracing">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <a href="datasets/Cargo+2000+Freight+Tracking+and+Tracing">
     Cargo 2000 Freight Tracking and Tracing
    </a>
    </b>
    <!-- <td>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).   -->
  Multivariate, Sequential
  Classification, Regression
  Integer
```

```
3942
  98
  2016
  <!-- <td>Business&nbsp; -->
 <a href="datasets/Cervical+cancer+%28Risk+Factors%29">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Cervical+cancer+%28Risk+Factors%29">
    Cervical cancer (Risk Factors)
    </a>
   </b>
   <!-- <td>This dataset focuses on the prediction of indicators/diagnosis of cervical cancer
. The features cover demographic information, habits, and historic medical records. 
  Multivariate
  Classification
  Integer, Real
  858
  >
```

```
36
  2017
  <!-- <td>Life&nbsp; -->
 <a href="datasets/Quality+Assessment+of+Digital+Colposcopies">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Quality+Assessment+of+Digital+Colposcopies">
    Quality Assessment of Digital Colposcopies
   </b>
   <!-- <td>This dataset explores the subjective quality assessment of digital colposcopies.
  -->
  Multivariate
  Classification
  Real
  287
  69
  2017
```

```
<!-- <td>Life&nbsp; -->
 <a href="datasets/KASANDR">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/KASANDR">
    KASANDR
    </a>
   </b>
   <!-- <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.  
 -->
  Multivariate
  Causal-Discovery
  Integer
  17764280
  2158859
  2017
  <!-- <td>Life&nbsp; -->
```

```
>
    <a href="datasets/FMA%3A+A+Dataset+For+Music+Analysis">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <b>
    <a href="datasets/FMA%3A+A+Dataset+For+Music+Analysis">
     FMA: A Dataset For Music Analysis
    </b>
    <!-- <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
.  -->
  Multivariate, Time-Series
   Classification, Clustering
   Real
   106574
   518
   2017
   <!-- <td>Computer&nbsp; -->
  <a href="datasets/Air+quality">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
```

くいひろ

```
</a>
   <b>
    <a href="datasets/Air+quality">
    Air quality
    </a>
    </b>
   <!-- <td> Contains the responses of a gas multisensor device deployed on the field in a
n Italian city.   -->
  Multivariate, Time-Series
  Regression
  Real
  9358
  15
  2016
  <!-- <td>Other&nbsp; -->
  <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>
    </b>
   <!-- <td>This dataset is a pre-processed and re-structured/reshaped version of a very c
ommonly used dataset featuring epileptic seizure detection.   -->
  Multivariate, Time-Series
  Classification, Clustering
  Integer, Real
  11500
  179
  2017
  <!-- <td>Life&nbsp; -->
  <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>
    </b>
```

```
<!-- <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
%).   -->
  Classification
   Integer
   92000
   2016
   <!-- <td>Computer&nbsp; -->
  <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>
    </b>
    <!-- <td>The data set of performances of weighted scoring stock portfolios are obtained
with mixture design from the US stock market historical database.   -->
  Multivariate
```

```
Regression
  Real
  315
  12
  2016
  <!-- <td>Business&nbsp; -->
  <a href="datasets/MoCap+Hand+Postures">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/MoCap+Hand+Postures">
    MoCap Hand Postures
    </a>
    </b>
   <!-- <td>5 types of hand postures from 12 users were recorded using unlabeled marker
s attached to fingers of a glove in a motion capture environment. Due to resolution and occlusion, missing value
s are common.  -->
  Multivariate
  Classification, Clustering
```

```
Integer, Real
    78095
    38
    2016
    <!-- <td>Computer&nbsp; -->
   <a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech">datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech</a>
      <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
     </a>
     <a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speec">datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speec</a>
h">
       Early biomarkers of Parkinson so disease based on natural connected speech
      </a>
      </b>
     <!-- <td>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 disease
se.  -->
   Multivariate
    Classification, Regression
    Integer, Real
```

```
130
  65
  2017
  <!-- <td>Life&nbsp; -->
  <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
    </b>
    <!-- <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  -->
  Sequential, Time-Series
  Classification
  Integer, Real
  74
  class="normal">
```

```
102
  2017
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/PM2.5+Data+of+Five+Chinese+Cities">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/PM2.5+Data+of+Five+Chinese+Cities">
    PM2.5 Data of Five Chinese Cities
    </a>
    </b>
   <!-- <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. 
  Multivariate, Time-Series
  Regression
  Integer, Real
  52854
  86
  2017
```

```
<!-- <td>Physical&nbsp; -->
  <a href="datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <b>
    <a href="datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet">
     Parkinson Disease Spiral Drawings Using Digitized Graphics Tablet
    </b>
    <!-- <td>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; -->
  Multivariate
  Classification, Regression, Clustering
  Integer
  77
  2017
  <!-- <td>Computer&nbsp; -->
```

```
<lable>
   <a href="datasets/Sales_Transactions_Dataset_Weekly">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Sales_Transactions_Dataset_Weekly">
     Sales_Transactions_Dataset_Weekly
    </a>
    </b>
    <!-- <td>Contains weekly purchased quantities of 800 over products over 52 weeks. Nor
malised values are provided too.  -->
  Multivariate, Time-Series
  Clustering
  Integer, Real
  811
  53
  2017
  <!-- <td>&nbsp; -->
  <a href="datasets/Las+Vegas+Strip">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
```

```
<b>
    <a href="datasets/Las+Vegas+Strip">
     Las Vegas Strip
    </a>
    </b>
    <!-- <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> -->
  Classification, Regression
  Integer
  504
  20
  2017
  <!-- <td>Business&nbsp; -->
  <a href="datasets/Eco-hotel">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <a href="datasets/Eco-hotel">
     Eco-hotel
    </a>
    </b>
```

```
<!-- <td>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. 
  Text
  401
  2017
  <!-- <td>Business&nbsp; -->
  <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>
    </b>
   <!-- <td>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.
```

```
Multivariate
  Classification
  Integer, Real
  2856
  71
  2016
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Crowdsourced+Mapping">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Crowdsourced+Mapping">
    Crowdsourced Mapping
    </a>
    </b>
   <!-- <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).  
 -->
  Multivariate
  Classification
```

```
10546
  29
  2016
  <!-- <td>Physical&nbsp; -->
  <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>
    </b>
   <!-- <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.  -->
  Multivariate
  Classification, Clustering
  Real
```

```
801
  20531
  2016
  <!-- <td>Life&nbsp; -->
 eter">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
  meter">
   Hybrid Indoor Positioning Dataset from WiFi RSSI, Bluetooth and magnetometer
   </b>
   <!-- <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.   -->
  Multivariate, Sequential, Time-Series
  Classification
  Real
  1540
  </n>
```

```
65
  2016
  <!-- <td>Computer&nbsp; -->
 <a href="datasets/chestnut+%E2%80%93+LARVIC">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <a href="datasets/chestnut+%E2%80%93+LARVIC">
    chestnut - LARVIC
   </a>
   </b>
   <!-- <td>The research project presents this database, shows the images of chestnuts th
at will be processed to determine the presence or absence of defects 
  Classification, Clustering
  1451
  3
  2017
```

```
<!-- <td>Computer&nbsp; -->
         <a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%"><a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%"><a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%"><a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%"><a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%"><a href="datasets/Burst+Burst+Switching+%"><a href="datasets/Burst+Burst+Switching+%"><a href="datasets/Burst+Burst+Switching+%"><a href="datasets/Burst+Burst+Switching+%"><a href="datasets/Burst+Burst+Switching+%"><a href="datasets/Burst+Burst+Switching+%"><a href="datasets/Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burst+Burs
28OBS%29+Network">
                  <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
                 </a>
               <b>
                    <a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+">datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+</a>
%28OBS%29+Network">
                     Burst Header Packet (BHP) flooding attack on Optical Burst Switching (OBS) Network
                  </b>
                 <!-- <td>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; -->
           Text
            Classification
            Integer
            1075
            22
            2017
            <!-- <td>Computer&nbsp; -->
```

```
<a href="datasets/Motion+Capture+Hand+Postures">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Motion+Capture+Hand+Postures">
    Motion Capture Hand Postures
    </a>
    </b>
   <!-- <td>5 types of hand postures from 12 users were recorded using unlabeled marker
s on fingers of a glove in a motion capture environment. Due to resolution and occlusion, missing values are co
mmon.  -->
  Multivariate
  Classification, Clustering
  Real
  78095
  38
  2017
  <!-- <td>Computer&nbsp; -->
```

```
<a nrei= datasets/Anuran+Gais+%28MFGG5%29 >
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
    <a href="datasets/Anuran+Calls+%28MFCCs%29">
     Anuran Calls (MFCCs)
    </a>
    </b>
    <!-- <td>Acoustic features extracted from syllables of anuran (frogs) calls, including the
family, the genus, and the species labels (multilabel).   -->
  Multivariate
  Classification, Clustering
  Real
  7195
  22
  2017
  <!-- <td>Life&nbsp; -->
  <a href="datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
```

```
<a href="datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization">
     TTC-3600: Benchmark dataset for Turkish text categorization
     </b>
    <!-- <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.
  -->
  Text
   Classification, Clustering
   Integer
   3600
   4814
   2017
   <!-- <td>Computer&nbsp; -->
  <a href="datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy">
     <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
    <a href="datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy">
     Gastrointestinal Lesions in Regular Colonoscopy
     </a>
     </b>
```

```
<!-- <td>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.   -->
  Multivariate
  Classification
  Real
  76
  698
  2016
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Daily+Demand+Forecasting+Orders">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <a href="datasets/Daily+Demand+Forecasting+Orders">
     Daily Demand Forecasting Orders
    </a>
    </b>
   <!-- <td>The dataset was collected during 60 days, this is a real database of a brazilian
logistics company.  -->
```

```
Time-Series
  Regression
  Integer
  60
  13
  2017
  <!-- <td>Business&nbsp; -->
  <a href="datasets/Paper+Reviews">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Paper+Reviews">
    Paper Reviews
    </a>
   </b>
   <!-- <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.  -->
  Text
```

```
Classification, Regression
Integer
405
10
2017
<!-- <td>Computer&nbsp; -->
<a href="datasets/extention+of+Z-Alizadeh+sani+dataset">
 <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
 </a>
 <b>
  <a href="datasets/extention+of+Z-Alizadeh+sani+dataset">
  extention of Z-Alizadeh sani dataset
  </a>
 </b>
 <!-- <td>It was collected for CAD diagnosis.&nbsp; -->
Classification
Integer, Real
```

```
303
59
2017
<!-- <td>Life&nbsp; -->
<a href="datasets/Z-Alizadeh+Sani">
 <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
 </a>
 <a href="datasets/Z-Alizadeh+Sani">
  Z-Alizadeh Sani
 </a>
 </b>
 <!-- <td>It was collected for CAD diagnosis.&nbsp; -->
Classification
Integer, Real
303
56
```

```
2017
  <!-- <td>Life&nbsp; -->
  <a href="datasets/Dynamic+Features+of+VirusShare+Executables">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Dynamic+Features+of+VirusShare+Executables">
    Dynamic Features of VirusShare Executables
    </a>
    </b>
   <!-- <td>This dataset contains the dynamic features of 107,888 executables, collected b
y VirusShare from Nov/2010 to Jul/2014.  -->
  Multivariate, Time-Series
  Classification, Regression
  Integer
  107888
  482
  2017
  <!-- <td>Computer&nbsp; -->
```

```
<a href="datasets/IDA2016Challenge">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/IDA2016Challenge">
    IDA2016Challenge
    </a>
   </b>
   <!-- <td>The dataset consists of data collected from heavy Scania trucks in everyday us
age.   -->
  Multivariate
  Classification
  Integer
  76000
  171
  2017
  <!-- <td>Computer&nbsp; -->
 <a href="datasets/DSRC+Vehicle+Communications">
```

```
<iiiig border= 1 SiC= assets/iviLimages/SinaiiLargederauit.jpg />
   <b>
    <a href="datasets/DSRC+Vehicle+Communications">
     DSRC Vehicle Communications
    </b>
    <!-- <td>This set Provides data regarding wireless communications between vehicles an
d road side units. two separate data sets are provided (normal scenario) and in the presence of attacker (jamm
er).  -->
  Sequential, Text
  Clustering
  Real
  10000
  5
  2017
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Mturk+User-Perceived+Clusters+over+Images">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
```

```
<a nref="datasets/inturk+user-perceived+ulusters+over+images">
     Mturk User-Perceived Clusters over Images
    </a>
    </b>
    <!-- <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.  
 -->
  Multivariate, Text
  Clustering
  Integer
  180
  500
  2016
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Character+Font+Images">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <a href="datasets/Character+Font+Images">
     Character Font Images
    </a>
    </b>
```

```
<!-- <td>Character images from scanned and computer generated fonts.&nbsp;
           Multivariate
            Classification
            Integer, Real
            745000
            411
            2016
            <!-- <td>Computer&nbsp; -->
         <a href="datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+Instance+Learning+with+I
e+Labels">
                  <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
                 </a>
                <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>
                 <!-- <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 doc

```
uments.  -->
  Text
  Classification
  Integer
  12234
  8519
  2016
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++">
    Autistic Spectrum Disorder Screening Data for Children
    </a>
    </b>
   <!-- <td>Children screening data for autism suitable for classification and predictive task
s   -->
  Multivariate
```

```
Classification
  Integer
  292
  21
  2017
  <!-- <td>Life&nbsp; -->
  <a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++">
    Autistic Spectrum Disorder Screening Data for Adolescent
    </a>
    </b>
    <!-- <td>Autistic Spectrum Disorder Screening Data for Adolescent. This dataset is relat
ed to classification and predictive tasks.  -->
  Multivariate
  Classification
  Integer
```

```
104
  21
  2017
  <!-- <td>Life&nbsp; -->
  <a href="datasets/APS+Failure+at+Scania+Trucks">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/APS+Failure+at+Scania+Trucks">
    APS Failure at Scania Trucks
    </a>
    </b>
   <!-- <td>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.  -->
  >
  Multivariate
  Classification
  Integer, Real
  60000
  </n>
```

```
171
  2017
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Wireless+Indoor+Localization">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <a href="datasets/Wireless+Indoor+Localization">
    Wireless Indoor Localization
    </a>
    </b>
   <!-- <td>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.  
  Multivariate
  Classification
  Real
  2000
  ~td~
```

```
2017
  <!-- <td>Computer&nbsp; -->
 <a href="datasets/HCC+Survival">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/HCC+Survival">
    HCC Survival
    </a>
   </b>
   <!-- <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. 
  Multivariate
  Classification
  Integer, Real
  165
  49
  2017
  <!-- <td>Life&nbsp; -->
```

```
<a href="datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </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>
    </b>
    <!-- <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; -->
  Multivariate
  Classification, Regression
  Integer
  217
  12
  2017
  <!-- <td>Computer&nbsp; -->
```

Suz

```
<a nrei= datasets/University+01+Tenran+Question+Dataset+2016+%2801QD.2016%29 >
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
    <a href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29">
     University of Tehran Question Dataset 2016 (UTQD.2016)
    </b>
    <!-- <td>Persian questions gathered from a jeopardy game broadcasted on Iranian nati
onal television.   -->
  Text
  Classification
  1175
  3
  2017
  <!-- <td>Other&nbsp; -->
  <a href="datasets/Autism+Screening+Adult">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <a href="datasets/Autism+Screening+Adult">
```

```
Autism Screening Adult
                                     </a>
                                  </b>
                                <!-- <td>Autistic Spectrum Disorder Screening Data for Adult. This dataset is related to
classification and predictive tasks.  -->
                    Classification
                      Integer
                      704
                      21
                      2017
                      <!-- <td>Social&nbsp; -->
                 <a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sens">- a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+using+a+batteryless+wearable+using+a+batteryless+wearable+using+a+batteryless+wearable+using+a+batteryless+wearable+using+a+batteryless+wearable+using+a+batteryless+wearable+using+a+batteryless+wearable+using+a+batteryless+wearable+using+a+batteryless+wearable+using+a+batteryless+wearable+using+a+batteryless+wearable+using+a+batteryle
or">
                                  <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
                               </a>
                             <a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sen">- a href="datasets/Activity+a+batteryless+wearable+sen">- a href="datasets/Activity+a+b
sor">
                                        Activity recognition with healthy older people using a batteryless wearable sensor
                                    </a>
                                  </b>
```

```
<!-- <td>Sequential motion data from 14 healthy older people aged 66 to 86 years old u
sing a batteryless, wearable sensor on top of their clothing for the recognition of activities in clinical environmen
ts.  -->
  Sequential
  Classification
  Real
  75128
  9
  2016
  <!-- <td>Life&nbsp; -->
  <a href="datasets/Immunotherapy+Dataset">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <a href="datasets/Immunotherapy+Dataset">
     Immunotherapy Dataset
    </a>
    </b>
    <!-- <td>This dataset contains information about wart treatment results of 90 patients us
ing immunotherapy.  -->
```

```
Univariate
  Classification
  Integer, Real
  90
  8
  2018
  <!-- <td>Life&nbsp; -->
 <a href="datasets/Cryotherapy+Dataset+">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Cryotherapy+Dataset+">
    Cryotherapy Dataset
   </a>
   </b>
   <!-- <td>This dataset contains information about wart treatment results of 90 patients us
ing cryotherapy.  -->
  Univariate
  Classification
```

```
Integer, Real
  90
  2018
  <!-- <td>Life&nbsp; -->
  <a href="datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes">
    OCT data & DCT data & DCT land & Samp; Right Eyes
    </a>
    </b>
   <!-- <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.  
  Multivariate
  Classification
  Real
```

```
50
  2
  2016
  <!-- <td>Computer&nbsp; -->
 <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>
   </b>
   <!-- <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.
  -->
  Multivariate
  Classification
  71
```

11

```
2018
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/News+Popularity+in+Multiple+Social+Media+Platforms">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/News+Popularity+in+Multiple+Social+Media+Platforms">
     News Popularity in Multiple Social Media Platforms
    </a>
    </b>
   <!-- <td>Large data set of news items and their respective social feedback on multiple p
latforms: Facebook, Google+ and LinkedIn.  -->
  Multivariate, Time-Series, Text
  Regression
  Integer, Real
  93239
  11
  2018
  -/td.
```

```
<!-- <td>Computer&nbsp; -->
<a href="datasets/Ultrasonic+flowmeter+diagnostics">
 <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
 </a>
 <b>
  <a href="datasets/Ultrasonic+flowmeter+diagnostics">
  Ultrasonic flowmeter diagnostics
  </a>
 </b>
 <!-- <td>Fault diagnosis of four liquid ultrasonic flowmeters&nbsp; -->
Multivariate
Classification
Real
540
173
2018
<!-- <td>Computer&nbsp; -->
```

f IIdata ata /ICNALA - 0014 - Assaultad - Danara - Data - Catil

```
<a nite = ualasels/10/vilA+2014+Accepteu+Fapers+Data+set >
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <a href="datasets/ICMLA+2014+Accepted+Papers+Data+Set">
     ICMLA 2014 Accepted Papers Data Set
    </b>
    <!-- <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.
  -->
  Multivariate
   Classification, Clustering
   105
   5
   2018
   <!-- <td>Other&nbsp; -->
  <a href="datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
```

```
<a nrei="datasets/ble+R55i+Dataset+tor+Indoor+localization+and+Navigation">
     BLE RSSI Dataset for Indoor localization and Navigation
     </a>
    </b>
    <!-- <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 purpo
ses.  -->
  Multivariate, Sequential, Time-Series
   Classification, Clustering
   Integer
   6611
   15
   2018
   <!-- <td>Computer&nbsp; -->
  <a href="datasets/Container+Crane+Controller+Data+Set">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
    <a href="datasets/Container+Crane+Controller+Data+Set">
     Container Crane Controller Data Set
     </a>
    </b>
```

```
<!-- <td>A container crane has the function of transporting containers from one point to
another point.  -->
  Univariate, Domain-Theory
  Classification, Regression
  Real
  15
  3
  2018
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Residential+Building+Data+Set">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <a href="datasets/Residential+Building+Data+Set">
     Residential Building Data Set
    </a>
    </b>
    <!-- <td>Data set includes construction cost, sale prices, project variables, and economi
c variables corresponding to real estate single-family residential apartments in Tehran, Iran.
```

```
Multivariate
  Regression
  Real
  372
  105
  2018
  <!-- <td>Computer&nbsp; -->
 <a href="datasets/Health+News+in+Twitter">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Health+News+in+Twitter">
    Health News in Twitter
    </a>
   </b>
   <!-- <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.   -->
  Text
  Clustering
```

```
Real
  58000
  25000
  2018
  <!-- <td>Computer&nbsp; -->
 <a href="datasets/chipseq">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/chipseq">
    chipseq
    </a>
   </b>
   <!-- <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.  -->
  Sequential
  Classification
  Integer
```

```
4960
  2018
  <!-- <td>Life&nbsp; -->
 <a href="datasets/SGEMM+GPU+kernel+performance">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/SGEMM+GPU+kernel+performance">
    SGEMM GPU kernel performance
    </a>
   </b>
   <!-- <td>Running times for multiplying two 2048 x 2048 matrices using a GPU OpenCL
SGEMM kernel with varying parameters (using the library 'CLTune').  -->
  Multivariate
  Regression
  Integer
  241600
```

```
18
  2018
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Repeat+Consumption+Matrices">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Repeat+Consumption+Matrices">
    Repeat Consumption Matrices
    </a>
    </b>
   <!-- <td>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; -->
  Multivariate
  Clustering
  Real
  130000
  21000
```

```
<!-- <td>Computer&nbsp; -->
  <a href="datasets/detection_of_loT_botnet_attacks_N_BaloT">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
    <a href="datasets/detection_of_loT_botnet_attacks_N_BaloT">
     detection_of_loT_botnet_attacks_N_BaloT
    </a>
    </b>
    <!-- <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 BAS
HLITE.  -->
  Multivariate, Sequential
  Classification, Clustering
  Real
  7062606
  115
  2018
  <!-- <td>Computer&nbsp; -->
```

```
<a href="datasets/Absenteeism+at+work">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <a href="datasets/Absenteeism+at+work">
     Absenteeism at work
    </a>
    </b>
    <!-- <td>The database was created with records of absenteeism at work from July 2007
to July 2010 at a courier company in Brazil.  -->
  Multivariate, Time-Series
  Classification, Clustering
  Integer, Real
  740
  21
  2018
  <!-- <td>Business&nbsp; -->
  <a href="datasets/SCADI">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
```

```
</lu>
 <a href="datasets/SCADI">
  SCADI
  </a>
  </b>
 <!-- <td>First self-care activities dataset based on ICF-CY.&nbsp; -->
Multivariate
Classification, Clustering
70
206
2018
<!-- <td>Life&nbsp; -->
<a href="datasets/Condition+monitoring+of+hydraulic+systems">
  <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
 </a>
 <a href="datasets/Condition+monitoring+of+hydraulic+systems">
  Condition monitoring of hydraulic systems
  </a>
  </b>
```

```
<!-- <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 quantif
ication.  -->
  Multivariate, Time-Series
  Classification, Regression
  Real
  2205
  43680
  2018
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Carbon+Nanotubes">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
    <a href="datasets/Carbon+Nanotubes">
     Carbon Nanotubes
    </a>
    </b>
    <!-- <td>This dataset contains 10721 initial and calculated atomic coordinates of carbon
nanotubes.  -->
```

```
Univariate
  Regression
  Real
  10721
  8
  2018
  <!-- <td>Computer&nbsp; -->
  <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>
   </b>
   <!-- <td>This dataset contains 640 performance measurements from a simulation of 2-
Dimensional Multiprocessor Optical Interconnection Network.  
  Multivariate
```

```
Classification, Regression
  Integer, Real
  640
  10
  2018
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Sports+articles+for+objectivity+analysis">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
    <a href="datasets/Sports+articles+for+objectivity+analysis">
     Sports articles for objectivity analysis
    </a>
    </b>
    <!-- <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 provid
ed.  -->
  Multivariate, Text
  Classification
  Integer
```

```
1000
  59
  2018
  <!-- <td>Social&nbsp; -->
 <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>
   </b>
   <!-- <td>Clinical features were observed or measured for 64 patients with breast cancer
and 52 healthy controls.   -->
  Multivariate
  Classification
  Integer
  116
```

```
10
  2018
  <!-- <td>Life&nbsp; -->
  >
    <a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data">
     GNFUV Unmanned Surface Vehicles Sensor Data
    </a>
    </b>
   <!-- <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).   -->
  Multivariate, Time-Series
  Regression
  Real
  1672
  5
  <ht>td>
```

```
2018
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Dishonest+Internet+users+Dataset">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <a href="datasets/Dishonest+Internet+users+Dataset">
    Dishonest Internet users Dataset
    </b>
   <!-- <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. 
> -->
  Multivariate
  Classification, Clustering
  >
  322
  5
  2018
  <!-- <td>Computer&nbsp; -->
```

```
<a href="datasets/Victorian+Era+Authorship+Attribution">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
    <b>
     <a href="datasets/Victorian+Era+Authorship+Attribution">
     Victorian Era Authorship Attribution
     </a>
    </b>
    <!-- <td>To create the largest authorship attribution dataset, we extracted works of 50 w
ell-known authors. To have a non-exhaustive learning, in training there are 45 authors whereas, in the testing, i
t's 50  -->
  Text
   Classification
   93600
   1000
   2018
   <!-- <td>Computer&nbsp; -->
  <a href="datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
```

```
<b>
     <a href="datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set">
     Simulated Falls and Daily Living Activities Data Set
    </b>
    <!-- <td>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.  -->
  Time-Series
   Classification
   Integer
   3060
   138
   2018
   <!-- <td>Life&nbsp; -->
  <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>
    </b>
    <!-- <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.  -->
  Multivariate, Text
   Classification
   Integer
   5879
   2018
   <!-- <td>Social&nbsp; -->
  >
   <a href="datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
    <b>
     <a href="datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals">
     EEG Steady-State Visual Evoked Potential Signals
     </a>
    </b>
```

```
<!-- <td>This database consists on 30 subjects performing Brain Computer Interface for
Steady State Visual Evoked Potentials (BCI-SSVEP).   -->
  Multivariate, Time-Series
  Classification, Regression
  Integer
  9200
  16
  2018
  <!-- <td>Life&nbsp; -->
  <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>
    </b>
    <!-- <td>Roman Urdu (the scripting style for Urdu language) is one of the limited resour
ce languages. A data corpus comprising of more than 20000 records was collected.  
  Text
```

```
Classification
  20000
  2
  2018
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Avila">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Avila">
    Avila
    </a>
    </b>
   <!-- <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.&
nbsp; -->
  >
  Multivariate
  Classification
```

```
Real
  20867
  10
  2018
  <!-- <td>Computer&nbsp; -->
 <a href="datasets/PANDOR">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
   <a href="datasets/PANDOR">
    PANDOR
   </a>
   </b>
   <!-- <td>PANDOR is a novel and publicly available dataset for online recommendation p
rovided by Purch (http://www.purch.com/).   -->
  Multivariate
  Recommendation
  Categorical
```

```
>
  2018
  <!-- <td>Life&nbsp; -->
  <a href="datasets/Drug+Review+Dataset+%28Druglib.com%29">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Drug+Review+Dataset+%28Druglib.com%29">
    Drug Review Dataset (Druglib.com)
    </a>
    </b>
   <!-- <td>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.  -->
  Multivariate, Text
  Classification, Regression, Clustering
  Integer
  4143
  8
  >
```

```
2018
  <!-- <td>&nbsp; -->
  <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)
    </b>
   <!-- <td>The dataset provides patient reviews on specific drugs along with related condi
tions and a 10 star patient rating reflecting overall patient satisfaction. 
  Multivariate, Text
  Classification, Regression, Clustering
  Integer
  215063
  6
  2018
  <!-- <td>Life&nbsp; -->
```

```
<a href="datasets/Physical+Unclonable+Functions">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
     <a href="datasets/Physical+Unclonable+Functions">
     Physical Unclonable Functions
     </a>
    </b>
    <!-- <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.  -->
  Multivariate
   Classification
   Integer
   6000000
   129
   2018
   <!-- <td>Computer&nbsp; -->
  <a href="datasets/Superconductivty+Data">
      a barder "1" are "acceta/MI images/Cmall argodefault ing"/s
```

```
<iiiig bolder= 1 Sic= assets/iviLiniages/SinaliLargedeladit.jpg />
   <a href="datasets/Superconductivty+Data">
    Superconductivty Data
    </a>
    </b>
   <!-- <td>Two file s contain data on 21263 superconductors and their relevant features.&
nbsp; -->
  Multivariate
  Regression
  Real
  21263
  81
  2018
  <!-- <td>Physical&nbsp; -->
  <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>
    </b>
    <!-- <td>WESAD (Wearable Stress and Affect Detection) contains data of 15 subjects d
uring a stress-affect lab study, while wearing physiological and motion sensors.  -->
  Multivariate, Time-Series
  >
  Classification, Regression
  Real
  63000000
  12
  >
  2018
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
    <a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2">
     GNFUV Unmanned Surface Vehicles Sensor Data Set 2
    </a>
    </b>
```

```
<!-- <td>The data-set contains eight (2x4) data-sets of mobile sensor readings data (hu
midity, temperature) corresponding to a swarm of four Unmanned Surface Vehicles (USVs) in a test-bed, Athen
s, Greece.  -->
  Multivariate, Sequential, Time-Series
  Regression
  Real
  10190
  6
  2018
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Student+Academics+Performance">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <a href="datasets/Student+Academics+Performance">
     Student Academics Performance
    </a>
    </b>
    <!-- <td>The dataset tried to find the end semester percentage prediction based on diff
erent social, economic and academic attributes.   -->
  Multivariate
```

```
>
  Classification
  300
  22
  2018
  <!-- <td>Computer&nbsp; -->
  <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>
    </b>
   <!-- <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.  -->
  Multivariate
  Classification, Clustering
```

```
Integer, Real
  12330
  18
  2018
  <!-- <td>Business&nbsp; -->
 <a href="datasets/PMU-UD">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/PMU-UD">
    PMU-UD
   </a>
   </b>
   <!-- <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).   -->
  Univariate
  Classification
```

```
5180
  2018
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Parkinson%27s+Disease+Classification">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Parkinson%27s+Disease+Classification">
    Parkinson's Disease Classification
    </a>
    </b>
   <!-- <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). 
  Multivariate
  Classification
  Integer, Real
  756
```

754

```
2018
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/Electrical+Grid+Stability+Simulated+Data+">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>>
    <a href="datasets/Electrical+Grid+Stability+Simulated+Data+">
    Electrical Grid Stability Simulated Data
    </a>
    </b>
   <!-- <td>The local stability analysis of the 4-node star system (electricity producer is in t
he center) implementing Decentral Smart Grid Control concept.   -->
  Multivariate
  Classification, Regression
  Real
  10000
  14
  2018
```

```
<!-- <td>Physical&nbsp; -->
  <a href="datasets/Caesarian+Section+Classification+Dataset">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Caesarian+Section+Classification+Dataset">
    Caesarian Section Classification Dataset
    </a>
    </b>
   <!-- <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. 
  Univariate
  Classification
  Integer
  80
  5
  2018
  <!-- <td>Life&nbsp; -->
  -td
```

```
<a href="datasets/BAUM-1">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <b>
    <a href="datasets/BAUM-1">
     BAUM-1
    </a>
    </b>
    <!-- <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.  -->
  Time-Series
  Classification
  1184
  2018
  <!-- <td>Computer&nbsp; -->
  <a href="datasets/BAUM-2">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
        4 "Jalaka a aka /D A L IM O"
```

```
<a file = ualasels/dAUW-2 >
    BAUM-2
   </a>
   </b>
   <!-- <td>A multilingual audio-visual affective face database consisting of 1047 video clip
s of 286 subjects.   -->
  Time-Series
  Classification
  1047
  2018
  <!-- <td>Computer&nbsp; -->
 <a href="datasets/Audit+Data">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Audit+Data">
    Audit Data
   </a>
   </b>
```

```
<!-- <ta>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. 
  Multivariate
  Classification
  Real
  777
  18
  2018
  <!-- <td>Other&nbsp; -->
  <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>
    </b>
    <!-- <td>User interest information extracted from user reviews published in holidayiq.co
m about various types of point of interests in South India  -->
  Multivariate, Text
```

```
Classification, Clustering
  Real
  249
  2018
  <!-- <td>Other&nbsp; -->
  <a href="datasets/Real+estate+valuation+data+set">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/Real+estate+valuation+data+set">
    Real estate valuation data set
    </a>
    </b>
   <!-- <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.   -->
  Multivariate
  Regression
```

```
Integer, Real
            414
            7
            2018
            <!-- <td>Business&nbsp; -->
          d+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+connect">eca href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*-on+natural+connect*
ed+speech+Data+Set+">
                      Early biomarkers of Parkinson's disease based on natural connected speech Data Set
                    </a>
                   </b>
                 <!-- <td>.&nbsp; -->
           Multivariate
            Classification
            Real
```

```
2018
  <!-- <td>Life&nbsp; -->
 <a href="datasets/Somerville+Happiness+Survey">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
   <a href="datasets/Somerville+Happiness+Survey">
    Somerville Happiness Survey
   </a>
   </b>
   <!-- <td>A data extract of a non-federal dataset posted here https://catalog.data.gov/dat
aset/somerville-happiness-survey-responses-2011-2013-2015 
  Classification
  Integer
  143
  2018
```

```
<!-- <td>Life&nbsp; -->
 <a href="datasets/2.4+GHZ+Indoor+Channel+Measurements">
   <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <b>
    <a href="datasets/2.4+GHZ+Indoor+Channel+Measurements">
    2.4 GHZ Indoor Channel Measurements
    </a>
   </b>
   <!-- <td>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. 
  Multivariate
  Classification
  Real
  7840
  5
  2018
  <!-- <td>Computer&nbsp; -->
```

```
<a href="datasets/EMG+data+for+gestures">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   <b>
    <a href="datasets/EMG+data+for+gestures">
    EMG data for gestures
    </a>
    </b>
   <!-- <td>These are files of raw EMG data recorded by MYO Thalmic bracelet&nbsp;</p
> -->
  Time-Series
  Classification
  Real
  30000
  6
  2019
  <!-- <td>Life&nbsp; -->
  <a href="datasets/Parking+Birmingham">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   ∠td<
```

```
<a href="datasets/Parking+Birmingham">
     Parking Birmingham
     </a>
    </b>
    <!-- <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 
  Multivariate, Univariate, Sequential, Time-Series
   Classification, Regression, Clustering
   Real
   35717
   4
   2019
   <!-- <td>Computer&nbsp; -->
  <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>
    <b>
     <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>The database was created with records of behavior of the urban traffic of the ci
ty of Sao Paulo in Brazil.  -->
  Multivariate, Time-Series
  Classification, Regression
  Integer, Real
  135
  18
  2018
  <!-- <td>Computer&nbsp; -->
  >
   <a href="datasets/Travel+Reviews">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
   </a>
   <a href="datasets/Travel+Reviews">
    Travel Reviews
    </a>
    </b>
   <!-- <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.  -->
  Multivariate, Text
  Classification, Clustering
  Real
  980
  11
   2018
  <!-- <td>Other&nbsp; -->
  <a href="datasets/Tarvel+Review+Ratings">
    <img border="1" src="assets/MLimages/SmallLargedefault.jpg"/>
    </a>
   <b>
    <a href="datasets/Tarvel+Review+Ratings">
     Tarvel Review Ratings
    </a>
    </b>
    <!-- <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. 
  Multivariate, Text
```

```
Classification, Clustering
  Real
  5456
  25
  2018
  <!-- <td>Other&nbsp; -->
  <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>
    </b>
   <!-- <td>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.   -->
  Multivariate
  Classification
```

```
Integer
 120
 2019
 <!-- <td>Computer&nbsp; -->
 Supported By:
<img height="60" src="assets/nsfe.gif"/>
In Collaboration With:
<img src="assets/rexaSmall.jpg"/>
<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]:

[,

Center for Machine Learning and Intelligent Systems,

About,

Citation Policy,

Donate a Data Set,

Contact,

,

View ALL Data Sets,

Classification,

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

Clustering,

Other,

Categorical,

Numerical,

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

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

Univariate,

Sequential,

Time-Series,

Text,

Domain-Theory,

Other,

Life Sciences,

Physical Sciences,

<a href="datasets.php?format=&task=&att=&area=comp&numAtt=&numIns=&typ
e=&sort=nameUp&view=table">CS / Engineering,

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

 Rusinoss /a>

- Game,
- Other,
- Less than 10,
- <a href="datasets.php?format=&task=&att=&area=&numAtt=10to100&numIns=&
 type=&sort=nameUp&view=table">10 to 100,
- Greater than 100,
- Less than 100,
- 100 to 1000,
- Greater than 1000,
- Matrix,
- <a href="datasets.php?format=nonmat&task=&att=&area=&numAtt=&numIns=&t
 ype=&sort=nameUp&view=table">Non-Matrix,
- List View,
- Name,
- Data Types,
- Default Task,
- Attribute Types,
- # Instances/a>,
- # Attributes,
- Year,
- ,
- Abalone,
- ,
- Adult,
- ,
- Annealing,
- ,
- Anonymous Microsoft Web Data,
- ,
- Arrhythmia,
- ,
- Artificial Characters,
-
- Audiology (Original),
- ,
- Audiology (Standardized),
- ,
- Auto MPG,
- ,
- Automobile,
- ,
- Badges,
 - hyd Hatasta Dauges / Dalays Colleges August Hillars Handta (Milinary) (Crolling H. / /

- <iiiig border= 1 sic= assets/iviLimages/smailLarge12.jpg />,
- Balance Scale,
- ,
- Balloons,
- ,
- Breast Cancer,
- ,
- Breast Cancer Wisconsin (Original),
- ,
- $<\! a\ href="datasets/Breast+Cancer+Wisconsin+\%28Prognostic\%29">\! Breast\ Cancer\ Wisconsin\ (Prognostic)<\! /a>,$
- ,
- Breast Cancer Wisconsin (Diagnostic),
- ,
- Pittsburgh Bridges,
- ,
- Car Evaluation,
- ,
- Census Income,
- ,
- Chess (King-Rook vs. King-Knight),
- ,
- Chess (King-Rook vs. King-Pawn),
- ,
- Chess (King-Rook vs. King),
- ,
- Chess (Domain Theories),
- ,
- Bach Chorales,
- ,
- Connect-4,
- ,
- Credit Approval,
- .
- Japanese Credit Screening,
- ,
- Computer Hardware,
- ,
- Contraceptive Method Choice,
- ,
- Covertype,
- ,
- Cylinder Bands,
- ,
- Dermatology,
- ,
- Diabetes,
- ,
- DGP2 The Second Data Generation P rogram,
- .
- Document Understanding,

```
<a nref="datasets/EBL+Domain+Theories"><img border="1" src="assets/MLimages/5mailLargedefault.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/lonosphere"><img border="1" src="assets/MLimages/SmallLarge52.jpg"/></a>,
<a href="datasets/lonosphere">lonosphere</a>,
<a href="datasets/Iris"><img border="1" src="assets/MLimages/SmallLarge53.jpg"/></a>,
<a href="datasets/lris">lris</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>,
```

- Mobile Robots,
- ,
- Molecular Biology (Promoter Gene Sequences),
- ,
- Molecular Biology (Protein Secondary Structure),
- ,
- Molecular Biology (Splice-junction Gene Sequences),
- ,
- MONK's Problems,
- ,
- Moral Reasoner,
- ,
- Multiple Features,
- ,
- Mushroom,
- .
- Musk (Version 1),
- ,
- Musk (Version 2),
- ,
- Nursery,
- , /a>,
- Othello Domain Theory,
- ,
- Page Blocks Classification,
- ,
- Optical Recognition of Handwritten Digits,
- <img border="1" src="assets/MLimages/S
 mallLargedefault.jpg"/>,
- Pen-Based Recognition of Handwritten Digits,
- , a>,
- Post-Operative Patient,
- ,
- Primary Tumor,
- ,
- Prodigy,
- ,
- Qualitative Structure Activity Relationships
- ,
- Quadruped Mammals,
- ,
- Servo,
- ,
- Shuttle Landing Control,
- ,
- Solar Flare,

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

- ,
- Soybean (Small),
- ,
- 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 (Version 1),
- ,
- Waveform Database Generator (Version 2),
- ,
- Wine,
- ,
- Yeast,
- ,
- Zoo,
- ,
- Undocumented,
-
- Twenty Newsgroups,
- ,
- Australian Sign Language signs,

```
<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),
- ,
- US Census Data (1990),
- ,
- Census-Income (KDD),
- <img border="1" src="assets/MLimages/SmallLarge118.jpg"/
- Coil 1999 Competition Data,
- ,
- Corel Image Features,
- ,
- E. Coli Genes,
- ,
- EEG Database,
- ,
- El Nino,
- ge123.jpg"/>,
- Entree Chicago Recommendation Data,
- ,
- CMU Face Images,
- ,
- Insurance Company Benchmark (C OIL 2000).
- ,
- Internet Usage Data,
-
- IPUMS Census Database,
- ,
- Japanese Vowels,
- </a
- KDD Cup 1998 Data,
- </a
- KDD Cup 1999 Data,
- </a
- M. Tuberculosis Genes,
- ,
- Movie,
- <a href="datasets/MLimages/SmallLarge<a href="datasets/MLimages/SmallLarge<a href="datasets/MLimages/SmallLarge,
- MSNBC.com Anonymous Web Data,
- ,
- NSF Research Award Abstracts 1990-2003< /a>,
- ,
- Pioneer-1 Mobile Robot Data,
- ,
- Pseudo Periodic Synthetic Time Series,

- ,
- Reuters-21578 Text Categorization Collection,
- , a>,
- Robot Execution Failures,
- ,
- Synthetic Control Chart Time Series,
- ,
- Syskill and Webert Web Page Ratings,
- ,
- UNIX User Data,
- ,
- Volcanoes on Venus JARtool experiment, a>,
- ,
- Statlog (Australian Credit Approval),
- ,
- Statlog (German Credit Data),
- ,
- Statlog (Heart),
- ,
- Statlog (Landsat Satellite),
- ,
- Statlog (Image Segmentation),
- ,
- Statlog (Shuttle),
- ,
- Statlog (Vehicle Silhouettes),
- ,
- Connectionist Bench (Nettalk Corpus), ,
- Connectionist Bench (Sonar, Mines vs. Rocks),
- ,
- Connectionist Bench (Vowel Recognition Deterding Data),
- ,
- Economic Sanctions,
- ,
- Protein Data,
- ,
- Cloud,
- ,
- Callt2 Building People Counts,
- ,
- Dodgers Loop Sensor,
- ,
- Poker Hand

```
<a href="datasets/MAGIC+Gamma+Telescope"><img border="1" src="assets/MLimages/SmallLarge159.jpg"/>
<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.jp
g''/></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.i
pg"/></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/SmallLargedefa
ult.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/SmallLarge17" src="assets/MLimages/SmallArge17" src="assets/MLimages/SmallArge17" src="assets/MLimages/SmallArge17" src="assets/MLimages/SmallArge17" src="assets/MLimages/SmallArge17" src="assets/MLimages/Sm
6.\text{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/SmallLar"
ge160.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>,
```

 Acuta_Inflammations

```
<a href="datasets/Wine+Quality"><img border="1" src="assets/MLimages/SmallLarge186.jpg"/></a>,
```

- Wine Quality,
- ,
- URL Reputation,
- ,
- p53 Mutants,
- , /a>,
- Parkinsons Telemonitoring,
- ,
- Demospongiae,
- ,
- Opinosis Opinion/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,
- ,
 - sa brof-"datacots/EMG, Physical, Action, Data, Sot", EMG Physical Action Data Sot /a

- ,
- Vicon Physical Action Data Set,
- ,
- Amazon Commerce reviews set,
- .
- Amazon Access Samples,
- ,
- Reuter_50_50,
- ,
- Farm Ads,
- ,
- DBWorld e-mails,
- ,
- KEGG Metabolic Relation Network (Directed),
- ,
- KEGG Metabolic Reaction Network (Undirected),
- ,
- Bank Marketing,
- ,
- YouTube Comedy Slam Preference Data,
- ,
- Gas Sensor Array Drift Dataset,
- ,
- ILPD (Indian Liver Patient Dataset),
- ,
- OPPORTUNITY Activity Recognition,
- ,
- Nomao,
- , a>,
- SMS Spam Collection,
- ,
- Skin Segmentation,
- ,
- Planning Relax,
- ,
- PAMAP2 Physical Activity Monitoring,
- ,
- Restaurant & amp; consumer data,
- ,
- CNAE-9,
- ,
- Individual household electric power consumption,
- ,
- seeds,
- ,
- a law of II alasa a asa /N lawshi wii N lawshi w /a

```
<a href="datasets/Northix >Northix</a>,
<a href="datasets/QtyT40I10D100K"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
```

- 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 (North Jutland, Denmark),
- ,
- ISTANBUL STOCK EXCHANGE,
- ,
- Buzz in social media ,
- ,
- First-order theorem proving,
- ,
- Wearable Computing: Classification of Body Postures and Movements (PUC-Rio),
- ,
- Gas sensor arrays in open sampling settings,
- ,
- Climate Model Simulation Crashes,
- ,
- MicroMass,
- ,
- QSAR biodegradation,
- ,
- BLOGGER,
- ,
- Daily and Sports Activities,
- ,
- User Knowledge Modeling,
- $<\!\!a\ href="datasets/Reuters+RCV1+RCV2+Multilingual\%2C+Multiview+Text+Categorization+Test+collection"><\!\!i\ a\ href="datasets/Reuters+RCV1+RCV2+Multilingual\%2C+Multiview+Text+Categorization+Test+collection"><<\!\!i\ a\ href="datasets/Reuters+RCV1+RCV2+Multilingual\%2C+Multiview+Text+Categorization+Test+collection"><< \ href="datasets/Reuters+$

- mg border="1" src="assets/MLImages/SmallLargedefault.jpg"/>,
 Re uters RCV1 RCV2 Multilingual, Multiview Text Categorization Test collection,
- ,
- NYSK,
- .
- Turkiye Student Evaluation,
- ,
- ser Knowledge Modeling Data (Students' Knowledge Levels on DC Electrical Machines),
- ,
- EEG Eye State,
- ,
- Physicochemical Properties of Protein Tertiary Structure,
- ,
- seismic-bumps,
- ,
- banknote authentication,
- -/a>,
- USPTO Algorithm Challenge, run by NASA-Harvard Tournament Lab and TopCoder Problem: Pat.
- ,
- YouTube Multiview Video Games Dataset
- <img border="1" src="asset
 s/MLimages/SmallLargedefault.jpg"/>,
- Gas Sensor Array Drift Dataset at Different Concentrations
- <img border
 ="1" src="assets/MLimages/SmallLargedefault.jpg"/>,
- Activities of Daily Living (ADLs) 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,

```
<a href="datasets/SUSY"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
```

- SUSY,
- ,
- HIGGS,
- .
- Qualitative Bankruptcy,
- ,
- LSVT Voice Rehabilitation,
- ,
- Dataset for ADL Recognition on 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, alysis,
- ,
- Combined Cycle Power Plant,
- ,
- Urban Land Cover,
- ,
- Diabetes 130-US hospitals for years 19 99-2008,
- , a>,
- Bach Choral Harmony,
- ,
- StoneFlakes,
- ,
- Tennis Major Tournament Match Statistics
- ,
- Parkinson Spee

- ch Dataset with Multiple Types of Sound Recordings, <img border="1" src="assets/MLimages/SmallLargedefault.j"
- Gesture Phase Segmentation,
- ,
- Perfume Data,

pg"/>,

- ,
- BlogFeedback,
- ,
- REALDISP Activity Recognition Dataset,
- ,
- Newspaper and magazine images segmentation dataset,
- ,
- AAAI 2014 Accepted Papers,
- ,
- Gas sensor array under flow modulation,
- <img border="1" src="assets/MLi
 mages/SmallLargedefault.jpg"/>,
- Gas sensor array exposed to turb ulent gas mixtures,
- ,
- UJIIndoorLoc,
- , a>,
- Sentence Classification,
- ,
- Dow Jones Index,
- ,
- sEMG for Basic Hand movements,
- ,
- AAAI 2013 Accepted Papers,
- ,
- 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,

```
<a href="datasets/microblogPCU"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>,
```

- microblogPCU,
- ,
- Firm-Teacher_Clave-Direction_Classification .
- ,
- Dataset for Sensorless Drive Diagnosis,
- ,
- TV News Channel Commercial Detect ion 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 L earning based ZZAlpha Ltd. Stock Recommendations 2012-2014,
- ,
- Folio,
- ,
- Taxi Service Trajectory Prediction Challenge, ECML PKDD 2015,
- ,
- Cuff-Less Blood Pressure Estimation,
- ,
- Smartphone-Based Recognition of Human Activities and Postural Transitions,
- ,
- Mice Protein Expression.

```
<a href="datasets/UJIIndoorLoc-Mag"><img border="1" src="assets/MLimages/SmallLargedefault.jpg"/></a>, <a href="datasets/UJIIndoorLoc-Mag">UJIIndoorLoc-Mag</a>,
```

- ,
- Heterogeneity Activity Recognition,
- ,
- Educational Process Mining (EPM): A Learning Analytics Data Set,
- ,
- HEPMASS,
- ,
- Indoor User Movement Prediction from RSS data,
- ,
- Open University Learning Analytics dataset
- ,
- default of credit card clients,
- ,
- Mesothelioma's disease data set ,
- ,
- Online Retail,
- ,
- SIFT10M,
- ,
- GPS Trajectories,
- ,
- Detect Malacious Executable(AntiVirus),
- .
- Occupancy Detection ,
- ,
- Improved Spiral Test Using Digitized Graphics Tablet for Monitoring Parkinson's 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 (HAR) in Ambient Assisted Living (AAL),
- <imq border="1" src="assets/MI images/SmallI argede</p>

- fault.jpg"/>,
- Polish companies bankruptcy data,
- ,
- Activity Recognition system based on Multisensor data fusion (AReM),
- , a>,
 - Dota2 Games Results,
 - ,
 - Facebook metrics,
- ,
- UbiqLog (smartphone lifelogging),
- ,
- NIPS Conference Papers 1987-2015,
- ,
- HTRU2,
- ,
- Drug consumption (quantified),
- ,
- Appliances energy prediction,
- , a>,
- 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 (Risk Factors),
- ,
- Quality Assessment of Digital Colposcopies,
- ,
- KASANDR,
- ,
- FMA: A Dataset For Music Analysis,
- ,
- ~a hraf-"datacate/Air (quality"\Air quality//a>

- ,
- Epileptic Seizure Recognition,
- ,
- Devanagari Handwritten Character Dataset
- ,
- Stock portfolio performance,
- ,
- MoCap Hand Postures,
- <im g border="1" src="assets/MLimages/SmallLargedefault.jpg"/>,
- Early biomarkers of Parkinson so disease based on natural connected speech,
- ,
- Data for S oftware 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, Bluetooth and magnetometer,
- ,
- chestnut LARVIC,
- ,
- Burst Header Packet (BHP) flooding attack on Optical Burst Switching (OBS) Network,
- ,
- Motion Capture Hand Postures,
- ,
- Anuran Calls (MFCCs),
- <img border="1" src=""" sr

- -a href="datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization">TTC-3600: 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: A Data Set for Multi-Label Multi-Instance Learning with Instance Labels,
- ,
- Autistic Spectrum Disorder S creening Data for Children ,
- ,
- Autistic Spectrum Disorder+Screening+Data+for+Adolescent+++">Autistic Spectrum Disorder+Screening+Data+for+Adolescent++++">Autistic Spectrum Disorder+Screening+Data+for+Adolescent++++">Autistic Spectrum Disorder+Screening+Data+for+Adolescent++++*

 Autistic Spectrum Spec
- ,
- APS Failure at Scania Trucks,
- ,
- Wireless Indoor Localization,
- ,
- HCC Survival,
- ,
- CSM (Conventional and Social Media Movies) Dataset 2014 and 2015,
- ,
- University of Tehran Question Dataset 2016 (UTQD.2016),
- <

- /a>,
 Autism Screening Adult,
- <i mg border="1" src="assets/MLimages/SmallLargedefault.jpg"/>,
- Activity recognition with healthy older people using a batteryless wearable sensor,
- ,
- Immunotherapy Dataset,
- ,
- Cryotherapy Dataset ,
- ,
- OCT data & amp; Color Fundus Images of Left & amp; 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 I ocalization 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_loT_botnet_attacks_N_BaloT
-
- 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 (Druglib.com),
- ,
- Drug Review Dataset (Drugs.com),
- ,
- Physical Unclonable Functions,
- , a>,
- Superconductivty Data,
- ,
- WESAD (Wearable Stress and A ffect Detection),
- ,
- GNFUV Unmanned Surface V ehicles Sensor Data Set 2,
- <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/S
mallLargedefault.jpg"/></a>,
 <a href="datasets/Online+Shoppers+Purchasing+Intention+Dataset">Online Shoppers Purchasing Intention Da
taset</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/SmallLarg"
edefault.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/SmallLa"
rgedefault.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/SmallL"
argedefault.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/SmallLargedefaul
t.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+spe
ech+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+spe"><a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+spe"><a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+spe"><a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+spe"><a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+spe"><a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+spe"><a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+spe"><a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+spe"><a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+spe"><a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+spe"><a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+biomarkers+biomarkers+biomarkers+biomarkers+biomarkers+biomarkers+biomarkers+biomarkers+biomarkers+biomarkers+b
ech+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.j
pg"/></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/SmallL
argedefault.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 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/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/Pittsburgh+Bridges', 'datasets/Pittsburgh+Bridges', 'datasets/Car+Evaluation', 'datasets/Car+Evaluation', 'datasets/Census+Income', 'datasets/Census+Income',

```
'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/lonosphere',
'datasets/lonosphere',
'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/EI+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/Callt2+Building+People+Counts',
'datasets/Callt2+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'.
'datacate/Librac Mayamant'
```

```
datascts/ Libi as Tivio verificiti.
'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',
```

```
ualasels/neuler 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/UbigLog+%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+Ne
twork'.
'datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%28OBS%29+Ne
twork',
'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',
'datacete/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/chipseg',
'datasets/chipseq',
'datasets/SGEMM+GPU+kernel+performance',
'datasets/SGEMM+GPU+kernel+performance',
'datasets/Repeat+Consumption+Matrices',
'datasets/Repeat+Consumption+Matrices',
'datasets/detection_of_loT_botnet_attacks_N_BaloT',
'datasets/detection of IoT botnet attacks N BaloT',
'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'.
'dotoooto/DMILLID'
```

```
ualasels/i MU-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
'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'.
```

'dataaata/Annaal

```
ualasels/Alliealling,
'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/lonosphere',
'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/EI+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/Callt2+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++++Pr
oblem%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-Seg',
'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 BaloT',
'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', 'datas ets/Arrhythmia', 'datasets/Artificial+Characters', 'datasets/Audiology+%28Original%29', 'datasets/Audiology+%2 8Standardized%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/Breast+Cancer+Wisconsin+%28Diagnost-Cancer+Wiscons-Cancer+Wiscons-Cancer+Wiscons-C asets/Pittsburgh+Bridges', 'datasets/Car+Evaluation', 'datasets/Census+Income', 'datasets/Chess+%28King-Ro ok+vs.+King-Knight%29', 'datasets/Chess+%28King-Rook+vs.+King-Pawn%29', 'datasets/Chess+%28King-Roo k+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/C ontraceptive+Method+Choice', 'datasets/Covertype', 'datasets/Cylinder+Bands', 'datasets/Dermatology', 'datase ts/Diabetes', 'datasets/DGP2+-+The+Second+Data+Generation+Program', 'datasets/Document+Understanding' , 'datasets/EBL+Domain+Theories', 'datasets/Echocardiogram', 'datasets/Ecoli', 'datasets/Flags', 'datasets/Func tion+Finding', 'datasets/Glass+Identification', 'datasets/Haberman%27s+Survival', 'datasets/Hayes-Roth', 'datas ets/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', 'd atasets/Liver+Disorders', 'datasets/Logic+Theorist', 'datasets/Lung+Cancer', 'datasets/Lymphography', 'dataset s/Mechanical+Analysis', 'datasets/Meta-data', 'datasets/Mobile+Robots', 'datasets/Molecular+Biology+%28Prom oter+Gene+Sequences%29', 'datasets/Molecular+Biology+%28Protein+Secondary+Structure%29', 'datasets/M olecular+Biology+%28Splice-junction+Gene+Sequences%29', 'datasets/MONK%27s+Problems', 'datasets/Mora I+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+Clas sification', 'datasets/Optical+Recognition+of+Handwritten+Digits', 'datasets/Pen-Based+Recognition+of+Handwritten+Digits', 'datasets/Pen-Based+Recognition+Of+Pen-Based+Recognition+Of+Pen-Based+Recognition+Of+Pen-Based+Recognition+Of+Pen-Based+Recognition+Of+Pen-Based+Recognition+Of+Pen-Based+Recognition+Of+Pe itten+Digits', 'datasets/Post-Operative+Patient', 'datasets/Primary+Tumor', 'datasets/Prodigy', 'datasets/Qualitati ve+Structure+Activity+Relationships', 'datasets/Quadruped+Mammals', 'datasets/Servo', 'datasets/Shuttle+Land ing+Control', 'datasets/Solar+Flare', 'datasets/Soybean+%28Large%29', 'datasets/Soybean+%28Small%29', 'datasets/Soybean+%29Small%29', 'datasets/Soybean+%29Small%29', 'datasets/Soybean+%29Small%29', 'datasets/Soybean+%29Small%29', 'datasets/Soybean+%29Sma tasets/Challenger+USA+Space+Shuttle+O-Ring', 'datasets/Low+Resolution+Spectrometer', 'datasets/Spambas e', 'datasets/SPECT+Heart', 'datasets/SPECTF+Heart', 'datasets/Sponge', 'datasets/Statlog+Project', 'datasets/ Student+Loan+Relational', 'datasets/Teaching+Assistant+Evaluation', 'datasets/Tic-Tac-Toe+Endgame', 'datas ets/Thyroid+Disease', 'datasets/Trains', 'datasets/University', 'datasets/Congressional+Voting+Records', 'datase ts/Water+Treatment+Plant', 'datasets/Waveform+Database+Generator+%28Version+1%29', 'datasets/Wavefor m+Database+Generator+%28Version+2%29', 'datasets/Wine', 'datasets/Yeast', 'datasets/Zoo', 'datasets/Undoc umented', 'datasets/Twenty+Newsgroups', 'datasets/Australian+Sign+Language+signs', 'datasets/Australian+Si gn+Language+signs+%28High+Quality%29', 'datasets/US+Census+Data+%281990%29', 'datasets/Census-Inc ome+%28KDD%29', 'datasets/Coil+1999+Competition+Data', 'datasets/Corel+Image+Features', 'datasets/E.+C oli+Genes', 'datasets/EEG+Database', 'datasets/El+Nino', 'datasets/Entree+Chicago+Recommendation+Data', datasets/CMU+Face+Images', 'datasets/Insurance+Company+Benchmark+%28COIL+2000%29', 'datasets/Inte rnet+Usage+Data', 'datasets/IPUMS+Census+Database', 'datasets/Japanese+Vowels', 'datasets/KDD+Cup+19 98+Data', 'datasets/KDD+Cup+1999+Data', 'datasets/M.+Tuberculosis+Genes', 'datasets/Movie', 'datasets/MS NBC.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+Ca tegorization+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+Venu s+-+JARtool+experiment', 'datasets/Statlog+%28Australian+Credit+Approval%29', 'datasets/Statlog+%28Germa n+Credit+Data%29', 'datasets/Statlog+%28Heart%29', 'datasets/Statlog+%28Landsat+Satellite%29', 'datasets/S tatlog+%28Image+Segmentation%29', 'datasets/Statlog+%28Shuttle%29', 'datasets/Statlog+%28Vehicle+Silhou ettes%29', 'datasets/Connectionist+Bench+%28Nettalk+Corpus%29', 'datasets/Connectionist+Bench+%28Sona r%2C+Mines+vs.+Rocks%29', 'datasets/Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29', 'datasets/Economic+Sanctions', 'datasets/Protein+Data', 'datasets/Cloud', 'datasets/Callt2+Building+People+Co unts', 'datasets/Dodgers+Loop+Sensor', 'datasets/Poker+Hand', 'datasets/MAGIC+Gamma+Telescope', 'datase ts/UJI+Pen+Characters', 'datasets/Mammographic+Mass', 'datasets/Forest+Fires', 'datasets/Reuters+Transcrib ed+Subset', 'datasets/Bag+of+Words', 'datasets/Concrete+Compressive+Strength', 'datasets/Hill-Valley', 'datas ets/Arcene', 'datasets/Dexter', 'datasets/Dorothea', 'datasets/Gisette', 'datasets/Madelon', 'datasets/Ozone+Lev el+Detection', 'datasets/Abscisic+Acid+Signaling+Network', 'datasets/Parkinsons', 'datasets/Character+Trajector ies', 'datasets/Blood+Transfusion+Service+Center', 'datasets/UJI+Pen+Characters+%28Version+2%29', 'datase ts/Semeion+Handwritten+Digit', 'datasets/SECOM', 'datasets/Plants', 'datasets/Libras+Movement', 'datasets/Co ncrete+Slump+Test', 'datasets/Communities+and+Crime', 'datasets/Acute+Inflammations', 'datasets/Wine+Qual ity', 'datasets/URL+Reputation', 'datasets/p53+Mutants', 'datasets/Parkinsons+Telemonitoring', 'datasets/Demo spongiae', 'datasets/Opinosis+Opinion+%26frasl%3B+Review', 'datasets/Breast+Tissue', 'datasets/Cardiotocogr aphy', 'datasets/Wall-Following+Robot+Navigation+Data', 'datasets/Spoken+Arabic+Digit', 'datasets/Localization

```
+Data+for+Person+Activity, datasets/AutoOniv, datasets/Steel+Plates+Fautis, datasets/MiniBoonE+particle+i
dentification', 'datasets/YearPredictionMSD', 'datasets/PEMS-SF', 'datasets/OpinRank+Review+Dataset', 'datasets/PEMS-SF', 'datasets/PEMS-S
ets/Relative+location+of+CT+slices+on+axial+axis', 'datasets/Online+Handwritten+Assamese+Characters+Data
set', 'datasets/PubChem+Bioassay+Data', 'datasets/Record+Linkage+Comparison+Patterns', 'datasets/Commu
nities+and+Crime+Unnormalized', 'datasets/Vertebral+Column', 'datasets/EMG+Physical+Action+Data+Set', 'da
tasets/Vicon+Physical+Action+Data+Set', 'datasets/Amazon+Commerce+reviews+set', 'datasets/Amazon+Acce
ss+Samples', 'datasets/Reuter 50 50', 'datasets/Farm+Ads', 'datasets/DBWorld+e-mails', 'datasets/KEGG+Met
abolic+Relation+Network+%28Directed%29', 'datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%2
9', 'datasets/Bank+Marketing', 'datasets/YouTube+Comedy+Slam+Preference+Data', 'datasets/Gas+Sensor+Ar
ray+Drift+Dataset', 'datasets/ILPD+%28Indian+Liver+Patient+Dataset%29', 'datasets/OPPORTUNITY+Activity+
Recognition', 'datasets/Nomao', 'datasets/SMS+Spam+Collection', 'datasets/Skin+Segmentation', 'datasets/Plan
ning+Relax', 'datasets/PAMAP2+Physical+Activity+Monitoring', 'datasets/Restaurant+%26+consumer+data', 'datasets/Restaurant-%26+consumer-data', 'datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaurant-%26+consumer-datasets/Restaura
asets/CNAE-9', 'datasets/Individual+household+electric+power+consumption', 'datasets/seeds', 'datasets/Northi
x', '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', 'dat
asets/Climate+Model+Simulation+Crashes', 'datasets/MicroMass', 'datasets/QSAR+biodegradation', 'datasets/B
LOGGER', 'datasets/Daily+and+Sports+Activities', 'datasets/User+Knowledge+Modeling', 'datasets/Reuters+RC
V1+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+Tertia
ry+Structure', 'datasets/seismic-bumps', 'datasets/banknote+authentication', 'datasets/USPTO+Algorithm+Chall
enge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Problem%3A+Pat', 'datasets/YouTub
e+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', 'dat
asets/SML2010', 'datasets/Bike+Sharing+Dataset', 'datasets/Predict+keywords+activities+in+a+online+social+m
edia', 'datasets/Thoracic+Surgery+Data', 'datasets/EMG+dataset+in+Lower+Limb', 'datasets/SUSY', 'datasets/H
IGGS', 'datasets/Qualitative_Bankruptcy', 'datasets/LSVT+Voice+Rehabilitation', 'datasets/Dataset+for+ADL+Re
cognition+with+Wrist-worn+Accelerometer', 'datasets/Wilt', 'datasets/User+Identification+From+Walking+Activity
', 'datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer', 'datasets/Leaf', 'datasets/Dress'
es_Attribute_Sales', 'datasets/Tamilnadu+Electricity+Board+Hourly+Readings', 'datasets/Airfoil+Self-Noise', 'datasets/Airfoi
asets/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+1999-20
08', 'datasets/Bach+Choral+Harmony', 'datasets/StoneFlakes', 'datasets/Tennis+Major+Tournament+Match+St
atistics', 'datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings', 'datasets/Gestur
e+Phase+Segmentation', 'datasets/Perfume+Data', 'datasets/BlogFeedback', 'datasets/REALDISP+Activity+Rec
ognition+Dataset', 'datasets/Newspaper+and+magazine+images+segmentation+dataset', 'datasets/AAAI+2014
+Accepted+Papers', 'datasets/Gas+sensor+array+under+flow+modulation', 'datasets/Gas+sensor+array+expos
ed+to+turbulent+gas+mixtures', 'datasets/UJIIndoorLoc', 'datasets/Sentence+Classification', 'datasets/Dow+Jon
es+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', 'data
sets/Grammatical+Facial+Expressions', 'datasets/NoisyOffice', 'datasets/MHEALTH+Dataset', 'datasets/Student
+Performance', 'datasets/ElectricityLoadDiagrams20112014', 'datasets/Gas+sensor+array+under+dynamic+ga
s+mixtures', 'datasets/microblogPCU', 'datasets/Firm-Teacher_Clave-Direction_Classification', 'datasets/Datase
t+for+Sensorless+Drive+Diagnosis', 'datasets/TV+News+Channel+Commercial+Detection+Dataset', 'datasets/P
hishing+Websites', 'datasets/Greenhouse+Gas+Observing+Network', 'datasets/Diabetic+Retinopathy+Debrece
n+Data+Set', 'datasets/HIV-1+protease+cleavage', 'datasets/Sentiment+Labelled+Sentences', 'datasets/Online
+News+Popularity', 'datasets/Forest+type+mapping', 'datasets/wiki4HE', 'datasets/Online+Video+Characteristic
s+and+Transcoding+Time+Dataset', 'datasets/Chronic Kidney Disease', 'datasets/Machine+Learning+based+
ZZAlpha+Ltd.+Stock+Recommendations+2012-2014', 'datasets/Folio', 'datasets/Taxi+Service+Trajectory+-+Pre
diction+Challenge%2C+ECML+PKDD+2015', 'datasets/Cuff-Less+Blood+Pressure+Estimation', 'datasets/Smart
phone-Based+Recognition+of+Human+Activities+and+Postural+Transitions', 'datasets/Mice+Protein+Expressio
n', 'datasets/UJIIndoorLoc-Mag', 'datasets/Heterogeneity+Activity+Recognition', 'datasets/Educational+Process
+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set', 'datasets/HEPMASS', 'datasets/Indoor+User+Mo
vement+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', 'd
atasets/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+ga s+sensor+arrays', 'datasets/Gas+sensors+for+home+activity+monitoring', 'datasets/Facebook+Comment+Volu me+Dataset', 'datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assi sted+Living+%28AAL%29', 'datasets/Polish+companies+bankruptcy+data', 'datasets/Activity+Recognition+syste m+based+on+Multisensor+data+fusion+%28AReM%29', 'datasets/Dota2+Games+Results', 'datasets/Facebook +metrics', 'datasets/UbiqLog+%28smartphone+lifelogging%29', 'datasets/NIPS+Conference+Papers+1987-201 5', 'datasets/HTRU2', 'datasets/Drug+consumption+%28quantified%29', 'datasets/Appliances+energy+predictio n', '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+200 0+Freight+Tracking+and+Tracing', 'datasets/Cervical+cancer+%28Risk+Factors%29', 'datasets/Quality+Assess ment+of+Digital+Colposcopies', 'datasets/KASANDR', 'datasets/FMA%3A+A+Dataset+For+Music+Analysis', 'dat asets/Air+quality', 'datasets/Epileptic+Seizure+Recognition', 'datasets/Devanagari+Handwritten+Character+Dat aset', 'datasets/Stock+portfolio+performance', 'datasets/MoCap+Hand+Postures', 'datasets/Early+biomarkers+o f+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/Park inson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet', 'datasets/Sales_Transactions_Dataset_W eekly', 'datasets/Las+Vegas+Strip', 'datasets/Eco-hotel', 'datasets/MEU-Mobile+KSD', 'datasets/Crowdsourced+ Mapping', 'datasets/gene+expression+cancer+RNA-Seg', 'datasets/Hybrid+Indoor+Positioning+Dataset+from+ WiFi+RSSI%2C+Bluetooth+and+magnetometer', 'datasets/chestnut+%E2%80%93+LARVIC', 'datasets/Burst+H eader+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+Benchm ark+dataset+for+Turkish+text+categorization', 'datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy', 'd atasets/Daily+Demand+Forecasting+Orders', 'datasets/Paper+Reviews', 'datasets/extention+of+Z-Alizadeh+sa ni+dataset', 'datasets/Z-Alizadeh+Sani', 'datasets/Dynamic+Features+of+VirusShare+Executables', 'datasets/ID A2016Challenge', '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-Ins tance+Learning+with+Instance+Labels', 'datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++ ', 'datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++', 'datasets/APS+Failure+at+Scani a+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+o lder+people+using+a+batteryless+wearable+sensor', 'datasets/Immunotherapy+Dataset', 'datasets/Cryotherap y+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+fl owmeter+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+B uilding+Data+Set', 'datasets/Health+News+in+Twitter', 'datasets/chipseq', 'datasets/SGEMM+GPU+kernel+perf ormance', 'datasets/Repeat+Consumption+Matrices', 'datasets/detection of IoT botnet attacks N BaloT', 'da tasets/Absenteeism+at+work', 'datasets/SCADI', 'datasets/Condition+monitoring+of+hydraulic+systems', 'datase ts/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/Simulat ed+Falls+and+Daily+Living+Activities+Data+Set', 'datasets/Multimodal+Damage+Identification+for+Humanitaria n+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 g+Review+Dataset+%28Drugs.com%29', 'datasets/Physical+Unclonable+Functions', 'datasets/Superconductivt y+Data', 'datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29', 'datasets/GNFUV+Unmanned+S urface+Vehicles+Sensor+Data+Set+2', 'datasets/Student+Academics+Performance', 'datasets/Online+Shopper s+Purchasing+Intention+Dataset', 'datasets/PMU-UD', 'datasets/Parkinson%27s+Disease+Classification', 'datas ets/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+va luation+data+set', 'datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connec ted+speech+Data+Set+', 'datasets/Somerville+Happiness+Survey', 'datasets/2.4+GHZ+Indoor+Channel+Measu rements', '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/Tarvel+Review+Ratings', 'datasets/Tarvel+Reviews', 'datasets/Tarvel+Review+Ratings', 'datasets/Tarvel+Reviews', 'datasets/Tarvel+Review+Ratings', 'datasets/Tarvel+Review+Ratings', 'datasets/Tarvel+Reviews', 'datasets/Tarvel+Review+Ratings', 'datasets/Tarvel+Review+Review+Ratings', 'datasets/Tarvel+Review+ tasets/Rice+Leaf+Diseases']"

```
In [45]:

name_database = re.findall("V(\S+)\',",str_database)
name_database

Out[45]:

['Abalone',
   'Adult',
   'Annealing',
   'Anonymous+Microsoft+Web+Data',
   'Arrhythmia',
   'Artificial+Characters',
```

'Audiology+%28Original%29', 'Audiology+%28Standardized%29',

'Breast+Cancer+Wisconsin+%28Original%29', 'Breast+Cancer+Wisconsin+%28Prognostic%29', 'Breast+Cancer+Wisconsin+%28Diagnostic%29',

'Chess+%28King-Rook+vs.+King-Knight%29', 'Chess+%28King-Rook+vs.+King-Pawn%29',

'DGP2+-+The+Second+Data+Generation+Program',

'Chess+%28King-Rook+vs.+King%29', 'Chess+%28Domain+Theories%29',

'Japanese+Credit+Screening',

'Document+Understanding', 'EBL+Domain+Theories',

'Contraceptive+Method+Choice',

'Computer+Hardware',

'Auto+MPG', 'Automobile', 'Badges',

'Balloons',

'Balance+Scale',

'Breast+Cancer',

'Pittsburgh+Bridges', 'Car+Evaluation', 'Census+Income',

'Bach+Chorales',
'Connect-4',
'Credit+Approval',

'Covertype',

'Cylinder+Bands', 'Dermatology', 'Diabetes',

'Echocardiogram',

'Function+Finding',
'Glass+Identification',
'Haberman%27s+Survival',

'Image+Segmentation', 'Internet+Advertisements',

'Hayes-Roth', 'Heart+Disease',

'lonosphere',

'Labor+Relations',
'LED+Display+Domain'.

'Hepatitis',
'Horse+Colic',

'ICU',

'Iris', 'ISOLET', 'Kinship',

'Ecoli', 'Flags',



```
'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',
'F +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',
'Callt2+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'.
'Canarata , Slump , Tact'
```

```
Condictor Clamp + 1 cot,
'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',
                                Hisatian, of Dady, Dagtimas, and Ma
```

```
vvearable+Computing %3A+Classincation+or+Douy+Fostures+and+iviovements+%2oFoC-nio %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%3
A+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,
'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-Seg',
'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_loT_botnet_attacks_N_BaloT',
'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/lonosphere',
```

'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://grahiyo.jog.ugi.adu/ml/datagats/LIS. Consus. Data 19/ 2010009/ 201
```

```
11(tp3.//arcinvc.io3.doi.cdd/111/ad(d3c(3/00+0c13d3+Dd(d+/0201330/023.
'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/Callt2+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',
```

```
nitips://archive.ics.uci.edu/m/datasets/semelon+handwitten+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/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+Move
ments+%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+Se
nsors',
'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+Recordin
gs',
'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',
```

nttps://arcnive.ics.uci.edu/mi/datasets/3D+Road+Network+%28Nortn+Jutiand%2C+Denmark%29,

```
'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+20
12-2014',
'https://archive.ics.uci.edu/ml/datasets/Folio',
'https://archive.ics.uci.edu/ml/datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+20
15',
'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+Tr
ansitions',
'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+Monitorin
g+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+i
n+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+%28
AReM%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+conn
ected+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-Seg',
'https://archive.ics.uci.edu/ml/datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+a
nd+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/chipseg',
'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_loT_botnet_attacks_N BaloT',
'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 Iso 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:,
   <b>Default Task</b> ,
   <a href="datasets.php?format=&amp;task=cla&amp;att=&amp;area=&amp;numAtt=&amp;n
  umIns=&type=&sort=nameUp&view=table">Classification</a> <font color="red">(350)</font><br/>br/
  ><a href="datasets.php?format=&amp;task=reg&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type"
  =&sort=nameUp&view=table">Regression</a> <font color="red">(96)</font><br/><a href="datasets.p"
  hp?format=&task=clu&att=&area=&numAtt=&numIns=&type=&sort=nameUp
  &view=table">Clustering</a> <font color="red">(84)</font><br/><a href="datasets.php?format=&amp;task"
  =other&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table">
  Other </a> < font color = "red" > (55) < /font > ,
   <b>Attribute Type</b> ,
   <a href="datasets.php?format=&amp;task=&amp;att=cat&amp;area=&amp;numAtt=&amp;n
  umIns=&type=&sort=nameUp&view=table">Categorical</a> <font color="red">(38)</font><br/><
  a href="datasets.php?format=&task=&att=num&area=&numAtt=&numIns=&type=
  &sort=nameUp&view=table">Numerical</a> <font color="red">(307)</font><br/><a href="datasets.ph">(307)</font><br/><a href="datasets.ph">(307)</font><br/><a href="datasets.ph">(307)</a>
  p?format=&task=&att=mix&area=&numAtt=&numIns=&type=&sort=nameUp
  & view=table">Mixed</a> <font color="red">(55)</font> ,
   <b>Data Type</b> ,
   <a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&a
  Ins=&type=mvar&sort=nameUp&view=table">Multivariate</a> <font color="red">(357)</font><br/>br/
  ><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;type=uv
  ar&sort=nameUp&view=table">Univariate</a> <font color="red">(23)</font><br/><a href="datasets.p"
  hp?format=&task=&att=&area=&numAtt=&numIns=&type=seq&sort=nameU
  p&view=table">Sequential</a> <font color="red">(47)</font><br/><a href="datasets.php?format=&amp;tas"
  k=&att=&area=&numAtt=&numIns=&type=ts&sort=nameUp&view=table">Ti
  me-Series</a> <font color="red">(91)</font><br/><a href="datasets.php?format=&amp;task=&amp;att=&amp;ar
  aa-&amp:numAtt-&amp:numIns-&amp:typa-tayt&amp:sort-namal In&amp:yiaw-tahla">Tayt-/a> >font color-"
```

red">(53)
chrotrological red">(54)
chrotrological red">(54)
chrotrolo

mp;view=table">Social Sciences (26)
 Business (29)
Game (10)
Other (73) ,

- # Attributes ,
 Less than 10 (113)
br/>10 to 100 (210)
Greater than 100 (84) ,
- # Instances,
- Less than 100 (27)
100 to 1000 (162)
Greater than 1000 (246) ,
- Format Type ,
- Matrix (324)
>ca h ref="datasets.php?format=nonmat&task=&att=&area=&numAtt=&numIns=&type= &sort=nameUp&view=table">Non-Matrix (145) ,
- 469 Data Sets,
- Table View List View,Name,
- Data Types/a>,
- Default Task,
- Attribute Types,
- # Instances,
- # Attributes/a>,
- Year,,
- Abalone,
- Multivariate ,
- Classification ,
- Categorical, Integer, Real ,
- 4177 ,
- 8 ,
- 1005

```
 1333 ,
<b><a href="datasets/Adult">Adult</a></b>,
Multivariate ,
Classification ,
Categorical, Integer ,
48842 ,
14 ,
1996 ,
<b><a href="datasets/Annealing">Annealing</a></b>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
798 ,
38 ,
 ,
<b><a href="datasets/Anonymous+Microsoft+Web+Data">Anonymous Microsoft Web Data
</a></b>,
 ,
Recommender-Systems ,
Categorical ,
37711 ,
294 ,
1998 ,
<b><a href="datasets/Arrhythmia">Arrhythmia</a></b>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
452 ,
279 ,
1998 ,
<b><a href="datasets/Artificial+Characters">Artificial Characters</a></b>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
6000 ,
7 ,
1992 ,
<b><a href="datasets/Audiology+%28Original%29">Audiology (Original)</a></b>,
Multivariate ,
Classification ,
Categorical ,
226 ,
 ,
1987 ,
<b><a href="datasets/Audiology+%28Standardized%29">Audiology (Standardized)</a></b>
,
Multivariate ,
Classification ,
Categorical ,
226 ,
69 ,
1992 ,
<b><a href="datasets/Auto+MPG">Auto MPG</a></b>,
Multivariate ,
Regression ,
Categorical, Real ,
398 ,
8 ,
1993 ,
<b><a href="datasets/Automobile">Automobile</a></b>,
Multivariate ,
Regression ,
```

```
Calegoricai, integer, Reai ,
205 ,
26 ,
1987 ,
<b><a href="datasets/Badges">Badges</a></b>,
Univariate, Text ,
Classification ,
 ,
294 ,
1 ,
1994 ,
<b><a href="datasets/Balance+Scale">Balance Scale</a></b>,
Multivariate ,
Classification ,
Categorical ,
625 ,
4 ,
1994 ,
<b><a href="datasets/Balloons">Balloons</a>,
Multivariate ,
Classification ,
Categorical ,
16 ,
4 ,
 ,
<b><a href="datasets/Breast+Cancer">Breast Cancer</a></b>,
Multivariate ,
Classification ,
Categorical ,
286 ,
9 ,
1988 .
<b><a href="datasets/Breast+Cancer+Wisconsin+%28Original%29">Breast Cancer Wiscon
sin (Original)</a></b>,
Multivariate ,
Classification ,
Integer ,
699 ,
10 ,
1992 ,
<b><a href="datasets/Breast+Cancer+Wisconsin+%28Prognostic%29">Breast Cancer Wisc
onsin (Prognostic)</a></b>,
Multivariate ,
Classification, Regression ,
Real ,
198 ,
34 ,
1995 ,
<b><a href="datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29">Breast Cancer Wisc
onsin (Diagnostic)</a></b>,
Multivariate ,
Classification ,
Real ,
569 ,
32 ,
1995 ,
<b><a href="datasets/Pittsburgh+Bridges">Pittsburgh Bridges</a></b>,
Multivariate ,
Classification ,
Categorical, Integer ,
108 ,
13 ,
```

```
1990 ,
<b><a href="datasets/Car+Evaluation">Car Evaluation</a></b>,
Multivariate ,
Classification ,
Categorical ,
1728 ,
6 ,
1997 ,
<b><a href="datasets/Census+Income">Census Income</a></b>,
Multivariate ,
Classification ,
Categorical, Integer ,
48842 ,
14 ,
1996 ,
<b><a href="datasets/Chess+%28King-Rook+vs.+King-Knight%29">Chess (King-Rook vs.
King-Knight)</a></b>,
Multivariate, Data-Generator ,
Classification ,
Categorical, Integer ,
 ,
22 ,
1988 ,
<b><a href="datasets/Chess+%28King-Rook+vs.+King-Pawn%29">Chess (King-Rook vs. Ki
ng-Pawn)</a></b>,
Multivariate ,
Classification ,
Categorical ,
3196 ,
36 ,
1989 ,
<b><a href="datasets/Chess+%28King-Rook+vs.+King%29">Chess (King-Rook vs. King)</a>/a
></b>,
Multivariate ,
Classification ,
Categorical, Integer ,
28056 ,
6 ,
1994 ,
<b><a href="datasets/Chess+%28Domain+Theories%29">Chess (Domain Theories)</a>
b>,
Domain-Theory ,
 ,
 ,
 ,
 ,
 ,
<b><a href="datasets/Bach+Chorales">Bach Chorales</a></b>,
Univariate, Time-Series ,
 ,
Categorical, Integer ,
100 ,
6 ,
 ,
<b><a href="datasets/Connect-4">Connect-4</a></b>,
Multivariate, Spatial ,
Classification ,
Categorical ,
67557 ,
42 ,
1995 ,
<b><a href="datasets/Credit+Approval">Credit Approval</a></b>,
```

```
Multivariate ,
Classification ,
Categorical, Integer, Real ,
690 ,
15 ,
 ,
<b><a href="datasets/Japanese+Credit+Screening">Japanese Credit Screening</a></b></
Multivariate, Domain-Theory ,
Classification ,
Categorical, Real, Integer ,
125 ,
 ,
1992 ,
<b><a href="datasets/Computer+Hardware">Computer Hardware</a></b>,
Multivariate ,
Regression ,
Integer ,
209 ,
9 ,
1987 ,
<b><a href="datasets/Contraceptive+Method+Choice">Contraceptive Method Choice</a></</pre>
b>,
Multivariate ,
Classification ,
Categorical, Integer ,
1473 ,
9 ,
1997 ,
<b><a href="datasets/Covertype">Covertype</a></b>,
Multivariate ,
Classification ,
Categorical, Integer ,
581012 ,
54 ,
1998 ,
<b><a href="datasets/Cylinder+Bands">Cylinder Bands</a></b>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
512 ,
39 ,
1995 ,
<b><a href="datasets/Dermatology">Dermatology</a></b>,
Multivariate ,
Classification ,
Categorical, Integer ,
366 ,
33 ,
1998 ,
<b><a href="datasets/Diabetes">Diabetes</a></b>,
Multivariate, Time-Series ,
 ,
Categorical, Integer ,
 ,
20 ,
 ,
<b><a href="datasets/DGP2+-+The+Second+Data+Generation+Program">DGP2 - The Sec
ond Data Generation Program</a></b>,
Data-Generator ,
 ,
Real ,
```

```
 ,
 ,
 ,
<b><a href="datasets/Document+Understanding">Document Understanding</a></b>,
 ,
 ,
 ,
 ,
 ,
1994 ,
<b><a href="datasets/EBL+Domain+Theories">EBL Domain Theories</a></b>,
 ,
 ,
 ,
 ,
 ,
 ,
<b><a href="datasets/Echocardiogram">Echocardiogram</a></b>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
132 ,
12 ,
1989 ,
<b><a href="datasets/Ecoli">Ecoli</a></b>,
Multivariate ,
Classification ,
Real ,
336 ,
8 ,
1996 ,
<b><a href="datasets/Flags">Flags</a></b>,
Multivariate ,
Classification ,
Categorical, Integer ,
194 ,
30 ,
1990 ,
<b><a href="datasets/Function+Finding">Function Finding</a></b>,
 ,
Function-Learning ,
Real ,
352 ,
 ,
1990 ,
<b><a href="datasets/Glass+Identification">Glass Identification</a></b>,
Multivariate ,
Classification ,
Real ,
214 ,
10 ,
1987 ,
<b><a href="datasets/Haberman%27s+Survival">Haberman's Survival</a></b>,
Multivariate ,
Classification ,
Integer ,
306 ,
3 ,
1999 ,
<b><a href="datasets/Hayes-Roth">Hayes-Roth</a></b>,
Multivariate ,
Classification ,
```

```
Categorical ,
160 ,
5 ,
1989 ,
<b><a href="datasets/Heart+Disease">Heart Disease</a></b>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
303 ,
75 ,
1988 ,
<b><a href="datasets/Hepatitis">Hepatitis</a></b>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
155 ,
19 ,
1988 ,
<b><a href="datasets/Horse+Colic">Horse Colic</a></b>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
368 ,
27 ,
1989 ,
<b><a href="datasets/ICU">ICU</a></b>,
Multivariate, Time-Series ,
 ,
Real ,
 ,
 ,
 ,
<b><a href="datasets/Image+Segmentation">Image Segmentation</a></b>,
Multivariate ,
Classification ,
Real ,
2310 ,
19 ,
1990 ,
<b><a href="datasets/Internet+Advertisements">Internet Advertisements</a></b>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
3279 ,
1558 ,
1998 ,
<b><a href="datasets/lonosphere">lonosphere</a></b>,
Multivariate ,
Classification ,
Integer, Real ,
351 ,
34 ,
1989 ,
<b><a href="datasets/Iris">Iris</a></b>,
Multivariate ,
Classification ,
Real ,
150 ,
4 ,
1988 ,
<b><a href="datasets/ISOLET">ISOLET</a></b>,
Multivariate .
```

```
Classification ,
Real ,
7797 ,
617 ,
1994 ,
<b><a href="datasets/Kinship">Kinship</a></b>,
Relational ,
Relational-Learning ,
Categorical ,
104 ,
12 ,
1990 ,
<b><a href="datasets/Labor+Relations">Labor Relations</a></b>,
Multivariate ,
 ,
Categorical, Integer, Real ,
57 ,
16 ,
1988 ,
<b><a href="datasets/LED+Display+Domain">LED Display Domain</a></b>,
Multivariate, Data-Generator ,
Classification ,
Categorical .
 ,
7 ,
1988 ,
<b><a href="datasets/Lenses">Lenses</a></b>,
Multivariate ,
Classification ,
Categorical ,
24 ,
4 ,
1990 ,
<b><a href="datasets/Letter+Recognition">Letter Recognition</a></b>,
Multivariate ,
Classification ,
Integer ,
20000 ,
16 ,
1991 ,
<b><a href="datasets/Liver+Disorders">Liver Disorders</a></b>,
Multivariate ,
 ,
Categorical, Integer, Real ,
345 ,
7 ,
1990 ,
<b><a href="datasets/Logic+Theorist">Logic Theorist</a></b>,
Domain-Theory ,
 ,
 ,
 ,
 ,
 ,
<b><a href="datasets/Lung+Cancer">Lung Cancer</a></b>,
Multivariate ,
Classification ,
Integer ,
32 ,
56 ,
1992 ,
<n class="normal"><h><a href="datasets/l ymphography">| ymphography</a></h></n>
```

```
Multivariate ,
Classification ,
Categorical ,
148 ,
18 ,
1988 ,
<b><a href="datasets/Mechanical+Analysis">Mechanical Analysis</a></b>,
Multivariate .
Classification ,
Categorical, Integer, Real ,
209 ,
8 ,
1990 ,
<b><a href="datasets/Meta-data">Meta-data</a></b>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
528 ,
22 ,
1996 ,
<b><a href="datasets/Mobile+Robots">Mobile Robots</a></b>,
Domain-Theory ,
 ,
Categorical, Integer, Real ,
 ,
 ,
1995 ,
<b><a href="datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29">Molecular
Biology (Promoter Gene Sequences)</a></b>,
Sequential, Domain-Theory ,
Classification ,
Categorical ,
106 ,
58 ,
1990 ,
<b><a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%29">Molecula
r Biology (Protein Secondary Structure)</a></b>,
Sequential ,
Classification ,
Categorical ,
128 ,
 ,
 ,
<b><a href="datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29">Mole
cular Biology (Splice-junction Gene Sequences)</a>,
Sequential, Domain-Theory ,
Classification ,
Categorical ,
3190 ,
61 ,
1992 ,
<b><a href="datasets/MONK%27s+Problems">MONK's Problems</a></b>,
Multivariate ,
Classification ,
Categorical ,
432 ,
7 ,
1992 ,
<b><a href="datasets/Moral+Reasoner">Moral Reasoner</a></b>,
Domain-Theory ,
 ,
n clace_"normal"> /n
```

```
\p \( \lambda \) \( \lambda
202 ,
 ,
1994 ,
<b><a href="datasets/Multiple+Features">Multiple Features</a></b>,
Multivariate ,
Classification ,
Integer, Real ,
2000 ,
649 ,
 ,
<b><a href="datasets/Mushroom">Mushroom</a>,
Multivariate ,
Classification ,
Categorical ,
8124 ,
22 ,
1987 ,
<b><a href="datasets/Musk+%28Version+1%29">Musk (Version 1)</a></b>,
Multivariate ,
Classification ,
Integer ,
476 ,
168 ,
1994 ,
<b><a href="datasets/Musk+%28Version+2%29">Musk (Version 2)</a></b>,
Multivariate ,
Classification ,
Integer ,
6598 ,
168 ,
1994 ,
<b><a href="datasets/Nursery">Nursery</a></b>,
Multivariate ,
Classification ,
Categorical ,
12960 ,
8 ,
1997 ,
<b><a href="datasets/Othello+Domain+Theory">Othello Domain Theory</a></b>,
Domain-Theory ,
 ,
 ,
 ,
 ,
1991 ,
<b><a href="datasets/Page+Blocks+Classification">Page Blocks Classification</a></b>,
Multivariate ,
Classification ,
Integer, Real ,
5473 ,
10 ,
1995 ,
<b><a href="datasets/Optical+Recognition+of+Handwritten+Digits">Optical Recognition of
Handwritten Digits</a></b>,
Multivariate ,
Classification ,
Integer ,
5620 ,
64 ,
1998 ,
<b><a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits">Pen-Based Recogni
```

```
11011 01 Halluwillen Digits</a></b>
Multivariate ,
Classification ,
Integer ,
10992 ,
16 ,
1998 ,
<b><a href="datasets/Post-Operative+Patient">Post-Operative Patient</a></b>,
Multivariate ,
Classification ,
Categorical, Integer ,
90 ,
8 ,
1993 ,
<b><a href="datasets/Primary+Tumor">Primary Tumor</a></b>,
Multivariate ,
Classification ,
Categorical ,
339 ,
17 ,
1988 ,
<b><a href="datasets/Prodigy">Prodigy</a></b>,
Domain-Theory ,
 ,
 ,
 ,
 ,
 ,
<b><a href="datasets/Qualitative+Structure+Activity+Relationships">Qualitative Structure A
ctivity Relationships</a></b>,
Domain-Theory ,
 ,
 ,
 ,
 ,
 ,
<b><a href="datasets/Quadruped+Mammals">Quadruped Mammals</a></b>,
Multivariate, Data-Generator ,
Classification ,
Real ,
 ,
72 ,
1992 ,
<b><a href="datasets/Servo">Servo</a></b>,,
Multivariate ,
Regression ,
Categorical, Integer ,
167 ,
4 ,
1993 ,
<b><a href="datasets/Shuttle+Landing+Control">Shuttle Landing Control</a></b>,
Multivariate ,
Classification ,
Categorical ,
15 ,
6 ,
1988 ,
<b><a href="datasets/Solar+Flare">Solar Flare</a></b>,
Multivariate ,
Regression ,
Categorical ,
1389 ,
```

```
10 ,
1989 ,
<b><a href="datasets/Soybean+%28Large%29">Soybean (Large)</a></b>,
Multivariate ,
Classification ,
Categorical ,
307 ,
35 ,
1988 ,
<b><a href="datasets/Soybean+%28Small%29">Soybean (Small)</a></b>,
Multivariate ,
Classification ,
Categorical ,
47 ,
35 ,
1987 ,
<b><a href="datasets/Challenger+USA+Space+Shuttle+O-Ring">Challenger USA Space Sh
uttle O-Ring</a></b>,
Multivariate ,
Regression ,
Integer ,
23 ,
4 ,
1993 ,
<b><a href="datasets/Low+Resolution+Spectrometer">Low Resolution Spectrometer</a></</pre>
b>,
Multivariate ,
Classification ,
Integer, Real ,
531 ,
102 ,
1988 ,
<b><a href="datasets/Spambase">Spambase</a></b>,
Multivariate ,
Classification ,
Integer, Real ,
4601 ,
57 ,
1999 ,
<b><a href="datasets/SPECT+Heart">SPECT Heart</a></b>,
Multivariate ,
Classification ,
Categorical ,
267 ,
22 ,
2001 ,
<b><a href="datasets/SPECTF+Heart">SPECTF Heart</a></b>,
Multivariate ,
Classification ,
Integer ,
267 ,
44 ,
2001 ,
<b><a href="datasets/Sponge">Sponge</a></b>,
Multivariate ,
Clustering ,
Categorical, Integer ,
76 ,
45 ,
 ,
<b><a href="datasets/Statlog+Project">Statlog Project</a></b>,
 ,
```

```
 ,
 ,
 ,
 ,
1992 ,
<b><a href="datasets/Student+Loan+Relational">Student Loan Relational</a></b>,
Domain-Theory ,
 ,
 ,
1000 ,
 ,
1993 ,
<b><a href="datasets/Teaching+Assistant+Evaluation">Teaching Assistant Evaluation</a>
b 
Multivariate ,
Classification ,
Categorical, Integer ,
151 ,
5 ,
1997 ,
<b><a href="datasets/Tic-Tac-Toe+Endgame">Tic-Tac-Toe Endgame</a></b>,
Multivariate ,
Classification ,
Categorical ,
958 ,
9 ,
1991 ,
<b><a href="datasets/Thyroid+Disease">Thyroid Disease</a></b>,
Multivariate, Domain-Theory ,
Classification ,
Categorical, Real ,
7200 ,
21 ,
1987 ,
<b><a href="datasets/Trains">Trains</a></b>,
Multivariate ,
Classification ,
Categorical ,
10 ,
32 ,
1994 ,
<b><a href="datasets/University">University</a></b>,
Multivariate ,
Classification ,
Categorical, Integer ,
285 ,
17 ,
1988 ,
<b><a href="datasets/Congressional+Voting+Records">Congressional Voting Records</a>
</b>,
Multivariate .
Classification ,
Categorical ,
435 ,
16 ,
1987 ,
<b><a href="datasets/Water+Treatment+Plant">Water Treatment Plant</a></b>,
Multivariate ,
Clustering ,
Integer, Real ,
527 ,
38 ,
```

```
1993 ,
<b><a href="datasets/Waveform+Database+Generator+%28Version+1%29">Waveform Dat
abase Generator (Version 1)</a></b>,
Multivariate, Data-Generator ,
Classification ,
Real ,
5000 ,
21 ,
1988 ,
<b><a href="datasets/Waveform+Database+Generator+%28Version+2%29">Waveform Dat
abase Generator (Version 2)</a></b>,
Multivariate, Data-Generator ,
Classification ,
Real ,
5000 ,
40 ,
1988 ,
<b><a href="datasets/Wine">Wine</a></b>,,
Multivariate ,
Classification ,
Integer, Real ,
178 ,
13 ,
1991 ,
<b><a href="datasets/Yeast">Yeast</a></b>,
Multivariate ,
Classification ,
Real ,
1484 ,
8 ,
1996 ,
<b><a href="datasets/Zoo">Zoo</a></b>,
Multivariate ,
Classification ,
Categorical, Integer ,
101 ,
17 ,
1990 ,
<b><a href="datasets/Undocumented">Undocumented</a></b>,
 ,
 ,
 ,
 ,
 ,
 ,
<b><a href="datasets/Twenty+Newsgroups">Twenty Newsgroups</a></b>,
Text ,
 ,
 ,
20000 ,
 ,
1999 ,
<b><a href="datasets/Australian+Sign+Language+signs">Australian Sign Language signs
a > </b > ,
Multivariate, Time-Series ,
Classification ,
Categorical, Real ,
6650 ,
15 ,
1999 ,
Sign Language signs (High Quality)</a></b>,
```

```
Multivariate, Time-Series ,
Classification ,
Real ,
2565 ,
22 ,
2002 ,
<b><a href="datasets/US+Census+Data+%281990%29">US Census Data (1990)</a></b>
/p>,
Multivariate ,
Clustering ,
Categorical ,
2458285 ,
68 ,
 ,
<b><a href="datasets/Census-Income+%28KDD%29">Census-Income (KDD)</a></b>,
Multivariate ,
Classification ,
Categorical, Integer ,
299285 ,
40 ,
2000 ,
<b><a href="datasets/Coil+1999+Competition+Data">Coil 1999 Competition Data</a></b>
/p>,
Multivariate ,
 ,
Categorical, Real ,
340 ,
17 ,
1999 ,
<b><a href="datasets/Corel+Image+Features">Corel Image Features</a></b>,
Multivariate ,
 ,
Real ,
68040 ,
89 ,
1999 ,
<b><a href="datasets/E.+Coli+Genes">E. Coli Genes</a></b>,
Relational ,
 ,
 ,
 ,
 ,
2001 ,
<b><a href="datasets/EEG+Database">EEG Database</a></b>,
Multivariate, Time-Series ,
 ,
Categorical, Integer, Real ,
122 ,
4 ,
1999 ,
<b><a href="datasets/EI+Nino">EI Nino</a></b>,
Spatio-temporal ,
 ,
Integer, Real ,
178080 ,
12 ,
1999 ,
<b><a href="datasets/Entree+Chicago+Recommendation+Data">Entree Chicago Recomm
endation Data</a></b>,
Transactional, Sequential ,
Recommender-Systems ,
Categorical .
```

```
50672 ,
 ,
2000 ,
<b><a href="datasets/CMU+Face+Images">CMU Face Images</a></b>,
Image ,
Classification ,
Integer ,
640 ,
 ,
1999 ,
<b><a href="datasets/Insurance+Company+Benchmark+%28COIL+2000%29">Insurance C
ompany Benchmark (COIL 2000)</a></b>,
Multivariate ,
Regression, Description ,
Categorical, Integer ,
9000 ,
86 ,
2000 ,
<b><a href="datasets/Internet+Usage+Data">Internet Usage Data</a></b>,
Multivariate ,
 ,
Categorical, Integer ,
10104 ,
72 ,
1999 ,
<b><a href="datasets/IPUMS+Census+Database">IPUMS Census Database</a></b>,
Multivariate ,
 ,
Categorical, Integer ,
256932 ,
61 ,
1999 ,
<b><a href="datasets/Japanese+Vowels">Japanese Vowels</a>,
Multivariate, Time-Series ,
Classification ,
Real ,
640 ,
12 ,
 ,
<b><a href="datasets/KDD+Cup+1998+Data">KDD Cup 1998 Data</a></b>,
Multivariate ,
Regression ,
Categorical, Integer ,
191779 ,
481 ,
1998 ,
<b><a href="datasets/KDD+Cup+1999+Data">KDD Cup 1999 Data</a></b>,
Multivariate ,
Classification ,
Categorical, Integer ,
4000000 ,
42 ,
1999 ,
<b><a href="datasets/M.+Tuberculosis+Genes">M. Tuberculosis Genes</a></b>,
Relational ,
 ,
 ,
 ,
 ,
2001 ,
<b><a href="datasets/Movie">Movie</a></b>,,
Multivariate Relational
```

```
 ,
 ,
10000 ,
 ,
1999 ,
<b><a href="datasets/MSNBC.com+Anonymous+Web+Data">MSNBC.com Anonymous We
b Data</a></b>,
Sequential ,
 ,
Categorical ,
989818 ,
 ,
 ,
<b><a href="datasets/NSF+Research+Award+Abstracts+1990-2003">NSF Research Award
Abstracts 1990-2003</a></b>,
Text ,
 ,
 ,
129000 ,
 ,
2003 ,
<b><a href="datasets/Pioneer-1+Mobile+Robot+Data">Pioneer-1 Mobile Robot Data</a>
Multivariate, Time-Series ,
 ,
Categorical, Real ,
 ,
 ,
1999 ,
<b><a href="datasets/Pseudo+Periodic+Synthetic+Time+Series">Pseudo Periodic Synthetic
Time Series</a></b>,
Univariate, Time-Series ,
 ,
 ,
100000 ,
 ,
1999 ,
<b><a href="datasets/Reuters-21578+Text+Categorization+Collection">Reuters-21578 Text+Categorization+Collection">Reuters-21578 Text+Categorization+Collection">Reuters-21578 Text+Categorization+Collection</a>
t Categorization Collection</a></b>,
Text ,
Classification ,
Categorical ,
21578 ,
5 ,
1997 ,
<b><a href="datasets/Robot+Execution+Failures">Robot Execution Failures</a></b>,
Multivariate, Time-Series ,
Classification ,
Integer ,
463 ,
90 ,
1999 ,
<b><a href="datasets/Synthetic+Control+Chart+Time+Series">Synthetic Control Chart Time
Series</a></b>,
Time-Series ,
Classification, Clustering ,
Real ,
600 ,
 ,
1999 ,
<b><a href="datasets/Syskill+and+Webert+Web+Page+Ratings">Syskill and Webert Web P
and Ratings/as/hs/ns
```

```
ago riamigo \\ a \\ b \\ p \,
Multivariate, Text ,
Classification ,
Categorical ,
332 ,
5 ,
1998 ,
<b><a href="datasets/UNIX+User+Data">UNIX User Data</a></b>,
Text, Sequential ,
 ,
 ,
 ,
 ,
 ,
<b><a href="datasets/Volcanoes+on+Venus+-+JARtool+experiment">Volcanoes on Venus -
JARtool experiment</a></b>,
Image ,
...]
In [74]:
```

1998 ,

```
data_2 = data_[25:] #los elementos que me interesan
data 2
Out[74]:
[<b><a href="datasets/Abalone">Abalone</a></b>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
4177 ,
8 ,
1995 ,
<b><a href="datasets/Adult">Adult</a></b>,,
Multivariate ,
Classification ,
Categorical, Integer ,
48842 ,
14 ,
1996 ,
<b><a href="datasets/Annealing">Annealing</a></b>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
798 ,
38 ,
 ,
<b><a href="datasets/Anonymous+Microsoft+Web+Data">Anonymous Microsoft Web Data
</a></b>,
 ,
Recommender-Systems ,
Categorical ,
37711 ,
294 ,
1998 ,
<b><a href="datasets/Arrhythmia">Arrhythmia</a></b>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
452 ,
279 ,
```

Artificial Characters,

```
Multivariate ,
Classification ,
Categorical, Integer, Real ,
6000 ,
7 ,
1992 ,
<b><a href="datasets/Audiology+%28Original%29">Audiology (Original)</a></b>
Multivariate ,
Classification ,
Categorical ,
226 ,
 ,
1987 ,
<b><a href="datasets/Audiology+%28Standardized%29">Audiology (Standardized)</a></b>
,
Multivariate ,
Classification ,
Categorical ,
226 ,
69 ,
1992 ,
<b><a href="datasets/Auto+MPG">Auto MPG</a></b>,
Multivariate ,
Regression ,
Categorical, Real ,
398 ,
8 ,
1993 ,
<b><a href="datasets/Automobile">Automobile</a></b>,
Multivariate ,
Regression ,
Categorical, Integer, Real ,
205 ,
26 ,
1987 ,
<b><a href="datasets/Badges">Badges</a></b>,
Univariate, Text ,
Classification ,
 ,
294 ,
1 ,
1994 ,
<b><a href="datasets/Balance+Scale">Balance Scale</a></b>,
Multivariate ,
Classification ,
Categorical ,
625 ,
4 ,
1994 ,
<b><a href="datasets/Balloons">Balloons</a>,
Multivariate ,
Classification ,
Categorical ,
16 ,
4 ,
 ,
<b><a href="datasets/Breast+Cancer">Breast Cancer</a></b>,
Multivariate ,
Classification ,
Categorical ,
286 ,
9 ,
```

```
1988 ,
<b><a href="datasets/Breast+Cancer+Wisconsin+%28Original%29">Breast Cancer Wiscon
sin (Original)</a></b>,
Multivariate ,
Classification ,
Integer ,
699 ,
10 ,
1992 ,
<b><a href="datasets/Breast+Cancer+Wisconsin+%28Prognostic%29">Breast Cancer Wisc
onsin (Prognostic)</a></b>,
Multivariate ,
Classification, Regression ,
Real ,
198 ,
34 ,
1995 ,
<b><a href="datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29">Breast Cancer Wisc
onsin (Diagnostic)</a></b>,
Multivariate ,
Classification ,
Real ,
569 ,
32 ,
1995 ,
<b><a href="datasets/Pittsburgh+Bridges">Pittsburgh Bridges</a></b>,
Multivariate ,
Classification ,
Categorical, Integer ,
108 ,
13 ,
1990 ,
<b><a href="datasets/Car+Evaluation">Car Evaluation</a></b>,
Multivariate ,
Classification ,
Categorical ,
1728 ,
6 ,
1997 ,
<b><a href="datasets/Census+Income">Census Income</a></b>,
Multivariate ,
Classification ,
Categorical, Integer ,
48842 ,
14 ,
1996 ,
<b><a href="datasets/Chess+%28King-Rook+vs.+King-Knight%29">Chess (King-Rook vs.
King-Knight)</a></b>,
Multivariate, Data-Generator ,
Classification ,
Categorical, Integer ,
 ,
22 ,
1988 ,
<b><a href="datasets/Chess+%28King-Rook+vs.+King-Pawn%29">Chess (King-Rook vs. Ki
ng-Pawn)</a></b>,
Multivariate ,
Classification ,
Categorical ,
3196 ,
36 ,
1989 ,
```

```
<b><a href="datasets/Chess+%28King-Rook+vs.+King%29">Chess (King-Rook vs. King)</a>/a
></b>,
Multivariate ,
Classification ,
Categorical, Integer ,
28056 ,
6 ,
1994 ,
<b><a href="datasets/Chess+%28Domain+Theories%29">Chess (Domain Theories)</a></
Domain-Theory ,
 ,
 ,
 ,
 ,
 ,
<b><a href="datasets/Bach+Chorales">Bach Chorales</a></b>,
Univariate, Time-Series ,
 ,
Categorical, Integer ,
100 ,
6 ,
 ,
<b><a href="datasets/Connect-4">Connect-4</a></b>,
Multivariate, Spatial ,
Classification ,
Categorical ,
67557 ,
42 ,
1995 ,
<b><a href="datasets/Credit+Approval">Credit Approval</a>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
690 ,
15 ,
 ,
<b><a href="datasets/Japanese+Credit+Screening">Japanese Credit Screening</a></b></
p>,
Multivariate, Domain-Theory ,
Classification ,
Categorical, Real, Integer ,
125 ,
 ,
1992 ,
<b><a href="datasets/Computer+Hardware">Computer Hardware</a></b>,
Multivariate ,
Regression ,
Integer ,
209 ,
9 ,
1987 ,
<b><a href="datasets/Contraceptive+Method+Choice">Contraceptive Method Choice</a></</pre>
Multivariate ,
Classification ,
Categorical, Integer ,
1473 ,
9 ,
1997 ,
<b><a href="datasets/Covertype">Covertype</a></b>,
Multivariate .
```

```
Classification ,
Categorical, Integer ,
581012 ,
54 ,
1998 ,
<b><a href="datasets/Cylinder+Bands">Cylinder Bands</a></b>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
512 ,
39 ,
1995 ,
<b><a href="datasets/Dermatology">Dermatology</a></b>,
Multivariate ,
Classification ,
Categorical, Integer ,
366 ,
33 ,
1998 ,
<b><a href="datasets/Diabetes">Diabetes</a></b>,
Multivariate, Time-Series ,
 ,
Categorical, Integer ,
 ,
20 ,
 ,
<b><a href="datasets/DGP2+-+The+Second+Data+Generation+Program">DGP2 - The Sec
ond Data Generation Program</a></b>,
Data-Generator ,
 ,
Real ,
 ,
 ,
 ,
<b><a href="datasets/Document+Understanding">Document Understanding</a></b>,
 ,
 ,
 ,
 ,
 ,
1994 ,
<b><a href="datasets/EBL+Domain+Theories">EBL Domain Theories</a></b>,
 ,
 ,
 ,
 ,
 ,
 ,
<b><a href="datasets/Echocardiogram">Echocardiogram</a></b>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
132 ,
12 ,
1989 ,
<b><a href="datasets/Ecoli">Ecoli</a></b>,
Multivariate ,
Classification ,
Real ,
336 ,
8 ,
cn class="normal">1996
```

```
<b><a href="datasets/Flags">Flags</a></b>,,
Multivariate ,
Classification ,
Categorical, Integer ,
194 ,
30 ,
1990 ,
<b><a href="datasets/Function+Finding">Function Finding</a></b>,
 ,
Function-Learning ,
Real ,
352 ,
 ,
1990 ,
<b><a href="datasets/Glass+Identification">Glass Identification</a></b>
Multivariate ,
Classification ,
Real ,
214 ,
10 ,
1987 ,
<b><a href="datasets/Haberman%27s+Survival">Haberman's Survival</a></b>,
Multivariate ,
Classification ,
Integer ,
306 ,
3 ,
1999 ,
<b><a href="datasets/Hayes-Roth">Hayes-Roth</a></b>,
Multivariate ,
Classification ,
Categorical ,
160 ,
5 ,
1989 ,
<b><a href="datasets/Heart+Disease">Heart Disease</a></b>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
303 ,
75 ,
1988 ,
<b><a href="datasets/Hepatitis">Hepatitis</a></b>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
155 ,
19 ,
1988 ,
<b><a href="datasets/Horse+Colic">Horse Colic</a></b>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
368 ,
27 ,
1989 ,
<b><a href="datasets/ICU">ICU</a></b>,,
Multivariate, Time-Series ,
 ,
Real ,
 ,
n clace-"normal"> /n
```

```
\p \(\alpha \alpha \bar{\p} \)
 ,
<b><a href="datasets/Image+Segmentation">Image Segmentation</a></b>,
Multivariate ,
Classification ,
Real ,
2310 ,
19 ,
1990 ,
<b><a href="datasets/Internet+Advertisements">Internet Advertisements</a></b>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
3279 ,
1558 ,
1998 ,
<b><a href="datasets/lonosphere">lonosphere</a></b>,
Multivariate ,
Classification ,
Integer, Real ,
351 ,
34 ,
1989 ,
<b><a href="datasets/Iris">Iris</a></b>,
Multivariate ,
Classification ,
Real ,
150 ,
4 ,
1988 ,
<b><a href="datasets/ISOLET">ISOLET</a></b>,
Multivariate ,
Classification ,
Real ,
7797 ,
617 ,
1994 ,
<b><a href="datasets/Kinship">Kinship</a></b>,
Relational ,
Relational-Learning ,
Categorical ,
104 ,
12 ,
1990 ,
<b><a href="datasets/Labor+Relations">Labor Relations</a></b>,
Multivariate ,
 ,
Categorical, Integer, Real ,
57 ,
16 ,
1988 ,
<b><a href="datasets/LED+Display+Domain">LED Display Domain</a></b>,
Multivariate, Data-Generator ,
Classification ,
Categorical ,
 ,
7 ,
1988 ,
<b><a href="datasets/Lenses">Lenses</a></b>,
Multivariate ,
Classification ,
Categorical ,
```

```
24 
4 ,
1990 ,
<b><a href="datasets/Letter+Recognition">Letter Recognition</a></b>,
Multivariate ,
Classification ,
Integer ,
20000 ,
16 ,
1991 ,
<b><a href="datasets/Liver+Disorders">Liver Disorders</a></b>,
Multivariate ,
 ,
Categorical, Integer, Real ,
345 ,
7 ,
1990 ,
<b><a href="datasets/Logic+Theorist">Logic Theorist</a></b>,
Domain-Theory ,
 ,
 ,
 ,
 ,
 ,
<b><a href="datasets/Lung+Cancer">Lung Cancer</a></b>,
Multivariate ,
Classification ,
Integer ,
32 ,
56 ,
1992 ,
<b><a href="datasets/Lymphography">Lymphography</a></b>,
Multivariate ,
Classification ,
Categorical ,
148 ,
18 ,
1988 ,
<b><a href="datasets/Mechanical+Analysis">Mechanical Analysis</a></b>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
209 ,
8 ,
1990 ,
<b><a href="datasets/Meta-data">Meta-data</a></b>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
528 ,
22 ,
1996 ,
<b><a href="datasets/Mobile+Robots">Mobile Robots</a></b>,
Domain-Theory ,
 ,
Categorical, Integer, Real ,
 ,
 ,
1995 ,
<b><a href="datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29">Molecular
Biology (Promoter Gene Sequences)</a></b>,
Sequential, Domain-Theory ,
```

```
Classification ,
Categorical ,
106 ,
58 ,
1990 ,
<b><a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular+Biology+%28Protein+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+Secondary+
r Biology (Protein Secondary Structure)</a></b>,
Sequential ,
Classification ,
Categorical ,
128 ,
 ,
 ,
<b><a href="datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29">Mole
cular Biology (Splice-junction Gene Sequences)</a></b>,
Sequential, Domain-Theory ,
Classification ,
Categorical ,
3190 ,
61 ,
1992 ,
<b><a href="datasets/MONK%27s+Problems">MONK's Problems</a></b>,
Multivariate ,
Classification ,
Categorical ,
432 ,
7 ,
1992 ,
<b><a href="datasets/Moral+Reasoner">Moral Reasoner</a></b>,
Domain-Theory ,
 ,
 ,
202 .
 ,
1994 ,
<b><a href="datasets/Multiple+Features">Multiple Features</a></b>,
Multivariate ,
Classification ,
Integer, Real ,
2000 ,
649 ,
 ,
<b><a href="datasets/Mushroom">Mushroom</a></b>,,
Multivariate ,
Classification ,
Categorical ,
8124 ,
22 ,
1987 ,
<b><a href="datasets/Musk+%28Version+1%29">Musk (Version 1)</a></b>,
Multivariate ,
Classification ,
Integer ,
476 ,
168 ,
1994 ,
<b><a href="datasets/Musk+%28Version+2%29">Musk (Version 2)</a></b>,
Multivariate ,
Classification ,
Integer ,
6598 ,
168 ,
```

```
1994 ,
<b><a href="datasets/Nursery">Nursery</a></b>,
Multivariate ,
Classification ,
Categorical ,
12960 ,
8 ,
1997 ,
<b><a href="datasets/Othello+Domain+Theory">Othello Domain Theory</a></b>,
Domain-Theory ,
 ,
 ,
 ,
 ,
1991 ,
<b><a href="datasets/Page+Blocks+Classification">Page Blocks Classification</a></b>,
Multivariate ,
Classification ,
Integer, Real ,
5473 ,
10 ,
1995 ,
<b><a href="datasets/Optical+Recognition+of+Handwritten+Digits">Optical Recognition of
Handwritten Digits</a></b>,
Multivariate ,
Classification ,
Integer ,
5620 ,
64 ,
1998 ,
<b><a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits">Pen-Based Recogni
tion of Handwritten Digits</a></b>,
Multivariate ,
Classification ,
Integer ,
10992 ,
16 ,
1998 ,
<b><a href="datasets/Post-Operative+Patient">Post-Operative Patient</a></b>,
Multivariate ,
Classification ,
Categorical, Integer ,
90 ,
8 ,
1993 ,
<b><a href="datasets/Primary+Tumor">Primary Tumor</a></b>,
Multivariate ,
Classification ,
Categorical ,
339 ,
17 ,
1988 ,
<b><a href="datasets/Prodigy">Prodigy</a></b>,
Domain-Theory ,
 ,
 ,
 ,
 ,
 ,
<b><a href="datasets/Qualitative+Structure+Activity+Relationships">Qualitative Structure A
ctivity Relationships</a></b>,
Domain-Theory ,
```

```
 ,
 ,
 ,
 ,
 ,
<b><a href="datasets/Quadruped+Mammals">Quadruped Mammals</a></b>,
Multivariate, Data-Generator ,
Classification ,
Real ,
 ,
72 ,
1992 ,
<b><a href="datasets/Servo">Servo</a></b>,
Multivariate ,
Regression ,
Categorical, Integer ,
167 ,
4 ,
1993 ,
<b><a href="datasets/Shuttle+Landing+Control">Shuttle Landing Control</a></b>,
Multivariate ,
Classification ,
Categorical ,
15 ,
6 ,
1988 ,
<b><a href="datasets/Solar+Flare">Solar Flare</a></b>,
Multivariate ,
Regression ,
Categorical ,
1389 ,
10 ,
1989 ,
<b><a href="datasets/Soybean+%28Large%29">Soybean (Large)</a></b>,
Multivariate ,
Classification ,
Categorical ,
307 ,
35 ,
1988 ,
<b><a href="datasets/Soybean+%28Small%29">Soybean (Small)</a></b>,
Multivariate ,
Classification ,
Categorical ,
47 ,
35 ,
1987 ,
<b><a href="datasets/Challenger+USA+Space+Shuttle+O-Ring">Challenger USA Space Sh
uttle O-Ring</a></b>,
Multivariate ,
Regression ,
Integer ,
23 ,
4 ,
1993 ,
<b><a href="datasets/Low+Resolution+Spectrometer">Low Resolution Spectrometer</a></</pre>
b>,
Multivariate ,
Classification ,
Integer, Real ,
531 ,
102 ,
```

```
1988 ,
<b><a href="datasets/Spambase">Spambase</a></b>,
Multivariate ,
Classification ,
Integer, Real ,
4601 ,
57 ,
1999 ,
<b><a href="datasets/SPECT+Heart">SPECT Heart</a></b>,
Multivariate ,
Classification ,
Categorical ,
267 ,
22 ,
2001 ,
<b><a href="datasets/SPECTF+Heart">SPECTF Heart</a></b>,
Multivariate ,
Classification ,
Integer ,
267 ,
44 ,
2001 ,
<b><a href="datasets/Sponge">Sponge</a></b>,
Multivariate ,
Clustering ,
Categorical, Integer ,
76 ,
45 ,
 ,
<b><a href="datasets/Statlog+Project">Statlog Project</a>,
 ,
 ,
 ,
 ,
 ,
1992 ,
<b><a href="datasets/Student+Loan+Relational">Student Loan Relational</a></b>,
Domain-Theory ,
 ,
 ,
1000 ,
 ,
1993 ,
<b><a href="datasets/Teaching+Assistant+Evaluation">Teaching Assistant Evaluation</a><
b 
Multivariate ,
Classification ,
Categorical, Integer ,
151 ,
5 ,
1997 ,
<b><a href="datasets/Tic-Tac-Toe+Endgame">Tic-Tac-Toe Endgame</a></b>,
Multivariate ,
Classification ,
Categorical ,
958 ,
9 ,
1991 ,
<b><a href="datasets/Thyroid+Disease">Thyroid Disease</a></b>,
Multivariate, Domain-Theory ,
Classification ,
Categorical, Real ,
```

```
7200 ,
21 ,
1987 ,
<b><a href="datasets/Trains">Trains</a></b>,
Multivariate ,
Classification ,
Categorical ,
10 ,
32 ,
1994 ,
<b><a href="datasets/University">University</a></b>,
Multivariate ,
Classification ,
Categorical, Integer ,
285 ,
17 ,
1988 ,
<b><a href="datasets/Congressional+Voting+Records">Congressional Voting Records</a>
</b>,
Multivariate ,
Classification ,
Categorical ,
435 ,
16 ,
1987 ,
<b><a href="datasets/Water+Treatment+Plant">Water Treatment Plant</a></b>,
Multivariate ,
Clustering ,
Integer, Real ,
527 ,
38 ,
1993 ,
<b><a href="datasets/Waveform+Database+Generator+%28Version+1%29">Waveform Dat
abase Generator (Version 1)</a></b>,
Multivariate, Data-Generator ,
Classification ,
Real ,
5000 ,
21 ,
1988 ,
<b><a href="datasets/Waveform+Database+Generator+%28Version+2%29">Waveform Dat
abase Generator (Version 2)</a></b>,
Multivariate, Data-Generator ,
Classification ,
Real ,
5000 ,
40 ,
1988 ,
<b><a href="datasets/Wine">Wine</a></b>,,
Multivariate ,
Classification ,
Integer, Real ,
178 ,
13 ,
1991 ,
<b><a href="datasets/Yeast">Yeast</a></b>,,
Multivariate ,
Classification ,
Real ,
1484 ,
8 ,
1996
```

```
<b><a href="datasets/Zoo">Zoo</a></b>,
Multivariate ,
Classification ,
Categorical, Integer ,
101 ,
17 ,
1990 ,
<b><a href="datasets/Undocumented">Undocumented</a></b>,
 ,
 ,
 ,
 ,
 ,
 ,
<b><a href="datasets/Twenty+Newsgroups">Twenty Newsgroups</a></b>,
Text ,
 ,
 ,
20000 ,
 ,
1999 ,
<b><a href="datasets/Australian+Sign+Language+signs">Australian Sign Language signs
a></b>,
Multivariate, Time-Series ,
Classification ,
Categorical, Real ,
6650 ,
15 ,
1999 ,
Sign Language signs (High Quality)</a></b>,
Multivariate, Time-Series ,
Classification ,
Real ,
2565 ,
22 ,
2002 ,
<b><a href="datasets/US+Census+Data+%281990%29">US Census Data (1990)</a></b>
Multivariate ,
Clustering ,
Categorical ,
2458285 ,
68 ,
 ,
<b><a href="datasets/Census-Income+%28KDD%29">Census-Income (KDD)</a></b>,
Multivariate ,
Classification ,
Categorical, Integer ,
299285 ,
40 ,
2000 ,
<b><a href="datasets/Coil+1999+Competition+Data">Coil 1999 Competition Data</a></b>
Multivariate ,
 ,
Categorical, Real ,
340 ,
17 ,
1999 ,
<b><a href="datasets/Corel+Image+Features">Corel Image Features</a></b>,
n class="normal">Multivariate //n>
```

```
 ,
Real ,
68040 ,
89 ,
1999 ,
<b><a href="datasets/E.+Coli+Genes">E. Coli Genes</a></b>,
Relational ,
 ,
 ,
 ,
 ,
2001 ,
<b><a href="datasets/EEG+Database">EEG Database</a></b>,
Multivariate, Time-Series ,
 ,
Categorical, Integer, Real ,
122 ,
4 ,
1999 ,
<b><a href="datasets/EI+Nino">EI Nino</a></b>,
Spatio-temporal ,
 ,
Integer, Real ,
178080 ,
12 ,
1999 ,
<b><a href="datasets/Entree+Chicago+Recommendation+Data">Entree Chicago Recomm
endation Data</a></b>,
Transactional, Sequential ,
Recommender-Systems ,
Categorical ,
50672 ,
 ,
2000 ,
<b><a href="datasets/CMU+Face+Images">CMU Face Images</a></b>,
Image ,
Classification ,
Integer ,
640 ,
 ,
1999 ,
<b><a href="datasets/Insurance+Company+Benchmark+%28COIL+2000%29">Insurance C
ompany Benchmark (COIL 2000)</a></b>,
Multivariate ,
Regression, Description ,
Categorical, Integer ,
9000 ,
86 ,
2000 ,
<b><a href="datasets/Internet+Usage+Data">Internet Usage Data</a></b>,
Multivariate ,
 ,
Categorical, Integer ,
10104 ,
72 ,
1999 ,
<b><a href="datasets/IPUMS+Census+Database">IPUMS Census Database</a>,
Multivariate ,
 ,
Categorical, Integer ,
256932 ,
 alaca "narmal", 61 /n
```

```
01 ,
1999 ,
<b><a href="datasets/Japanese+Vowels">Japanese Vowels</a>,
Multivariate, Time-Series ,
Classification ,
Real ,
640 ,
12 ,
 ,
<b><a href="datasets/KDD+Cup+1998+Data">KDD Cup 1998 Data</a></b>,
Multivariate ,
Regression ,
Categorical, Integer ,
191779 ,
481 ,
1998 ,
<b><a href="datasets/KDD+Cup+1999+Data">KDD Cup 1999 Data</a></b>,
Multivariate ,
Classification ,
Categorical, Integer ,
4000000 ,
42 ,
1999 ,
<b><a href="datasets/M.+Tuberculosis+Genes">M. Tuberculosis Genes</a></b>,
Relational ,
 ,
 ,
 ,
 ,
2001 ,
<b><a href="datasets/Movie">Movie</a>,
Multivariate, Relational ,
 ,
 ,
10000 ,
 ,
1999 ,
<b><a href="datasets/MSNBC.com+Anonymous+Web+Data">MSNBC.com Anonymous We
b Data</a></b>,
Sequential ,
 ,
Categorical ,
989818 ,
 ,
 ,
<b><a href="datasets/NSF+Research+Award+Abstracts+1990-2003">NSF Research Award
Abstracts 1990-2003</a></b>,
Text ,
 ,
 ,
129000 ,
 ,
2003 ,
<b><a href="datasets/Pioneer-1+Mobile+Robot+Data">Pioneer-1 Mobile Robot Data</a>
Multivariate, Time-Series ,
 ,
Categorical, Real ,
 ,
 ,
1999 ,
<b><a href="datasets/Pseudo+Periodic+Synthetic+Time+Series">Pseudo Periodic Synthetic
```

```
nine Series</a></b>,
Univariate, Time-Series ,
 ,
 ,
100000 ,
 ,
1999 ,
<b><a href="datasets/Reuters-21578+Text+Categorization+Collection">Reuters-21578 Tex
t Categorization Collection</a></b>,
Text ,
Classification ,
Categorical ,
21578 ,
5 ,
1997 ,
<b><a href="datasets/Robot+Execution+Failures">Robot Execution Failures</a></b>,
Multivariate, Time-Series ,
Classification ,
Integer ,
463 ,
90 ,
1999 ,
<b><a href="datasets/Synthetic+Control+Chart+Time+Series">Synthetic Control Chart Time
Series</a></b>,
Time-Series ,
Classification, Clustering ,
Real ,
600 ,
 ,
1999 ,
<b><a href="datasets/Syskill+and+Webert+Web+Page+Ratings">Syskill and Webert Web P
age Ratings</a></b>,
Multivariate, Text ,
Classification ,
Categorical ,
332 ,
5 ,
1998 ,
<b><a href="datasets/UNIX+User+Data">UNIX User Data</a></b>,
Text, Sequential ,
 ,
 ,
 ,
 ,
 ,
<b><a href="datasets/Volcanoes+on+Venus+-+JARtool+experiment">Volcanoes on Venus -
JARtool experiment</a></b>,
Image ,
Classification ,
 ,
 ,
 ,
 ,
<b><a href="datasets/Statlog+%28Australian+Credit+Approval%29">Statlog (Australian Cre
dit Approval)</a></b>,
Multivariate ,
Classification ,
Categorical, Integer, Real ,
690 ,
14 ,
 ,
<b><a href="datasets/Statlog+%28German+Credit+Data%29">Statlog (German Credit Data
```

```
)</a></b>

class="normal">Multivariate ,

class="normal">Classification ,

class="normal">Categorical, Integer ,

class="normal">1000 ,

class="normal">20 ,

class="normal">1994 ,

class="normal">20 ,

class="normal">20 ,

class="normal">

class="normal">

class="normal">Multivariate ,

class="normal">

class="normal">

Categorical, Real ,

class="normal">

class="normal">
```

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, Data-Generator ,
Multivariate ,
Multivariate ,
Domain-Theory ,
Univariate, Time-Series ,
```

Multivariate, Spatial ,

Multivariate, Domain-Theory ,

Multivariate, Time-Series ,

Data-Generator ,

,
 ,

Multivariate ,

Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,
Multivariate ,

```
iviullivariale 
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 ,
 ,
Domain-Theory ,
Multivariate ,
Multivariate ,
```

```
Multivariate, Domain- i neory ,
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 ,
 ,
Domain-Theory ,
 ,
Multivariate ,
Multivariate, Time-Series ,
Multivariate, Time-Series ,
Multivariate ,
Multivariate ,
Multivariate, Sequential ,
Multivariate ,
Multivariate ,
Text ,
```

```
 I ext ,
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]:

```
[Classification ,
Classification ,
Classification ,
Recommender-Systems ,
Classification ,
Classification ,
Classification .
Classification ,
Regression ,
Regression ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification, Regression ,
Classification ,
 ,
 ,
Classification ,
Classification .
Classification ,
Regression ,
Classification ,
Classification ,
Classification ,
Classification ,
 ,
 ,
 ,
 ,
Classification ,
Classification ,
Classification ,
Function-Learning ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
 ,
Classification ,
Classification ,
Classification ,
Classification ,
Classification ,
```

Relational-Learning ,

```
 ,
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 ,
Regression ,
Classification ,
Clustering ,
Regression ,
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 ,
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, 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 ,
<n class="normal">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 ,
n clace_"normal"> Claccification //n
```

```
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, Regression ,
Classification ,
Regression ,
Classification ,
Classification, Clustering ,
Classification, Clustering ,
Classification, Regression ,
Regression ,
Clustering ,
Classification ,
Regression ,
```

```
Glustering ,
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 ,
n clace="normal">Real /n>
```

```
Real ,
Real ,
 ,
Real ,
Integer ,
Integer, Real ,
 ,
Real ,
 ,
 ,
Integer, Real ,
Integer, Real ,
Real ,
 ,
Real ,
Integer, 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 ,
alaga "narmal", Intag
```

```
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, Reai ,
 ,
 ,
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]:

```
798 .
37711 ,
452 ,
6000 ,
226 ,
226 ,
398 ,
205 ,
294 ,
625 ,
16 ,
286 ,
699 ,
198 ,
569 ,
108 ,
1728 ,
48842 ,
 ,
3196 ,
28056 ,
 ,
100 ,
67557 ,
690 ,
cn class="normal">125
```

[4177 ,
 48842 ,

```
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 ,
n clace_"normal"> 22 /n
```

```
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 ,

- <p class="normal">13910 </p>,
- 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 ,
- c| c|ass="normal">2584 ,
- 1372 ,
- 306 ,
- 120000 ,
- 13910 ,
- 2747 ,
- 3395 ,
- 39242 ,
- 4137 ,
- 17389 ,
- 51 ,
- 470 ,
- 132 ,
- 5000000 ,
- 11000000 ,
- 250 ,
- 126 ,
- ,
- 4889 ,
- class="normal"> ,

```
 ,
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 ,
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 ,
- to class= Hormal >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 ,
- 10 ,
 35 ,
- 35 ,
- 4 ,
- 102 ,
- 57 ,
- 22 ,
- 44 ,
- 45 ,
- ,
- ,
- 5 ,
- 9 ,
- 21 ,
- 32 ,
- 17 ,
- 16 ,
 38 ,
- 21 ,
- 2 ι ,
- 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 ,
class="normal">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]:

```
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 ,
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 ,
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 ,
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 ,
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 ,
2015 ,
2014 ,
2014 ,
2015 ,
2015 ,
n class="normal">2015 //n>
```

```
2015 ,
2015 ,
2015 ,
2015 ,
2015 ,
2014 ,
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 ,
alaca "narmal", 2017 //
```

```
201/ ,
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 ,
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', "),
('Domain', "),
('Univariate', "),
('Multivariate', ''),
('Multivariate', "),
('Data', "),
(", '\xa0'),
(", '\xa0'),
('Multivariate', "),
('Multivariate', "),
('Multivariate', "),
(", '\xa0'),
('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', "),
(", '\xa0'),
('Domain', "),
('Multivariate', "),
(", '\xa0'),
('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', "),
(", '\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', "),
('Text', "),
('Multivariate', "),
('Multivariate', "),
('Multivariate', "),
('Multivariate', "),
('Univariate', "),
('Multivariate', "),
('N Audtivariata'
```

```
iviuitivariate, j,
('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', "),
('Multivariate', "),
('Multivariate', "),
('Multivariate', "),
('Multivariate', ''),
('Multivariate', "),
('Multivariate', "),
('Multivariate', "),
```

```
( Mullivariale, ),
('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', "),
```

```
('lext', ''),
('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', "),
('Time', "),
('Multivariate', "),
('Multivariate', ''),
('Multivariate', "),
('Multivariate', "),
('Multivariate', "),
('Text', "),
('Multivariate', "),
('Multivariate', ''),
('Multivariate', "),
('Text', "),
('Multivariate', "),
('Multivariate', "),
('Multivariate', "),
('Multivariate', ''),
('Text', "),
('Multivariate', "),
('Multivariate', "),
('Multivariate', "),
('Multivariate', "),
('Multivariate', "),
('Multivariate', ''),
('Multivariate', "),
('Multivariate', "),
(", '\xa0'),
('Multivariate', "),
('Multivariate', "),
('Multivariate', ''),
('Sequential', "),
('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',
'Data',
'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',
'Domain',
'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',
'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', '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', 'Multivariata'

with the same of 'Multivariate', 'Time', 'Multivariate', 'Multivariate', 'Multivariate', 'Multivariate', 'Multivariate', 'Text', 'Multivariate', 'Multivariate', 'Multivariate', 'Text', 'Multivariate', 'Multivariate', 'Multivariate', 'Multivariate', 'Text', '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',

```
iviuitivariate,
'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', "),
(", '\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', "),
(", '\xa0'),
('Classification', "),
('Classification', "),
(", '\xa0'),
(", '\xa0'),
(", '\xa0'),
(", '\xa0'),
('', '\xa0'),
('Classification', "),
('Classification', "),
('Classification', "),
('Classification' ")
```

```
('Regression', "),
('Classification', ''),
('Clustering', "),
('Regression', "),
('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', "),
('Regression', "),
('Classification', "),
(", '\xa0'),
('Regression', "),
('Classification', ''),
('Classification', "),
('Classification', "),
('Regression', "),
('Classification', "),
('Classification', ''),
('Classification', "),
('Classification', "),
('Regression', "),
('Classification', "),
('Classification', "),
('Classification', "),
('Classification', ''),
('Classification', "),
('Classification', "),
('Classification', ''),
('Classification', "),
('Classification', "),
('Classification', "),
('Classification', "),
   localfication!
```

```
Ciassilication, j,
('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', "),
('Classification', "),
('Clustering', "),
('Classification', "),
('Classification', "),
('Classification', "),
('Regression', "),
('Classification', "),
('Classification', "),
('Classification', "),
('Classification', "),
('Classification', "),
('Classification', "),
('Regression', "),
('Classification', "),
('Regression', "),
('Regression', "),
(", '\xa0'),
('Classification', "),
(", '\xa0'),
('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', "),
('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', "),
('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', "),
('Recommendation', "),
('Classification', "),
('Classification', "),
('Classification', "),
('Regression', "),
('Classification', "),
('Regression', "),
('Classification', "),
('Classification', "),
('Classification', ''),
('Classification', "),
('Regression', "),
('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',
'Regression',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
'Function',
'Classification',
'Classification',
'Classification',
'Classification',
'Classification',
```

'Classification',

'Classification',

```
Ciassilication
'Classification',
'Classification',
'Classification',
'Relational',
'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',
'Classification',
'Regression',
'Classification',
'Regression',
'Classification',
'Regression',
'Classification',
'Clustering',
'Regression',
'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',
'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', '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', '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', 'Regression', 'Classification', 'Classification', 'Classification', 'Classification', 'Classification', 'Clustering', 'Clustering', '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 [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', "),
(", '\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', "),
```

```
iiilegei, ),
(", '\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',
'Integer',
'Real',
'Real',
'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',
۳,
'Real',
'Real',
'Real',
'Real',
'Real',
'Categorical',
'Integer',
'Real',
'Real',
'Real',
'Real',
'Integer',
'Integer',
'Real',
'Real',
'Real',
'Real',
'Real',
'Roal'
```

'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',

'Dool'

```
near,
",
'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', ''),
('425' ")
```

```
( TOO, j,
('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, ),
('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]:

```
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',
'<del>750</del>',
'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',
'7110'
```

```
'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', "),
```

```
('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', ''),
```

```
( ' ' ' , ' / ',
('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, ),
('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', "),
```

```
(81, "),
('12', "),
('6', "),
('22', "),
('18', "),
('754', "),
('14', "),
('5', "),
(", 'xa0'),
('18', "),
('7', "),
('7', "),
('7', "),
('7', "),
('6', "),
('4', "),
('18', "),
('11', "),
('11', "),
('11', "),
('11', "),
('11', "),
('11', "),
('15', "),
('11', "),
('11', "),
('11', "),
('11', "),
('15', "),
('15', "),
('11', "),
('11', "),
('11', "),
('15', "),
('15', "),
('15', "),
('11', "),
('15', "),
('11', "),
('15', "),
('11', "),
('15', "),
('15', "),
('11', "),
('15', "),
('11', "),
('15', "),
('15', "),
('11', "),
('15', "),
('15', "),
('15', "),
('11', "),
('15', "),
('15', "),
('15', "),
('11', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "),
('15', "
```

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',
'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',
'17',
",
'72',
'4',
'35',
'4',
'102'

'57', '22', '44', '45', '5',
'9',
'21',
'32',
'17',
'16',
'38',
'21',
'40',
'13',
'15',
'22',
'68',
'40',
'17', '89', '5',
",
'14',
'20',
'13', '36',
'19',
'9',
'18',
'4',
'60',
'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',
14 71
```

```
'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',
'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', "),

```
, \\\au /,
('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', "),
```

```
( , \xa∪),
('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'),
('1989', "),
('2006', "),
('2006', ''),
('2007', "),
('2007', "),
('2007', "),
('2007', ''),
('2008', "),
('2008', "),
('2008', "),
('2007', "),
('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', "),
('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', "),
('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', "),
('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', "),
('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', "),
('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', "),
('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', "),
('2018', "),
('2018', "),
('2018', "),
('2018', "),
('2018', "),
('2018', "),
('2018', "),
('2018', "),
('2018', "),
('2018', "),
('2018', "),
('2018', "),
('2018', "),
('2018', "),
('2019', "),
('2019', "),
('2018', "),
('2018', "),
('2018', "),
('2019', ")]
In [159]:
year_database_1=[]
```

Out[159]:

for element in year_database:

year_database_1.append(i)

i = element[0]

year_database_1

['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',
'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',		
'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',		
'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',		
'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',		
'2015',		
'2014',		
'2014',	,	
201 7 .		
'2015', '2015',		

'2015', '2015', '2015', '2015', '2015', '2014', '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',

```
2010,
'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]:
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-Book+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',
13661
```

```
'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',
 1401
```

```
ıυ,
'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'),
('lonosphere',
'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',
```

```
wullivariale,
'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',
'Toyt'
```

```
ι υλι,
'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',
```

```
wullivarial<del>e</del>
'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'),
('Callt2+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',
```

```
'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,
'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',
'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',
11001
```

```
100,
'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',
```

```
wullivarial<del>e</del>
'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',
'2158850'
```

```
'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',
'E00E4'
```

```
JZ0J4,
'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-Seg',
'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'),
```

2 . II. . 0/00ME00 . 0/00I

```
(Anuran+Galis+%28MFGGS%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 BaloT',
'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',
```

```
'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')]
```

"# Attributes", 'Year'])

df

Out[165]:

	Name	Data Type	Default Task	Attribute Type
0	Abalone	Multivariate	Classification	Multivariate
1	Adult	Multivariate	Classification	Multivariat€
2	Annealing	Multivariate	Classification	Multivariat€
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	Multivariat€
15	Breast+Cancer+Wisconsin+%28Prognostic%29	Multivariate	Classification	Multivariate
16	Breast+Cancer+Wisconsin+%28Diagnostic%29	Multivariate	Classification	Multivariat€
17	Pittsburgh+Bridges	Multivariate	Classification	Multivariate
18	Car+Evaluation	Multivariate	Classification	Multivariat€
19	Census+Income	Multivariate	Classification	Multivariat€
20	Chess+%28King-Rook+vs.+King-Knight%29	Multivariate	Classification	Multivariat€
21	Chess+%28King-Rook+vs.+King-Pawn%29	Multivariate	Classification	Multivariat€
22	Chess+%28King-Rook+vs.+King%29	Multivariate	Classification	Multivariat€
23	Chess+%28Domain+Theories%29	Domain		Domair
24	Bach+Chorales	Univariate		Univariat€
25	Connect-4	Multivariate	Classification	Multivariat€
26	Credit+Approval	Multivariate	Classification	Multivariat€
27	Japanese+Credit+Screening	Multivariate	Classification	Multivariat€
28	Computer+Hardware	Multivariate	Regression	Multivariat€
29	Contraceptive+Method+Choice	Multivariate	Classification	Multivariat€
438	Multimodal+Damage+Identification+for+Humanitar	Multivariate	Classification	Multivariat€
439	EEG+Steady-State+Visual+Evoked+Potential+Signals	Multivariate	Classification	Multivariate
440	Roman+Urdu+Data+Set	Text	Classification	Text
441	Avila	Multivariate	Classification	Multivariate

442	PANDOR Na me	Multivariate Data Type	Recommendation Default Task	Multivariate Attribute
443	Drug+Review+Dataset+%28Druglib.com%29	Multivariate	Classification	Multivar yat
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	${\sf GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2}$	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	Multivariat€

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]:

```
n<body>\n\n<!-- SITE HEADER (INCLUDES LOGO AND SEARCH BOX) -->\n\n<!-- SITE HEADER (INCLUDE
S LOGO AND SEARCH BOX) -->\n\n\n\n\t\n\t\t<span class="no"
rmal"><a href="index.html" alt="Home"><img src="assets/logo.gif" border=0></img></a><br/>br>&nbsp;&nbsp;&nbsp
p;      <a href="http://cml.ics.uci.edu"><font color="FFDD33">Center for M
achine Learning and Intelligent Systems</font></a></span>\n\t\n\t<td width=100% valign=top align="right"
>\n\t\t<span class="whitetext">\n\t\t<a href="about.html">About</a>&nbsp;\n\t\t<a href="citation_policy.html">Cit
ation Policy</a>&nbsp;\n\t\t<a href="donation_policy.html">Donate a Data Set</a>&nbsp;\n\t\t<a href="contact."
html">Contact</a>\n\t\t</pr>\n\t\t<br>\n\t\t<br>\n\t\t<!-- Search Google -->\n\n\t\t<FORM method=GET acti
on=http://www.google.com/custom onsubmit="return checkform(this);">\n\t\t<INPUT TYPE=text name=q size=3
0 maxlength=255 value="">\n\t\t<INPUT type=submit name=sa VALUE="Search">\n\t\t<INPUT type=hidden na
me=cof VALUE="AH:center;LH:130;L:http://archive.ics.uci.edu/assets/logo.gif;LW:384;AWFID:869c0b2eaa8d51
8e;">\n\t\t<input type=hidden name=domains value="ics.uci.edu">\n\t\t<br>\n\t\t<input type=radio name=sitesea
rch value="ics.uci.edu" checked> <span class="whitetext"><font size="1">Repository</font></span>\n\t\t<input t
ype=radio name=sitesearch value=""> <span class="whitetext"> <font size="1"> Web </font> </span> \n\t\t&nbsp;&
nbsp;       \n\t\t<A HREF=http://www.google.com/search><IM
G SRC=http://www.google.com/logos/Logo_25blk.gif border=0 ALT=Google align=middle height=27></A>\n\t\t<
br>\n\t\</FORM>\n\t\<!-- Search Google -->\n\n\n\t\<span class="whitetext"><a href="datasets.php"><font size
="3" color="#FFDD33"><b>View ALL Data Sets</b></font></a></span>\n\t\c/td>\n\n\n\n<b
r />\n\n\n\t\n\t\t\n\n\t\t\tBr
owse Through:\n\n\n\n\n\t\t<td bgcolor="#"
003366"><b>Default Task</b>\t\n\t\t\n\t<\tr>\n\t\n\t\t<p class=
"normal"><a href=\'datasets.php?format=&task=cla&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view
=list\'>Classification</a> <font color=red>(350)</font><br><a href=\'datasets.php?format=&task=reg&att=&area
=&numAtt=&numIns=&type=&sort=nameUp&view=list\'>Regression</a> <font color=red>(96)</font><br><a href
=\'datasets.php?format=&task=clu&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list\'>Clustering
</a> <font color=red>(84)</font><br><a href=\'datasets.php?format=&task=other&att=&area=&numAtt=&numIn
s=&type=&sort=nameUp&view=list\'>Other</a> <font color=red>(55)</font>\t\n\t\t\n\t
gcolor="#003366"><b>Attribute Type</b> \t\n\t\t\n\t\n\t\n\t<td valign=to
p><a href=\'datasets.php?format=&task=&att=cat&area=&numAtt=&numIns=&type=&sort=na
meUp&view=list\'>Categorical</a> <font color=red>(38)</font><br/>br><a href=\'datasets.php?format=&task=&att=n
um&area=&numAtt=&numIns=&type=&sort=nameUp&view=list\'>Numerical</a> <font color=red>(307)</font><br/>><br/>
r><a href=\'datasets.php?format=&task=&att=mix&area=&numAtt=&numIns=&type=&sort=nameUp&view=list\'>
\label{linear_model} \begin{tabular}{ll} Mixed </a > < font color = red > (55) </font > t  \n t  \n t  gcolor = "#003366" >  (55) 
<math display="block">\begin{tabular}{ll} Mixed </ds > (55)  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t  t < < td > t  t < < td > t  t < < td > t  t < < td > t  t < < td > t  t < < td > t < < t < < td > t < < td > t < < t < < t < < td > t < < 
xt"><b>Data Type</b> t/n/t</t>/n/t/n/t<a href=\'datasets.
php?format=&task=&att=&area=&numAtt=&numIns=&type=mvar&sort=nameUp&view=list\'>Multivariate</a> <fo
nt color=red>(357)</font><br><a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=uva
r&sort=nameUp&view=list\'>Univariate</a> <font color=red>(23)</font><br><a href=\'datasets.php?format=&tas
k=&att=&area=&numAtt=&numIns=&type=seq&sort=nameUp&view=list\'>Sequential</a> <font color=red>(47)</f
ont><br><a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=ts&sort=nameUp&view=li
st\'>Time-Series</a> <font color=red>(91)</font><br><a href=\'datasets.php?format=&task=&att=&area=&numA
tt=&numIns=&type=text&sort=nameUp&view=list\'>Text</a> <font color=red>(53)</font><br><a href=\'datasets.
php?format=&task=&att=&area=&numAtt=&numIns=&type=dt&sort=nameUp&view=list\'>Domain-Theory</a> <f
ont color=red>(23)</font><br><a href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=oth
er&sort=nameUp&view=list\'>Other</a> <font color=red>(21)</font><br>\t\n\t\n\t\n\t<td bgcol
or="#003366"><b>Area</b>\t\n\t\t\n\t<\tr>\n\t<p class="n
ormal"><a href=\'datasets.php?format=&task=&att=&area=life&numAtt=&numIns=&type=&sort=nameUp&view=li
st\'>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><br><
a href=\'datasets.php?format=&task=&att=&area=comp&numAtt=&numIns=&type=&sort=nameUp&view=list\'>C
S / Engineering</a> <font color=red>(170)</font><br><a href=\'datasets.php?format=&task=&att=&area=soc&n
umAtt=&numIns=&type=&sort=nameUp&view=list\'>Social Sciences</a> <font color=red>(26)</font><br><a href
=\'datasets.php?format=&task=&att=&area=bus&numAtt=&numIns=&type=&sort=nameUp&view=list\'>Business
</a> <font color=red>(29)</font><br><a href=\'datasets.php?format=&task=&att=&area=game&numAtt=&numIn
s=&type=&sort=nameUp&view=list\'>Game</a> <font color=red>(10)</font><br><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/ht/tont<<td>bgcolor="#003366"><b># Attributes</b>/t/n/t/t
\n\t\n\t\n\t<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><br/>ont>
<a href=\'datasets.php?format=&task=&att=&area=&numAtt=10to100&numIns=&type=&sort=nameUp&view=list\</p>
'>10 to 100</a> <font color=red>(210)</font><br><a href=\'datasets.php?format=&task=&att=&area=&numAtt=g
reater100&numIns=&type=&sort=nameUp&view=list\'>Greater than 100</a> <font color=red>(84)</font>\t\n
t/t/n/t<b># Instances</b>/p>/n/t//td>/n/t
```

```
\t\n\t\t\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><br/>ort><a href
=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=100to1000&type=&sort=nameUp&view=list\'>10
0 to 1000</a> <font color=red>(162)</font><br><a href=\'datasets.php?format=&task=&att=&area=&numAtt=&n
umIns=greater1000&type=&sort=nameUp&view=list\'>Greater than 1000</a> <font color=red>(246)</font>\t
\n\t\n\t\t<a href=\'datasets.php?format=mat&task=&att=&area=&num
Att=&numIns=&type=&sort=nameUp&view=list\'>Matrix</a> <font color=red>(324)</font><br><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\n\t\n\t\n\n\t\n\t
\n\t\t\n\t\t<b>469</b> Data Sets\n\t\t<a
href=\'datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table\'>Table
View</a>&nbsp;&nbsp;<font color=gray>List View</font>\n\t\t\n\t<tr
><hr>1. <b><a href="datasets/2.4+GHZ+Indoor+Channel+Measurements">2.4 GHZ Ind
oor Channel Measurements</a></b>: Measurement of the S21,consists of 10 sweeps, each sweep contains 60
1 frequency points with spacing of 0.167MHz to cover a 100MHz band centered at 2.4GHz.<p class="norm"
al">2. <b><a href="datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29">3D Road Network (No
rth Jutland, Denmark)</a></b>: 3D road network with highly accurate elevation information (+-20cm) from Den
mark used in eco-routing and fuel/Co2-estimation routing algorithms.3. <b><a href="dat
asets/AAAI+2013+Accepted+Papers">AAAI 2013 Accepted Papers</a></b>: This data set compromises the me
tadata for the 2013 AAAI conference\'s accepted papers (main track only), including paper titles, abstracts, and
keywords of varying granularity.4. <b><a href="datasets/AAAI+2014+Accepted+Papers"</p>
>AAAI 2014 Accepted Papers</a></b>: This data set compromises the metadata for the 2014 AAAI conference\
's accepted papers, including paper titles, authors, abstracts, and keywords of varying granularity.<p class=
"normal">5. <b><a href="datasets/Abalone">Abalone</a></b>: Predict the age of abalone from physical measu
rements6. <b><a href="datasets/Abscisic+Acid+Signaling+Network">Abscisic Acid Signa
ling Network</a></b>: 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 simulatio
ns using an asynchronous update scheme. 7. <b><a href="datasets/Absenteeism+at+w">4. <b href="datasets/Absenteeism+at+w">4. <
ork">Absenteeism at work</a></b>: The database was created with records of absenteeism at work from July 2
007 to July 2010 at a courier company in Brazil.8. <b><a href="datasets/Activities+of+D">b<a href="datasets/Activities+of+D">b<a href="datasets/Activities+of+D">datasets/Activities+of+D</a>
aily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors">Activities of Daily Living (ADLs) Recognition U
sing Binary Sensors</a></b>: This dataset comprises information regarding the ADLs performed by two users o
n a daily basis in their \r\nown homes. 
9. <b><a href="datasets/Activity+Recognition+fro">4 href="datasets/Activity+Recognition+fro
m+Single+Chest-Mounted+Accelerometer">Activity Recognition from Single Chest-Mounted Accelerometer</a>
</b>
The dataset collects data from a wearable accelerometer mounted on the chest. The dataset is intended f
or Activity Recognition research purposes.
class="normal">10. <b><a href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognitio">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognition">href="datasets/Activity+Recognit
n+system+based+on+Multisensor+data+fusion+%28AReM%29">Activity Recognition system based on Multisen
sor data fusion (AReM)</a></b>: This dataset contains temporal data from a Wireless Sensor Network worn by
an actor performing the activities: bending, cycling, lying down, sitting, standing, walking.<p class="normal"
>11. <b><a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sen">11. <b><a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sen">11. <b><a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sen">11. <a href="datasets/Activity+with+healthy+older+with+healthy+older+with+healthy+older+with+healthy+older+with+healthy+older+with+healthy+older+with+healthy+older+with+healthy+older+with+healthy+older+with+healthy+older+with+healthy+older+with+healthy+older+with+healthy+older+with+healthy+older+with+healthy+older+with+healthy+older+with+healthy+older+with+healthy+older+with+healthy+ol
sor">Activity recognition with healthy older people using a batteryless wearable sensor</a></b>: Sequential mot
ion data from 14 healthy older people aged 66 to 86 years old using a batteryless, wearable sensor on top of th
eir clothing for the recognition of activities in clinical environments.
12. <b><a href="data"><b><a href="data"><b><a href="data"><a 
sets/Acute+Inflammations">Acute Inflammations</a></b>: 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 sy
stem.\r\n13. <b><a href="datasets/Adult">Adult</a></b>: Predict whether income excee
ds $50K/yr based on census data. Also known as "Census Income" dataset.14. <b><a
href="datasets/Air+Quality">Air Quality</a></b>: 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. <b><a href="datasets/Air+quality">Air quality</a></b>: Co
ntains the responses of a gas multisensor device deployed on the field in an Italian city. 
16. <b><a href="datasets/Airfoil+Self-Noise">Airfoil Self-Noise</a></b>: 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. <b><a href="datasets/Amazon+Access+Samples">Amazon Access Sam
ples</a></b>: Amazon\'s InfoSec is getting smarter about the way Access data is leveraged. This is an anonymi
zed sample of access provisioned within the company.18. <b><a href="datasets/Amazo"><a href="datase
n+Commerce+reviews+set">Amazon Commerce reviews set</a></b>: The dataset is used for authorship identi
fication in online Writeprint which is a new research field of pattern recognition. 19. <b>
<a href="datasets/Annealing">Annealing</a></b>: Steel annealing data20. <b><a href="
datasets/Anonymous+Microsoft+Web+Data">Anonymous Microsoft Web Data</a></b>: Log of anonymous user
s of www.microsoft.com; predict areas of the web site a user visited based on data on other areas the user visit
```

```
ed.21. <b><a href="datasets/Anuran+Calls+%28MFCCs%29">Anuran Calls (MFCCs)
a></b>: Acoustic features extracted from syllables of anuran (frogs) calls, including the family, the genus, and t
he species labels (multilabel). 22. <b><a href="datasets/Appliances+energy+prediction"
>Appliances energy prediction</a></b>: Experimental data used to create regression models of appliances ene
rgy use in a low energy building.23. <b><a href="datasets/APS+Failure+at+Scania+Truc">4.5 cania+Truc</a>
ks">APS Failure at Scania Trucks</a></b>: The datasets\' positive class consists of component failures for a sp
ecific component of the APS system. The negative class consists of trucks with failures for components not relat
ed to the APS.24. <b><a href="datasets/Arcene">Arcene</a></b>: ARCENE\'s task is to
distinguish cancer versus normal patterns from mass-spectrometric data. This is a two-class classification probl
em with continuous input variables. This dataset is one of 5 datasets of the NIPS 2003 feature selection challen
ge.25. <b><a href="datasets/Arrhythmia">Arrhythmia</a></b>: Distinguish between the
presence and absence of cardiac arrhythmia and classify it in one of the 16 groups.
class="normal">26.
<b><a href="datasets/Artificial+Characters">Artificial Characters</a></b>: Dataset artificially generated by using
first order theory which describes structure of ten capital letters of English alphabet27. <</p>
b><a href="datasets/Audiology+%28Original%29">Audiology (Original)</a></b>: Nominal audiology dataset fro
m Baylor28. <b><a href="datasets/Audiology+%28Standardized%29">Audiology (Stand
ardized)</a></b>: Standardized version of the original audiology database29. <b><a hr
ef="datasets/Audit+Data">Audit Data</a></b>: Exhaustive one year non-confidential data in the year 2015 to 20
16 of firms is collected from the Auditor Office of India to build a predictor for classifying suspicious firms.
class="normal">30. <b><a href="datasets/Australian+Sign+Language+signs">Australian Sign Language signs</a>
a></b>: This data consists of sample of Auslan (Australian Sign Language) signs. Examples of 95 signs were c
ollected from five signers with a total of 6650 sign samples.
class="normal">31. <b><a href="datasets/Au</p>
stralian+Sign+Language+signs+%28High+Quality%29">Australian Sign Language signs (High Quality)</a></b>:
This data consists of sample of Auslan (Australian Sign Language) signs. 27 examples of each of 95 Auslan sign
ns were captured from a native signer using high-quality position trackers32. <b><a hre
f="datasets/Autism+Screening+Adult">Autism Screening Adult</a></b>: Autistic Spectrum Disorder Screening D
ata for Adult. This dataset is related to classification and predictive tasks.33. <b><a href
="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++">Autistic Spectrum Disorder Scre
ening Data for Adolescent </a></b>: Autistic Spectrum Disorder Screening Data for Adolescent. This dataset is
related to classification and predictive tasks.34. <b><a href="datasets/Autistic+Spectru">4. <br/>
**Boundard of the control of the co
m+Disorder+Screening+Data+for+Children++">Autistic Spectrum Disorder Screening Data for Children </a></b
>: Children screening data for autism suitable for classification and predictive tasks 35.
<b><a href="datasets/Auto+MPG">Auto MPG</a></b>: Revised from CMU StatLib library, data concerns city-cy
cle fuel consumption36. <b><a href="datasets/Automobile">Automobile</a></b>: From
1985 Ward\'s Automotive Yearbook37. <b><a href="datasets/AutoUniv">AutoUniv">AutoUniv</a></
b>: AutoUniv is an advanced data generator for classifications tasks. The aim is to reflect the nuances and hete
rogeneity of real data. Data can be generated in .csv, ARFF or C4.5 formats.38. <b><a
href="datasets/Avila">Avila</a></b>: 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 copyi
st.39. <b><a href="datasets/Bach+Choral+Harmony">Bach Choral Harmony</a></b>: T
he 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. <b
><a href="datasets/Bach+Chorales">Bach Chorales</a></b>: Time-series data based on chorales; challenge is
to learn generative grammar; data in Lisp41. <b><a href="datasets/Badges">Badges</a
></b>: Badges labeled with a "+" or "-" as a function of a person\'s name42. <b><a href
="datasets/Bag+of+Words">Bag of Words</a></b>: This data set contains five text collections in the form of ba
gs-of-words.43. <b><a href="datasets/Balance+Scale">Balance Scale</a></b>: Balance
scale weight & distance database44. <b><a href="datasets/Balloons">Balloons</a></b>
: Data previously used in cognitive psychology experiment; 4 data sets represent different conditions of an expe
riment45. <b><a href="datasets/Bank+Marketing">Bank Marketing</a></b>: The data is
related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification go
al is to predict if the client will subscribe a term deposit (variable y).46. <b><a href="dat
asets/banknote+authentication">banknote authentication</a></b>: Data were extracted from images that were t
aken for the evaluation of an authentication procedure for banknotes.
class="normal">47. <b><a href="d"><a 
atasets/BAUM-1">BAUM-1</a></b>: 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 men
tal states.48. <b><a href="datasets/BAUM-2">BAUM-2</a></b>: A multilingual audio-vis
ual affective face database consisting of 1047 video clips of 286 subjects. 
class="normal">49. <b><a hr</p>
ef="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></b>: The database was created with records of behavior of the urban traffi
c of the city of Sao Paulo in Brazil.50. <b><a href="datasets/Beijing+PM2.5+Data">Beijin
g PM2.5 Data</a></b>: 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. <b><a h
ref="datasets/Bike+Sharing+Dataset">Bike Sharing Dataset</a></b>: 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. <b><a href="datasets/BLE+RSSI+Dataset+for+Indoor+lo">4.52. <b href="datasets/BLE+RSSI+Dataset+for+lo">4.52. <b href="dataset-for+lo">4.52. <b href="dataset-for+lo">4.52. <b href="dataset-for+lo">4.52. <b href="dataset-for+lo">4.52. <b href="dataset-for-lo">4.52. <b href="dataset
calization+and+Navigation">BLE RSSI Dataset for Indoor localization and Navigation</a>
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.
class="normal">53. <b><a href="data"><b><a href="data"><b><a href="data"><a href="
sets/BlogFeedback">BlogFeedback</a></b>: 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.<p class="norm
al">54. <b><a href="datasets/BLOGGER">BLOGGER</a></b>: 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\nIran55.
<b><a href="datasets/Blood+Transfusion+Service+Center">Blood Transfusion Service Center</a></b>: Data ta
ken from the Blood Transfusion Service Center in Hsin-Chu City in Taiwan -- this is a classification problem. 
>56. <b><a href="datasets/Breast+Cancer">Breast Cancer</a></b>: Breast Cancer Data (R
estricted Access)57. <b><a href="datasets/Breast+Cancer+Coimbra">Breast Cancer Co
imbra</a></b>: Clinical features were observed or measured for 64 patients with breast cancer and 52 healthy
controls. 58. <b><a href="datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29">Bre
ast Cancer Wisconsin (Diagnostic)</a></b>: Diagnostic Wisconsin Breast Cancer Database<p class="norm"
al">59. <b><a href="datasets/Breast+Cancer+Wisconsin+%28Original%29">Breast Cancer Wisconsin (Original)
</a></b>: Original Wisconsin Breast Cancer Database60. <b><a href="datasets/Breast+">60. <b href=
Cancer+Wisconsin+%28Prognostic%29">Breast Cancer Wisconsin (Prognostic)</a></b>: Prognostic Wisconsin
Breast Cancer Database
class="normal">61. <b><a href="datasets/Breast+Tissue">Breast Tissue</a>
b>: Dataset with electrical impedance measurements of freshly excised tissue samples from the breast.
class="normal">62. <b><a href="datasets/BuddyMove+Data+Set">BuddyMove Data Set</a></b>: User interest
information extracted from user reviews published in holidayiq.com about various types of point of interests in S
outh India63. <b><a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attac">63. <b href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attac</b href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attac</b>
k+on+Optical+Burst+Switching+%28OBS%29+Network">Burst Header Packet (BHP) flooding attack on Optical
Burst Switching (OBS) Network</a></b>: 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. <b><a href="datasets/Buzz+in+social+media+">Buzz in social media </a></b>: 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. <b
><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. <b><a href="datasets/"
Callt2+Building+People+Counts">Callt2 Building People Counts</a></b>: This data comes from the main door
of the Callt2 building at UCI.67. <b><a href="datasets/Car+Evaluation">Car Evaluation</a>
/a></b>: Derived from simple hierarchical decision model, this database may be useful for testing constructive i
nduction and structure discovery methods.68. <b><a href="datasets/Carbon+Nanotube">b<a href="datasets/Carbon+Nanotube">datasets/Carbon+Nanotube</a>
s">Carbon Nanotubes</a></b>: This dataset contains 10721 initial and calculated atomic coordinates of carbon
nanotubes.69. <b><a href="datasets/Cardiotocography">Cardiotocography</a></b>: Th
e dataset consists of measurements of fetal heart rate (FHR) and uterine contraction (UC) features on cardioto
cograms classified by expert obstetricians.70. <b><a href="datasets/Cargo+2000+Freig">2000+Freig</a>
ht+Tracking+and+Tracing">Cargo 2000 Freight Tracking and Tracing</a></b>: 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). 
class="normal">71. <b><a href="datasets/Census+In">71. <b</a>
come">Census Income</a></b>: Predict whether income exceeds $50K/yr based on census data. Also known
as "Adult" dataset.72. <b><a href="datasets/Census-Income+%28KDD%29">Census-In
come (KDD)</a></b>: 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. <b><a href="datasets/Cerv"><b style="color: blue;">1.5 class="normal">73. <b><a href="datasets/Cerv"><a href="datasets
ical+cancer+%28Risk+Factors%29">Cervical cancer (Risk Factors)</a></b>: This dataset focuses on the predic
tion of indicators/diagnosis of cervical cancer. The features cover demographic information, habits, and historic
medical records.74. <b><a href="datasets/Challenger+USA+Space+Shuttle+O-Ring">C
hallenger USA Space Shuttle O-Ring</a></b>: 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
class="normal">75.
ref="datasets/Character+Font+Images">Character Font Images</a></b>: Character images from scanned and
computer generated fonts.76. <b><a href="datasets/Character+Trajectories">Character
Trajectories</a></b>: 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. <b><a href="datasets/Chess+%28Domain+Th
eories%29">Chess (Domain Theories)</a></b>: 6 different domain theories for generating legal moves of ches
s78, <b><a href="datasets/Chess+%28King-Book+vs+King%29">Chess (King-Book vs
```

```
King)</a></b>: Chess Endgame Database for White King and Rook against Black King (KRK).<p class="nor
mal">79. <b><a href="datasets/Chess+%28King-Rook+vs.+King-Knight%29">Chess (King-Rook vs. King-Knight
)</a></b>: Knight Pin Chess End-Game Database Creator80. <b><a href="datasets/Ch"><b><a href="datasets/Ch"><b style="color: blue;">b><a href="datasets/Ch">class="normal">80. <b><a href="datasets/Ch">b><a href="datasets/Ch">b</a>
ess+%28King-Rook+vs.+King-Pawn%29">Chess (King-Rook vs. King-Pawn)</a></b>: King+Rook versus King+
Pawn on a7 (usually abbreviated KRKPA7).81. <b><a href="datasets/chestnut+%E2%8">81. <b><a href="datasets/chestnut+%E2%8">81. <b><a href="datasets/chestnut+%E2%8">81. <b style="color: blue;">81. <b style=
0%93+LARVIC">chestnut – LARVIC</a></b>: The research project presents this database, shows the images o
f chestnuts that will be processed to determine the presence or absence of defects82. <
b><a href="datasets/chipseg">chipseg</a></b>: ChIP-seg experiments characterize protein modifications or bin
ding at\r\nspecific genomic locations in specific samples. The machine learning\r\nproblem in these data is stru
ctured binary classification.83. <b><a href="datasets/Chronic_Kidney_Disease">Chroni
c_Kidney_Disease</a></b>: This dataset can be used to predict the chronic kidney disease and it can be collec
ted from the hospital nearly 2 months of period.84. <b><a href="datasets/Climate+Mod
el+Simulation+Crashes">Climate Model Simulation Crashes</a></b>: Given Latin hypercube samples of 18 cli
mate model input parameter values, predict climate model simulation crashes and determine the parameter val
ue combinations that cause the failures.85. <b><a href="datasets/Cloud">Cloud</a></b
>: Little Documentation86. <b><a href="datasets/CMU+Face+Images">CMU Face Images</a>
es</a></b>: This data consists of 640 black and white face images of people taken with varying pose (straight, I
eft, right, up), expression (neutral, happy, sad, angry), eyes (wearing sunglasses or not), and size<p class=
"normal">87. <b><a href="datasets/CNAE-9">CNAE-9</a></b>: This is a data set containing 1080 documents o
f free text business descriptions of Brazilian companies categorized into a\r\nsubset of 9 categories<p class
="normal">88. <b><a href="datasets/Coil+1999+Competition+Data">Coil 1999 Competition Data</a></b>: This
data set is from the 1999 Computational Intelligence and Learning (COIL) competition. The data contains meas
urements of river chemical concentrations and algae densities.89. <b><a href="dataset">89. <b><a href="dataset">89. <b><a href="dataset">89. <b><a href="dataset">89. <b href="d
s/Combined+Cycle+Power+Plant">Combined Cycle Power Plant</a></b>: The dataset contains 9568 data point
s collected from a Combined Cycle Power Plant over 6 years (2006-2011), when the plant was set to work with
full load. 90. <b><a href="datasets/Communities+and+Crime">Communities and Crime
</a></b>: Communities within the United States. The data combines socio-economic data from the 1990 US Ce
nsus, law enforcement data from the 1990 US LEMAS survey, and crime data from the 1995 FBI UCR.
lass="normal">91. <b><a href="datasets/Communities+and+Crime+Unnormalized">Communities and Crime Un
normalized</a></b>: Communities in the US. Data combines socio-economic data from the \'90 Census, law en
forcement data from the 1990 Law Enforcement Management and Admin Stats survey, and crime data from the
e 1995 FBI UCR92. <b><a href="datasets/Computer+Hardware">Computer Hardware</
a></b>: Relative CPU Performance Data, described in terms of its cycle time, memory size, etc.<p class="n
ormal">93. <b><a href="datasets/Concrete+Compressive+Strength">Concrete Compressive Strength</a></b>:
Concrete is the most important material in civil engineering. The concrete compressive strength is a highly nonli
near function of age and ingredients. 94. <b><a href="datasets/Concrete+Slump+Test"
>Concrete Slump Test</a></b>: 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.<p class="normal"
>95. <b><a href="datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants">Condition Based Mai
ntenance of Naval Propulsion Plants</a></b>: Data have been generated from a sophisticated simulator of a G
as Turbines (GT), mounted on a Frigate characterized by a COmbined Diesel eLectric And Gas (CODLAG) pro
pulsion plant type.96. <b><a href="datasets/Condition+monitoring+of+hydraulic+system">96. <b><a href="datasets/Condition+monitoring+of+hydraulic+system">96. <b><a href="datasets/Condition+monitoring+of+hydraulic+system">96. <b><a href="datasets/Condition+monitoring+of+hydraulic+system">96. <b><a href="datasets/Condition+monitoring+of+hydraulic+system">96. <a href="datasets/Condition+monitoring+system">96. <a href="datasets/Condition+m
s">Condition monitoring of hydraulic systems</a></b>: The data set addresses the condition assessment of a h
ydraulic test rig based on multi sensor data. Four fault types are superimposed with several severity grades imp
eding selective quantification.97. <b><a href="datasets/Congressional+Voting+Records"
>Congressional Voting Records</a></b>: 1984 United Stated Congressional Voting Records; Classify as Repub
lican or Democrat98. <b><a href="datasets/Connect-4">Connect-4</a></b>: Contains c
onnect-4 positions
99. <b><a href="datasets/Connectionist+Bench+%28Nettalk+Corpus">98. <b href="datasets/Connect
%29">Connectionist Bench (Nettalk Corpus)</a></b>: The file "nettalk.data" contains a list of 20,008 English wo
rds, along with a phonetic transcription for each word. The task is to train a network to produce the proper phon
emes100. <b><a href="datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Roc
ks%29">Connectionist Bench (Sonar, Mines vs. Rocks)</a></b>: The task is to train a network to discriminate b
etween sonar signals bounced off a metal cylinder and those bounced off a roughly cylindrical rock.<p clas
s="normal">101. <b><a href="datasets/Connectionist+Bench+%28Vowel+Recognition+-+Deterding+Data%29">
Connectionist Bench (Vowel Recognition - Deterding Data)</b>
Speaker independent recognition of the element 
even steady state vowels of British English using a specified training set of lpc derived log area ratios.
ass="normal">102. <b><a href="datasets/Container+Crane+Controller+Data+Set">Container Crane Controller
Data Set</a></b>: A container crane has the function of transporting containers from one point to another point
 .103. <b><a href="datasets/Contraceptive+Method+Choice">Contraceptive Method Cho
ice</a></b>: Dataset is a subset of the 1987 National Indonesia Contraceptive Prevalence Survey.<p class
="normal">104. <b><a href="datasets/Corel+Image+Features">Corel Image Features</a></b>: This dataset co
 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-occurence105. <b><a h
ref="datasets/Covertype">Covertype</a></b>: Forest CoverType dataset106. <b><a hr
ef="datasets/Credit+Approval">Credit Approval</a></b>: This data concerns credit card applications; good mix
of attributes107. <b><a href="datasets/Crowdsourced+Mapping">Crowdsourced Mappi
ng</a></b>: Crowdsourced data from OpenStreetMap is used to automate the classification of satellite images i
nto different land cover classes (impervious, farm, forest, grass, orchard, water). 108. 
b><a href="datasets/Cryotherapy+Dataset+">Cryotherapy Dataset </a></b>: This dataset contains information
about wart treatment results of 90 patients using cryotherapy.109. <b><a href="dataset"><a href="d
s/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015">CSM (Conventional and
Social Media Movies) Dataset 2014 and 2015</a></b>: 12 features categorized as conventional and social med
ia features. Both conventional features, collected from movies databases on Web as well as social media featur
es(YouTube,Twitter).110. <b><a href="datasets/Cuff-Less+Blood+Pressure+Estimation"">110. <b><a href="datasets/Cuff-Less+Blood+Pressure+Estimation">110. <b href="datasets/Cuff-Less+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Blood+Bl
>Cuff-Less Blood Pressure Estimation</a></b>: This Data set provides preprocessed and cleaned vital signals
which can be used in designing algorithms for cuff-less estimation of the blood pressure.
class="normal"
>111. <b><a href="datasets/Cylinder+Bands">Cylinder Bands</a></b>: Used in decision tree induction for mitig
ating process delays known as "cylinder bands" in rotogravure printing112. <b><a href=
"datasets/Daily+and+Sports+Activities">Daily and Sports Activities</a></b>: The dataset comprises motion sens
or data of 19 daily and sports activities each performed by 8 subjects in their own style for 5 minutes. Five Xsen
s MTx units are used on the torso, arms, and legs.\r\n113. <b><a href="datasets/Daily+
Demand+Forecasting+Orders">Daily Demand Forecasting Orders</a></b>: The dataset was collected during 6
0 days, this is a real database of a brazilian logistics company.114. <b><a href="dataset"><a href
s/Daphnet+Freezing+of+Gait">Daphnet Freezing of Gait</a></b>: This dataset contains the annotated reading
s of 3 acceleration sensors at the hip and leg of Parkinson\'s disease patients that experience freezing of gait (
FoG) during walking tasks.\r\n115. <b><a href="datasets/Data+for+Software+Engineeri">115. <b href="datasets/Data+for+Bottware+Engineeri">115. <b href="datasets/Data+for+Bottware+Engin
ng+Teamwork+Assessment+in+Education+Setting">Data for Software Engineering Teamwork Assessment in E
ducation Setting</a></b>: 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,
FAU116. <b><a href="datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accele">+Recognition+with+Wrist-worn+Accele</a>
rometer">Dataset for ADL Recognition with Wrist-worn Accelerometer</a></b>: Recordings of 16 volunteers pe
rforming 14 Activities of Daily Living (ADL) while carrying a single wrist-worn tri-axial accelerometer.
s="normal">117. <b><a href="datasets/Dataset+for+Sensorless+Drive+Diagnosis">Dataset for Sensorless Driv
e Diagnosis</a></b>: Features are extracted from motor current. The motor has intact and defective componen
ts. This results in 11 different classes with different conditions. 118. <b><a href="datase"><a href="d
ts/DBWorld+e-mails">DBWorld e-mails</a></b>: It contains 64 e-mails which I have manually collected from DB
World mailing list. They are classified in: \'announces of conferences\' and \'everything else\'.<p class="nor
mal">119. <b><a href="datasets/default+of+credit+card+clients">default of credit card clients</a></b>: This res
earch aimed at the case of customers' default payments in Taiwan and compares the predictive accuracy of pr
obability of default among six data mining methods.
120. <b><a href="datasets/Delicious">><a href="datas
MIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Instance+Labels">DeliciousMIL: A Data S
et for Multi-Label Multi-Instance Learning with Instance Labels</a></b>: This dataset includes 1) 12234 docum
ents (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. <b><a href="datasets/Demospong">121. <b >121. <b >121
iae">Demospongiae</a></b>: Marine sponges of the Demospongiae class classification domain.<p class="
normal">122. <b><a href="datasets/Dermatology">Dermatology</a></b>: Aim for this dataset is to determine the
e type of Eryhemato-Squamous Disease.123. <b><a href="datasets/Detect+Malacious+">123. <b href="datasets/Detect+M
Executable%28AntiVirus%29">Detect Malacious Executable(AntiVirus)</a></b>: I extract features from malacio
us and non-malacious and create and training dataset to teach svm classifier. Dataset made of unknown execut
able to detect if it is virus or normal safe executable.124. <b><a href="datasets/detectio">+<a href="datasets
n_of_loT_botnet_attacks_N_BaloT">detection_of_loT_botnet_attacks_N_BaloT</a></b>: This dataset address
es the lack of public botnet datasets, especially for the IoT. It suggests *real* traffic data, gathered from 9 com
mercial IoT devices authentically infected by Mirai and BASHLITE.125. <b><a href="dat
asets/Devanagari+Handwritten+Character+Dataset">Devanagari Handwritten Character Dataset</a></b>: This
is an image database of Handwritten Devanagari characters. There are 46 classes of characters with 2000 exa
mples each. The dataset is split into training set(85%) and testing set(15%). 126. <b><a
href="datasets/Dexter">Dexter</a></b>: DEXTER is a text classification problem in a bag-of-word representation
n. This is a two-class classification problem with sparse continuous input variables. This dataset is one of five d
atasets of the NIPS 2003 feature selection challenge.\r\n127. <b><a href="datasets/DG"><br/>
P2+-+The+Second+Data+Generation+Program">DGP2 - The Second Data Generation Program</a></b>: Gen
erates application domains based on specific parameters, number of features, and proportion of positive to neg
ative examples128. <b><a href="datasets/Diabetes">Diabetes</a></b>: This diabetes d
ataset is from AIM \'94129. <b><a href="datasets/Diabetes+130-US+hospitals+for+year"
        1000 2000" Dishetos 120 LIC hasnitals for years 1000 2009 Jay Jby : This data has been n
```

```
5+1999-2000 >Diabetes 190-05 hospitals for years 1999-2000</a></b>
ze factors related to readmission as well as other \r\n\r\noutcomes pertaining to patients with diabetes.<p cl
ass="normal">130. <b><a href="datasets/Diabetic+Retinopathy+Debrecen+Data+Set">Diabetic Retinopathy De
brecen Data Set</a></b>: This dataset contains features extracted from the Messidor image set to predict whet
her an image contains signs of diabetic retinopathy or not. 131. <b><a href="datasets/D">131. <b hr
iscrete+Tone+Image+Dataset">Discrete Tone Image Dataset</a></b>: 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
p class="normal">132. <b><a href="datasets/Dishonest+Internet+users+Dataset">Dishonest Internet users Dat
aset</a></b>: The dataset was used to test an architecture based on a trust model capable to cope with the ev
aluation of the trustworthiness of users interacting in pervasive environments.
class="normal">133.
a href="datasets/Document+Understanding">Document Understanding</a></b>: Five concepts, expressed as
predicates, to be learned134. <b><a href="datasets/Dodgers+Loop+Sensor">Dodgers
Loop Sensor</a></b>: Loop sensor data was collected for the Glendale on ramp for the 101 North freeway in L
os Angeles135. <b><a href="datasets/Dorothea">Dorothea</a></b>: DOROTHEA is a d
rug 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.
/p>136. <b><a href="datasets/Dota2+Games+Results">Dota2 Games Results</a></b>: Dot
a 2 is a popular computer game with two teams of 5 players. At the start of the game each player chooses a uni
que hero with different strengths and weaknesses.
class="normal">137. <b><a href="datasets/Dow+Jon">137. <b><a href="datasets/Dow+Jon">137. <b><a href="datasets/Dow+Jon">137. <b><a href="datasets/Dow+Jon">137. <b><a href="datasets/Dow+Jon">137. <b href="datasets/Dow+Jon">137. <b
es+Index">Dow Jones Index</a></b>: This dataset contains weekly data for the Dow Jones Industrial Index. It
has been used in computational investing research.138. <b><a href="datasets/Dresses">4 href="datasets/Dresses"/datasets/Dresses"/datasets/Dresses
 Attribute Sales">Dresses Attribute Sales</a>></b>: This dataset contain Attributes of dresses and their recom
mendations according to their sales. Sales are monitor on the basis of alternate days. 1
39. <b><a href="datasets/DrivFace">DrivFace</a></b>: 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. <b><a href="datasets/Drug+consumption+%28qu">140. <b href="datasets/Drug+consum
antified%29">Drug consumption (quantified)</a></b>: Classify type of drug consumer by personality data
p class="normal">141. <b><a href="datasets/Drug+Review+Dataset+%28Druglib.com%29">Drug Review Datas
et (Druglib.com)</a></b>: 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. <b><a href="datasets/Drug+Review+Dataset+%28Drugs.com%29">Drug Review Da
taset (Drugs.com)</a></b>: The dataset provides patient reviews on specific drugs along with related condition
s and a 10 star patient rating reflecting overall patient satisfaction.
143. <b><a href="dat"><a href="dat
asets/DSRC+Vehicle+Communications">DSRC Vehicle Communications</a></b>: This set Provides data regar
ding wireless communications between vehicles and road side units. two separate data sets are provided (norm
al scenario) and in the presence of attacker (jammer).144. <b><a href="datasets/Dyna">+ datasets/Dyna</a>
mic+Features+of+VirusShare+Executables">Dynamic Features of VirusShare Executables</a></b>: This datas
et contains the dynamic features of 107,888 executables, collected by VirusShare from Nov/2010 to Jul/2014.</
p>145. <b><a href="datasets/E.+Coli+Genes">E. Coli Genes</a></b>: Data giving character
istics of each ORF (potential gene) in the E. coli genome. Sequence, homology (similarity to other genes) and s
tructural information, and function (if known) are provided.
class="normal">146. <b><a href="datasets/E">
arly+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connected+speech+Data+Set+">Ear
ly biomarkers of Parkinson's disease based on natural connected speech Data Set </a></b>: .<p class="no
rmal">147. <b><a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected">147. <b><a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected">147. <b><a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected">147. <b><a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected">147. <a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+biomarkers+of+Parkinson%92s+disease+biomarkers+of+Parkinson%92s+disease+biomarkers+of+Parkinson%92s+disease+biomarkers+of+Parkinson%92s+disease+biomarkers+of+Parkinson%92s+disease+biomarkers+of+Parkinson%92s+disease+biomarkers+of+Parkinson%92s+disease+biomarkers+of+Parkinson%92s+disease+biomarkers+of+Parkinson%92s+disease+biomarkers+of+Parkinson%92s+disease+biomarkers+of+Parkinso
+speech">Early biomarkers of Parkinson solutions based on natural connected speech</a></b>: Predict a pa
ttern of neurodegeneration in the dataset of speech features obtained from patients with early untreated Parkin
son's disease and patients at high risk developing Parkinson's disease.148. <b><a href
="datasets/EBL+Domain+Theories">EBL Domain Theories</a></b>: Assorted small-scale domain theories
149. <b><a href="datasets/Echocardiogram">Echocardiogram</a></b>: Data for classifying
if patients will survive for at least one year after a heart attack150. <b><a href="dataset"><a hr
s/Eco-hotel">Eco-hotel</a></b>: 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.151. <
b><a href="datasets/Ecoli">Ecoli</a></b>: This data contains protein localization sites1
52. <b><a href="datasets/Economic+Sanctions">Economic Sanctions</a></b>: Domain Theory on Economic S
anctions; Undocumented153. <b><a href="datasets/Educational+Process+Mining+%28"
EPM%29%3A+A+Learning+Analytics+Data+Set">Educational Process Mining (EPM): A Learning Analytics Data
Set</a></b>: 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. <b><a href="data"><b><a href="data"><a href="da
sets/EEG+Database">EEG Database</a></b>: This data arises from a large study to examine EEG correlates o
f genetic predisposition to alcoholism. It contains measurements from 64 electrodes placed on the scalp sample
d at 256 Hz155. <b><a href="datasets/EEG+Eye+State">EEG Eye State</a></b>: The d
ata set consists of 14 EEG values and a value indicating the eye state.156. <b><a href=
```

```
datasets/EEG+5teady-5tate+visual+Evoked+Potential+5tgnats >EEG 5teady-5tate visual Evoked Potential 5tg
nals</a></b>: This database consists on 30 subjects performing Brain Computer Interface for Steady State Visu
al Evoked Potentials (BCI-SSVEP). 157. <b><a href="datasets/El+Nino">El Nino</a></b
>: The data set contains oceanographic and surface meteorological readings taken from a series of buoys posit
ity+Simulated+Data+">Electrical Grid Stability Simulated Data </a></b>: The local stability analysis of the 4-nod
e star system (electricity producer is in the center) implementing Decentral Smart Grid Control concept. 
class="normal">159. <b><a href="datasets/ElectricityLoadDiagrams20112014">ElectricityLoadDiagrams20112014">ElectricityLoadDiagrams20112014">ElectricityLoadDiagrams20112014">ElectricityLoadDiagrams20112014">ElectricityLoadDiagrams20112014">ElectricityLoadDiagrams20112014">ElectricityLoadDiagrams20112014">ElectricityLoadDiagrams20112014">ElectricityLoadDiagrams20112014">ElectricityLoadDiagrams20112014">ElectricityLoadDiagrams20112014">ElectricityLoadDiagrams20112014">ElectricityLoadDiagrams20112014</br>
14</a></b>: This data set contains electricity consumption of 370 points/clients.\r\n160.
<b><a href="datasets/EMG+data+for+gestures">EMG data for gestures</a></b>: These are files of raw EMG d
ata recorded by MYO Thalmic bracelet161. <b><a href="datasets/EMG+dataset+in+Lo"
wer+Limb">EMG dataset in Lower Limb</a></b>: 3 different exercises: sitting, standing and walking in the mus
cles: biceps femoris, vastus medialis, rectus femoris and semitendinosus addition to goniometry in the exercise
s.162. <b><a href="datasets/EMG+Physical+Action+Data+Set">EMG Physical Action Da
ta Set</a></b>: The Physical Action Data Set includes 10 normal and 10 aggressive physical actions that meas
ure the human activity. The data have been collected by 4 subjects using the Delsys EMG wireless apparatus.<
/p>163. <b><a href="datasets/Energy+efficiency">Energy efficiency</a></b>: This study loo
ked into assessing the heating load and cooling load requirements of buildings (that is, energy efficiency) as a f
unction of building parameters.
164. <b><a href="datasets/Entree+Chicago+Recommen">164. <b > <a href="datasets/Entree+Chicago+Recommen">164. <a href="data
dation+Data">Entree Chicago Recommendation Data</a></b>: This data contains a record of user interactions
with the Entree Chicago restaurant recommendation system.165. <b><a href="datasets"><a href="datasets"><a
/Epileptic+Seizure+Recognition">Epileptic Seizure Recognition</a></b>: This dataset is a pre-processed and re
-structured/reshaped version of a very commonly used dataset featuring epileptic seizure detection. <p cla
ss="normal">166. <b><a href="datasets/extention+of+Z-Alizadeh+sani+dataset">extention of Z-Alizadeh sani d
ataset</a></b>: It was collected for CAD diagnosis.167. <b><a href="datasets/Faceboo"><a href="datasets/Fac
k+Comment+Volume+Dataset">Facebook Comment Volume Dataset</a></b>: Instances in this dataset contain
features extracted from facebook posts. The task associated with the data is to predict how many comments th
e post will receive.168. <b><a href="datasets/Facebook+metrics">Facebook metrics</a
></b>: Facebook performance metrics of a renowned cosmetic\'s brand Facebook page.<p class="normal"
>169. <b><a href="datasets/Farm+Ads">Farm Ads</a></b>: This data was collected from text ads found on tw
elve websites that deal with various farm animal related topics. The binary labels are based on whether or not t
he content owner approves of the ad.170. <b><a href="datasets/Fertility">Fertility</a></
b>: 100 volunteers provide a semen sample analyzed according to the WHO 2010 criteria. Sperm concentratio
n are related to socio-demographic data, environmental factors, health status, and life habits
class="nor"
mal">171. <b><a href="datasets/Firm-Teacher_Clave-Direction_Classification">Firm-Teacher_Clave-Direction_
Classification</a></b>: The data are binary attack-point vectors and their clave-direction class(es) according to
the partido-alto-based paradigm.
172. <b><a href="datasets/First-order+theorem+provin">172. <b href="datasets/
g">First-order theorem proving</a></b>: Given a theorem, predict which of five heuristics will give the fastest pr
oof when used by a first-order prover. A sixth prediction declines to attempt a proof, should the theorem be too
difficult.173. <b><a href="datasets/Flags">Flags</a></b>: From Collins Gem Guide to F
lags, 1986174. <b><a href="datasets/FMA%3A+A+Dataset+For+Music+Analysis">FMA:
A Dataset For Music Analysis</a></b>: FMA features 106,574 tracks and includes song title, album, artist, genr
es; play counts, favorites, comments; description, biography, tags; together with audio (343 days, 917 GiB) and
features.175. <b><a href="datasets/Folio">Folio</a></b>: 20 photos of leaves for each
of 32 different species.176. <b><a href="datasets/Forest+Fires">Forest Fires</a></b>:
This is a difficult regression task, where the aim is to predict the burned area of forest fires, in the northeast regi
on of Portugal, by using meteorological and other data (see details at: http://www.dsi.uminho.pt/~pcortez/forestf
ires).177. <b><a href="datasets/Forest+type+mapping">Forest type mapping</a></b>:
Multi-temporal remote sensing data of a forested area in Japan. The goal is to map different forest types using
spectral data.178. <b><a href="datasets/Function+Finding">Function Finding</a></b>:
Cases collected mostly from investigations in physical science; intention is to evaluate function-finding algorithm
s179. <b><a href="datasets/Gas+Sensor+Array+Drift+Dataset">Gas Sensor Array Drift
Dataset</a></b>: This archive contains 13910 measurements from 16 chemical sensors utilized in simulations f
or drift compensation in a discrimination task of 6 gases at various levels of concentrations.
al">180. <b><a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations">Gas Sensor Arr
ay Drift Dataset at Different Concentrations</a></b>: This archive contains 13910 measurements from 16 chem
ical sensors exposed to 6 different gases at various concentration levels.181. <b><a hr
ef="datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures">Gas sensor array exposed to turbulent
gas mixtures</a></b>: A chemical detection platform composed of 8 chemoresistive gas sensors was exposed t
o turbulent gas mixtures generated naturally in a wind tunnel. The acquired time series of the sensors are provi
ded.182. <b><a href="datasets/Gas+sensor+array+under+dynamic+gas+mixtures">Ga
s sensor array under dynamic gas mixtures</a></b>: 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. <b><a href="datasets/Gas+sensor+array+under+flow">183. <b href="datasets/Gas+sensor+array+un
+modulation">Gas sensor array under flow modulation</a></b>: 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. <b><a href="datasets/Gas+sensor+array">184. <b href="datasets/Gas+sensor-array">184. <b href="datasets/Gas
s+in+open+sampling+settings">Gas sensor arrays in open sampling settings</a></b>: 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 substances185. <b><a href="datasets"><a href="datasets">
/Gas+sensors+for+home+activity+monitoring">Gas sensors for home activity monitoring</a></b>: 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\n186.
<b><a href="datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy">Gastrointestinal Lesions in Regular
Colonoscopy</a></b>: 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. 
p class="normal">187. <b><a href="datasets/gene+expression+cancer+RNA-Seq">gene expression cancer RN
A-Seq</a></b>: 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.
p class="normal">188. <b><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></b>: A multisource and multivariate dataset for indoor localisation methods based on W
LAN and Geo-Magnetic field fingerprinting189. <b><a href="datasets/Geographical+Ori
ginal+of+Music">Geographical Original of Music</a></b>: 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</
p>190. <b><a href="datasets/Gesture+Phase+Segmentation">Gesture Phase Segmentation"
</a></b>: 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.<p class="
normal">191. <b><a href="datasets/Gisette">Gisette</a></b>: 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
class="normal">192. <b><a href="datasets/Glass+Identific">192. <b href="datasets/Glass
ation">Glass Identification</a></b>: From USA Forensic Science Service; 6 types of glass; defined in terms of t
heir oxide content (i.e. Na, Fe, K, etc)193. <b><a href="datasets/GNFUV+Unmanned+S">193. <b href="datasets/GNFUV+U
urface+Vehicles+Sensor+Data">GNFUV Unmanned Surface Vehicles Sensor Data</a></b>: 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. <b><a href="
datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2">GNFUV Unmanned Surface Vehicles S
ensor Data Set 2</a></b>: 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. <b><a href="datasets/GPS+Trajectories">GPS Trajectories</a></b>: 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. <b><a href="datasets/Grammatical+Facial+">
Expressions">Grammatical Facial Expressions</a></b>: This dataset supports the development of models that
make possible to interpret Grammatical Facial Expressions from Brazilian Sign Language (Libras).
="normal">197. <b><a href="datasets/Greenhouse+Gas+Observing+Network">Greenhouse Gas Observing Net
work</a></b>: 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. <b><a href="datasets/Haberman%27s+Survival">Haberman\'s Survival</a></b>: Dataset contains cases fr
om study conducted on the survival of patients who had undergone surgery for breast cancer<p class="nor
mal">199. <b><a href="datasets/Hayes-Roth">Hayes-Roth</a></b>: Topic: human subjects study<p class=
"normal">200. <b><a href="datasets/HCC+Survival">HCC Survival</a></b>: 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. <b><a href="datasets/Health+News+in+Twitter">Health News in
Twitter</a></b>: 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. <b><a href="
datasets/Heart+Disease">Heart Disease</a></b>: 4 databases: Cleveland, Hungary, Switzerland, and the VA L
ong Beach203. <b><a href="datasets/Hepatitis">Hepatitis</a></b>: From G.Gong: CMU
; Mostly Boolean or numeric-valued attribute types; Includes cost data (donated by Peter Turney)<p class=
"normal">204. <b><a href="datasets/HEPMASS">HEPMASS</a></b>: 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.
class="normal">205. <b><a href="datasets/Heterogeneity+Activity+Rec">4
ognition">Heterogeneity Activity Recognition</a></b>: 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.
class="normal">206. <b><a href="data"><b><a href="data"><b><a href="data"><a href="
sets/HIGGS">HIGGS</a></b>: This is a classification problem to distinguish between a signal process which pr
oduces Higgs bosons and a background process which does not. 207. <b><a href="dat
asets/Hill-Valley">Hill-Valley</a></b>: 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 $\phi\)bump$ in the te
rrain) or a Valley (a �dip� in the terrain).208. <b><a href="datasets/HIV-1+protease+c">4-protease+c</a>
leavage">HIV-1 protease cleavage</a></b>: 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. <b><a href="datasets/Horse+Colic">Horse Colic</a></b>: Well documented attribute
s; 368 instances with 28 attributes (continuous, discrete, and nominal); 30% missing values<p class="norm"
al">210. <b><a href="datasets/HTRU2">HTRU2</a></b>: 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. <b><a href="datasets/Human+Activity+Recognition">211. <b href="datasets/Huma
+Using+Smartphones">Human Activity Recognition Using Smartphones</a></b>: 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. <b><a href="datasets/Hybrid">212. <b href="datasets
+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer">Hybrid Indoor Positioning
Dataset from WiFi RSSI, Bluetooth and magnetometer</a></b>: 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. 
class="normal">213. <b><a href="datasets/ICMLA+2014+Accepte">4
d+Papers+Data+Set">ICMLA 2014 Accepted Papers Data Set</a></b>: 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. <b><a href="datasets/ICU">ICU</a></b>:
Data set prepared for the use of participants for the 1994 AAAI Spring Symposium on Artificial Intelligence in M
edicine.215. <b><a href="datasets/IDA2016Challenge">IDA2016Challenge</a></b>: 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.
class="normal">217. <b><a href="datasets/Image+Segmentation"</pre>
>Image Segmentation</a></b>: Image data described by high-level numeric-valued attributes, 7 classes
class="normal">218. <b><a href="datasets/lmmunotherapy+Dataset">lmmunotherapy Dataset</a></b>: This d
ataset contains information about wart treatment results of 90 patients using immunotherapy.<p class="nor
mal">219. <b><a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Park">219. <b><a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Park">219. <b><a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Park">219. <b><a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Park">219. <b style="color: blue;">219. <a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Park">219. <a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Park">219. <a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Park">219. <a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Park">219. <a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Park">219. <a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Park">219. <a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Digitized+Graphics+Tablet+for+Monitoring+Digitized+Graphics+Tablet+for+Monitoring+Digitized+Graphics+Tablet+for+Monitoring+Digitized+Graphics+Tablet+for+Monitoring+Digitized+Graphics+Tablet+for+Monitoring+Digitized+Graphics+Tablet+for+Monitoring+Digitized+Graphics+Tablet+for+Monitoring+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digitized+Graphics+Digiti
inson%E2%80%99s+Disease">Improved Spiral Test Using Digitized Graphics Tablet for Monitoring Parkinson's
Disease</a></b>: 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.
ormal">220. <b><a href="datasets/Individual+household+electric+power+consumption">Individual household el
ectric power consumption</a></b>: 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. <b><a href="datasets/Indoor+User+Movement+Prediction+from+">1. <b><a href="datasets/Indoor+User+Movement+Prediction+from+">1. <a href="datasets/Indoor+User+Movement+Frediction+from+">1. <a href="datasets/Indoor+User+Movement+Prediction+from+">1. <a href="datasets/Indoor+User+Movement+Frediction+frediction+frediction+frediction+frediction+frediction+frediction+frediction+fredicti
RSS+data">Indoor User Movement Prediction from RSS data</a></b>: 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. <b><a href="datasets/Insurance+Co
mpany+Benchmark+%28COIL+2000%29">Insurance Company Benchmark (COIL 2000)</a>
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 data223. <</p>
b><a href="datasets/Internet+Advertisements">Internet Advertisements</a></b>: This dataset represents a set
of possible advertisements on Internet pages.224. <b><a href="datasets/Internet+Usag">4. <b href="datasets/Internet+Usag">4. 
e+Data">Internet Usage Data</a></b>: This data contains general demographic information on internet users i
n 1997.
class="normal">225. <b><a href="datasets/lonosphere">lonosphere</a></b>: Classification of r
adar returns from the ionosphere226. <b><a href="datasets/IPUMS+Census+Database">226. <b href="datasets/IPUMS+Cens
">IPUMS Census Database</a></b>: 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. <b><a href="datase">227. <b href=
ts/Iris">Iris</a></b>: Famous database; from Fisher, 1936228. <b><a href="datasets/IS"
OLET">ISOLET</a></b>: Goal: Predict which letter-name was spoken--a simple classification task.<p class
s="normal">229. <b><a href="datasets/ISTANBUL+STOCK+EXCHANGE">ISTANBUL STOCK EXCHANGE</a>
</b>: 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.
class="normal">230. <br/><br/><br/>
><a href="datasets/Japanese+Credit+Screening">Japanese Credit Screening</a></b>: Includes domain theory
(generated by talking to Japanese domain experts); data in Lisp231. <b><a href="datas">231. <b href="datas"
ets/Japanese+Vowels">Japanese Vowels</a></b>: This dataset records 640 time series of 12 LPC cepstrum c
```

```
oefficients taken from nine male speakers.232. <b><a href="datasets/KASANDR">KASA
NDR</a></b>: 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.
<b><a href="datasets/KDC-4007+dataset+Collection">KDC-4007 dataset Collection</a></b>: KDC-4007 datase
t Collection is the Kurdish Documents Classification text used in categories regarding Kurdish Sorani news and
articles.234. <b><a href="datasets/KDD+Cup+1998+Data">KDD Cup 1998 Data</a></b
>: This is the data set used for The Second International Knowledge Discovery and Data Mining Tools Competit
ion, which was held in conjunction with KDD-98235. <b><a href="datasets/KDD+Cup+1">235. <b href="datasets/KDD+Cup+1">235. <b
999+Data">KDD Cup 1999 Data</a></b>: This is the data set used for The Third International Knowledge Disc
overy and Data Mining Tools Competition, which was held in conjunction with KDD-99
36. <b><a href="datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29">KEGG Metabolic Reaction
n Network (Undirected)</a></b>: KEGG Metabolic pathways modeled as un-directed reaction network. Variety
Network+%28Directed%29">KEGG Metabolic Relation Network (Directed)</a></b>: KEGG Metabolic pathways
modeled as directed relation network. Variety of graphical features presented.238. <b>
<a href="datasets/Kinship">Kinship</a></b>: Relational dataset239. <b><a href="datas
ets/Labor+Relations">Labor Relations</a></b>: From Collective Bargaining Review240.
<b><a href="datasets/Las+Vegas+Strip">Las Vegas Strip</a></b>: This dataset includes quantitative and categ
orical features from online reviews from 21 hotels located in Las Vegas Strip, extracted from TripAdvisor (http://
www.tripadvisor.com).241. <b><a href="datasets/Leaf">Leaf</a></b>: This dataset con
sists in a collection of shape and texture features extracted from digital images of leaf specimens originating fro
m a total of 40 different plant species.242. <b><a href="datasets/LED+Display+Domain"
>LED Display Domain</a></b>: From Classification and Regression Trees book; We provide here 2 C program
s for generating sample databases243. <b><a href="datasets/Legal+Case+Reports">Le
gal Case Reports</a></b>: A textual corpus of 4000 legal cases for automatic summarization and citation analy
sis. For each document we collect catchphrases, citations sentences, citation catchphrases and citation classes
.244. <b><a href="datasets/Lenses">Lenses</a></b>: Database for fitting contact lense
s245. <b><a href="datasets/Letter+Recognition">Letter Recognition</a></b>: Database
of character image features; try to identify the letter246. <b><a href="datasets/Libras+M">4.5 href="datasets/
ovement">Libras Movement</a></b>: The data set contains 15 classes of 24 instances each. Each class refere
nces to a hand movement type in LIBRAS (Portuguese\r\nname \'L\varPhgua BRAsileira de Sinais\', oficial brazilia
n signal language).247. <b><a href="datasets/Liver+Disorders">Liver Disorders</a></b
>: BUPA Medical Research Ltd. database donated by Richard S. Forsyth248. <b><a hr
ef="datasets/Localization+Data+for+Person+Activity">Localization Data for Person Activity</a></b>: Data contai
ns recordings of five people performing different activities. Each person wore four sensors (tags) while performi
ng the same scenario five times. 249. <b><a href="datasets/Logic+Theorist">Logic The
orist</a></b>: All code for Logic Theorist250. <b><a href="datasets/Low+Resolution+Sp">250. <b href="datasets/Low-Resolution+Sp">250. <b href="datasets/Low+Resolution+Sp">250. <b href="datasets/Low-Resolution+Sp">250. <b href="dat
ectrometer">Low Resolution Spectrometer</a></b>: From IRAS data -- NASA Ames Research Center<p cl
ass="normal">251. <b><a href="datasets/LSVT+Voice+Rehabilitation">LSVT Voice Rehabilitation</a></b>: 126
samples from 14 participants, 309 features. Aim: assess whether voice rehabilitation treatment lead to phonatio
ns considered \'acceptable\' or \'unacceptable\' (binary class classification problem).252
. <b><a href="datasets/Lung+Cancer">Lung Cancer</a></b>: Lung cancer data; no attribute definitions
class="normal">253. <b><a href="datasets/Lymphography">Lymphography</a></b>: This lymphography domai
n was obtained from the University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia. (Restricted ac
cess)254. <b><a href="datasets/M.+Tuberculosis+Genes">M. Tuberculosis Genes</a>
</b>: Data giving characteristics of each ORF (potential gene) in the M. tuberculosis bacterium. Sequence, ho
mology (similarity to other genes) and structural information, and function (if known) are provided<p class=
"normal">255. <b><a href="datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012"
-2014">Machine Learning based ZZAlpha Ltd. Stock Recommendations 2012-2014</a></b>: The data here ar
e 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. 
class="normal">256. <b><a href="datas">256. <b hre
ets/Madelon">Madelon</a></b>: MADELON is an artificial dataset, which was part of the NIPS 2003 feature sel
ection challenge. This is a two-class classification problem with continuous input variables. The difficulty is that t
he problem is multivariate and highly non-linear. 257. <b><a href="datasets/MAGIC+Ga">45. <br/>45. <br/>45. <br/>46. The problem is multivariate and highly non-linear. 
mma+Telescope">MAGIC Gamma Telescope</a></b>: Data are MC generated to simulate registration of high
energy gamma particles in an atmospheric Cherenkov telescope258. <b><a href="data"><b><a href="data"><a 
sets/Mammographic+Mass">Mammographic Mass</a></b>: Discrimination of benign and malignant mammogr
aphic masses based on BI-RADS attributes and the patient\'s age.259. <b><a href="dat
asets/Mechanical+Analysis">Mechanical Analysis</a></b>: Fault diagnosis problem of electromechanical devic
es; also PUMPS DATA SET is newer version with domain theory and results260. <b><a
href="datasets/Mesothelioma%E2%80%99s+disease+data+set+">Mesothelioma's disease data set </a></b>: M
esothelioma'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.
class="nor black patient data">"nor black patient data mal">261. Meta-data: Meta-Data was used in order to give advice a bout which classification method is appropriate for a particular dataset (taken from results of Statlog project). >262. MEU-Mobile KSD: This dataset co ntains keystroke dynamics data collected on a touch mobile device (Nexus 7). The dataset contains 2856 recor ds, 51 records per subject for 56 subjects. 263. 40. datasets/MHEALTH+Datas et">MHEALTH Dataset: The MHEALTH (Mobile Health) dataset is devised to benchmark techniques d ealing with human behavior analysis based on multimodal body sensing.264. <a hree f="datasets/Mice+Protein+Expression">Mice Protein Expression: Expression levels of 77 proteins mea sured in the cerebral cortex of 8 classes of control and Down syndrome mice exposed to context fear conditioni ng, a task used to assess associative learning.265. <b style="color: blue;">Lassess associative learning.265. <b style="color: blue;"><b style="color: blue;"><b style="color: blue;">265. <b style="color: blue;"><b style="color: blue;">265. <b style="color: blue;"><b style="color: blue;">265. <b style="color: blue; U">microblogPCU: MicroblogPCU data is crawled from sina weibo microblog[http://weibo.com/]. This d ata can be used to study machine learning methods as well as do some social network research. ="normal">266. MicroMass: A dataset to explore machine learning approaches for the identification of microorganisms from mass-spectrometry data.267. MiniBooNE particle identification: This datase t is taken from the MiniBooNE experiment and is used to distinguish electron neutrinos (signal) from muon neut rinos (background).268. Miskolc IIS Hy brid IPS: The dataset was created for the comparison and evaluation of hybrid indoor positioning meth ods. The dataset presented contains data from W-LAN and Bluetooth interfaces, and Magnetometer. ass="normal">269. Mobile Robots: Learning concepts from sen sor data of a mobile robot; set of data sets270. MoCap Hand Postures: 5 types of hand postures from 12 users were recorded using unlabele d markers attached to fingers of a glove in a motion capture environment. Due to resolution and occlusion, miss ing values are common.271. 28Promoter+G ene+Sequences%29">Molecular Biology (Promoter Gene Sequences): E. Coli promoter gene sequenc es (DNA) with partial domain theory272. 272. <b href="datasets/Molecul Protein+Secondary+Structure%29">Molecular Biology (Protein Secondary Structure): From CMU conn ectionist bench repository; Classifies secondary structure of certain globular proteins
p class="normal">27 3. Molecular Biology (Spl ice-junction Gene Sequences): Primate splice-junction gene sequences (DNA) with associated imperfe ct domain theory274. MONK\'s Problem s: A set of three artificial domains over the same attribute space; Used to test a wide range of inductio n algorithms275. Moral Reasoner: Ho rn-clause model that qualitatively simulates moral reasoning; Theory includes negated literals mal">276. Motion Capture Hand Postures: 5 t ypes 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. Movie: This data set contains a list of over 10000 films including many of der, odd, and cult films. There is information on actors, casts, directors, producers, studios, etc.<p class="n ormal">278. MSNBC.com Anonymous Web Data : This data describes the page visits of users who visited msnbc.com on September 28, 1999. Visits ar e recorded at the level of URL category (see description) and are recorded in time order.<p class="normal" >279. Mturk User-Perceived Clusters ov er Images: This dataset was collected by Shan-Hung Wu and DataLab members at NTHU, Taiwan. Th ere\'re 325 user-perceived clusters from 100 users and their corresponding descriptions.<p class="normal" >280. Multimodal Da mage Identification for Humanitarian Computing: 5879 captioned images (image and text) from social media related to damage during natural disasters/wars, and belong to 6 classes: Fires, Floods, Natural landsca pe, Infrastructural, Human, Non-damage.281. Multiple Features: This dataset consists of features of handwritten numerals (`0\'--`9\') extracted from a collection of Dutch utility maps282. Mushroom : From Audobon Society Field Guide; mushrooms described in terms of physical characteristics; classificati on: poisonous or edible283. Musk (Ve rsion 1): The goal is to learn to predict whether new molecules will be musks or non-musks<p class s="normal">284. Musk (Version 2): The goal is to lea rn to predict whether new molecules will be musks or non-musks285. 285. <b href="data">285. sets/News+Aggregator">News Aggregator: References to news pages collected from an web aggregat or in the period from 10-March-2014 to 10-August-2014. The resources are grouped into clusters that represen t pages discussing the same story.286. 286. <b href="datasets/News+Popularity+in+Multi">286. < ple+Social+Media+Platforms">News Popularity in Multiple Social Media Platforms: Large data set of ne ws items and their respective social feedback on multiple platforms: Facebook, Google+ and LinkedIn.

```
ass="normal">287. <b><a href="datasets/Newspaper+and+magazine+images+segmentation+dataset">Newspa
per and magazine images segmentation dataset</a></b>: Dataset is well suited for segmentation tasks. It cont
ains 101 scanned pages from different newspapers and magazines in Russian with ground truth pixel-based m
asks.288. <b><a href="datasets/NIPS+Conference+Papers+1987-2015">NIPS Conference+Papers+1987-2015">NIPS Conference+Paper
nce Papers 1987-2015</a></b>: This data set contains the distribution of words in the full text of the NIPS conf
erence papers published from 1987 to 2015.289. <b><a href="datasets/NoisyOffice">N
oisyOffice</a></b>: Corpus intended to do cleaning (or binarization) and enhancement of noisy grayscale printe
d text images using supervised learning methods. Noisy images and their corresponding ground truth provided.
290. <b><a href="datasets/Nomao">Nomao</a></b>: Nomao collects data about places
(name, phone, localization...) from many sources.\r\nDeduplication consists in detecting what data refer to the s
hix">Northix</a></b>: Northix is designed to be a schema matching benchmark problem for data integration of t
wo entity relationship databases. 292. <b><a href="datasets/NSF+Research+Award+Ab">495. <b ><a href="datasets/NSF+Research+Award+Ab">495. <b ><a href="datasets/NSF+Research+Award+Ab">495. <b ><a href="datasets/NSF+Research+Award+Ab">495. <a href="datasets/NSF+Research+Ab">495. <a href="datasets/NSF+Research+Ab">495. <a href="datasets/NSF-Research+Ab">495. <a href="datasets/NSF-
stracts+1990-2003">NSF Research Award Abstracts 1990-2003</a></b>: 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-word293. <b><a href="datasets/Nursery">N
ursery</a></b>: Nursery Database was derived from a hierarchical decision model originally developed to rank
applications for nursery schools.294. <b><a href="datasets/NYSK">NYSK</a></b>: NYS
K (New York v. Strauss-Kahn) is a collection of English news articles about the case relating to allegations of se
xual assault against the former IMF director Dominique Strauss-Kahn (May 2011).
295.
<b><a href="datasets/Occupancy+Detection+">Occupancy Detection </a></b>: Experimental data used for bina
ry classification (room occupancy) from Temperature, Humidity, Light and CO2. Ground-truth occupancy was obt
ained from time stamped pictures that were taken every minute.296. <b><a href="datas">296. <ba><a href="datas">
ets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes">OCT data & Color Fundus Images of L
eft & Right Eyes</a></b>: This dataset contains OCT data (in mat format) and color fundus data (in jpg format)
of left & right eyes of 50 healthy persons.297. <b><a href="datasets/One-hundred+plant">297. <b href="da
+species+leaves+data+set">One-hundred plant species leaves data set</a></b>: Sixteen samples of leaf each
of one-hundred plant species. For each sample, a shape descriptor, fine scale margin and texture histogram ar
e given.298. <b><a href="datasets/Online+Handwritten+Assamese+Characters+Datase"
t">Online Handwritten Assamese Characters Dataset</a></b>: This is a dataset of 8235 online handwritten ass
amese characters. The "online" process involves capturing of data as text is written on a digitizing tablet with an
electronic pen.
299. <b><a href="datasets/Online+News+Popularity">Online News Popul
arity</a></b>: This dataset summarizes a heterogeneous set of features about articles published by Mashable i
n a period of two years. The goal is to predict the number of shares in social networks (popularity).
="normal">300. <b><a href="datasets/Online+Retail">Online Retail</a></b>: This is a transnational data set whi
ch contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered n
on-store online retail.301. <b><a href="datasets/Online+Shoppers+Purchasing+Intentio">hoppers+Purchasing+Intentio</a>
n+Dataset">Online Shoppers Purchasing Intention Dataset</a></b>: Of the 12,330 sessions in the dataset,\r\n8
4.5% (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. <b><a href="datasets/Online+Video+Chara">40. <br/>class\r\nsamples ending with shopping.
cteristics+and+Transcoding+Time+Dataset">Online Video Characteristics and Transcoding Time Dataset</a>
b>: The dataset contains a million randomly sampled video instances listing 10 fundamental video characteristic
s along with the YouTube video ID. 303. <b><a href="datasets/Open+University+Learni">40. <br/> class="normal">40. <br/> clas
ng+Analytics+dataset">Open University Learning Analytics dataset</a></b>: Open University Learning Analytics
Dataset contains data about courses, students and their interactions with Virtual Learning Environment for seve
n selected courses and more than 30000 students.
class="normal">304. <b><a href="datasets/Opinosis">304. <b><a href="datasets/Opinosis">4
+Opinion+%26frasl%3B+Review">Opinosis Opinion ⁄ Review</a></b>: This dataset contains sentences ex
tracted from user reviews on a given topic. Example topics are "performance of Toyota Camry" and "sound qua
lity of ipod nano". 305. <b><a href="datasets/OpinRank+Review+Dataset">OpinRank R
eview Dataset</a></b>: This data set contains user reviews of cars and and hotels collected from Tripadvisor (
~259,000 \r\nreviews) and Edmunds (~42,230 reviews). 306. <b><a href="datasets/O"
PPORTUNITY+Activity+Recognition">OPPORTUNITY Activity Recognition</a></b>: The OPPORTUNITY Datas
et for Human Activity Recognition from Wearable, Object, and Ambient Sensors is a dataset devised to benchm
ark human activity recognition algorithms (classification, automatic data segmentation, sensor fusion, feature ex
traction, etc).307. <b><a href="datasets/Optical+Interconnection+Network+">Optical Int
erconnection Network </a></b>: This dataset contains 640 performance measurements from a simulation of 2-
Dimensional Multiprocessor Optical Interconnection Network. 
class="normal">308. <b><a href="dataset">4
s/Optical+Recognition+of+Handwritten+Digits">Optical Recognition of Handwritten Digits</a></b>: Two versions
y">Othello Domain Theory</a></b>: Used in research to generate features for an inductive learning system
310. <b><a href="datasets/Ozone+Level+Detection">Ozone Level Detection</a></b>: Two
ground ozone level data sets are included in this collection. One is the eight hour neak 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, Galve
ston and Brazoria area.311. <b><a href="datasets/p53+Mutants">p53 Mutants</a></b>
: The goal is to model mutant p53 transcriptional activity (active vs inactive) based on data extracted from bioph
ysical simulations.\r\n312. <b><a href="datasets/Page+Blocks+Classification">Page Blo
cks Classification</a></b>: The problem consists of classifying all the blocks of the page layout of a document t
hat has been detected by a segmentation process.313. <b><a href="datasets/PAMAP2">4 href="datasets/
+Physical+Activity+Monitoring">PAMAP2 Physical Activity Monitoring</a></b>: The PAMAP2 Physical Activity M
onitoring dataset contains data of 18 different physical activities, performed by 9 subjects wearing 3 inertial mea
surement units and a heart rate monitor.
class="normal">314. <b><a href="datasets/PANDOR">PANDOR
R</a></b>: PANDOR is a novel and publicly available dataset for online recommendation provided by Purch (htt
p://www.purch.com/). 315. <b><a href="datasets/Paper+Reviews">Paper Reviews</a><
/b>: This sentiment analysis data set contains scientific paper reviews from an international conference on com
puting and informatics. The task is to predict the orientation or the evaluation of a review.<p class="normal"
>316. <b><a href="datasets/Parking+Birmingham">Parking Birmingham</a></b>: Data collected from car parks
in Birmingham that are operated by NCP from \r\nBirmingham City Council. UK Open Government Licence (OG
L).\r\nhttps://data.birmingham.gov.uk/dataset/birmingham-parking317. <b><a href="data">>317. <b><a href="data">>><a href="data">>><a href="data">>><a href="data">>><a href="data">>><a href="data">>><a href="data">>><a href="data">><a href="data">>><a href="data">><a href="data">>><a href="data">><a href="data">>><a href="d
asets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet">Parkinson Disease Spiral Drawi
ngs Using Digitized Graphics Tablet</a></b>: 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. <b><a href="datasets/Parkinson+Speech+Dataset+with++Multiple+Type">4.5 class="normal">4.5 
s+of+Sound+Recordings">Parkinson Speech Dataset with Multiple Types of Sound Recordings</a></b>: The tr
aining data belongs to 20 Parkinson\'s Disease (PD) patients and 20 healthy subjects. From all subjects, multiple
e types of sound recordings (26) are taken.
319. <b><a href="datasets/Parkinson%27s+"
Disease+Classification">Parkinson\'s Disease Classification</a></b>: 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).
="normal">320. <b><a href="datasets/Parkinsons">Parkinsons</a></b>: Oxford Parkinson\'s Disease Detection
Dataset321. <b><a href="datasets/Parkinsons+Telemonitoring">Parkinsons Telemonito
ring</a></b>: Oxford Parkinson\'s Disease Telemonitoring Dataset322. <b><a href="dat
asets/PEMS-SF">PEMS-SF</a></b>: 15 months worth of daily data (440 daily records) that describes the occu
pancy rate, between 0 and 1, of different car lanes of the San Francisco bay area freeways across time.
class="normal">323. <b><a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits">Pen-Based Recognition+of+Handwritten+Digits">Pen-Based Recognition+of+Handwritten+Digits">Pen-Based Recognition+of+Handwritten+Digits</a>
nition of Handwritten Digits</a></b>: Digit database of 250 samples from 44 writers324.
<b><a href="datasets/Perfume+Data">Perfume Data</a></b>: This data consists of odors of 20 different perfu
mes. Data was obtained by using a handheld odor meter (OMX-GR sensor) per second for 28 seconds period.
325. <b><a href="datasets/Phishing+Websites">Phishing Websites</a></b>: This datas
et collected mainly from: PhishTank archive, MillerSmiles archive, Google's searching operators.
normal">326. <b><a href="datasets/Physical+Unclonable+Functions">Physical Unclonable Functions</a></b>:
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
. <b><a href="datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure">Physicochemical Propertie
s of Protein Tertiary Structure</a></b>: This is a data set of Physicochemical Properties of Protein Tertiary Stru
cture. The data set is taken from CASP 5-9. There are 45730 decoys and size varying from 0 to 21 armstrong.
/p>328. <b><a href="datasets/Pioneer-1+Mobile+Robot+Data">Pioneer-1 Mobile Robot Dat
a</a></b>: 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<p clas
s="normal">329. <b><a href="datasets/Pittsburgh+Bridges">Pittsburgh Bridges</a></b>: Bridges database that
has original and numeric-discretized datasets
class="normal">330. <b><a href="datasets/Planning+Rela"><br/>class="normal">330. <b><a href="datasets/Planning+Rela"><br/>class="normal"><br/>class="normal">330. <b><a href="datasets/Planning+Rela"><br/>class="normal"><br/>class="normal">330. <b><a href="datasets/Planning+Rela"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="normal"><br/>class="
x">Planning Relax</a></b>: The dataset concerns with the classification of two mental stages from recorded EE
G signals: Planning (during imagination of motor act) and Relax state. 
class="normal">331. <b><a href=</p>
"datasets/Plants">Plants</a></b>: Data has been extracted from the USDA plants database. It contains all plant
s (species and genera) in the database and the states of USA and Canada where they occur.<p class="nor
mal">332. <b><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.
ts/PMU-UD">PMU-UD</a></b>: 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). 334. <b><a href="datasets/Poker+Hand">Poker Hand</a></b>: Purpose is to pr
edict poker hands335. <b><a href="datasets/Polish+companies+bankruptcy+data">Poli
sh companies bankruptcy data</a></b>: The dataset is about bankruptcy prediction of Polish companies. The b
ankrupt companies were analyzed in the period 2000-2012, while the still operating companies were evaluated
from 2007 to 2013.336. <b><a href="datasets/Post-Operative+Patient">Post-Operative
Patient //ax //bx: Datacet of nationt features //ax on class="normal" 227 obside the professional Product Reviewers
```

```
aliente at a vibri. Dataset of patient realistes proposed in the vibries at a vibri
s+activities+in+a+online+social+media">Predict keywords activities in a online social media</a></b>: The data f
rom Twitter was collected during 360 consecutive days. It was done by querying 1497 English keywords sample
d from Wikipedia. This dataset is proposed in a Learning to rank setting.
338. <b><a hre="https://doi.org/10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.jps-10.2006/j.ps-10.2006/j.jps-1
f="datasets/Primary+Tumor">Primary Tumor</a></b>: From Ljubljana Oncology Institute<p class="normal"
>339. <b><a href="datasets/Prodigy">Prodigy</a></b>: Assorted domains like blocksworld, eightpuzzle, and sc
hedworld.340. <b><a href="datasets/Protein+Data">Protein Data</a></b>: Undocument
ed341. <b><a href="datasets/Pseudo+Periodic+Synthetic+Time+Series">Pseudo Period
ic Synthetic Time Series</a></b>: This data set is designed for testing indexing schemes in time series databas
es. The data appears highly periodic, but never exactly repeats itself.342. <b><a href="
datasets/PubChem+Bioassay+Data">PubChem Bioassay Data</a></b>: These highly imbalanced bioassay dat
asets are from the differing types of screening that can be performed using HTS technology. 21 datasets were
created from 12 bioassays.343. <b><a href="datasets/QSAR+biodegradation">QSAR bi
odegradation</a></b>: Data set containing values for 41 attributes (molecular descriptors) used to classify 105
5 chemicals into 2 classes (ready and not ready biodegradable).
class="normal">344. <b><a href="datas">344. <b href="datas">344. <b
ets/QtyT40I10D100K">QtyT40I10D100K</a></b>: Since there is no numerical sequential data stream available
in standard data sets, this data set is generated from the original T40I10D100K data set
class="normal">
345. <b><a href="datasets/Quadruped+Mammals">Quadruped Mammals</a></b>: The file animals.c is a data
generator of structured instances representing quadruped animals346. <b><a href="dat
asets/Qualitative+Structure+Activity+Relationships">Qualitative Structure Activity Relationships</a></b>: Two se
ts of datasets are given: pyrimidines and triazines
347. <b><a href="datasets/Qualitative">47. <b hre
 _Bankruptcy">Qualitative_Bankruptcy</a></b>: Predict the Bankruptcy from Qualitative parameters from expert
s.348. <b><a href="datasets/Quality+Assessment+of+Digital+Colposcopies">Quality As
sessment of Digital Colposcopies</a></b>: This dataset explores the subjective quality assessment of digital col
poscopies.349. <b><a href="datasets/Real+estate+valuation+data+set">Real estate val
uation data set</a></b>: 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. <b>
<a href="datasets/REALDISP+Activity+Recognition+Dataset">REALDISP Activity Recognition Dataset</a></b>:
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 <p class="normal
">351. <b><a href="datasets/Record+Linkage+Comparison+Patterns">Record Linkage Comparison Patterns</
a></b>: 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.
52. <b><a href="datasets/Relative+location+of+CT+slices+on+axial+axis">Relative location of CT slices on axial
axis</a></b>: 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.
class="normal">353.
<b><a href="datasets/Repeat+Consumption+Matrices">Repeat Consumption Matrices</a></b>: 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. <b><a href="datasets/"
Residential+Building+Data+Set">Residential Building Data Set</a></b>: 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. <b><a href="datasets/Restaurant+%26+consumer+data">Re
staurant & consumer data</a></b>: The dataset was obtained from a recommender system prototype. The tas
k was to generate a top-n list of restaurants according to the consumer preferences. 35
6. <b><a href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collectio">to the collection of the c
n">Reuters RCV1 RCV2 Multilingual, Multiview Text Categorization Test collection</a></b>: 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. <b><a href="datasets/Reuters+Transcribed+S">4557. <b href="datasets/Reuters+T
ubset">Reuters Transcribed Subset</a></b>: 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.358. <b><a href="datasets/Reuters-21578+Text+Categorization+Collection">Reut
ers-21578 Text Categorization Collection</a></b>: This is a collection of documents that appeared on Reuters
newswire in 1987. The documents were assembled and indexed with categories.359. <
b><a href="datasets/Reuter 50 50">Reuter 50 50</a></b>: The dataset is used for authorship identification i
n online Writeprint which is a new research field of pattern recognition. 360. <b><a href
="datasets/Rice+Leaf+Diseases">Rice Leaf Diseases</a></b>: There are three classes/diseases: Bacterial leaf
blight, Brown spot, and Leaf smut, each having 40 images. The format of all images is jpg. <p class="norm
al">361. <b><a href="datasets/Robot+Execution+Failures">Robot Execution Failures</a></b>: This dataset con
tains force and torque measurements on a robot after failure detection. Each failure is characterized by 15 force
e/torque samples collected at regular time intervals362. <b><a href="datasets/Roman+">4
Urdu+Data+Set">Roman Urdu Data Set</a></b>: Roman Urdu (the scripting style for Urdu language) is one of t
he limited resource languages. A data corpus comprising of more than 20000 records was collected. <p clas
```

```
s= normal >303. <b><a href="datasets/3ates_11atisactions_Dataset_weekly">3ates_11atisactions_Dataset_w
eekly</a></b>: Contains weekly purchased quantities of 800 over products over 52 weeks. Normalised values a
re provided too.364. <b><a href="datasets/SCADI">SCADI</a></b>: First self-care activ
ities dataset based on ICF-CY.365. <b><a href="datasets/SECOM">SECOM</a></b>: D
ata from a semi-conductor manufacturing process366. <b><a href="datasets/seeds">se
eds</a></b>: 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.
ss="normal">367. <b><a href="datasets/seismic-bumps">seismic-bumps</a></b>: The data describe the probl
em of high energy (higher than 10<sup>4</sup> J) seismic bumps forecasting in a coal \r\nmine. Data come from two of lo
ngwalls located in a Polish coal mine.368. <b><a href="datasets/Semeion+Handwritten+">here = "datasets/Semeion+Handwritten+">here = "datasets/Semeion+">here = "datasets/S
Digit">Semeion Handwritten Digit</a></b>: 1593 handwritten digits from around 80 persons were scanned, stre
tched in a rectangular box 16x16 in a gray scale of 256 values.
369. <b><a href="datase">369. <b href="
ts/sEMG+for+Basic+Hand+movements">sEMG for Basic Hand movements</a></b>: The "sEMG for Basic Han
d movements" includes 2 databases of surface electromyographic signals of 6 hand movements using Delsys\'
EMG System. Healthy subjects conducted six daily life grasps.
370. <b><a href="dataset"><a href="d
s/Sentence+Classification">Sentence Classification</a></b>: Contains sentences from the abstract and introdu
ction of 30 articles annotated with a modified Argumentative Zones annotation scheme. These articles come fro
m biology, machine learning and psychology.371. <b><a href="datasets/Sentiment+Lab">4.5 class="normal">371. <b><a href="datasets/Sentiment+Lab">4.5 class="normal">4.5 class="normal"
elled+Sentences">Sentiment Labelled Sentences</a></b>: The dataset contains sentences labelled with positiv
e or negative sentiment.372. <b><a href="datasets/ser+Knowledge+Modeling+Data+%2">48. <b style="color: blue;">48. <b sty
8Students%27+Knowledge+Levels+on+DC+Electrical+Machines%29">ser Knowledge Modeling Data (Students\
'Knowledge Levels on DC Electrical Machines)</a></b>: 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-cour
ses and reported in my Ph.D. Thesis.373. <b><a href="datasets/Servo">Servo</a></b>
U+kernel+performance">SGEMM GPU kernel performance</a></b>: Running times for multiplying two 2048 x
2048 matrices using a GPU OpenCL SGEMM kernel with varying parameters (using the library \'CLTune\').
375. <b><a href="datasets/Shuttle+Landing+Control">Shuttle Landing Control</a></b>: Tiny
database; all nominal values376. <b><a href="datasets/SIFT10M">SIFT10M</a></b>: I
n SIFT10M, each data point is a SIFT feature which is extracted from Caltech-256 by the open source VLFeat li
brary. The corresponding patches of the SIFT features are provided.377. <b><a href="
datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set">Simulated Falls and Daily Living Activities Dat
a Set</a></b>: 20 falls and 16 daily living activities were performed by 17 volunteers with 5 repetitions while we
aring 6 sensors (3.060 instances) that attached to their head, chest, waist, wrist, thigh and ankle.<p class="
normal">378. <b><a href="datasets/SkillCraft1+Master+Table+Dataset">SkillCraft1 Master Table Dataset</a>
b>: This data was used in Thompson et al. (2013). A list of possible game actions is discussed in Thompson, Bl
air, Chen, & Henrey (2013).379. <b><a href="datasets/Skin+Segmentation">Skin Segm
entation</a></b>: The Skin Segmentation dataset is constructed over B, G, R color space. Skin and Nonskin da
taset is generated using skin textures from face images of diversity of age, gender, and race people.
ss="normal">380. <b><a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29">380. <b><a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29">380. <b><a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29">380. <b><a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29">380. <b><a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29">380. <a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29">380. <a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29">380. <a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29">380. <a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29">380. <a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29">380. <a href="dataset-for+Human+Activity+Recognition+%28HAR%29">380. <a href="dataset-for+H
+in+Ambient+Assisted+Living+%28AAL%29">Smartphone Dataset for Human Activity Recognition (HAR) in Am
bient Assisted Living (AAL)</a></b>: This data is an addition to an existing dataset on UCI. We collected more d
ata to improve the accuracy of our human activity recognition algorithms applied in the domain of Ambient Assis
ted Living. 381. <b><a href="datasets/Smartphone-Based+Recognition+of+Human+Act">ted Living. class="normal">381. <b><a href="datasets/Smartphone-Based+Recognition+of+Human+Act">ted Living. 
ivities+and+Postural+Transitions">Smartphone-Based Recognition of Human Activities and Postural Transitions
</a></b>: Activity recognition data set built from the recordings of 30 subjects performing basic activities and po
stural transitions while carrying a waist-mounted smartphone with embedded inertial sensors.\r\n<p class="
normal">382. <b><a href="datasets/SML2010">SML2010</a></b>: This dataset is collected from a monitor sys
tem mounted in a domotic house. It corresponds to approximately 40 days of monitoring data.
class="no">"no
rmal">383. <b><a href="datasets/SMS+Spam+Collection">SMS Spam Collection</a></b>: The SMS Spam Coll
ection is a public set of SMS labeled messages that have been collected for mobile phone spam research.
384. <b><a href="datasets/Solar+Flare">Solar Flare</a></b>: Each class attribute counts th
e number of solar flares of a certain class that occur in a 24 hour period385. <b><a href
="datasets/Somerville+Happiness+Survey">Somerville Happiness Survey</a></b>: A data extract of a non-fede
ral dataset posted here https://catalog.data.gov/dataset/somerville-happiness-survey-responses-2011-2013-20
15386. <b><a href="datasets/Soybean+%28Large%29">Soybean (Large)</a></b>: Mic
halski\'s famous soybean disease database387. <b><a href="datasets/Soybean+%28S">285
mall%29">Soybean (Small)</a></b>: Michalski\'s famous soybean disease database38
8. <b><a href="datasets/Spambase">Spambase</a></b>: Classifying Email as Spam or Non-Spam<p class
="normal">389. <b><a href="datasets/SPECT+Heart">SPECT Heart</a></b>: Data on cardiac Single Proton E
mission Computed Tomography (SPECT) images. Each patient classified into two categories: normal and abno
rmal.390. <b><a href="datasets/SPECTF+Heart">SPECTF Heart</a></b>: Data on card
```

```
iac Single Proton Emission Computed Tomography (SPECT) images. Each patient classified into two categories
: normal and abnormal.391. <b><a href="datasets/Spoken+Arabic+Digit">Spoken Arabi
c Digit</a></b>: This dataset contains timeseries of mel-frequency cepstrum coefficients (MFCCs) corresponding
g to spoken Arabic digits. Includes data from 44 male and 44 female native Arabic speakers.
class="nor">"nor
mal">392. <b><a href="datasets/Sponge">Sponge</a></b>: Data on sponges; Attributes in spanish<p class
s="normal">393. <b><a href="datasets/Sports+articles+for+objectivity+analysis">Sports articles for objectivity a
nalysis</a></b>: 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.
="normal">394. <b><a href="datasets/Statlog+%28Australian+Credit+Approval%29">Statlog (Australian Credit
Approval)</a></b>: This file concerns credit card applications. This database exists elsewhere in the repository
g+%28German+Credit+Data%29">Statlog (German Credit Data)</a></b>: 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 matrix396. <b><a href="datasets/Statlog+%28Heart%29">Statlog (Heart)</a></b
>: 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 form397. <b><a href="datasets/Statlog+%28Ima">28Ima
ge+Segmentation%29">Statlog (Image Segmentation)</a></b>: 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. <b><a href="datasets/Statlog+%28Landsat+Satellite%29">Statlog (Lands
at Satellite)</a></b>: Multi-spectral values of pixels in 3x3 neighbourhoods in a satellite image, and the classific
ation associated with the central pixel in each neighbourhood399. <b><a href="datasets"><a href="datasets
/Statlog+%28Shuttle%29">Statlog (Shuttle)</a></b>: The shuttle dataset contains 9 attributes all of which are n
umerical. Approximately 80% of the data belongs to class 1400. <b><a href="datasets/S"><b
tatlog+%28Vehicle+Silhouettes%29">Statlog (Vehicle Silhouettes)</a></b>: 3D objects within a 2D image by ap
plication of an ensemble of shape feature extractors to the 2D silhouettes of the objects.
class="normal"
>401. <b><a href="datasets/Statlog+Project">Statlog Project</a></b>: Various Databases: Vehicle silhouttes, L
andsat Sattelite, Shuttle, Australian Credit Approval, Heart Disease, Image Segmentation, German Credit
p class="normal">402. <b><a href="datasets/Steel+Plates+Faults">Steel Plates Faults</a></b>: 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\n403. <b><a href="datasets/Stock+portfolio+performance">Stock portfolio+performance
io performance</a></b>: The data set of performances of weighted scoring stock portfolios are obtained with m
ixture design from the US stock market historical database.
404. <b><a href="datasets/S"><br/>
toneFlakes">StoneFlakes</a></b>: 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. <b><a href="datasets/Student+Academics+Performance">Student Academ
ics Performance</a></b>: The dataset tried to find the end semester percentage prediction based on different s
ocial, economic and academic attributes. 406. <b><a href="datasets/Student+Loan+Rel">406. <b href="datasets/Stu
ational">Student Loan Relational</a></b>: Student Loan Relational Domain407. <b><a
href="datasets/Student+Performance">Student Performance</a></b>: Predict student performance in secondar
y education (high school). 408. <b><a href="datasets/Superconductivty+Data">Superco
nductivty Data</a></b>: Two file s contain data on 21263 superconductors and their relevant features.
ass="normal">409. <b><a href="datasets/SUSY">SUSY</a></b>: This is a classification problem to distinguish
between a signal process which produces supersymmetric particles and a background process which does not.
410. <b><a href="datasets/Synthetic+Control+Chart+Time+Series">Synthetic Control C
hart Time Series</a></b>: This data consists of synthetically generated control charts.4
11. <b><a href="datasets/Syskill+and+Webert+Web+Page+Ratings">Syskill and Webert Web Page Ratings</a>
</b>: 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)
s="normal">412. <b><a href="datasets/Tamilnadu+Electricity+Board+Hourly+Readings">Tamilnadu Electricity B
oard Hourly Readings</a></b>: This data can be effectively produced the result to fewer parameter of the Load
profile can be reduced in the Database 
class="normal">413. <b><a href="datasets/Tarvel+Review+Rati">413. <b href="datasets/Tarvel+Review+Review+Rati">413. <b href="datasets/Tarvel+Review+Review+Review+Revi
ngs">Tarvel Review Ratings</a></b>: 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.<p class
="normal">414. <b><a href="datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+20"
15">Taxi Service Trajectory - Prediction Challenge, ECML PKDD 2015</a></b>: An accurate dataset describing
trajectories performed by all the 442 taxis running in the city of Porto, in Portugal.\r\n41
5. <b><a href="datasets/Teaching+Assistant+Evaluation">Teaching Assistant Evaluation</a></b>: 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></b>: 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. <b><a href="datasets/Thoracic+Surgery+Data">Thoracic Surgery Data</a></b>: 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. <b><a href="datasets/Thyroid+Dise">418. <b href="
ase">Thyroid Disease</a></b>: 10 separate databases from Garavan Institute419. <b>
<a href="datasets/Tic-Tac-Toe+Endgame">Tic-Tac-Toe Endgame</a></b>: Binary classification task on possible
e configurations of tic-tac-toe game420. <b><a href="datasets/Trains">Trains</a></b>:
2 data formats (structured, one-instance-per-line)
p class="normal">421. <b><a href="datasets/Travel+Re">421. <b href="datasets/Travel
views">Travel Reviews</a></b>: Reviews on destinations in 10 categories mentioned across East Asia. Each tr
aveler rating is mapped as Excellent(4), Very Good(3), Average(2), Poor(1), and Terrible(0) and average rating
is used.422. <b><a href="datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+te"
xt+categorization">TTC-3600: Benchmark dataset for Turkish text categorization</a></b>: The TTC-3600 data
set is a collection of Turkish news and articles including categorized 3,600 documents from 6 well-known portal
s in Turkey. It has 4 different forms in ARFF Weka format.
class="normal">423. <b><a href="datasets/Tu">
rkiye+Student+Evaluation">Turkiye Student Evaluation</a></b>: 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 ques
tions and additional 5 attributes.
424. <b><a href="datasets/TV+News+Channel+Comme">424. <b href="datasets/TV+News+Channel+Channel+Comme">424. <b href="datasets/TV+News+Channel+Channel+Channel+Channel+Channel+Channel+Channel+Channel+Channel+Channel+Cha
rcial+Detection+Dataset">TV News Channel Commercial Detection Dataset</a></b>: TV Commercials data set
consists of standard audio-visual features of video shots extracted from 150 hours of TV news broadcast of 3 I
ndian and 2 international news channels (30 Hours each). 
class="normal">425. <b><a href="datasets/"</pre>
Twenty+Newsgroups">Twenty Newsgroups</a></b>: This data set consists of 20000 messages taken from 20
newsgroups.426. <b><a href="datasets/Twin+gas+sensor+arrays">Twin gas sensor arr
ays</a></b>: 5 replicates of an 8-MOX gas sensor array were exposed to different gas conditions (4 volatiles at
10 concentration levels each).
427. <b><a href="datasets/Twitter+Data+set+for+Arabic+">427. <b href="dataset-for+Arabic+">427. <b href="datasets/
Sentiment+Analysis">Twitter Data set for Arabic Sentiment Analysis</a></b>: This problem of Sentiment Analysis
is (SA) has been studied well on the English language but not Arabic one. Two main approaches have been de
vised: corpus-based and lexicon-based. 428. <b><a href="datasets/UbiqLog+%28smart">428. <base>428. <br/>428. <base>428. <br/>428. <base>428. <br/>428. <br/>428
phone+lifelogging%29">UbiqLog (smartphone lifelogging)</a></b>: UbiqLog is the smartphone lifelogging tool t
hat runs on the smartphone of 35 users for about 2 months. \r\n429. <b><a href="datas"
ets/UJI+Pen+Characters">UJI Pen Characters</a></b>: Data consists of written characters in a UNIPEN-like for
mat430. <b><a href="datasets/UJI+Pen+Characters+%28Version+2%29">UJI Pen Char
acters (Version 2)</a></b>: A pen-based database with more than 11k isolated handwritten characters<p c
lass="normal">431. <b><a href="datasets/UJIIndoorLoc">UJIIndoorLoc</a></b>: The UJIIndoorLoc is a Multi-B
uilding Multi-Floor indoor localization database to test Indoor Positioning System that rely on WLAN/WiFi fingerp
rint.432. <b><a href="datasets/UJIIndoorLoc-Mag">UJIIndoorLoc-Mag</a></b>: The UJ
IIndoorLoc-Mag is an indoor localization database to test Indoor Positioning System that rely on Earth\'s magne
tic field variations.
433. <b><a href="datasets/Ultrasonic+flowmeter+diagnostics">Ultrasonic+flowmeter+diagnostics">Ultrasonic+flowmeter+diagnostics">Ultrasonic+flowmeter+diagnostics
nic flowmeter diagnostics</a></b>: Fault diagnosis of four liquid ultrasonic flowmeters43
4. <b><a href="datasets/Undocumented">Undocumented</a></b>: Various datasets without documentation (fe
el free to explore!)435. <b><a href="datasets/University">University</a></b>: Data in ori
ginal (LISP-readable) form436. <b><a href="datasets/University+of+Tehran+Question+">436. <b href="datasets/University+of+Tehran+">436. <b href="datasets/University+of+Tehran+"
Dataset+2016+%28UTQD.2016%29">University of Tehran Question Dataset 2016 (UTQD.2016)</a>
ian questions gathered from a jeopardy game broadcasted on Iranian national television. <p class="normal
">437. <b><a href="datasets/UNIX+User+Data">UNIX User Data</a></b>: 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 ye
ars.438. <b><a href="datasets/Urban+Land+Cover">Urban Land Cover</a></b>: Classi
fication of urban land cover using high resolution aerial imagery. Intended to assist sustainable urban planning
efforts.439. <b><a href="datasets/URL+Reputation">URL Reputation</a></b>: Anonymi
zed 120-day subset of the ICML-09 URL data containing 2.4 million examples and 3.2 million features.
ass="normal">440. <b><a href="datasets/US+Census+Data+%281990%29">US Census Data (1990)</a></b>:
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. <b><a href="datasets/Use"><b style="color: blue;">441. <b style="color: 
r+Identification+From+Walking+Activity">User Identification From Walking Activity</a></b>: 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\n442. <b><a href="datasets/User+Knowledge+Modeling">User Kno
wledge Modeling</a></b>: It is the real dataset about the students\' knowledge status about the subject of Elect
rical DC Machines. The dataset had been obtained from Ph.D. Thesis.
class="normal">443. <b><a href=</pre>
"datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Pr
oblem%3A+Pat">USPTO Algorithm Challenge, run by NASA-Harvard Tournament Lab and TopCoder Proble
m: Pat</a></b>: Data used for USPTO Algorithm Competition. Contains drawing pages from US patents with m
anually labeled figure and part labels.444. <b><a href="datasets/Vertebral+Column">Ve
rtebral Column</a></b>: 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).
s="normal">445. <b><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 hum
an activity. The data have been collected by 10 subjects using the Vicon 3D tracker.446
. <b><a href="datasets/Victorian+Era+Authorship+Attribution">Victorian Era Authorship Attribution</a></b>: To
create the largest authorship attribution dataset, we extracted works of 50 well-known authors. To have a non-e
xhaustive learning, in training there are 45 authors whereas, in the testing, it\'s 50447. <
b><a href="datasets/Volcanoes+on+Venus+-+JARtool+experiment">Volcanoes on Venus - JARtool experiment
</a></b>: The JARtool project was a pioneering effort to develop an automatic system for cataloging small volc
anoes in the large set of Venus images returned by the Magellan spacecraft.448. <b><
a href="datasets/Wall-Following+Robot+Navigation+Data">Wall-Following Robot Navigation Data</a></b>: The
data were collected as the SCITOS G5 robot navigates through the room following the wall in a clockwise directi
on, for 4 rounds, using 24 ultrasound sensors arranged circularly around its \'waist\'.449
. <b><a href="datasets/Water+Treatment+Plant">Water Treatment Plant</a></b>: Multiple classes predict plant
state450. <b><a href="datasets/Waveform+Database+Generator+%28Version+1%29">
Waveform Database Generator (Version 1)</a></b>: CART book\'s waveform domains
451. <b><a href="datasets/Waveform+Database+Generator+%28Version+2%29">Waveform Database Genera
tor (Version 2)</a></b>: CART book\'s waveform domains452. <b><a href="datasets/W"
earable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29">Wearable Co
mputing: Classification of Body Postures and Movements (PUC-Rio)</a></b>: A dataset with 5 classes (sitting-d
own, standing-up, standing, walking, and sitting) collected on 8 hours of activities of 4 healthy subjects. We also
established a baseline performance index.453. <b><a href="datasets/Website+Phishing">453. <b href="datasets/Website+Phishing</b href="datasets/Website+P
">Website Phishing</a></b>: \r\n\r\n454. <b><a href="datasets/Weight+Lifting+Exercise"
s+monitored+with+Inertial+Measurement+Units">Weight Lifting Exercises monitored with Inertial Measurement
Units</a></b>: Six young health subjects were asked to perform 5 variations of the biceps curl weight lifting exe
rcise. One of the variations is the one predicted by the health professional.
class="normal">455.
ref="datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29">WESAD (Wearable Stress and Affect
Detection)</a></b>: WESAD (Wearable Stress and Affect Detection) contains data of 15 subjects during a stres
s-affect lab study, while wearing physiological and motion sensors.456. <b><a href="dat
asets/Wholesale+customers">Wholesale customers</a></b>: The data set refers to clients of a wholesale distri
butor. It includes the annual spending in monetary units (m.u.) on diverse product categories
class="nor">"nor
mal">457. <b><a href="datasets/wiki4HE">wiki4HE</a></b>: Survey of faculty members from two Spanish univ
ersities on teaching uses of Wikipedia458. <b><a href="datasets/Wilt">Wilt</a></b>: Hig
h-resolution Remote Sensing data set (Quickbird). Small number of training samples of diseased trees, large n
umber for other land cover. Testing data set from stratified random sample of image.45
9. <b><a href="datasets/Wine">Wine</a></b>: Using chemical analysis determine the origin of wines<p cla
ss="normal">460. <b><a href="datasets/Wine+Quality">Wine Quality</a></b>: Two datasets are included, relat
ed to red and white vinho verde wine samples, from the north of Portugal. The goal is to model wine quality bas
ed on physicochemical tests (see [Cortez et al., 2009], http://www3.dsi.uminho.pt/pcortez/wine/).<p class="
normal">461. <b><a href="datasets/Wireless+Indoor+Localization">Wireless Indoor Localization</a></b>: Colle
cted 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. <b><a href="datasets/Yacht+Hydrodynamics">Y
acht Hydrodynamics</a></b>: Delft data set, used to predict the hydodynamic performance of sailing yachts fro
m dimensions and velocity.463. <b><a href="datasets/YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD">YearPredictionMSD</YearPredictionMSD</YearPredictionMSD</YearPredictionMSD</YearPredictionMSD</YearPredictionMSD</YearPredictionMSD</YearPredictionMSD</YearPredictionMSD</YearPredictionMSD</YearPredictionMSD</YearPredictionMSD</YearPredictionMSD</YearPredictionMSD</YearPredictionMSD</YearPredictionMSD</YearPredictionMSD</YearPredictionMSD</YearPredictionMSD</YearPredictionMSD</YearPredictionMSD</YearPredictionMSD</YearPredictionM
ionMSD</a></b>: Prediction of the release year of a song from audio features. Songs are mostly western, com
mercial tracks ranging from 1922 to 2011, with a peak in the year 2000s.464. <b><a hr
ef="datasets/Yeast">Yeast</a></b>: Predicting the Cellular Localization Sites of Proteins<p class="normal"
>465. <b><a href="datasets/YouTube+Comedy+Slam+Preference+Data">YouTube Comedy Slam Preference
Data</a></b>: 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.
ass="normal">466. <b><a href="datasets/YouTube+Multiview+Video+Games+Dataset">YouTube Multiview Vid
eo Games Dataset</a></b>: This dataset contains about 120k instances, each described by 13 feature types,
with class information, specially useful for exploring multiview topics (cotraining, ensembles, clustering,..).
p class="normal">467. <b><a href="datasets/YouTube+Spam+Collection">YouTube Spam Collection</a></b>: I
t is a public set of comments collected for spam research. It has five datasets composed by 1,956 real messag
es extracted from five videos that were among the 10 most viewed on the collection period.<p class="norm
al">468. <b><a href="datasets/Z-Alizadeh+Sani">Z-Alizadeh Sani</a></b>: It was collected for CAD diagnosis.<
/p>469. <b><a href="datasets/Zoo">Zoo</a></b>: Artificial, 7 classes of animals
d>< hr>\n\t/n/table> \n\n\t/r> \n \n\n\n\n \n\n\n\n \n\n\n\n \n\n\n\n
r valign=center>\n\t\tSupported By:\n
                                                                                                             <img src="assets/nsfe.gif" heigh
                             In Collaboration With:\n
                                                                                                                                <img src="assets/r"
exaSmall.jpg" />\n//table>\n\n<center>\n<span class="normal">\n<a href="about.html">About</a>&nb
sp; || \n<a href="citation_policy.html">Citation Policy</a>&nbsp;&nbsp;||&nbsp;\n<a href="donation"
 policy.html">Donation Policy</a>&nbsp;&nbsp;||&nbsp;\n<a href="contact.html">Contact</a>&nbsp;&nbsp;||&n
```

 $bsp; \n < a href="http://cml.ics.uci.edu">CML \n </center> \n \n </bdy> \n </html> \n'$

```
In [173]:
```

```
soup_2 = BeautifulSoup(html_2)
soup_2
```

```
Out[173]:
```

```
<!DOCTYPE HTML>
<a href="https://www.enalth.com/style="color: blue;">httml><body>"-//W3C//DTD HTML 4.01 Transitional//EN\"&gt;</a>
<title>UCI Machine Learning Repository: Data Sets</title>
<!-- Stylesheet link -->
k 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) -->
<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>
<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/>
<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"/>
<input checked="" name="sitesearch" type="radio" value="ics.uci.edu"/> <span class="whitetext"><font size="1"</pre>
>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/>
<br/>br/>
Browse Through:
 <b>Default Task</b> 
<a href="datasets.php?format=&amp;task=cla&amp;att=&amp;area=&amp;
numAtt=&numIns=&type=&sort=nameUp&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;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&amp;numAtt=&a
ns=&type=&sort=nameUp&view=list">Regression</a> <font color="red">(96)</font><br/>><a href
="datasets.php?format=&task=clu&att=&area=&numAtt=&numIns=&type=&s
```

```
ort=nameop&view=list >clustering</a> <iont color= red >(84)</iont><a nrei= datasets.pnp?iormat=&
amp;task=other&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view
=list">Other</a> <font color="red">(55)</font> 
 <b>Attribute Type</b> 
<a href="datasets.php?format=&amp;task=&amp;att=cat&amp;area=&amp;
numAtt=&numIns=&type=&sort=nameUp&view=list">Categorical</a> <font color="red">(38)
</font><br/><a href="datasets.php?format=&amp;task=&amp;att=num&amp;area=&amp;numAtt=&amp;numIns"
=&type=&sort=nameUp&view=list">Numerical</a> <font color="red">(307)</font><br/>><a href="d
atasets.php?format=&task=&att=mix&area=&numAtt=&numIns=&type=&sort
=nameUp&view=list">Mixed</a> <font color="red">(55)</font> 
 <b>Data Type</b> 
<a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;nu"
mAtt=&numIns=&type=mvar&sort=nameUp&view=list">Multivariate</a> <font color="red">(
357)</font><br/>a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns"
=&type=uvar&sort=nameUp&view=list">Univariate</a> <font color="red">(23)</font><br/>><a href
="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=seq&
sort=nameUp&view=list">Sequential</a> <font color="red">(47)</font><br/><a href="datasets.php?format">format
=&task=&att=&area=&numAtt=&numIns=&type=ts&sort=nameUp&view
=list">Time-Series</a> <font color="red">(91)</font><br/><a href="datasets.php?format=&amp;task=&amp;att="
&area=&numAtt=&numIns=&type=text&sort=nameUp&view=list">Text</a> <font c
olor="red">(53)</font><br/><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&a
mp;numIns=&type=dt&sort=nameUp&view=list">Domain-Theory</a> <font color="red">(23)</fon
t><br/><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=&amp;ty"
pe=other&sort=nameUp&view=list">Other</a> <font color="red">(21)</font><br/> 
<b>Area</b> 
<a href="datasets.php?format=&amp;task=&amp;att=&amp;area=life&amp;
numAtt=&numIns=&type=&sort=nameUp&view=list">Life Sciences</a> <font color="red">(1
07)</font><br/><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=phys&amp;numAtt=&amp;nu
mIns=&type=&sort=nameUp&view=list">Physical Sciences</a> <font color="red">(49)</font><br/>>c/font>
><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=comp&amp;numAtt=&amp;numIns=&amp;ty
pe=&sort=nameUp&view=list">CS / Engineering</a> <font color="red">(170)</font><br/><a href="dat
asets.php?format=&task=&att=&area=soc&numAtt=&numIns=&type=&sort=
nameUp&view=list">Social Sciences</a> <font color="red">(26)</font><br/>ohr/><a href="datasets.php?format">red">(26)</font><br/>ohr/><a href="datasets.php?format">red">(26)</a><br/>ohr/><a href="datasets.php?format">(26)</a><br/>ohr/><a href="datasets.php?format">(2
=&task=&att=&area=bus&numAtt=&numIns=&type=&sort=nameUp&vie
w=list">Business</a> <font color="red">(29)</font><br/><a href="datasets.php?format=&amp;task=&amp;att=&
amp;area=game&numAtt=&numIns=&type=&sort=nameUp&view=list">Game</a> <fon
t color="red">(10)</font><br/><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=other&amp;nu
mAtt=&numIns=&type=&sort=nameUp&view=list">Other</a> <font color="red">(73)</font>
<b># Attributes</b> 
<a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;nu"
mAtt=less10&numIns=&type=&sort=nameUp&view=list">Less than 10</a> <font color="red"
```

```
>(113)</font><br/>ca href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=10to100&a
mp;numIns=&type=&sort=nameUp&view=list">10 to 100</a> <font color="red">(210)</font><br/>br/
><a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=greater100&amp;numIns=&a
mp;type=&sort=nameUp&view=list">Greater than 100</a> <font color="red">(84)</font> 
<b># Instances</b>
<a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIns=less100&amp;t
ype=&sort=nameUp&view=list">Less than 100</a> <font color="red">(27)</font><br/><a href="datase"
ts.php?format=&task=&att=&area=&numAtt=&numIns=100to1000&type=&so
rt=nameUp&view=list">100 to 1000</a> <font color="red">(162)</font><br/><a href="datasets.php?format">rt=nameUp&amp;view=list">100 to 1000</a> <font color="red">(162)</font><br/><a href="datasets.php?format">rt=nameUp&amp;view=list">100 to 1000</a> <font color="red">(162)</font><br/><a href="datasets.php?format">100 to 1000</a> <font color="red">(162)</font><br/><a href="datasets.php?format">100 to 1000</a> <font color="red">(162)</font><br/><a href="datasets.php?format">100 to 1000</a> <br/><a href="datase
=&task=&art=&area=&numAtt=&numIns=greater1000&type=&sort=nameUp
&view=list">Greater than 1000</a> <font color="red">(246)</font> 
<b>Format Type</b> 
<a href="datasets.php?format=mat&amp;task=&amp;att=&amp;area=&amp
;numAtt=&numIns=&type=&sort=nameUp&view=list">Matrix</a> <font color="red">(324)</f
ont><br/><a href="datasets.php?format=nonmat&amp;task=&amp;att=&amp;area=&amp;numAtt=&amp;numIn
s=&type=&sort=nameUp&view=list">Non-Matrix</a> <font color="red">(145)</font> 
<b>469</b> Data Sets
<a href="datasets.php?format=&amp;task=&amp;att=&amp;area=&amp;nu
mAtt=&numIns=&type=&sort=nameUp&view=table">Table View</a> <font color="gray">Lis
t View</font>
<hr/>1. <b><a href="datasets/2.4+GHZ+Indoor+Cha">
nnel+Measurements">2.4 GHZ Indoor Channel Measurements</a></b>: Measurement of the S21,consists of 1
0 sweeps, each sweep contains 601 frequency points with spacing of 0.167MHz to cover a 100MHz band cente
red at 2.4GHz.2. <b><a href="datasets/3D+Road+Network+%28North+Jutland%2C+De"
nmark%29">3D Road Network (North Jutland, Denmark)</a></b>: 3D road network with highly accurate elevati
on information (+-20cm) from Denmark used in eco-routing and fuel/Co2-estimation routing algorithms.
class="normal">3. <b><a href="datasets/AAAI+2013+Accepted+Papers">AAAI 2013 Accepted Papers</a></b>:
This data set compromises the metadata for the 2013 AAAI conference's accepted papers (main track only), in
cluding paper titles, abstracts, and keywords of varying granularity.4. <b><a href="datas">4. <b href=
ets/AAAI+2014+Accepted+Papers">AAAI 2014 Accepted Papers</a></b>: This data set compromises the meta
data for the 2014 AAAI conference's accepted papers, including paper titles, authors, abstracts, and keywords
of varying granularity.5. <b><a href="datasets/Abalone">Abalone</a></b>: Predict the a
ge of abalone from physical measurements
class="normal">6. <b><a href="datasets/Abscisic+Acid+Sign">6.</a>
aling+Network">Abscisic Acid Signaling Network</a></b>: 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 separat
e boolean pseudodynamic simulations using an asynchronous update scheme. 7. <b><
a href="datasets/Absenteeism+at+work">Absenteeism at work</a></b>: The database was created with record
s of absenteeism at work from July 2007 to July 2010 at a courier company in Brazil.8.
<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>: This dataset comprises information regarding
g the ADLs performed by two users on a daily basis in their
own homes. 9. <b><a href="datasets/Activity+Recognition+from+Single+Chest-Mounte">+Chest-Mounte
d+Accelerometer">Activity Recognition from Single Chest-Mounted Accelerometer</a></b>: The dataset collect
```

s data from a wearable accelerometer mounted on the chest. The dataset is intended for Activity Recognition re search purposes.
class="normal">10.
class="normal">10.
class="datasets/Activity+Recognition+system+based+on+Mu

Itisensor+data+fusion+%28AReM%29">Activity Recognition system based on Multisensor data fusion (AReM)/b>: This dataset contains temporal data from a Wireless Sensor Network worn by an actor performing the activities: bending, cycling, lying down, sitting, standing, walking.

p>11. Activity recognition with healthy older people using a batteryless wearable sensor: 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>12. Acute Inflammations
The data was created by a medical expert as a data set to test the expert syst amount of the sensor of the sensor

which will perform the presumptive diagnosis of two diseases of the urinary system.

13. Adult: Predict whether income exceeds \$50K/ yr based on census data. Also known as "Census Income" dataset.14. Air Quality: Contains the responses of a gas multisensor device deployed on the fiel d in an Italian city. Hourly responses averages are recorded along with gas concentrations references from a c ertified analyzer. 15. Air quality: Contains th e responses of a gas multisensor device deployed on the field in an Italian city. 16. Airfoil Self-Noise: NASA data set, obtained from a series of aerod vnamic and acoustic tests of two and three-dimensional airfoil blade sections conducted in an anechoic wind tu nnel.17. Amazon Access Samples a>: Amazon's InfoSec is getting smarter about the way Access data is leveraged. This is an anonymized sa mple of access provisioned within the company.18. Amazon Commerce reviews set: The dataset is used for authorship identificatio n in online Writeprint which is a new research field of pattern recognition. 19. <a hre f="datasets/Annealing">Annealing: Steel annealing data20. Anonymous Microsoft Web Data: 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.< /p>21. Anuran Calls (MFCCs) b>: Acoustic features extracted from syllables of anuran (frogs) calls, including the family, the genus, and the sp ecies labels (multilabel). 22. Appli ances energy prediction: Experimental data used to create regression models of appliances energy us e in a low energy building.23. A PS Failure at Scania Trucks: The datasets' positive class consists of component failures for a specific c omponent of the APS system. The negative class consists of trucks with failures for components not related to t he APS.24. Arcene: ARCENE's task is to disting uish cancer versus normal patterns from mass-spectrometric data. This is a two-class classification problem wit h continuous input variables. This dataset is one of 5 datasets of the NIPS 2003 feature selection challenge. >25. Arrhythmia: Distinguish between the prese nce and absence of cardiac arrhythmia and classify it in one of the 16 groups.
class="normal">26. <a</p> href="datasets/Artificial+Characters">Artificial Characters: Dataset artificially generated by using first or der theory which describes structure of ten capital letters of English alphabet
class="normal">27. <a</p> href="datasets/Audiology+%28Original%29">Audiology (Original): Nominal audiology dataset from Bayl or28. Audiology (Standardized): Standardized version of the original audiology database29. Audit Data: Exhaustive one year non-confidential data in the year 2015 to 2016 of fir ms is collected from the Auditor Office of India to build a predictor for classifying suspicious firms.<p class=" normal">30. Australian Sign Language signs : 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.
class="normal">31. a href="datas Sign+Language+signs+%28High+Quality%29">Australian Sign Language signs (High Quality): This dat a 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 trackers32. 4>44 ets/Autism+Screening+Adult">Autism Screening Adult: Autistic Spectrum Disorder Screening Data for Adult. This dataset is related to classification and predictive tasks.
33. 45. ets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++">Autistic Spectrum Disorder Screening D ata for Adolescent : Autistic Spectrum Disorder Screening Data for Adolescent. This dataset is related to classification and predictive tasks.
34. 4. <b href="datasets/ der+Screening+Data+for+Children++">Autistic Spectrum Disorder Screening Data for Children : Childr en screening data for autism suitable for classification and predictive tasks 35. <a hr ef="datasets/Auto+MPG">Auto MPG: Revised from CMU StatLib library, data concerns city-cycle fuel c onsumption36. Automobile: From 1985 Wa rd's Automotive Yearbook37. AutoUniv">AutoUniv: Auto Univ 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.
p class="normal">38. Avila: 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 class="normal">39. Bach Choral Harmony: 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. Bach Chorales : Time-series data based on chorales; challenge is to learn generative grammar; data in Lisp<p class="nor mal">41. Badges: Badges labeled with a "+" or "-" as a function of a pe rson's name42. Bag of Words: This dat a set contains five text collections in the form of bags-of-words.
class="normal">43

class="normal">43

class="normal">43

class="normal">43 s/Balance+Scale">Balance Scale: Balance scale weight & amp; distance database<p class="norma" l">44. Balloons: Data previously used in cognitive psychology experim ent; 4 data sets represent different conditions of an experiment45. 45. <b href="d s/Bank+Marketing">Bank Marketing: 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. banknote authentication">banknote authentication n: Data were extracted from images that were taken for the evaluation of an authentication procedure f or banknotes.47. BAUM-1: BAUM-1 dataset c ontains 1184 multimodal facial video clips collected from 31 subjects. The 1184 video clips contain spontaneou s facial expressions and speech of 13 emotional and mental states.
48. <b<<b<BAUM-2: A multilingual audio-visual affective face database consisting of 1047 video c lips of 286 subjects. 49. 49. <b href="datasets/Behavior+of+the+urban+traffic+of+the+cit">49. <b href="datasets/Behavior+of+the+urban+traffic+of+the+cit">49. <b href="datasets/Behavior+of+the+urban+traffic+of+the+cit">49. <b href="datasets/Behavior+of+the+urban+traffic+ y+of+Sao+Paulo+in+Brazil">Behavior of the urban traffic of the city of Sao Paulo in Brazil: The databas e was created with records of behavior of the urban traffic of the city of Sao Paulo in Brazil.
class="norm" language of the city o al">50. Beijing PM2.5 Data: This hourly data set contains t he PM2.5 data of US Embassy in Beijing. Meanwhile, meteorological data from Beijing Capital International Airp ort are also included. 51. Bike Sharing D ataset: This dataset contains the hourly and daily count of rental bikes between years 2011 and 2012 i n Capital bikeshare system with the corresponding weather and seasonal information. 2. BLE RSSI Dataset for In door localization and Navigation: 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.
class="normal">53. BlogFeedback">BlogFeedback: Instan ces in this dataset contain features extracted from blog posts. The task associated with the data is to predict ho w many comments the post will receive.54. BLOGG ER: In this paper, we look for to recognize the causes of users tend

to cyber space in Kohkiloye and Boyer Ahmad Province in Iran55. Blood Transfusion S ervice Center: Data taken from the Blood Transfusion Service Center in Hsin-Chu City in Taiwan -- this is a classification problem. 56. Breast Cancer: Breast Cancer Data (Restricted Access)57.
 +Coimbra">Breast Cancer Coimbra: Clinical features were observed or measured for 64 patients with breast cancer and 52 healthy controls.
class="normal">58. 68.
 nsin+%28Diagnostic%29">Breast Cancer Wisconsin (Diagnostic): Diagnostic Wisconsin Breast Cancer Database59. Breas t Cancer Wisconsin (Original): Original Wisconsin Breast Cancer Database60. Breast Cancer Wisconsin (Prognostic) a>: Prognostic Wisconsin Breast Cancer Database61. 4 href="dataset +Tissue">Breast Tissue: Dataset with electrical impedance measurements of freshly excised tissue sa mples from the breast.
class="normal">62. BuddyMove D ata Set: User interest information extracted from user reviews published in holidayiq.com about variou s types of point of interests in South India63. 63. 63. et+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%28OBS%29+Network">Burst Header Packet (BHP) flooding attack on Optical Burst Switching (OBS) Network: One of the primary challenges in identi fying the risks of the Burst Header Packet (BHP) flood attacks in Optical Burst Switching networks (OBS) is the s carcity of reliable historical data. 64. Bu zz in social media : 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.</ p>65. Caesarian Section Classification Dataset: 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.<p class="normal

```
">66. <b><a href="datasets/Callt2+Building+People+Counts">Callt2 Building People Counts</a></b>: This data
comes from the main door of the Callt2 building at UCI.67. <b><a href="datasets/Car+E">comes from the main door of the Callt2 building at UCI.67. <b><a href="datasets/Car+E">comes from the main door of the Callt2 building at UCI.67. <b><a href="datasets/Car+E">comes from the main door of the Callt2 building at UCI.class="normal">67. <b><a href="datasets/Car+E">comes from the main door of the Callt2 building at UCI.</a>
valuation">Car Evaluation</a></b>: Derived from simple hierarchical decision model, this database may be usef
ul for testing constructive induction and structure discovery methods.68. <b><a href="d
atasets/Carbon+Nanotubes">Carbon Nanotubes</a></b>: This dataset contains 10721 initial and calculated at
omic coordinates of carbon nanotubes.
class="normal">69. <b><a href="datasets/Cardiotocography">Ca
rdiotocography</a></b>: The dataset consists of measurements of fetal heart rate (FHR) and uterine contraction
n (UC) features on cardiotocograms classified by expert obstetricians.70. <b><a href="d
atasets/Cargo+2000+Freight+Tracking+and+Tracing">Cargo 2000 Freight Tracking and Tracing</a></b>: Sanit
ized 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. <b><a
href="datasets/Census+Income">Census Income</a></b>: Predict whether income exceeds $50K/yr based on
census data. Also known as "Adult" dataset.72. <b><a href="datasets/Census-Income+
%28KDD%29">Census-Income (KDD)</a></b>: This data set contains weighted census data extracted from the
e 1994 and 1995 current population surveys conducted by the U.S. Census Bureau.73.
<b><a href="datasets/Cervical+cancer+%28Risk+Factors%29">Cervical cancer (Risk Factors)</a></b>: This da
taset focuses on the prediction of indicators/diagnosis of cervical cancer. The features cover demographic infor
mation, habits, and historic medical records.
class="normal">74. <b><a href="datasets/Challenger+USA"</pre>
+Space+Shuttle+O-Ring">Challenger USA Space Shuttle O-Ring</a></b>: 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
class="normal">75. <b><a href="datasets/Character+Font+Images">Character Font Images</a></b>: Characte
r images from scanned and computer generated fonts.
class="normal">76. <b><a href="datasets/Chara"><a href="data
cter+Trajectories">Character Trajectories</a></b>: Multiple, labelled samples of pen tip trajectories recorded w
hilst 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. <b><a href="data"><b><a href="data"><b><a href="data"><a href="data
sets/Chess+%28Domain+Theories%29">Chess (Domain Theories)</a></b>: 6 different domain theories for ge
nerating legal moves of chess78. <b><a href="datasets/Chess+%28King-Rook+vs.+Kin
g%29">Chess (King-Rook vs. King)</a></b>: Chess Endgame Database for White King and Rook against Black
King (KRK).79. <b><a href="datasets/Chess+%28King-Rook+vs.+King-Knight%29">Che
ss (King-Rook vs. King-Knight)</a></b>: Knight Pin Chess End-Game Database Creator
80. <b><a href="datasets/Chess+%28King-Rook+vs.+King-Pawn%29">Chess (King-Rook vs. King-Pawn)</a></
b>: King+Rook versus King+Pawn on a7 (usually abbreviated KRKPA7).
"datasets/chestnut+%E2%80%93+LARVIC">chestnut – LARVIC</a></b>: The research project presents this da
tabase, shows the images of chestnuts that will be processed to determine the presence or absence of defects
82. <b><a href="datasets/chipseq">chipseq</a></b>: ChIP-seq experiments characteriz
e protein modifications or binding at
specific genomic locations in specific samples. The machine learning
problem in these data is structured binary classification.
83. <b><a href="datasets/Chron"><a href="datasets/Chro"><a href="datasets/Chro"><a href="datasets/Chro"><a href="da
ic_Kidney_Disease">Chronic_Kidney_Disease</a></b>: 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. <b><a h
ref="datasets/Climate+Model+Simulation+Crashes">Climate Model Simulation Crashes</a></b>: Given Latin hy
percube samples of 18 climate model input parameter values, predict climate model simulation crashes and det
ermine the parameter value combinations that cause the failures.85. <b><a href="datas"><b><a href="datas"><a href="datas
ets/Cloud">Cloud</a></b>: Little Documentation86. <b><a href="datasets/CMU+Face+I
mages">CMU Face Images</a></b>: This data consists of 640 black and white face images of people taken wit
h varying pose (straight, left, right, up), expression (neutral, happy, sad, angry), eyes (wearing sunglasses or n
ot), and size87. <b><a href="datasets/CNAE-9">CNAE-9</a></b>: This is a data set co
ntaining 1080 documents of free text business descriptions of Brazilian companies categorized into a
subset of 9 categories88. <b><a href="datasets/Coil+1999+Competition+Data">Coil 199
9 Competition Data</a></b>: This data set is from the 1999 Computational Intelligence and Learning (COIL) co
mpetition. The data contains measurements of river chemical concentrations and algae densities.
class=
"normal">89. <b><a href="datasets/Combined+Cycle+Power+Plant">Combined Cycle Power Plant</a></b>: Th
e dataset contains 9568 data points collected from a Combined Cycle Power Plant over 6 years (2006-2011), w
hen the plant was set to work with full load. 90. <b><a href="datasets/Communities+and">40. <br/>
+Crime">Communities and Crime</a></b>: Communities within the United States. The data combines socio-ec
onomic data from the 1990 US Census, law enforcement data from the 1990 US LEMAS survey, and crime dat
a from the 1995 FBI UCR.91. <b><a href="datasets/Communities+and+Crime+Unnorm">+ Unnorm</a>
alized">Communities and Crime Unnormalized</a></b>: Communities in the US. Data combines socio-economi
c data from the '90 Census, law enforcement data from the 1990 Law Enforcement Management and Admin St
ats survey, and crime data from the 1995 FBI UCR92. <b><a href="datasets/Computer">+ datasets/Computer</a>
+Hardware">Computer Hardware</a></b>: Relative CPU Performance Data, described in terms of its cycle tim
```

e, memory size, etc.
class="normal">93, Concrete+Compressive+Strength

ete Compressive Strength: 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: 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. 4.5. ulsion+Plants">Condition Based Maintenance of Naval Propulsion Plants: 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.
class="normal">96. b<class="normal">b n+monitoring+of+hydraulic+systems">Condition monitoring of hydraulic systems: 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. >67. <b href="d ets/Congressional+Voting+Records">Congressional Voting Records: 1984 United Stated Congression al Voting Records; Classify as Republican or Democrat
p class="normal">98. Connect-4: Contains connect-4 positions99. Connectionist Bench (Nettalk Corpus): 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 phonemes100. +B ench+%28Sonar%2C+Mines+vs.+Rocks%29">Connectionist Bench (Sonar, Mines vs. Rocks): 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. 28Vowel +Recognition+-+Deterding+Data%29">Connectionist Bench (Vowel Recognition - Deterding Data) 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. Container Crane Controller Data Set: A container crane has the function of transporting contain ners from one point to another point.
class="normal">103. 103. <b href="datasets/C +Choice">Contraceptive Method Choice: Dataset is a subset of the 1987 National Indonesia Contrace ptive Prevalence Survey.104. Corel Ima ge Features: 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 ce105. Covertype: Forest CoverType datas et106. Credit Approval: This data conc erns credit card applications; good mix of attributes107. Crowdsourced Mapping: 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. Cryotherapy Dataset : This dataset contains information about wart treatment results of 90 patients using cryotherapy.<p cla ss="normal">109. 109. 109. <b style="color: blue;">109. < 4+and+2015">CSM (Conventional and Social Media Movies) Dataset 2014 and 2015: 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. /Cuff-Less+Blood+Pressure+Estimation">Cuff-Less Blood Pressure Estimation: 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. Cylinder Bands: U sed in decision tree induction for mitigating process delays known as "cylinder bands" in rotogravure printing</p >112. Daily and Sports Activities: 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. Daily Demand Foreca sting Orders: The dataset was collected during 60 days, this is a real database of a brazilian logistics c ompany.114. Daphnet Freezing of Gait: 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. 115. <b href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in">115. <b href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in">115. <b href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in">115. <b href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in">115. <b href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in">115. <b href="datasets/Data+for+Bottware+Engineering+Teamwork+Assessment+in">115. <b href="datasets/Data+for+Bottware+Engineering+Teamwork+Assessment+in">115. <b href="datasets/Data+for+Bottware+Engineering+Teamwork+Bottware+Engineering+Teamwork+Bottware+Engineering+Teamwork+Bottware+Engineering+Teamwork+Bottware+Engineering+Teamwork+Bottware+Engineering+Teamwork+Bottware+Bottware+Engineering+Teamwork+Bottware +Education+Setting">Data for Software Engineering Teamwork Assessment in Education Setting: 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, FAU11 6. Dataset for ADL Rec ognition with Wrist-worn Accelerometer: Recordings of 16 volunteers performing 14 Activities of Daily L iving (ADL) while carrying a single wrist-worn tri-axial accelerometer.117. Dataset for Sensorless Drive Diagnosis: Features

are extracted from motor current. The motor has intact and defective components. This results in 11 different classes with different conditions.

e-mails: 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. default of credit card clients: 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 bel+Multi-Instance+Learning+with+Instance+Labels">DeliciousMIL: A Data Set for Multi-Label Multi-Instance Le arning with Instance Labels: 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. Demospongiae: M arine sponges of the Demospongiae class classification domain.122. 122. <b href="data ets/Dermatology">Dermatology: Aim for this dataset is to determine the type of Eryhemato-Squamous Disease.123. D etect Malacious Executable(AntiVirus): I extract features from malacious and non-malacious and creat e and training dataset to teach sym classifier. Dataset made of unknown executable to detect if it is virus or nor mal safe executable.124. detection of IoT botnet attacks N BaIoT: 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. 125. <b href="datasets/Devanagari+Handwritten+C">125. < haracter+Dataset">Devanagari Handwritten Character Dataset: 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. 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. DGP2 - The Second Data Generation Program: Generates application domains based on specific parameters, number of features, and proportion of positive to negative examples128. Diabetes: This diabetes dataset is from AIM '94129. Diabetes 130-US hospitals for years 1999-2008: This data has been prepared to analyze factors related to readmission as well as other

outcomes pertaining to patients with diabetes.
class="normal">130</pr>
class="normal">130
class="normal">130
class="normal">130
class="normal">130
class="normal">130
class="normal">131
class="norm

132. Dishonest Internet user s Dataset: 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.
class="normal">133. Document Understanding: Five concepts, expresse d as predicates, to be learned134. Dodg ers Loop Sensor: Loop sensor data was collected for the Glendale on ramp for the 101 North freeway i n Los Angeles135. Dorothea: 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. Dota2 Games Results: 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.
class="normal">137. 4 ones+Index">Dow Jones Index: This dataset contains weekly data for the Dow Jones Industrial Index. It has been used in computational investing research.
class="normal">138. a href="datasets/Dresse" s Attribute Sales">Dresses Attribute Sales: This dataset contain Attributes of dresses and their reco mmendations according to their sales. Sales are monitor on the basis of alternate days. 139. DrivFace: 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. 28qu antified%29">Drug consumption (quantified): Classify type of drug consumer by personality data p class="normal">141. Drug Review Datas et (Druglib.com): 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. Drug Review Da taget (Drugg com) //ax //bx: The dataget provides nationt reviews on specific drugg along with related condition

taset (brags.com) \(\alpha \times \cdot \ s and a 10 star patient rating reflecting overall patient satisfaction.143. DSRC Vehicle Communications: This set Provides data regar ding wireless communications between vehicles and road side units. two separate data sets are provided (norm al scenario) and in the presence of attacker (jammer).144. 144. <b href="dat mic+Features+of+VirusShare+Executables">Dynamic Features of VirusShare Executables: This datas et contains the dynamic features of 107,888 executables, collected by VirusShare from Nov/2010 to Jul/2014.</ p>145. E. Coli Genes: Data giving character istics of each ORF (potential gene) in the E. coli genome. Sequence, homology (similarity to other genes) and s tructural information, and function (if known) are provided.146. Ear ly biomarkers of Parkinson's disease based on natural connected speech Data Set : .<p class="no rmal">147. 147. 147. 147. 147. Early biomarkers of Parkinson solutions based on natural connected speech: Predict a pa ttern of neurodegeneration in the dataset of speech features obtained from patients with early untreated Parkin son's disease and patients at high risk developing Parkinson's disease.
148. <a href</p> ="datasets/EBL+Domain+Theories">EBL Domain Theories: Assorted small-scale domain theories 149. Echocardiogram: Data for classifying if patients will survive for at least one year after a heart attack150. <a hr s/Eco-hotel">Eco-hotel: 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.151. < b>Ecoli: This data contains protein localization sites1 52. Economic Sanctions: Domain Theory on Economic S anctions; Undocumented153. 28. 48. <b href="datasets/Educational EPM%29%3A+A+Learning+Analytics+Data+Set">Educational Process Mining (EPM): A Learning Analytics Data Set: 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. EEG Database: This data arises from a large study to examine EEG correlates o f genetic predisposition to alcoholism. It contains measurements from 64 electrodes placed on the scalp sample d at 256 Hz155. EEG Eye State: The d ata set consists of 14 EEG values and a value indicating the eye state.156. EEG Steady-State Visual Evoked Potential Sig nals: This database consists on 30 subjects performing Brain Computer Interface for Steady State Visu al Evoked Potentials (BCI-SSVEP). 157. El Nino: The data set contains oceanographic and surface meteorological readings taken from a series of buoys posit ity+Simulated+Data+">Electrical Grid Stability Simulated Data : The local stability analysis of the 4-nod e star system (electricity producer is in the center) implementing Decentral Smart Grid Control concept. class="normal">159. ElectricityLoadDiagrams201120 14: This data set contains electricity consumption of 370 points/clients. 160. EMG data for gestures:

These are files of raw EMG data recorded by MYO Thalmic bracelet
class="normal">161. <a href="d</pre> atasets/EMG+dataset+in+Lower+Limb">EMG dataset in Lower Limb: 3 different exercises: sitting, stan ding and walking in the muscles: biceps femoris, vastus medialis, rectus femoris and semitendinosus addition to goniometry in the exercises.
class="normal">162. hr t">EMG Physical Action Data Set: 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. Energy efficiency cy: This study looked into assessing the heating load and cooling load requirements of buildings (that i s, energy efficiency) as a function of building parameters.164. Entree Chicago Recommendation Data: This data contains a r ecord of user interactions with the Entree Chicago restaurant recommendation system. 165. Epileptic Seizure Recognition: This dataset is a pre-processed and re-structured/reshaped version of a very commonly used dataset featuring epileptic seiz ure detection. 166. exten tion of Z-Alizadeh sani dataset: It was collected for CAD diagnosis.167. Facebook Comment Volume Dataset: Instanc es in this dataset contain features extracted from facebook posts. The task associated with the data is to predic t how many comments the post will receive.168. +metri cs">Facebook metrics: Facebook performance metrics of a renowned cosmetic's brand Facebook pag e.169. Farm Ads: This data was collected fr om text ads found on twelve websites that deal with various farm animal related topics. The binary labels are b

ased on whether of not the content owner approves of the ad. s/Fertility">Fertility: 100 volunteers provide a semen sample analyzed according to the WHO 2010 crit eria. Sperm concentration are related to socio-demographic data, environmental factors, health status, and life habits171. Firm-T eacher_Clave-Direction_Classification: The data are binary attack-point vectors and their clave-direction n class(es) according to the partido-alto-based paradigm.
class="normal">172. <b style="color: blue;">172. <b style="color: blue;">172. <b style="color: blue;">172. <b style="color: blue;">172. <b s st-order+theorem+proving">First-order theorem proving: Given a theorem, predict which of five heurist ics will give the fastest proof when used by a first-order prover. A sixth prediction declines to attempt a proof, sh ould the theorem be too difficult.173. Flags: Fro m Collins Gem Guide to Flags, 1986174. 174. <b href="dataset-F">174. < or+Music+Analysis">FMA: A Dataset For Music Analysis: FMA features 106,574 tracks and includes so ng title, album, artist, genres; play counts, favorites, comments; description, biography, tags; together with audi o (343 days, 917 GiB) and features.175. Folio: 20 photos of leaves for each of 32 different species.
176. <b style="color: blue;">b>Forest Fires: 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. Fore st type mapping: Multi-temporal remote sensing data of a forested area in Japan. The goal is to map d ifferent forest types using spectral data.178. F unction Finding: Cases collected mostly from investigations in physical science; intention is to evaluate function-finding algorithms179. 179. <b href="datasets/Gas+Sensor+Britan+Datasets/Gas+Britan+Dataset ">Gas Sensor Array Drift Dataset: This archive contains 13910 measurements from 16 chemical senso rs utilized in simulations for drift compensation in a discrimination task of 6 gases at various levels of concentrat ions.180. Gas Sensor Array Drift Dataset at Different Concentrations: This archive contains 13910 mea surements from 16 chemical sensors exposed to 6 different gases at various concentration levels. ="normal">181. Gas sensor ar ray exposed to turbulent gas mixtures: A chemical detection platform composed of 8 chemoresistive g as sensors was exposed to turbulent gas mixtures generated naturally in a wind tunnel. The acquired time serie s of the sensors are provided.182. 182. 182. 182. 182. 182. 182. Gas sensor array under dynamic gas mixtures: The data set contains the recordi ngs of 16 chemical sensors exposed to two dynamic gas mixtures at varying concentrations. For each mixture, signals were acquired continuously during 12 hours.183. 483. <b href="datasets/Gas+se" nsor+array+under+flow+modulation">Gas sensor array under flow modulation: The data set contains 5 8 time series acquired from 16 chemical sensors under gas flow modulation conditions. The sensors were expo sed to different gaseous binary mixtures of acetone and ethanol. sets/Gas+sensor+arrays+in+open+sampling+settings">Gas sensor arrays in open sampling settings: T he 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. Gas sensors for home activity monitoring</ a>: 100 recordings of a sensor array under different conditions in a home setting: background, wine and ba nana presentations. The array includes 8 MOX gas sensors, and humidity and temperature sensors. 186. Gastroi ntestinal Lesions in Regular Colonoscopy: This dataset contains features extracted from colonoscopy videos used to detect gastrointestinal lesions. It contains 76 lesions: 15 serrated adenomas, 21 hyperplastic lesi ons and 40 adenoma. 187. gene expression cancer RNA-Seq: This collection of data is part of the RNA-Seq (HiSeq) PANCAN dat a set, it is a random extraction of gene expressions of patients having different types of tumor: BRCA, KIRC, C aset+for+indoor+localisation+from+wristband+and+smartphone">Geo-Magnetic field and WLAN dataset for ind oor localisation from wristband and smartphone: A multisource and multivariate dataset for indoor local isation methods based on WLAN and Geo-Magnetic field fingerprinting189. Geographical Original of Music: Instances in this dataset contain audio features extracted from 1059 wave files. The task associated with the data is to predict the geogr aphical origin of music. 190. Gesture Phase Segmentation">

190. Gesture Phase Segmentationon: The dataset is composed by features extracted from 7 videos with people gesticulating, aiming at s tudying Gesture Phase Segmentation. It contains 50 attributes divided into two files for each video.class="normal">191. Gisette: GISETTE is a handwritten digit recognition pro blem. 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.

 $<\!/p\!><\!p\ class="normal">192.\ <\!b><\!a\ href="datasets/Glass+Identification">Glass\ Identification<\!/b>: From\ US$

A Forensic Science Service; 6 types of glass; defined in terms of their oxide content (i.e. Na, Fe, K, etc) class="normal">193. GNFUV Un manned Surface Vehicles Sensor Data: 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. -bed in Athens (Greece). cles+Sensor+Data+Set+2">GNFUV Unmanned Surface Vehicles Sensor Data Set 2: The data-set cont ains 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. GPS Trajectories: 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). p class="normal">196. Grammatical Facial Expressions">Grammatical Facial Expressions ns: This dataset supports the development of models that make possible to interpret Grammatical Faci al Expressions from Brazilian Sign Language (Libras).
197. Greenhouse Gas Observing Network: Design an observing network to monitor emissions of a greenhouse gas (GHG) in California given time series of synthetic observations and tr acers from weather model simulations.

198. Haberman's Survival: D ataset contains cases from study conducted on the survival of patients who had undergone surgery for breast c ancer199. Hayes-Roth: Topic: human subj ects study200. HCC Survival: Hepatocell ular Carcinoma dataset (HCC dataset) was collected at a University Hospital in Portugal. It contains real clinical data of 165 patients diagnosed with HCC.201. 401. <base>401.
401. <base>401.
401.
401.
401.
401.
401.
401.
401. <base>401.
401. <b witter">Health News in Twitter: 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. <p class="norm" al">202. Heart Disease: 4 databases: Cleveland, Hungary, Swit zerland, and the VA Long Beach
class="normal">203. Hepatitis: From G.Gong: CMU; Mostly Boolean or numeric-valued attribute types; Includes cost data (donated by Peter Turney)204. HEPMASS: The search for exo tic particles requires sorting through a large number of collisions to find the events of interest. This data set chal lenges one to detect a new particle of unknown mass.205. Heterogeneity Activity Recognition: The Heterogeneity Human Activity R ecognition (HHAR) dataset from Smartphones and Smartwatches is a dataset devised to benchmark human act ivity recognition algorithms (classification, automatic data segmentation, sensor fusion, feature extraction, etc.) i n real-world contexts; specifically, the dataset is gathered with a variety of different device models and use-sce narios, in order to reflect sensing heterogeneities to be expected in real deployments.
class="normal">2 06. HIGGS: This is a classification problem to distinguish between a sig nal process which produces Higgs bosons and a background process which does not. 2 07. Hill-Valley: Each record represents 100 points on a two-dimension nal graph. When plotted in order (from 1 through 100) as the Y co-ordinate, the points will create either a Hill (a ♦bump♦ in the terrain) or a Valley (a ♦dip♦ in the terrain).208. HIV-1 protease cleavage: The data contains lists of octamers (8 amino ac ids) and a flag (-1 or 1) depending on whether HIV-1 protease will cleave in the central position (between amino acids 4 and 5).
class="normal">209. Horse Colic: Well docu mented attributes; 368 instances with 28 attributes (continuous, discrete, and nominal); 30% missing values >210. HTRU2: Pulsar candidates collected during t he 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. 211. <b href="datasets/H tivity+Recognition+Using+Smartphones">Human Activity Recognition Using Smartphones: Human Acti vity Recognition database built from the recordings of 30 subjects performing activities of daily living (ADL) while e carrying a waist-mounted smartphone with embedded inertial sensors.
212. 212. <b hr f="datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer">Hybrid I ndoor Positioning Dataset from WiFi RSSI, Bluetooth and magnetometer: 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. ICMLA 2014 Accepted Papers Data Set: This data set compromi ses the metadata for the 2014 ICMLA conference's accepted papers, including ID, paper titles, author's keywor ds, abstracts and sessions in which they were exposed.214. ICU: Data set prepared for the use of participants for the 1994 AAAI Spring Symposium on Artificial In telligence in Medicine.215. IDA2016Challenge">IDA2016Challenge e: The dataset consists of data collected from heavy Scania trucks in everyday usage. <p class="n ormal">216. ILPD (Indian Liver Patient Dat aset): 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. <b><a href="datasets/Image+
Segmentation">Image Segmentation</a></b>: Image data described by high-level numeric-valued attributes, 7
classes218. <b><a href="datasets/lmmunotherapy+Dataset">lmmunotherapy+Dataset">lmmunotherapy+Dataset</a>
a></b>: This dataset contains information about wart treatment results of 90 patients using immunotherapy.
>219. <b><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 Monito
ring Parkinson's Disease</a></b>: Handwriting database consists of 25 PWP(People with Parkinson) and 15 he
althy individuals. Three types of recordings (Static Spiral Test, Dynamic Spiral Test and Stability Test) are taken.
220. <b><a href="datasets/Individual+household+electric+power+consumption">Individual+household+electric+power+consumption">Individual+household+electric+power+consumption</a>
ual household electric power consumption</a></b>: Measurements of electric power consumption in one house
hold with a one-minute sampling rate over a period of almost 4 years. Different electrical quantities and some s
rediction+from+RSS+data">Indoor User Movement Prediction from RSS data</a></b>: This dataset contains te
mporal data from a Wireless Sensor Network deployed in real-world office environments. The task is intended a
s real-life benchmark in the area of Ambient Assisted Living.
class="normal">222
class="normal">222
nsurance+Company+Benchmark+%28COIL+2000%29">Insurance Company Benchmark (COIL 2000)</a></b>:
This data set used in the CoIL 2000 Challenge contains information on customers of an insurance company. Th
e data consists of 86 variables and includes product usage data and socio-demographic data<p class="nor
mal">223. <b><a href="datasets/Internet+Advertisements">Internet Advertisements</a></b>: This dataset repr
esents a set of possible advertisements on Internet pages.224. <b><a href="datasets/In
ternet+Usage+Data">Internet Usage Data</a></b>: This data contains general demographic information on int
ernet users in 1997.225. <b><a href="datasets/lonosphere">lonosphere</a></b>: Class
ification of radar returns from the ionosphere
class="normal">226. <b><a href="datasets/IPUMS+Census">4
+Database">IPUMS Census Database</a></b>: This data set contains unweighted PUMS census data from the
Los Angeles and Long Beach areas for the years 1970, 1980, and 1990.
f="datasets/Iris">Iris</a></b>: Famous database; from Fisher, 1936228. <b><a href="d
atasets/ISOLET">ISOLET</a></b>: Goal: Predict which letter-name was spoken--a simple classification task.</
p>229. <b><a href="datasets/ISTANBUL+STOCK+EXCHANGE">ISTANBUL STOCK EXCHA
NGE</a></b>: 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. <b><a href="datasets/Japanese+Credit+Screening">Japanese Credit Screening</a></b>: Includes dom
ain theory (generated by talking to Japanese domain experts); data in Lisp231. <b><a h
ref="datasets/Japanese+Vowels">Japanese Vowels</a></b>: This dataset records 640 time series of 12 LPC c
epstrum coefficients taken from nine male speakers.232. <b><a href="datasets/KASAN"><a href="datase
DR">KASANDR</a></b>: KASANDR is a novel, publicly available collection for recommendation systems that re
cords the behavior of customers of the European leader in e-Commerce advertising, Kelkoo. <p class="nor
mal">233. <b><a href="datasets/KDC-4007+dataset+Collection">KDC-4007 dataset Collection</a></b>: KDC-4
007 dataset Collection is the Kurdish Documents Classification text used in categories regarding Kurdish Sorani
news and articles.
class="normal">234. <b><a href="datasets/KDD+Cup+1998+Data">KDD Cup 1998 D
ata</a></b>: This is the data set used for The Second International Knowledge Discovery and Data Mining Tool
s Competition, which was held in conjunction with KDD-98235. <b><a href="datasets/K">100.000 class="normal">235. <b><a href="datasets/K">235. <b href="datasets/
DD+Cup+1999+Data">KDD Cup 1999 Data</a></b>: This is the data set used for The Third International Knowl
edge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-99<p class="n
ormal">236. <b><a href="datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29">KEGG Metabolic
Reaction Network (Undirected)</a></b>: KEGG Metabolic pathways modeled as un-directed reaction network.
elation+Network+%28Directed%29">KEGG Metabolic Relation Network (Directed)</a></b>: KEGG Metabolic pa
thways modeled as directed relation network. Variety of graphical features presented.2
38. <b><a href="datasets/Kinship">Kinship</a></b>: Relational dataset239. <b><a href="datasets/kinship">Kinship</a></b>: Relational dataset
="datasets/Labor+Relations">Labor Relations</a></b>: From Collective Bargaining Review<p class="norm"
al">240. <b><a href="datasets/Las+Vegas+Strip">Las Vegas Strip</a></b>: This dataset includes quantitative a
nd categorical features from online reviews from 21 hotels located in Las Vegas Strip, extracted from TripAdvis
or (http://www.tripadvisor.com).241. <b><a href="datasets/Leaf">Leaf</a></b>: This dat
aset consists in a collection of shape and texture features extracted from digital images of leaf specimens origin
ating from a total of 40 different plant species.242. <b><a href="datasets/LED+Display+"
Domain">LED Display Domain</a></b>: From Classification and Regression Trees book; We provide here 2 C
programs for generating sample databases243. <b><a href="datasets/Legal+Case+Re
ports">Legal Case Reports</a></b>: A textual corpus of 4000 legal cases for automatic summarization and cita
tion analysis. For each document we collect catchphrases, citations sentences, citation catchphrases and citatio
n classes.244. <b><a href="datasets/Lenses">Lenses</a></b>: Database for fitting cont
act lenses245. <b><a href="datasets/Letter+Recognition">Letter Recognition</a></b>:
Database of character image features; try to identify the letter246. <b><a href="dataset"><a href=
```

s/Libras+Movement">Libras Movement: The data set contains 15 classes of 24 instances each. Each class references to a hand movement type in LIBRAS (Portuguese

name 'L�ngua BRAsileira de Sinais', oficial brazilian signal language).247. Liver Disorders: BUPA Medical Research Ltd. database donated by Richar d S. Forsyth248. Localiz ation Data for Person Activity: Data contains recordings of five people performing different activities. E ach person wore four sensors (tags) while performing the same scenario five times. 24 9. Logic Theorist: All code for Logic Theorist<p class="nor mal">250. Low Resolution Spectrometer: From IRAS data -- NASA Ames Research Center251. LSVT Voice Rehabilitation: 126 samples from 14 participants, 309 features. Aim: assess w hether voice rehabilitation treatment lead to phonations considered 'acceptable' or 'unacceptable' (binary class classification problem).252. Lung Cancer : Lung cancer data; no attribute definitions253. L ymphography: This lymphography domain was obtained from the University Medical Centre, Institute o f Oncology, Ljubljana, Yugoslavia. (Restricted access)
class="normal">254. c uberculosis+Genes">M. Tuberculosis Genes: Data giving characteristics of each ORF (potential gene) in the M. tuberculosis bacterium. Sequence, homology (similarity to other genes) and structural information, an d function (if known) are provided255. 255. 455. <b href="datasets/Machine+Le +ZZAlpha+Ltd.+Stock+Recommendations+2012-2014">Machine Learning based ZZAlpha Ltd. Stock Recomme ndations 2012-2014: The data here are the ZZAlpha® machine learning recommendations made for v arious US traded stock portfolios the morning of each day during the 3 year period Jan 1, 2012 - Dec 31, 2014. 256. Madelon: MADELON is an artificial datas et, which was part of the NIPS 2003 feature selection challenge. This is a two-class classification problem with c ontinuous input variables. The difficulty is that the problem is multivariate and highly non-linear.
class="n">"n ormal">257. MAGIC Gamma Telescope: Data are MC generated to simulate registration of high energy gamma particles in an atmospheric Cherenkov telescope 258. Mammographic Mass: Disc rimination of benign and malignant mammographic masses based on BI-RADS attributes and the patient's age. 259. Mechanical Analysis: Fault d iagnosis problem of electromechanical devices; also PUMPS DATA SET is newer version with domain theory an d results260. Mesothelioma's disease data set : 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. class="normal">261. Meta-data: Meta-Data was used in order to giv e advice about which classification method is appropriate for a particular dataset (taken from results of Statlog project).262. MEU-Mobile KSD: 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. 263. 263. 263. 263. 263. 263. 263. LTH+Dataset">MHEALTH Dataset: The MHEALTH (Mobile Health) dataset is devised to benchmark te chniques dealing with human behavior analysis based on multimodal body sensing.264. Mice Protein Expression: Expression levels of 77 pro teins 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. microblogPCU: 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. MicroMass: A dataset to explore machine I earning approaches for the identification of microorganisms from mass-spectrometry data.<p c l">267. MiniBooNE particle identification: Thi s dataset is taken from the MiniBooNE experiment and is used to distinguish electron neutrinos (signal) from m uon neutrinos (background).268. Misko lc IIS Hybrid IPS: The dataset was created for the comparison and evaluation of hybrid indoor positioni ng methods. The dataset presented contains data from W-LAN and Bluetooth interfaces, and Magnetometer. </ p>269. Mobile Robots: Learning concepts fr om sensor data of a mobile robot; set of data sets270. 270. 270. 270. 270. 270. <b and+Postures">MoCap Hand Postures: 5 types of hand postures from 12 users were recorded using u nlabeled markers attached to fingers of a glove in a motion capture environment. Due to resolution and occlusi on, missing values are common.271. 28Pro moter+Gene+Sequences%29">Molecular Biology (Promoter Gene Sequences): E. Coli promoter gene sequences (DNA) with partial domain theory272. 40.000 gy+%28Protein+Secondary+Structure%29">Molecular Biology (Protein Secondary Structure): From C

MU connectionist bench repository; Classifies secondary structure of certain globular proteins mal">273. Molecular Biol ogy (Splice-junction Gene Sequences): Primate splice-junction gene sequences (DNA) with associated imperfect domain theory274. MONK's Pr oblems: A set of three artificial domains over the same attribute space; Used to test a wide range of in duction algorithms275. Moral Reasoner</ b>: Horn-clause model that qualitatively simulates moral reasoning; Theory includes negated literals s="normal">276. Motion Capture Hand Postures /b>: 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. al">277. Movie: This data set contains a list of over 10000 films including g many older, odd, and cult films. There is information on actors, casts, directors, producers, studios, etc. p class="normal">278. MSNBC.com Anonymous Web Data: This data describes the page visits of users who visited msnbc.com on September 28, 199 9. Visits are recorded at the level of URL category (see description) and are recorded in time order. s="normal">279. Mturk User-Perceived Clusters over Images: 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. s="normal">280. Mult imodal Damage Identification for Humanitarian Computing: 5879 captioned images (image and text) fr om social media related to damage during natural disasters/wars, and belong to 6 classes: Fires, Floods, Natur al landscape, Infrastructural, Human, Non-damage.
class="normal">281. 281. <b href="datasets/Multiple Features">Multiple Features: This dataset consists of features of handwritten numerals (`0'--`9') extrac ted from a collection of Dutch utility maps282. Mushr oom: From Audobon Society Field Guide; mushrooms described in terms of physical characteristics; cl assification: poisonous or edible283. Musk (Version 1): The goal is to learn to predict whether new molecules will be musks or non-musks</ p>284. Musk (Version 2): The go al is to learn to predict whether new molecules will be musks or non-musks285. News Aggregator: References to news pages collected from an we b aggregator in the period from 10-March-2014 to 10-August-2014. The resources are grouped into clusters the at represent pages discussing the same story.286. 4 href="datasets/New rity+in+Multiple+Social+Media+Platforms">News Popularity in Multiple Social Media Platforms: Large d ata set of news items and their respective social feedback on multiple platforms: Facebook, Google+ and Linke dln.287. 287. <b href="datasets/Newspaper+and+magazine+images+segmentation+data">287. <b href="datasets/Newspaper+and+magazine+images+segmentation+data">287. <b href="datasets/Newspaper-and+magazine+images+segmentation+data">287. <b href="datasets/Newspaper-and+magazine+images+segmentation+data">287. <b href="datasets/Newspaper-and+magazine+images+segmentation+datasets/Newspaper-and+magazine+images+segmentation+datasets/Newspaper-and+magazine+images+segmentation+datasets/Newspaper-and+magazine+images+segmentation+datasets/Newspaper-and+magazine+images+segmentation+datasets/Newspaper-and+magazine+images+segmentation+datasets/Newspaper-and+magazine+images+segmentation+datasets/Newspaper-and+magazine+images+segmentation+datasets/Newspaper-and+magazine+images+segmentation+datasets/Newspaper-and+magazine+images+segmentation+datasets/Newspaper-and+magazine+images+segmentation+datasets/Newspaper-and+magazine+images+segmentation+datasets/Newspaper-and+magazine+images+segmentation+datasets/Newspaper-and+magazine+images+segmentation+datasets/Newspaper-and+magazine+images+segmentation+datasets/Newspaper-and+magazine+images-and+magazine+images-and+magazine+images-and+magazine+images-and+magazine+images-and+magazine+images-and+magazine+images-and+magazine+images-and+magazine+images-and+magazine+images-and+images-and+magazine+images-and+magazine+images-and+magazine+images-and+magazine+images-and+magazine+images-and+magazine+images-and+magazine+images-and+magazine+images-and+magazine+images-and set">Newspaper and magazine images segmentation dataset: Dataset is well suited for segmentation t asks. It contains 101 scanned pages from different newspapers and magazines in Russian with ground truth pix el-based masks.288. NI PS Conference Papers 1987-2015: This data set contains the distribution of words in the full text of the NIPS conference papers published from 1987 to 2015.
p class="normal">289. NoisyOffice: Corpus intended to do cleaning (or binarization) and enhancement of noisy grays cale printed text images using supervised learning methods. Noisy images and their corresponding ground trut h provided.290. Nomao: Nomao collects data a bout 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. Northix a>: Northix is designed to be a schema matching benchmark problem for data integration of two entity relat ionship databases. 292. NSF Research Award Abstracts 1990-2003: This data set consists of (a) 129,000 abstracts des cribing NSF awards for basic research, (b) bag-of-word data files extracted from the abstracts, (c) a list of word s used for indexing the bag-of-word293. Nursery</ b>: Nursery Database was derived from a hierarchical decision model originally developed to rank applications for nursery schools.294. NYSK: 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). datasets/Occupancy+Detection+">Occupancy Detection : Experimental data used for binary classificati on (room occupancy) from Temperature, Humidity, Light and CO2. Ground-truth occupancy was obtained from ti a+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes">OCT data & amp; Color Fundus Images of Left & am p; Right Eyes: This dataset contains OCT data (in mat format) and color fundus data (in jpg format) of I eft & amp; right eyes of 50 healthy persons.
class="normal">297. 4 healthy persons. ant+species+leaves+data+set">One-hundred plant species leaves data set: Sixteen samples of leaf ea

end with shopping, and the rest (1908) were positive class samples ending with shopping.302. 400. 400. 400. 400. 400. 400. 400. 400 +and+Transcoding+Time+Dataset">Online Video Characteristics and Transcoding Time Dataset: The dataset contains a million randomly sampled video instances listing 10 fundamental video characteristics along with the YouTube video ID. 303. 40. datasets/Open+University+Learning+Anal ytics+dataset">Open University Learning Analytics dataset: 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. 40. <base="datasets/Opinosis+Opinio">40. <base="datasets/Opinio">40. <base="data n+%26frasl%3B+Review">Opinosis Opinion / Review: 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. OpinRank Review Datas et: This data set contains user reviews of cars and and hotels collected from Tripadvisor (~259,000 +Activity+Recognition">OPPORTUNITY Activity Recognition: 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. Optical Interconnection Network : This dataset contains 640 performance measurements from a simulation of 2-Dimensional Multiprocessor Optical Interconnection Network. 308. Optical Recognition of Handwritten Digits: Two versions of this datab ase available; see folder309. Othello D omain Theory: Used in research to generate features for an inductive learning system<p class="n ormal">310. Ozone Level Detection: 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. p53 Mutants: The goal is to model mutant p53 transcriptional activity (active vs inactive) based on data extracted from biophysical simula tions.

312. Page Blocks Classification: The problem consists of classifying all the blocks of the page layout of a document that has been detected by a segmentation process.313. PAMAP2+Physical+Activity+Monitoring">PAMAP2 Physical Activity Monitoring: 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. PANDORPANDOR315. PANDOR315. Ab>Paper ReviewsPaper ReviewsPaper Reviews: This sentiment a nalysis 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.Paper Reviews="normal">316. Ab>316. Ab><a

Birmingham City Council. UK Open Government Licence (OGL).

https://data.birmingham.gov.uk/dataset/birmingham-parking
Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet">Parkinson Disease Spiral Drawings U sing Digitized Graphics Tablet
Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet">Parkinson Disease Spiral Drawings U sing Digitized Graphics Tablet
Parkinson Tablet
Parkinson) and 15 healthy individuals. Three types of recordings (Static Spiral Test, Dynamic Spiral Test and Stability Test) are taken.

Parkinson Speech Datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of Sound+Recordings
Parkinson Speech Dataset with Multiple Types of Sound Recordings

Parkinson's Disease (PD) patients and 20 healthy subjects. From all subjects, multiple ty pes of sound recordings (26) are taken.

Parkinson's Disease Classification

Parkinson's Disease Classification

The data used in this study were gathered from

```
m 188 patients with PD (107 men and 81 women) with ages ranging from 33 to 87 (65.1±10.9).
ormal">320. <b><a href="datasets/Parkinsons">Parkinsons</a></b>: Oxford Parkinson's Disease Detection Dat
aset321. <b><a href="datasets/Parkinsons+Telemonitoring">Parkinsons Telemonitoring</a>
</a></b>: Oxford Parkinson's Disease Telemonitoring Dataset322. <b><a href="dataset">322. <b><a href="dataset">4 href="d
s/PEMS-SF">PEMS-SF</a></b>: 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.
="normal">323. <b><a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits">Pen-Based Recognition
of Handwritten Digits</a></b>: Digit database of 250 samples from 44 writers324. <b><
a href="datasets/Perfume+Data">Perfume Data</a></b>: 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. <b><a href="datasets/Phishing+Websites">Phishing Websites</a></b>: This dataset coll
ected mainly from: PhishTank archive, MillerSmiles archive, Google's searching operators.
I">326. <b><a href="datasets/Physical+Unclonable+Functions">Physical Unclonable Functions</a></b>: 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. <b>
<a href="datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure">Physicochemical Properties of
Protein Tertiary Structure</a></b>: 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. <b><a href="datasets/Pioneer-1+Mobile+Robot+Data">Pioneer-1 Mobile Robot Data</
a></b>: 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
"normal">329. <b><a href="datasets/Pittsburgh+Bridges">Pittsburgh Bridges</a></b>: Bridges database that h
as original and numeric-discretized datasets330. <b><a href="datasets/Planning+Relax"
>Planning Relax</a></b>: The dataset concerns with the classification of two mental stages from recorded EEG
signals: Planning (during imagination of motor act) and Relax state. 331. <b><a href="d
atasets/Plants">Plants</a></b>: 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 class="norm"
al">332. <b><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. <b><a href="datasets"><a href="dataset
/PMU-UD">PMU-UD</a></b>: 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. <b><a href="datasets/Poker+Hand">Poker Hand</a></b>: Purpose is to predi
ct poker hands335. <b><a href="datasets/Polish+companies+bankruptcy+data">Polish
companies bankruptcy data</a></b>: 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. <b><a href="datasets/Post-Operative+Patient">Post-Operative Pa
tient</a></b>: Dataset of patient features337. <b><a href="datasets/Predict+keywords+"
activities+in+a+online+social+media">Predict keywords activities in a online social media</a></b>: 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. <b><a href=
"datasets/Primary+Tumor">Primary Tumor</a></b>: From Ljubljana Oncology Institute3
39. <b><a href="datasets/Prodigy">Prodigy</a></b>: Assorted domains like blocksworld, eightpuzzle, and sche
dworld.340. <b><a href="datasets/Protein+Data">Protein Data</a></b>: Undocumented
341. <b><a href="datasets/Pseudo+Periodic+Synthetic+Time+Series">Pseudo Periodic
Synthetic Time Series</a></b>: This data set is designed for testing indexing schemes in time series databases
. The data appears highly periodic, but never exactly repeats itself.
class="normal">342
class="normal"
asets/PubChem+Bioassay+Data">PubChem Bioassay Data</a></b>: 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. <b><a href="datasets/QSAR+biodegradation">QSAR biodegradation">QSAR biodegradation
gradation</a></b>: Data set containing values for 41 attributes (molecular descriptors) used to classify 1055 ch
emicals into 2 classes (ready and not ready biodegradable).
class="normal">344. <b><a href="datasets/"><a href="dataset
QtyT40I10D100K">QtyT40I10D100K</a></b>: 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
class="normal">34
5. <b><a href="datasets/Quadruped+Mammals">Quadruped Mammals</a></b>: The file animals.c is a data g
enerator of structured instances representing quadruped animals346. <b><a href="data"><b><a href="data"><a href="data">
sets/Qualitative+Structure+Activity+Relationships">Qualitative Structure Activity Relationships</a></b>: Two set
s of datasets are given: pyrimidines and triazines347. <b><a href="datasets/Qualitative">41. <b><a href="datasets/Qualitative">42. <b ><a href="datasets/Qualitative">43. <a href="datasets/Qualitative">4
 Bankruptcy">Qualitative Bankruptcy</a></b>: Predict the Bankruptcy from Qualitative parameters from expert
s.348. <b><a href="datasets/Quality+Assessment+of+Digital+Colposcopies">Quality As
sessment of Digital Colposcopies</a></b>: This dataset explores the subjective quality assessment of digital col
noccopies //n-/n class="normal"-340 /h-/a href-"datasets/Real estate valuation (data : set"-Real estate val
```

uation data set: 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. REALDISP Activity Recognition Dataset: 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 <p class="normal ">351. 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.
class="normal">3 52. Relative location of CT slices on axial axis: 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.
class="normal">353. Repeat Consumption Matrices: 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. Residential Building Data Set: 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. Re staurant & amp; consumer data: 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. <p class="normal ">356. Reuters RCV1 RCV2 Multilingual, Multiview Text Categorization Test collection: This test collection tion contains feature characteristics of documents originally written in five different languages and their translati ons, over a common set of 6 categories. 357. 4.500 class="normal">357. 4.500 class="normal">4.500 class="normal">4.50 ed+Subset">Reuters Transcribed Subset: This dataset is created by reading out 200 files from the 10 I argest Reuters

classes and using an Automatic Speech Recognition system to create

corresponding transcriptions.358. Reuters-21578 Text Categorization Collection: This is a collection of documents tha t appeared on Reuters newswire in 1987. The documents were assembled and indexed with categories. class="normal">359. Reuter_50_50: The dataset is used for aut horship identification in online Writeprint which is a new research field of pattern recognition. mal">360. Rice Leaf Diseases: 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. Robot Execution Failures : 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 intervals362. <a hr ef="datasets/Roman+Urdu+Data+Set">Roman Urdu Data Set: 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. Sales_Tr ansactions Dataset Weekly: Contains weekly purchased quantities of 800 over products over 52 wee ks. Normalised values are provided too.364. SCADI : First self-care activities dataset based on ICF-CY.365. 4 href="datasets/SEC"> OM">SECOM: Data from a semi-conductor manufacturing process366. seeds: 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. seismic-bumps: T he data describe the problem of high energy (higher than 10⁴ 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: 1593 handwritten digits from around 80 persons were scanned, stretched in a rectangular box 16x16 in a gray scale of 256 values. rmal">369. sEMG for Basic Hand movements : 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. mal">370. Sentence Classification: 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: The dataset contain s sentences labelled with positive or negative sentiment.372. ser Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machines%29">ser Knowledge+Levels+on+DC+Electrical+Machines%29">ser Knowledge+Levels+on+DC+Electrical+Machines%29 wledge Modeling Data (Students' Knowledge Levels on DC Electrical Machines): The dataset is about t he users' learning activities and knowledge levels on subjects of DC Electrical Machines. The dataset had been tained from anline web sources, and reported in my Dh.D. Thesis, the the class "normal", 272

obtained from online web-courses, and reported in my r n.b. Thesis. atasets/Servo">Servo: Data was from a simulation of a servo system374. < a href="datasets/SGEMM+GPU+kernel+performance">SGEMM GPU kernel performance: Running tim es for multiplying two 2048 x 2048 matrices using a GPU OpenCL SGEMM kernel with varying parameters (usi ng the library 'CLTune').375. Shuttle L FT10M">SIFT10M: 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.<p class= "normal">377. Simulated Falls a nd Daily Living Activities Data Set: 20 falls and 16 daily living activities were performed by 17 volunteer s with 5 repetitions while wearing 6 sensors (3.060 instances) that attached to their head, chest, waist, wrist, thi gh and ankle.378. SkillCraft1 Master Table Dataset: This data was used in Thompson et al. (2013). A list of possible game actions is discussed in Thompson, Blair, Chen, & href="datasets/">discussed in Thompson, Blair, Chen, & href="datasets/">datasets/</rr> Skin+Segmentation">Skin Segmentation: The Skin Segmentation dataset is constructed over B, G, R c olor space. Skin and Nonskin dataset is generated using skin textures from face images of diversity of age, gen der, and race people.380. Smartphone Dataset for Human A ctivity Recognition (HAR) in Ambient Assisted Living (AAL): This data is an addition to an existing datas et on UCI. We collected more data to improve the accuracy of our human activity recognition algorithms applied d+Recognition+of+Human+Activities+and+Postural+Transitions">Smartphone-Based Recognition of Human Act ivities and Postural Transitions: Activity recognition data set built from the recordings of 30 subjects per forming basic activities and postural transitions while carrying a waist-mounted smartphone with embedded iner

382. SML2010: This dataset is collected fro m a monitor system mounted in a domotic house. It corresponds to approximately 40 days of monitoring data. /p>383. SMS Spam Collection: The SMS Spam Collection is a public set of SMS labeled messages that have been collected for mobile phone spam research.384. Solar Flare: Each class attri bute counts the number of solar flares of a certain class that occur in a 24 hour period3 85. Somerville Happiness Survey: A data extrac t of a non-federal dataset posted here https://catalog.data.gov/dataset/somerville-happiness-survey-responses-2011-2013-2015386. Soybean (Large) : Michalski's famous soybean disease database387. 509 bean+%28Small%29">Soybean (Small): Michalski's famous soybean disease database<p class=" normal">388. Spambase: Classifying Email as Spam or Non-Spam 389. SPECT Heart: Data on cardiac Sing le Proton Emission Computed Tomography (SPECT) images. Each patient classified into two categories: norma I and abnormal.390. SPECTF Heart: D ata on cardiac Single Proton Emission Computed Tomography (SPECT) images. Each patient classified into tw o categories: normal and abnormal.391. S poken Arabic Digit: This dataset contains timeseries of mel-frequency cepstrum coefficients (MFCCs) c orresponding to spoken Arabic digits. Includes data from 44 male and 44 female native Arabic speakers. p class="normal">392. Sponge: Data on sponges; Attributes in spanish 393. Sports articles for objectivity analysis: 1000 sports articles were labeled using Amazon Mechanical Turk as objective or s ubjective. The raw texts, extracted features, and the URLs from which the articles were retrieved are provided.< /p>394. Statlog (Austral ian Credit Approval): This file concerns credit card applications. This database exists elsewhere in the repository (Credit Screening Database) in a slightly different form395. Statlog (German Credit Data): This dataset classifies peo ple described by a set of attributes as good or bad credit risks. Comes in two formats (one all numeric). Also co mes with a cost matrix396. Statlog (Hear t): This dataset is a heart disease database similar to a database already present in the repository (He art Disease databases) but in a slightly different form397. 40.5 +%28Image+Segmentation%29">Statlog (Image Segmentation): This dataset is an image segmentation n database similar to a database already present in the repository (Image segmentation database) but in a slig htly different form.398. Statlo g (Landsat Satellite): Multi-spectral values of pixels in 3x3 neighbourhoods in a satellite image, and the classification associated with the central pixel in each neighbourhood399. Statlog (Shuttle): The shuttle dataset contains 9 attributes all of whi

atasets/Statlog+%28Vehicle+Silhouettes%29">Statlog (Vehicle Silhouettes): 3D objects within a 2D ima ge by application of an ensemble of shape feature extractors to the 2D silhouettes of the objects.
rormal">401. Statlog Project: Various Databases: Vehicle silho uttes, Landsat Sattelite, Shuttle, Australian Credit Approval, Heart Disease, Image Segmentation, German Credit
rormal">402. Steel Plates Faults: A datas et of steel plates' faults, classified into 7 different types.

The goal was to train machine learning for automatic pattern recognition.

class="normal">403. Stock portfolio performance
a>: The data set of performances of weighted scoring stock portfolios are obtained with mixture design fro m the US stock market historical database.
class="normal">404. StoneFlakes">StoneFlakes
coneFlakes
coneFlakes
coneFlakes
coneFlakes
coneFlakes
coneFlakes
coneFl

the prehistoric era. The variables are means of geometric and

stylistic features of the flakes contained in different inventories.405. 405. <b href="datase"> ts/Student+Academics+Performance">Student Academics Performance: The dataset tried to find the e nd semester percentage prediction based on different social, economic and academic attributes. <p class= "normal">406. Student Loan Relational: Student Loa n Relational Domain
407. Student Performance mance: Predict student performance in secondary education (high school). 40 8. Superconductivty Data: Two file s contain data on 21 263 superconductors and their relevant features.
class="normal">409. SU SY: This is a classification problem to distinguish between a signal process which produces supersym metric particles and a background process which does not.410. Synthetic Control Chart Time Series: This data consists of synt hetically generated control charts.411. 411. 411. 411. 411. Syskill and Webert Web Page Ratings: 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 (Bandsrecording artists; Goats; Sheep; and BioMedical)412. Tamilnadu Electricity Board Hourly Readings: This data can be ef fectively produced the result to fewer parameter of the Load profile can be reduced in the Database ss="normal">413. Tarvel Review Ratings: Google revi ews on attractions from 24 categories across Europe are considered. Google user rating ranges from 1 to 5 an d average user rating per category is calculated.414. 414. 414. Taxi Service Trajectory - Prediction Challenge , ECML PKDD 2015: An accurate dataset describing trajectories performed by all the 442 taxis running in the city of Porto, in Portugal.

415. Teaching Assistant Evaluat ion: The data consist of evaluations of teaching performance; scores are "low", "medium", or "high"</p >416. Tennis Major To urnament Match Statistics: 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. Thoracic Surgery Data</ b>: The data is dedicated to classification problem related to the post-operative life expectancy in the lung canc er patients: class 1 - death within one year after surgery, class 2 - survival.418. <a h ref="datasets/Thyroid+Disease">Thyroid Disease: 10 separate databases from Garavan Institute p class="normal">419. Tic-Tac-Toe Endgame: Binary c lassification task on possible configurations of tic-tac-toe game420. 420. <b href="da ts/Trains">Trains: 2 data formats (structured, one-instance-per-line)421. a href="datasets/Travel+Reviews">Travel Reviews: Reviews on destinations in 10 categories mentione d across East Asia. Each traveler rating is mapped as Excellent(4), Very Good(3), Average(2), Poor(1), and Ter rible(0) and average rating is used.422. 425.

datasets/TTC-3600%3A+Benchm ark+dataset+for+Turkish+text+categorization">TTC-3600: Benchmark dataset for Turkish text categorization: The TTC-3600 data set is a collection of Turkish news and articles including categorized 3,600 documen ts from 6 well-known portals in Turkey. It has 4 different forms in ARFF Weka format.
class="normal">42 3. Turkiye Student Evaluation: This data set contai ns 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.
t V+News+Channel+Commercial+Detection+Dataset">TV News Channel Commercial Detection Dataset : 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. Twenty Newsgroups: This data set consists of 2000 0 messages taken from 20 newsgroups.426. 426. <b href="datasets/Twi

as conditions (4 volatiles at 10 concentration levels each).
class="normal">427. Twitter Data set for Arabic Sentiment Analysis
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.
class="normal">428. UbiqLog (smartphone lifelogging)
UbiqLog is the smartphone lifelogging tool that runs on the smartphone of 35 users for about 2 months.

429. UJI Pen Characters: Data co nsists of written characters in a UNIPEN-like format430. 430. <b href="datasets +Characters+%28Version+2%29">UJI Pen Characters (Version 2): A pen-based database with more t han 11k isolated handwritten characters
431. UJII ndoorLoc: The UJIIndoorLoc is a Multi-Building Multi-Floor indoor localization database to test Indoor Positioning System that rely on WLAN/WiFi fingerprint.432. 432. <b href="datase doorLoc-Mag">UJIIndoorLoc-Mag: The UJIIndoorLoc-Mag is an indoor localization database to test In door Positioning System that rely on Earth's magnetic field variations.
class="normal">433. datasets/Ultrasonic+flowmeter+diagnostics">Ultrasonic flowmeter diagnostics: Fault diagnosis of four li quid ultrasonic flowmeters434. Undocumented< /a>: Various datasets without documentation (feel free to explore!)435. University: Data in original (LISP-readable) form436. University of Tehran Question Dataset 2016 (UTQD.2016): Persian questions gathered from a jeopardy game broadcasted on Iranian national television. 437. UNIX User Data: 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.
class="normal">438. Urban Land Cover: Classification of urban land cover using high resolution aerial ima gery. Intended to assist sustainable urban planning efforts.439. RL+Reputation">URL Reputation: Anonymized 120-day subset of the ICML-09 URL data containing 2. 4 million examples and 3.2 million features.
440. 440. <b href="datasets/US+Census+Datasets/US+C a+%281990%29">US Census Data (1990): 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. User Identification n From Walking Activity: 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. User Knowledge Modeling

: 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. 443. <b 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: Data used for USPTO Algorithm Competition. Contains drawing pages from US patents with manually labeled figure and part labels.444. Vertebral Column</ 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. Vicon Physical Action Data Set: 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. 446. <b hr ts/Victorian+Era+Authorship+Attribution">Victorian Era Authorship Attribution: 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 50447. +447. <b href="datasets/V">+ olcanoes+on+Venus+-+JARtool+experiment">Volcanoes on Venus - JARtool experiment: 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. 448. <b hre ollowing+Robot+Navigation+Data">Wall-Following Robot Navigation Data: 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. Water Treatment Plant: Multiple classes predict plant state<p class="nor mal">450. Waveform Database G enerator (Version 1): CART book's waveform domains451. 451. 451. ets/Waveform+Database+Generator+%28Version+2%29">Waveform Database Generator (Version 2): CART book's waveform domains452. Wearable Computing: Classification of Body Postures and Movements (PUC-Rio): A dataset with 5 classes (sitting-down, standing-up, standing-up, standing-up) g, walking, and sitting) collected on 8 hours of activities of 4 healthy subjects. We also established a baseline pe rformance index.453. Website Phishing:

454. 454. 454. 454. 454. 454. 454. <baseline="datasets/weight+Lifting+Exercises+monitored+with+Inertial+Measure">454. 454. Weight Lifting Exercises monitored with Inertial Measurement Units: Six young health sub jects were asked to perform 5 variations of the biceps curl weight lifting exercise. One of the variations is the on e+Stress+and+Affect+Detection%29">WESAD (Wearable Stress and Affect Detection): WESAD (Wear able Stress and Affect Detection) contains data of 15 subjects during a stress-affect lab study, while wearing ph vsiological and motion sensors.456. Whol esale customers: The data set refers to clients of a wholesale distributor. It includes the annual spendi ng in monetary units (m.u.) on diverse product categories457. <b style="color: blue;">b class="normal">457. <b style="color: blue;">457. <b style="color: blue;">b class="normal">457. <b style="color: blue;">b class="normal">457. <b style="color: blue;">457. <b style="color: blue;"> i4HE">wiki4HE: Survey of faculty members from two Spanish universities on teaching uses of Wikipedi a458. Wilt: High-resolution Remote Sensing data s et (Quickbird). Small number of training samples of diseased trees, large number for other land cover. Testing data set from stratified random sample of image.
class="normal">459. Win e: Using chemical analysis determine the origin of wines460. Wine Quality: Two datasets are included, related to red and white vinho verde win e 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. 461. <b href="dataset s/Wireless+Indoor+Localization">Wireless Indoor Localization: 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. Yacht Hydrodynamics: Delf t data set, used to predict the hydodynamic performance of sailing yachts from dimensions and velocity. class="normal">463. YearPredictionMSD: 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.
class="normal">464. Yeast: Predicting the Cellular Localization Sites of Proteins465. 465. <b hre +Comedy+Slam+Preference+Data">YouTube Comedy Slam Preference Data: This dataset provides u ser 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. 466. <b href="da ts/YouTube+Multiview+Video+Games+Dataset">YouTube Multiview Video Games Dataset: This datas et contains about 120k instances, each described by 13 feature types, with class information, specially useful fo r exploring multiview topics (cotraining, ensembles, clustering,..).467. YouTube Spam Collection: It is a public set of comments collected fo r spam research. It has five datasets composed by 1,956 real messages extracted from five videos that were a mong the 10 most viewed on the collection period.468. 468. <b href="datasets/Z-Aliz h+Sani">Z-Alizadeh Sani: It was collected for CAD diagnosis.469. Zoo: Artificial, 7 classes of animals<hr/>

```
Supported By:
<img height="60" src="assets/nsfe.gif"/> 
 In Collaboration With:
<img src="assets/rexaSmall.jpg"/>
<center>
<span class="normal">
<a href="about.html">About</a> ||
<a href="citation">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 [179]:

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',

```
i rogriosiic ,
'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 ',
```

```
AIIII ,
Ί',
'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 ',
'lt ',
'lt ',
'Artificial,']
```