

Taller 8

Métodos Computacionales para Políticas Públicas - URosario

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

César Mateo Lancheros Cañón

cesar.lancheros@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.

In [1]:

```
import re
```

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 [2]:

```
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 [3]:

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

Out[3]:

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

2. [1 punto]

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

In [4]:

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

Out [4]:

```
['75%', '4%']
```

3. [2 puntos]

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

In []:

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

In [5]:

```
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) / 590.1 (Spring 2018): Ethics and AI. COMPSCI 590.2 (Spring 2017): Computation, Information, and Learning in Market Design. COMPSCI 590.4 (Spring 2016): Computational Microeconomics: Game Theory, Social Choice, and Mechanism Design. COMPSCI 290.4/590.4 (Spring 2015): Crowdsourcing Societal Tradeoffs. COMPSCI 570 (Fall 2014): Artificial Intelligence. COMPSCI 590.4 (Spring 2014): Computational Microeconomics: Game Theory, Social Choice, and Mechanism Design. COMPSCI 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: Game Theory, Social Choice, and Mechanism Design. 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 [6]:

```
re.findall("COMPSCI\s([1-9]+)", ob2)
```

Out [6]:

```
['27',
 '59',
 ...]
```

```
'223',
'57',
'59',
'59',
'59',
'29',
'57',
'59',
'59',
'173',
'296',
'296',
'173',
'196',
'17',
'27',
'196',
'196',
'296',
'296']
```

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.\n", "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.\n", "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.\n", "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.\n", "McDonnell, TE, Bail, CA, and Tavory, I. \"A Theory of Resonance.\" Sociological Theory 35.1 (March 1, 2017): 1-14. Full Text.\n", "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.\n", "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.\n", "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.\n", "Bail, CA. \"The cultural environment: Measuring culture with big data.\" Theory and Society 43.3 (January 1, 2014): 465-524. Full Text."

In [115]:

```
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.\n\", \"\"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.\n\", \"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.\n\", \"\"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.\n\", \"McDonnell, TE, Bail, CA, and Tavory, I. \"\"A Theory of Resonance.\" \"Sociological Theory 35.1 (March 1, 2017): 1-14. Full Text.\n\", \"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.\n\", \"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.\n\", \"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.\n\", \"Bail, CA. \"\"The cultural environment: Measuring culture with big data.\" \"Theory and Society 43.3 (January 1, 2014): 465-524. Full Text.\""
```

6. [10 puntos]

Vamos a hacer "scrapping" a esta página: <https://archive.ics.uci.edu/ml/datasets/nhn> que contiene un listado de 468 bases de datos

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=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list>)

Diviértase.

Web Scrapping Base de datos Repositorio Universidad de California

In [7]:

```
import requests
from bs4 import BeautifulSoup
import pandas as pd
import re
```

In [8]:

```
URL = 'https://archive.ics.uci.edu/ml/datasets.php'
```

In [9]:

```
HTML = requests.get(URL).text
```

In [10]:

```
database = BeautifulSoup(HTML, "lxml")
```

In [11]:

```
database
```

Out[11]:

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

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

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

```

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

// ** END **
return true ;
}

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

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

```

```

</tr>
</table>
<br/>
<table cellpadding="3">
<tr>
<td valign="top">
<table><tr><td><p align="center" class="big">Browse Through:</p></td></tr></table>
<table border="2" cellpadding="2" width="150">
<tr> <td bgcolor="#003366"><p class="whitetext"><b>Default Task</b> </p>
</td>
</tr>
<tr>
<td valign="top"><p class="normal"><a href="datasets.php?format=&task=cla&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=e">Classification</a> <font color="red">(350)</font><br/><a href="datasets.php?format=&task=reg&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=e">Regression</a> <font color="red">(96)</font><br/><a href="datasets.php?format=&task=clu&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=e">Clustering</a> <font color="red">(84)</font><br/><a href="datasets.php?format=&task=other&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=e">Other</a> <font color="red">(55)</font> </p>
</td>
</tr>
<tr> <td bgcolor="#003366"><p class="whitetext"><b>Attribute Type</b> </p>
</td>
</tr>
<tr>
<td valign="top"><p class="normal"><a href="datasets.php?format=&task=&att=cat&area=&numAtt=&numIns=&type=&sort=nameUp&view=e">Categorical</a> <font color="red">(38)</font><br/><a href="datasets.php?format=&task=&att=num&area=&numAtt=&numIns=&type=&sort=nameUp&view=e">Numerical</a> <font color="red">(307)</font><br/><a href="datasets.php?format=&task=&att=mix&area=&numAtt=&numIns=&type=&sort=nameUp&view=e">Mixed</a> <font color="red">(55)</font> </p>
</td>
</tr>
<tr> <td bgcolor="#003366"><p class="whitetext"><b>Data Type</b> </p>
</td>
</tr>
<tr>
<td valign="top"><p class="normal"><a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=mvar&sort=nameUp&viewle">Multivariate</a> <font color="red">(357)</font><br/><a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=uvar&sort=nameUp&viewle">Univariate</a> <font color="red">(23)</font><br/><a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=seq&sort=nameUp&viewe">Sequential</a> <font color="red">(47)</font><br/><a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=ts&sort=nameUp&view=t">Time-Series</a> <font color="red">(91)</font><br/><a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=text&sort=nameUp&viewle">Text</a> <font color="red">(53)</font><br/><a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=dt&sort=nameUp&view=t">Domain-Theory</a> <font color="red">(23)</font><br/><a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=other&sort=nameUp&viewle">Other</a> <font color="red">(21)</font><br/> </p>
</td>
</tr>
<tr><td bgcolor="#003366"><p class="whitetext"><b>Area</b> </p>
</td>
</tr>
<tr>
<td valign="top"><p class="normal"><a href="datasets.php?format=&task=&att=&area=life&numAtt=&numIns=&type=&sort=nameUp&viewle">Life Sciences</a> <font color="red">(107)</font><br/><a href="datasets.php?format=&task=&att=&area=phys&numAtt=&numIns=&type=&sort=nameUp&viewle">Physical Sciences</a> <font color="red">(49)</font><br/><a href="datasets.php?format=&task=&att=&area=comp&numAtt=&numIns=&type=&sort=nameUp&viewle">CS / Engineering</a> <font color="red">(170)</font><br/><a href="datasets.php?format=&task=&att=&area=soc&numAtt=&numIns=&type=&sort=nameUp&viewe">Social Sciences</a> <font color="red">(26)</font><br/><a href="datasets.php?format=&task=&att=&area=bus&numAtt=&numIns=&type=&sort=nameUp&viewe">Business</a> <font color="red">(29)</font><br/><a href="datasets.php?format=&task=&att=&area=game&numAtt=&numIns=&type=&sort=nameUp&viewle">Game</a> <font color="red">(10)</font><br/><a href="datasets.php?format=&task=&att=&area=other&numAtt=&numIns=&type=&sort=nameUp&viewle">Other</a> <font color="red">(73)</font> </p>
</td>

```

```

</tr>
<tr><td bgcolor="#003366"><p class="whitetext"><b># Attributes</b> </p>
</td>
</tr>
<tr>
<td valign="top"><p class="normal"><a href="datasets.php?format=&task=&att=&area=&numAtt=less10&numIns=&type=&sort=nameUp&visible">Less than 10</a> <font color="red">(113)</font><br/><a href="datasets.php?format=&task=&att=&area=&numAtt=10to100&numIns=&type=&sort=nameUp&visible">10 to 100</a> <font color="red">(210)</font><br/><a href="datasets.php?format=&task=&att=&area=&numAtt=greater100&numIns=&type=&sort=nameUp&view=table">Greater than 100</a> <font color="red">(84)</font> </p>
</td>
</tr>
<tr><td bgcolor="#003366"><p class="whitetext"><b># Instances</b></p></td>
</tr>
<tr>
<td valign="top"><p class="normal">
<a href="datasets.php?format=&task=&att=&area=&numAtt=less100&numIns=&type=&sort=nameUp&view=table">Less than 100</a> <font color="red">(27)</font><br/><a href="datasets.php?format=&task=&att=&area=&numAtt=100to1000&numIns=&type=&sort=nameUp&view=table">100 to 1000</a> <font color="red">(162)</font><br/><a href="datasets.php?format=&task=&att=&area=&numAtt=greater1000&numIns=&type=&sort=nameUp&view=table">Greater than 1000</a> <font color="red">(246)</font> </p>
</td>
</tr>
<tr><td bgcolor="#003366"><p class="whitetext"><b>Format Type</b> </p>
</td>
</tr>
<tr>
<td valign="top"><p class="normal"><a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=e">Matrix</a> <font color="red">(324)</font><br/><a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=nonmat">Non-Matrix</a> <font color="red">(145)</font> </p>
</td>
</tr></table>
</td>
<td valign="top">
<table width="100%">
<tr>
<td><p class="big"><b>469</b> Data Sets</p></td>
<td align="right"><p class="normal"><font color="gray">Table View</font> <a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list">List View</a></p></td>
</tr>
</table><table border="1" cellpadding="5"><tr bgcolor="#003366">
<td class="normal, whitetext"><p class="normal, whitetext"><a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameDown&view=e"><b>Name</b></a></p></td>
<!-- <td><p class="normal, whitetext"><b>Abstract</b></p></td> -->
<td><p class="normal, whitetext"><a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=typeUp&view=tak"><b>Data Types</b></a></p></td>
<td><p class="normal, whitetext"><a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=taskUp&view=tak"><b>Default Task</b></a></p></td>
<td><p class="normal, whitetext"><a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=attTypeUp&view=e"><b>Attribute Types</b></a></p></td>
<td><p class="normal, whitetext"><a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=instUp&view=tak"><b>Instances</b></a></p></td>
<td><p class="normal, whitetext"><a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=attUp&view=tabl"><b># Attributes</b></a></p></td>
<td><p class="normal, whitetext"><a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=dateUp&view=tak"><b>Year</b></a></p></td>
<!-- <td><p class="normal, whitetext"><b>Area</b></p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Abalone"></a> </td><td><p class="normal"><b><a href="datasets/Abalone">Abalone</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Predict the age of abalone from physical measurements<br/></p></td> -->

```

```

<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer, Real </p></td>
<td><p class="normal">4177 </p></td>
<td><p class="normal">8 </p></td>
<td><p class="normal">1995 </p></td>
<!-- <td><p class="normal">Life &nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Adult"></a> </td><td><p class="normal"><b><a href="datasets/Adult">Adult</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Predict whether income exceeds $50K/yr based on census data. Also known as "Census Income" dataset.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">48842 </p></td>
<td><p class="normal">14 </p></td>
<td><p class="normal">1996 </p></td>
<!-- <td><p class="normal">Social&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Annealing"></a> </td><td><p class="normal"><b><a href="datasets/Annealing">Annealing</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Steel annealing data&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer, Real </p></td>
<td><p class="normal">798 </p></td>
<td><p class="normal">38 </p></td>
<td><p class="normal"> </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Anonymous+Microsoft+Web+Data"></a> </td><td><p class="normal"><b><a href="datasets/Anonymous+Microsoft+Web+Data">Anonymous Microsoft Web Data</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbsp;</p></td> -->
<td><p class="normal"> </p></td>
<td><p class="normal">Recommender-Systems </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">37711 </p></td>
<td><p class="normal">294 </p></td>
<td><p class="normal">1998 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Arrhythmia"></a> </td><td><p class="normal"><b><a href="datasets/Arrhythmia">Arrhythmia</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Distinguish between the presence and absence of cardiac arrhythmia and classify it in one of the 16 groups.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer, Real </p></td>
<td><p class="normal">452 </p></td>
<td><p class="normal">279 </p></td>
<td><p class="normal">1998 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Artificial+Characters"></a> </td><td><p class="normal"><b><a href="datasets/Artificial+Characters">Artificial Characters</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Dataset artificially generated by using first order theory which describes structure of ten capital letters of English alphabet&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer, Real </p></td>
<td><p class="normal">6000 </p></td>
<td><p class="normal">7 </p></td>
<td><p class="normal">1992 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Audiology+%28Original%29"></a> </td><td><p class="normal"><b><a href="datasets/Audiology+%28Original%29">Audiology (Original)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Nominal audiology dataset from Ravelor&nbsp;</p></td> -->

```

.. class="normal">Normal & Nominal audiology, database from Baytak et al., 1997, 1998

<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">226 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">1987 </p></td>
<!-- <td><p class="normal">Life </p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">Audiology (Standardized)</p></td></tr></table></td>
--

</td>

<!-- <td><p class="normal">Standardized version of the original audiology database </p></td> -->

<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">226 </p></td>
<td><p class="normal">69 </p></td>
<td><p class="normal">1992 </p></td>
<!-- <td><p class="normal">Life </p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">Auto MPG</p></td></tr></table></td>
--

<!-- <td><p class="normal">Revised from CMU StatLib library, data concerns city-cycle fuel consumption </p></td> -->

<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Categorical, Real </p></td>
<td><p class="normal">398 </p></td>
<td><p class="normal">8 </p></td>
<td><p class="normal">1993 </p></td>
<!-- <td><p class="normal">Other </p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">Automobile</p></td></tr></table></td>

<!-- <td><p class="normal">From 1985 Ward's Automotive Yearbook </p></td> -->

<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Categorical, Integer, Real </p></td>
<td><p class="normal">205 </p></td>
<td><p class="normal">26 </p></td>
<td><p class="normal">1987 </p></td>
<!-- <td><p class="normal">Other </p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">Badges</p></td></tr></table></td>
--

<!-- <td><p class="normal">Badges labeled with a "+" or "-" as a function of a person's name </p></td> -->

<td><p class="normal">Univariate, Text </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">294 </p></td>
<td><p class="normal">1 </p></td>
<td><p class="normal">1994 </p></td>
<!-- <td><p class="normal">Other </p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">Balance Scale</p></td></tr></table></td>
--

<!-- <td><p class="normal">Balance scale weight & distance database </p></td> -->

<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">625 </p></td>
<td><p class="normal">4 </p></td>
<td><p class="normal">1994 </p></td>
<!-- <td><p class="normal">Social </p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">Balloons</p></td></tr></table></td>

<!-- <td><p class="normal">Data previously used in cognitive psychology experiment; 4 data sets represent different conditions of an experiment </p></td> -->

present different conditions of an experimental setup, etc. -->

<p class="normal">Multivariate</p>	<p class="normal">Classification</p>	<p class="normal">Categorical</p>			
<p class="normal">16</p>	<p class="normal">4</p>	<p class="normal"></p></td></p>			
<p class="normal"></p></td></p>	<p class="normal"></p></td> --></p>	<p class="normal"><!-- <td><p class="normal">Social&nbsp;:</p></td> --></p>			
<tr><td colspan="3" style="background-color: #DDEEFF;"><p class="normal"><td><table><tr><td> </td><td><p class="normal">Breast Cancer</p></td></tr></table></td></p></td></tr>			<p class="normal"><td><table><tr><td> </td><td><p class="normal">Breast Cancer</p></td></tr></table></td></p>		
<p class="normal"><td><table><tr><td> </td><td><p class="normal">Breast Cancer</p></td></tr></table></td></p>					
<p class="normal"><!-- <td><p class="normal">Breast Cancer Data (Restricted Access) &nbsp;:</p></td> --></p>					
<p class="normal"><td><p class="normal">Multivariate</p>	<p class="normal"></p></td></p>	<p class="normal"><td><p class="normal">Classification</p>			
<p class="normal"></p></td></p>	<p class="normal"><td><p class="normal">Categorical</p>	<p class="normal"></p></td></p>			
<p class="normal"><td><p class="normal">286</p>	<p class="normal"></p></td></p>	<p class="normal"><td><p class="normal">9</p>			
<p class="normal"></p></td></p>	<p class="normal"><td><p class="normal">1988</p>	<p class="normal"></p></td></p>			
<p class="normal"><!-- <td><p class="normal">Life&nbsp;:</p></td> --></p>	<p class="normal"></td></p>	<p class="normal"></td></p>			
<tr><td colspan="3" style="background-color: #DDEEFF;"><p class="normal"><td><table><tr><td> </td><td><p class="normal">Breast Cancer Wisconsin (Original)</p></td></tr></table></td></p></td></tr>			<p class="normal"><td><table><tr><td> </td><td><p class="normal">Breast Cancer Wisconsin (Original)</p></td></tr></table></td></p>		
<p class="normal"><td><table><tr><td> </td><td><p class="normal">Breast Cancer Wisconsin (Original)</p></td></tr></table></td></p>					
<p class="normal"><!-- <td><p class="normal">Original Wisconsin Breast Cancer Database &nbsp;:</p></td> --></p>					
<p class="normal"><td><p class="normal">Multivariate</p>	<p class="normal"></p></td></p>	<p class="normal"><td><p class="normal">Classification</p>			
<p class="normal"></p></td></p>	<p class="normal"><td><p class="normal">Integer</p>	<p class="normal"></p></td></p>			
<p class="normal"><td><p class="normal">699</p>	<p class="normal"></p></td></p>	<p class="normal"><td><p class="normal">10</p>			
<p class="normal"></p></td></p>	<p class="normal"><td><p class="normal">1992</p>	<p class="normal"></p></td></p>			
<p class="normal"><!-- <td><p class="normal">Life&nbsp;:</p></td> --></p>	<p class="normal"></td></p>	<p class="normal"></td></p>			
<tr><td colspan="3" style="background-color: #DDEEFF;"><p class="normal"><td><table><tr><td> </td><td><p class="normal">Breast Cancer Wisconsin (Prognostic)</p></td></tr></table></td></p></td></tr>			<p class="normal"><td><table><tr><td> </td><td><p class="normal">Breast Cancer Wisconsin (Prognostic)</p></td></tr></table></td></p>		
<p class="normal"><td><table><tr><td> </td><td><p class="normal">Breast Cancer Wisconsin (Prognostic)</p></td></tr></table></td></p>					
<p class="normal"><!-- <td><p class="normal">Prognostic Wisconsin Breast Cancer Database &nbsp;:</p></td> --></p>					
<p class="normal"><td><p class="normal">Multivariate</p>	<p class="normal"></p></td></p>	<p class="normal"><td><p class="normal">Classification, Regression</p>			
<p class="normal"></p></td></p>	<p class="normal"><td><p class="normal">Real</p>	<p class="normal"></p></td></p>			
<p class="normal"><td><p class="normal">198</p>	<p class="normal"></p></td></p>	<p class="normal"><td><p class="normal">34</p>			
<p class="normal"></p></td></p>	<p class="normal"><td><p class="normal">1995</p>	<p class="normal"></p></td></p>			
<p class="normal"><!-- <td><p class="normal">Life&nbsp;:</p></td> --></p>	<p class="normal"></td></p>	<p class="normal"></td></p>			
<tr><td colspan="3" style="background-color: #DDEEFF;"><p class="normal"><td><table><tr><td> </td><td><p class="normal">Breast Cancer Wisconsin (Diagnostic)</p></td></tr></table></td></p></td></tr>			<p class="normal"><td><table><tr><td> </td><td><p class="normal">Breast Cancer Wisconsin (Diagnostic)</p></td></tr></table></td></p>		
<p class="normal"><td><table><tr><td> </td><td><p class="normal">Breast Cancer Wisconsin (Diagnostic)</p></td></tr></table></td></p>					
<p class="normal"><!-- <td><p class="normal">Diagnostic Wisconsin Breast Cancer Database &nbsp;:</p></td> --></p>					
<p class="normal"><td><p class="normal">Multivariate</p>	<p class="normal"></p></td></p>	<p class="normal"><td><p class="normal">Classification</p>			
<p class="normal"></p></td></p>	<p class="normal"><td><p class="normal">Real</p>	<p class="normal"></p></td></p>			
<p class="normal"><td><p class="normal">569</p>	<p class="normal"></p></td></p>	<p class="normal"><td><p class="normal">32</p>			
<p class="normal"></p></td></p>	<p class="normal"><td><p class="normal">1995</p>	<p class="normal"></p></td></p>			
<p class="normal"><!-- <td><p class="normal">Life&nbsp;:</p></td> --></p>	<p class="normal"></td></p>	<p class="normal"></td></p>			
<tr><td colspan="3" style="background-color: #DDEEFF;"><p class="normal"><td><table><tr><td> </td><td><p class="normal">Pittsburgh Bridges</p></td></tr></table></td></p></td></tr>			<p class="normal"><td><table><tr><td> </td><td><p class="normal">Pittsburgh Bridges</p></td></tr></table></td></p>		
<p class="normal"><td><table><tr><td> </td><td><p class="normal">Pittsburgh Bridges</p></td></tr></table></td></p>					
<p class="normal"><!-- <td><p class="normal">Bridges database that has original and numeric-discretized datasets &nbsp;:</p></td> --></p>					
<p class="normal"><td><p class="normal">Multivariate</p>	<p class="normal"></p></td></p>	<p class="normal"><td><p class="normal">Classification</p>			
<p class="normal"></p></td></p>	<p class="normal"><td><p class="normal">Categorical, Integer</p>	<p class="normal"></p></td></p>			
<p class="normal"><td><p class="normal">108</p>	<p class="normal"></p></td></p>	<p class="normal"><td><p class="normal">13</p>			
<p class="normal"></p></td></p>	<p class="normal"><td><p class="normal">1990</p>	<p class="normal"></p></td></p>			
<p class="normal"><!-- <td><p class="normal">Other &nbsp;:</p></td> --></p>	<p class="normal"></td></p>	<p class="normal"></td></p>			
<tr><td colspan="3" style="background-color: #DDEEFF;"><p class="normal"><td><table><tr><td> </td><td><p class="normal">Car Evaluation</p></td></tr></table></td></p></td></tr>			<p class="normal"><td><table><tr><td> </td><td><p class="normal">Car Evaluation</p></td></tr></table></td></p>		
<p class="normal"><td><table><tr><td> </td><td><p class="normal">Car Evaluation</p></td></tr></table></td></p>					
<p class="normal"><!-- <td><p class="normal">Derived from simple hierarchical decision model, this database may be useful for testing constructive induction and structure discovery methods &nbsp;:</p></td> --></p>					

serve for testing constructive induction and structure discovery methods. [amsp](#), [~/p/n/](#), [tar](#) -->

<p class="normal">Multivariate</p>	<p class="normal">Classification</p>	<p class="normal">Categorical</p>	<p class="normal">1728</p>	<p class="normal">6</p>	<p class="normal">1997</p>	<p class="normal">&nbsp;</p>
------------------------------------	--------------------------------------	-----------------------------------	----------------------------	-------------------------	----------------------------	----------------------------------

--> Other || | | | | | | | | | --- | --- | --- | --- | --- | --- | --- | | Predict whether income exceeds \$50K/yr based on census data. Also known as "Adult" dataset. | | | | | | | | | | | | | |

<p class="normal">Multivariate</p>	<p class="normal">Classification</p>	<p class="normal">Categorical, Integer</p>	<p class="normal">48842</p>	<p class="normal">14</p>	<p class="normal">1996</p>	<p class="normal">&nbsp;</p>
------------------------------------	--------------------------------------	--	-----------------------------	--------------------------	----------------------------	----------------------------------

--> Social || | | | | | | | | | --- | --- | --- | --- | --- | --- | --- | | Knight Pin Chess End-Game Database Creator | | | | | | | | | | | | | |

<p class="normal">Multivariate, Data-Generator</p>	<p class="normal">Classification</p>	<p class="normal">Categorical, Integer</p>	<p class="normal"></p></td></p>	<p class="normal">22</p>	<p class="normal">1988</p>	<p class="normal">&nbsp;</p>
--	--------------------------------------	--	---	--------------------------	----------------------------	----------------------------------

--> Game || | | | | | | | | | --- | --- | --- | --- | --- | --- | --- | | King+Rook versus King+Pawn on a7 (usually abbreviated KRKP A7) | | | | | | | | | | | | | |

<p class="normal">Multivariate</p>	<p class="normal">Classification</p>	<p class="normal">Categorical</p>	<p class="normal">3196</p>	<p class="normal">36</p>	<p class="normal">1989</p>	<p class="normal">&nbsp;</p>
------------------------------------	--------------------------------------	-----------------------------------	----------------------------	--------------------------	----------------------------	----------------------------------

--> Game || | | | | | | | | | --- | --- | --- | --- | --- | --- | --- | | Chess Endgame Database for White King and Rook against Black King (KRK) | | | | | | | | | | | | | |

<p class="normal">Multivariate</p>	<p class="normal">Classification</p>	<p class="normal">Categorical, Integer</p>	<p class="normal">28056</p>	<p class="normal">6</p>	<p class="normal">1994</p>	<p class="normal">&nbsp;</p>
------------------------------------	--------------------------------------	--	-----------------------------	-------------------------	----------------------------	----------------------------------

--> Game || | | | | | | | | | --- | --- | --- | --- | --- | --- | --- | | 6 different domain theories for generating legal moves of chess | | | | | | | | | | | | | |

<p class="normal">Domain-Theory</p>	<p class="normal">&nbsp;</p>					
-------------------------------------	----------------------------------	----------------------------------	----------------------------------	----------------------------------	----------------------------------	----------------------------------

--> Game |

```

<tr><td><a href="datasets/Bach+Chorales"></a> </td><td><p class="normal"><b><a href="datasets/Bach+Chorales">Bach Chorales</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Time-series data based on chorales; challenge is to learn generative grammar; data in Lisp </p></td> -->
<td><p class="normal">Univariate, Time-Series </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">100 </p></td>
<td><p class="normal">6 </p></td>
<td><p class="normal"> </p></td>
<!-- <td><p class="normal">Other </p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Connect-4"></a> </td><td><p class="normal"><b><a href="datasets/Connect-4">Connect-4</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Contains connect-4 positions </p></td> -->
<td><p class="normal">Multivariate, Spatial </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">67557 </p></td>
<td><p class="normal">42 </p></td>
<td><p class="normal">1995 </p></td>
<!-- <td><p class="normal">Game </p></td> -->
</tr>
<td><table><tr><td><a href="datasets/Credit+Approval"></a> </td><td><p class="normal"><b><a href="datasets/Credit+Approval">Credit Approval</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This data concerns credit card applications; good mix of attributes </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer, Real </p></td>
<td><p class="normal">690 </p></td>
<td><p class="normal">15 </p></td>
<td><p class="normal"> </p></td>
<!-- <td><p class="normal">Financial </p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Japanese+Credit+Screening"></a> </td><td><p class="normal"><b><a href="datasets/Japanese+Credit+Screening">Japanese Credit Screening</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Includes domain theory (generated by talking to Japanese domain experts); data in Lisp </p></td> -->
<td><p class="normal">Multivariate, Domain-Theory </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Real, Integer </p></td>
<td><p class="normal">125 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">1992 </p></td>
<!-- <td><p class="normal">Financial </p></td> -->
</tr>
<td><table><tr><td><a href="datasets/Computer+Hardware"></a> </td><td><p class="normal"><b><a href="datasets/Computer+Hardware">Computer Hardware</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Relative CPU Performance Data, described in terms of its cycle time, memory size, etc. </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">209 </p></td>
<td><p class="normal">9 </p></td>
<td><p class="normal">1987 </p></td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Contraceptive+Method+Choice"></a> </td><td><p class="normal"><b><a href="datasets/Contraceptive+Method+Choice">Contraceptive Method Choice</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Dataset is a subset of the 1987 National Indonesia Contraceptive Prevalence Survey. </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">1473 </p></td>
<td><p class="normal">9 </p></td>
<td><p class="normal">1997 </p></td>

```

```

<!-- <ta><p class="normal">Life &nbsp;</p></ta> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Covertype"></a> </td><td><p class="normal"><b><a href="datasets/Covertype">Covertype</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Forest CoverType dataset &nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">581012 </p></td>
<td><p class="normal">54 </p></td>
<td><p class="normal">1998 </p></td>
<!-- <td><p class="normal">Life &nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Cylinder+Bands"></a> </td><td><p class="normal"><b><a href="datasets/Cylinder+Bands">Cylinder Bands</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Used in decision tree induction for mitigating process delays known as "cylinder bands" in rotogravure printing&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer, Real </p></td>
<td><p class="normal">512 </p></td>
<td><p class="normal">39 </p></td>
<td><p class="normal">1995 </p></td>
<!-- <td><p class="normal">Physical &nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Dermatology"></a> </td><td><p class="normal"><b><a href="datasets/Dermatology">Dermatology</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Aim for this dataset is to determine the type of Eryhemato-Squamous Disease.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">366 </p></td>
<td><p class="normal">33 </p></td>
<td><p class="normal">1998 </p></td>
<!-- <td><p class="normal">Life &nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Diabetes"></a> </td><td><p class="normal"><b><a href="datasets/Diabetes">Diabetes</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This diabetes dataset is from AIM '94 &nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">20 </p></td>
<td><p class="normal"> </p></td>
<!-- <td><p class="normal">Life &nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/DGP2+-+The+Second+Data+Generation+Program"></a> </td><td><p class="normal"><b><a href="datasets/DGP2+-+The+Second+Data+Generation+Program">DGP2 - The Second Data Generation Program</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Generates application domains based on specific parameters, number of features, and proportion of positive to negative examples&nbsp;</p></td> -->
<td><p class="normal">Data-Generator </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<!-- <td><p class="normal">Other &nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Document+Understanding"></a> </td><td><p class="normal"><b><a href="datasets/Document+Understanding">Document Understanding</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Five concepts, expressed as predicates, to be learned&nbsp;</p></td> -->
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">1994 </p></td>

```

```

<!-- <td><p class="normal">Other &nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/EBL+Domain+Theories"></a> </td><td><p class="normal"><b><a href="datasets/EBL+Domain+Theories">EBL Domain Theories</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Assorted small-scale domain theories&nbsp;</p></td> -->
<td><p class="normal"> </p></td>
<td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Echocardiogram"></a> </td><td><p class="normal"><b><a href="datasets/Echocardiogram">Echocardiogram</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Data for classifying if patients will survive for at least one year after a heart attack&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer, Real </p></td>
<td><p class="normal">132 </p></td>
<td><p class="normal">12 </p></td>
<td><p class="normal">1989 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Ecoli"></a> </td><td><p class="normal"><b><a href="datasets/Ecoli">Ecoli</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This data contains protein localization sites&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">336 </p></td>
<td><p class="normal">8 </p></td>
<td><p class="normal">1996 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Flags"></a> </td><td><p class="normal"><b><a href="datasets/Flags">Flags</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">From Collins Gem Guide to Flags, 1986&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">194 </p></td>
<td><p class="normal">30 </p></td>
<td><p class="normal">1990 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Function+Finding"></a> </td><td><p class="normal"><b><a href="datasets/Function+Finding">Function Finding</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Cases collected mostly from investigations in physical science; intention is to evaluate function-finding algorithms&nbsp;</p></td> -->
<td><p class="normal"> </p></td>
<td><p class="normal">Function-Learning </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">352 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">1990 </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Glass+Identification"></a> </td><td><p class="normal"><b><a href="datasets/Glass+Identification">Glass Identification</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">From USA Forensic Science Service; 6 types of glass; defined in terms of their oxide content (i.e. Na, Fe, K, etc)&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">214 </p></td>
<td><p class="normal">10 </p></td>
<td><p class="normal">1987 </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr>

```

<p></p> <p>Haberman's Survival</p> <p>Multivariate</p> <p>Classification</p> <p>Integer</p> <p>306</p> <p>3</p> <p>1999</p> <p>Life</p>
--

Dataset contains cases from study conducted on the survival of patients who had undergone surgery for breast cancer

Multivariate

Classification

Integer

306

3

1999

Life

Topic: human subjects study

Multivariate

Classification

Categorical

160

5

1989

Social

Heart+Disease

Heart Disease

4 databases: Cleveland, Hungary, Switzerland, and the VA Long Beach

Multivariate

Classification

Categorical, Integer, Real

303

75

1988

Life

Hepatitis

Hepatitis

From G.Gong: CMU; Mostly Boolean or numeric-valued attribute types; Includes cost data (donated by Peter Turney)

Multivariate

Classification

Categorical, Integer, Real

155

19

1988

Life

Horse+Colic

Horse Colic

Well documented attributes; 368 instances with 28 attributes (continuous, discrete, and nominal); 30% missing values

Multivariate

Classification

Categorical, Integer, Real

368

27

1989

Life

ICU

ICU

Data set prepared for the use of participants for the 1994 AAAI Spring Symposium on Artificial Intelligence in Medicine.

Multivariate, Time-Series

Real

</p></td>

Life

Image+Segmentation

Image+Segmentation

```

src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td><p class="normal"><b><a href="datasets/Image+Segmentation">Image Segmentation</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Image data described by high-level numeric-valued attributes, 7 classes&nbsp; </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">2310 </p></td>
<td><p class="normal">19 </p></td>
<td><p class="normal">1990 </p></td>
<!-- <td><p class="normal">Other&nbsp; </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Internet+Advertisements"></a> </td><td><p class="normal"><b><a href="datasets/Internet+Advertisements">Internet Advertisements</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This dataset represents a set of possible advertisements on Internet pages.&nbsp; </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer, Real </p></td>
<td><p class="normal">3279 </p></td>
<td><p class="normal">1558 </p></td>
<td><p class="normal">1998 </p></td>
<!-- <td><p class="normal">Computer&nbsp; </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Ionosphere"></a> </td><td><p class="normal"><b><a href="datasets/Ionosphere">Ionosphere</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Classification of radar returns from the ionosphere&nbsp; </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">351 </p></td>
<td><p class="normal">34 </p></td>
<td><p class="normal">1989 </p></td>
<!-- <td><p class="normal">Physical&nbsp; </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Iris"></a> </td><td><p class="normal"><b><a href="datasets/Iris">Iris</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Famous database; from Fisher, 1936&nbsp; </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">150 </p></td>
<td><p class="normal">4 </p></td>
<td><p class="normal">1988 </p></td>
<!-- <td><p class="normal">Life&nbsp; </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/ISOLET"></a> </td><td><p class="normal"><b><a href="datasets/ISOLET">ISOLET</a></b></p></td></tr></table></td>
<!-- <td><p class="normal"> Goal: Predict which letter-name was spoken--a simple classification task.&nbsp; </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">7797 </p></td>
<td><p class="normal">617 </p></td>
<td><p class="normal">1994 </p></td>
<!-- <td><p class="normal">Computer&nbsp; </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Kinship"></a> </td><td><p class="normal"><b><a href="datasets/Kinship">Kinship</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Relational dataset&nbsp; </p></td> -->
<td><p class="normal">Relational </p></td>
<td><p class="normal">Relational-Learning </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">104 </p></td>
<td><p class="normal">12 </p></td>
<td><p class="normal">1990 </p></td>
<!-- <td><p class="normal">Social&nbsp; </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Labor+Relations"></a> </td><td><p class="normal"><b><a href="datasets/Labor+Relations">Labor Relations</a></b></p></td></tr></table></td>

```

```

<!-- <td><p class="normal">From Collective Bargaining Review&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal" > </p></td>
<td><p class="normal">Categorical, Integer, Real </p></td>
<td><p class="normal">57 </p></td>
<td><p class="normal">16 </p></td>
<td><p class="normal">1988 </p></td>
<!-- <td><p class="normal">Social&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/LED+Display+Domain"></a> </td><td><p class="normal"><b><a
href="datasets/LED+Display+Domain">LED Display Domain</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">From Classification and Regression Trees book; We provide here 2 C
programs for generating sample databases&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Data-Generator </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal" > </p></td>
<td><p class="normal">7 </p></td>
<td><p class="normal">1988 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Lenses"></a> </td><td><p class="normal"><b><a
href="datasets/Lenses">Lenses</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Database for fitting contact lenses&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">24 </p></td>
<td><p class="normal">4 </p></td>
<td><p class="normal">1990 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Letter+Recognition"></a> </td><td><p class="normal"><b><a
href="datasets/Letter+Recognition">Letter Recognition</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Database of character image features; try to identify the letter&nbsp;<
/p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">20000 </p></td>
<td><p class="normal">16 </p></td>
<td><p class="normal">1991 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Liver+Disorders"></a> </td><td><p class="normal"><b><a
href="datasets/Liver+Disorders">Liver Disorders</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">BUPA Medical Research Ltd. database donated by Richard S. Forsyth&nbsp;
</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal" > </p></td>
<td><p class="normal">Categorical, Integer, Real </p></td>
<td><p class="normal">345 </p></td>
<td><p class="normal">7 </p></td>
<td><p class="normal">1990 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Logic+Theorist"></a> </td><td><p class="normal"><b><a
href="datasets/Logic+Theorist">Logic Theorist</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">All code for Logic Theorist&nbsp;</p></td> -->
<td><p class="normal">Domain-Theory </p></td>
<td><p class="normal" > </p></td>
<td><p class="normal" > </p></td>
<td><p class="normal" > </p></td>
<td><p class="normal" > </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Lung+Cancer"></a> </td><td><p class="normal"><b><a
href="datasets/Lung+Cancer">Lung Cancer</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Lung cancer data; no attribute definitions&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>

```

```

<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">32 </p></td>
<td><p class="normal">56 </p></td>
<td><p class="normal">1992 </p></td>
<!-- <td><p class="normal">Life &nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Lymphography"></a> </td><td><p class="normal"><b><a href="datasets/Lymphography">Lymphography</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This lymphography domain was obtained from the University Medical Centr
e, Institute of Oncology, Ljubljana, Yugoslavia. (Restricted access) &nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">148 </p></td>
<td><p class="normal">18 </p></td>
<td><p class="normal">1988 </p></td>
<!-- <td><p class="normal">Life &nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Mechanical+Analysis"></a> </td><td><p class="normal"><b><a href="datasets/Mechanical+Analysis">Mechanical Analysis</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Fault diagnosis problem of electromechanical devices; also PUMPS DATA S
ET is newer version with domain theory and results&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer, Real </p></td>
<td><p class="normal">209 </p></td>
<td><p class="normal">8 </p></td>
<td><p class="normal">1990 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Meta-data"></a> </td><td><p class="normal"><b><a href="datasets/Meta-data">Meta-data</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Meta-Data was used in order to give advice about which classification m
ethod is appropriate for a particular dataset (taken from results of Statlog project).&nbsp;</p></
td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer, Real </p></td>
<td><p class="normal">528 </p></td>
<td><p class="normal">22 </p></td>
<td><p class="normal">1996 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Mobile+Robots"></a> </td><td><p class="normal"><b><a href="datasets/Mobile+Robots">Mobile Robots</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Learning concepts from sensor data of a mobile robot; set of data sets&
nbsp;</p></td> -->
<td><p class="normal">Domain-Theory </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">Categorical, Integer, Real </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">1995 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29"></a> </td><td><p class="normal"><b><a href="datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29">Molecular Biology (Promoter Gene S
equences)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">E. Coli promoter gene sequences (DNA) with partial domain theory&nbsp;<
/p></td> -->
<td><p class="normal">Sequential, Domain-Theory </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">106 </p></td>
<td><p class="normal">58 </p></td>
<td><p class="normal">1990 </p></td>
<!-- <td><p class="normal">Life &nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%29"></a> </td><td><p class="normal"><b><a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular Biology (Protein Sec

```

ondary Structure)</p></td></tr></table></td>

<!-- <td><p class="normal">From CMU connectionist bench repository; Classifies secondary structure of certain globular proteins ;</p></td> -->

<td><p class="normal">Sequential </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Categorical </p></td>

<td><p class="normal">128 </p></td>

<td><p class="normal"> </p></td>

<td><p class="normal"> </p></td>

<!-- <td><p class="normal">Life ;</p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">Molecular Biology (Splice-junction Gene Sequences) </p></td></tr></table></td>

<!-- <td><p class="normal">Primate splice-junction gene sequences (DNA) with associated imperfect domain theory ;</p></td> -->

<td><p class="normal">Sequential, Domain-Theory </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Categorical </p></td>

<td><p class="normal">3190 </p></td>

<td><p class="normal">61 </p></td>

<td><p class="normal">1992 </p></td>

<!-- <td><p class="normal">Life ;</p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">MONK's Problems</p></td></tr></table></td>

<!-- <td><p class="normal">A set of three artificial domains over the same attribute space; Used to test a wide range of induction algorithms ;</p></td> -->

<td><p class="normal">Multivariate </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Categorical </p></td>

<td><p class="normal">432 </p></td>

<td><p class="normal">7 </p></td>

<td><p class="normal">1992 </p></td>

<!-- <td><p class="normal">Other ;</p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">Moral Reasoner</p></td></tr></table></td>

<!-- <td><p class="normal">Horn-clause model that qualitatively simulates moral reasoning; Theory includes negated literals ;</p></td> -->

<td><p class="normal">Domain-Theory </p></td>

<td><p class="normal"> </p></td>

<td><p class="normal"> </p></td>

<td><p class="normal">202 </p></td>

<td><p class="normal"> </p></td>

<td><p class="normal">1994 </p></td>

<!-- <td><p class="normal">Computer ;</p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">Multiple Features</p></td></tr></table></td>

<!-- <td><p class="normal">This dataset consists of features of handwritten numerals ('0'--'9') extracted from a collection of Dutch utility maps ;</p></td> -->

<td><p class="normal">Multivariate </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Integer, Real </p></td>

<td><p class="normal">2000 </p></td>

<td><p class="normal">649 </p></td>

<td><p class="normal"> </p></td>

<!-- <td><p class="normal">Computer ;</p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">Mushroom</p></td></tr></table></td>

<!-- <td><p class="normal">From Audobon Society Field Guide; mushrooms described in terms of physical characteristics; classification: poisonous or edible ;</p></td> -->

<td><p class="normal">Multivariate </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Categorical </p></td>

<td><p class="normal">8124 </p></td>

<td><p class="normal">22 </p></td>

<td><p class="normal">1987 </p></td>

<!-- <td><p class="normal">Life ;</p></td> -->

```

<td><table><tr><td><a href="datasets/Musk+Version+1%29"></a> </td><td><p class="normal"><b><a href="datasets/Musk+Version+1%29">Musk (Version 1)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The goal is to learn to predict whether new molecules will be musks or non-musks&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">476 </p></td>
<td><p class="normal">168 </p></td>
<td><p class="normal">1994 </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Musk+Version+2%29"></a> </td><td><p class="normal"><b><a href="datasets/Musk+Version+2%29">Musk (Version 2)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The goal is to learn to predict whether new molecules will be musks or non-musks&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">6598 </p></td>
<td><p class="normal">168 </p></td>
<td><p class="normal">1994 </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Nursery"></a> </td><td><p class="normal"><b><a href="datasets/Nursery">Nursery</a></b></p></td></tr></table></td>
<!-- <td><p class="normal"> Nursery Database was derived from a hierarchical decision model originally developed to rank applications for nursery schools.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">12960 </p></td>
<td><p class="normal">8 </p></td>
<td><p class="normal">1997 </p></td>
<!-- <td><p class="normal">Social&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Othello+Domain+Theory"></a> </td><td><p class="normal"><b><a href="datasets/Othello+Domain+Theory">Othello Domain Theory</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Used in research to generate features for an inductive learning system&nbsp;</p></td> -->
<td><p class="normal">Domain-Theory </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">1991 </p></td>
<!-- <td><p class="normal">Game&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Page+Blocks+Classification"></a> </td><td><p class="normal"><b><a href="datasets/Page+Blocks+Classification">Page Blocks Classification</a></b></p></td></tr>
</table></td>
<!-- <td><p class="normal">The problem consists of classifying all the blocks of the page layout of a document that has been detected by a segmentation process.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">5473 </p></td>
<td><p class="normal">10 </p></td>
<td><p class="normal">1995 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Optical+Recognition+of+Handwritten+Digits"></a> </td><td><p class="normal"><b><a href="datasets/Optical+Recognition+of+Handwritten+Digits">Optical Recognition of Handwritten Digits</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Two versions of this database available; see folder&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">5620 </p></td>
<td><p class="normal">64 </p></td>
<td><p class="normal">1998 </p></td>

```

```

<!-- <td><p class="normal">Computer </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits"></a> </td><td><p class="normal"><b><a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits">Pen-Based Recognition of Handwritten Digits</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Digit database of 250 samples from 44 writers </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">10992 </p></td>
<td><p class="normal">16 </p></td>
<td><p class="normal">1998 </p></td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Post-Operative+Patient"></a> </td><td><p class="normal"><b><a href="datasets/Post-Operative+Patient">Post-Operative Patient</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Dataset of patient features </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">90 </p></td>
<td><p class="normal">8 </p></td>
<td><p class="normal">1993 </p></td>
<!-- <td><p class="normal">Life </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Primary+Tumor"></a> </td><td><p class="normal"><b><a href="datasets/Primary+Tumor">Primary Tumor</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">From Ljubljana Oncology Institute </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">339 </p></td>
<td><p class="normal">17 </p></td>
<td><p class="normal">1988 </p></td>
<!-- <td><p class="normal">Life </p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Prodigy"></a> </td><td><p class="normal"><b><a href="datasets/Prodigy">Prodigy</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Assorted domains like blocksworld, eightpuzzle, and schedworld. </p></td> -->
<td><p class="normal">Domain-Theory </p></td>
<td><p class="normal"> </p></td>
<!-- <td><p class="normal">Other </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Qualitative+Structure+Activity+Relationships"></a> </td><td><p class="normal"><b><a href="datasets/Qualitative+Structure+Activity+Relationships">Qualitative Structure Activity Relationships</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Two sets of datasets are given: pyrimidines and triazines </p></td> -->
<td><p class="normal">Domain-Theory </p></td>
<td><p class="normal"> </p></td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Quadruped+Mammals"></a> </td><td><p class="normal"><b><a href="datasets/Quadruped+Mammals">Quadruped Mammals</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The file animals.c is a data generator of structured instances representing quadruped animals </p></td> -->
<td><p class="normal">Multivariate, Data-Generator </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">72 </p></td>
<td><p class="normal">1992 </p></td>

```

```

<!-- <td><p class="normal">Life </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Servo"></a> </td><td><p class="normal"><b><a
href="datasets/Servo">Servo</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Data was from a simulation of a servo system </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">167 </p></td>
<td><p class="normal">4 </p></td>
<td><p class="normal">1993 </p></td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Shuttle+Landing+Control"></a> </td><td><p class="normal"><b><a
href="datasets/Shuttle+Landing+Control">Shuttle Landing Control</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Tiny database; all nominal values </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">15 </p></td>
<td><p class="normal">6 </p></td>
<td><p class="normal">1988 </p></td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Solar+Flare"></a> </td><td><p class="normal"><b><a
href="datasets/Solar+Flare">Solar Flare</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Each class attribute counts the number of solar flares of a certain cla
ss that occur in a 24 hour period </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">1389 </p></td>
<td><p class="normal">10 </p></td>
<td><p class="normal">1989 </p></td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Soybean+%28Large%29"></a> </td><td><p class="normal"><b><a
href="datasets/Soybean+%28Large%29">Soybean (Large)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Michalski's famous soybean disease database </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">307 </p></td>
<td><p class="normal">35 </p></td>
<td><p class="normal">1988 </p></td>
<!-- <td><p class="normal">Life </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Soybean+%28Small%29"></a> </td><td><p class="normal"><b><a
href="datasets/Soybean+%28Small%29">Soybean (Small)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Michalski's famous soybean disease database </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">47 </p></td>
<td><p class="normal">35 </p></td>
<td><p class="normal">1987 </p></td>
<!-- <td><p class="normal">Life </p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Challenger+USA+Space+Shuttle+O-Ring"></a> </td><td><p class="normal"><b><a
href="datasets/Challenger+USA+Space+Shuttle+O-Ring">Challenger USA Space Shuttle O-Ring</a></b></p>
</td></tr></table></td>
<!-- <td><p class="normal">Task: predict the number of O-rings that experience thermal distress on
a flight at 31 degrees F given data on the previous 23 shuttle flights </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">23 </p></td>
<td><p class="normal">4 </p></td>
<td><p class="normal">1993 </p></td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr><tr>

```

```

<td><table><tr><td><a href="datasets/Low+Resolution+Spectrometer"></a> </td><td><p class="normal"><b><a href="datasets/Low+Resolution+Spectrometer">Low Resolution Spectrometer</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">From IRAS data -- NASA Ames Research Center&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">531 </p></td>
<td><p class="normal">102 </p></td>
<td><p class="normal">1988 </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Spambase"></a> </td><td><p class="normal"><b><a href="datasets/Spambase">Spambase</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Classifying Email as Spam or Non-Spam&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">4601 </p></td>
<td><p class="normal">57 </p></td>
<td><p class="normal">1999 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/SPECT+Heart"></a> </td><td><p class="normal"><b><a href="datasets/SPECT+Heart">SPECT Heart</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Data on cardiac Single Proton Emission Computed Tomography (SPECT) images. Each patient classified into two categories: normal and abnormal.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">267 </p></td>
<td><p class="normal">22 </p></td>
<td><p class="normal">2001 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/SPECTF+Heart"></a> </td><td><p class="normal"><b><a href="datasets/SPECTF+Heart">SPECTF Heart</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Data on cardiac Single Proton Emission Computed Tomography (SPECT) images. Each patient classified into two categories: normal and abnormal.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">267 </p></td>
<td><p class="normal">44 </p></td>
<td><p class="normal">2001 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Sponge"></a> </td><td><p class="normal"><b><a href="datasets/Sponge">Sponge</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Data on sponges; Attributes in spanish&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Clustering </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">76 </p></td>
<td><p class="normal">45 </p></td>
<td><p class="normal"> </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Statlog+Project"></a> </td><td><p class="normal"><b><a href="datasets/Statlog+Project">Statlog Project</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Various Databases: Vehicle silhouettes, Landsat Sattelite, Shuttle, Australian Credit Approval, Heart Disease, Image Segmentation, German Credit&nbsp;</p></td> -->
<td><p class="normal"> </p></td>
<td><p class="normal">1992 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Student+Loan+Relational"></a> </td><td><p class="normal"><b><a href="datasets/Student+Loan+Relational">Student Loan Relational</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Student Loan Relational Domain&nbsp;</p></td> -->
<td><p class="normal">Domain-Theory </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">1000 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">1993 </p></td>
<!-- <td><p class="normal">Social&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Teaching+Assistant+Evaluation"></a> </td><td><p class="normal"><b><a href="datasets/Teaching+Assistant+Evaluation">Teaching Assistant Evaluation</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The data consist of evaluations of teaching performance; scores are "low", "medium", or "high"&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">151 </p></td>
<td><p class="normal">5 </p></td>
<td><p class="normal">1997 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Tic-Tac-Toe+Endgame"></a> </td><td><p class="normal"><b><a href="datasets/Tic-Tac-Toe+Endgame">Tic-Tac-Toe Endgame</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Binary classification task on possible configurations of tic-tac-toe game&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">958 </p></td>
<td><p class="normal">9 </p></td>
<td><p class="normal">1991 </p></td>
<!-- <td><p class="normal">Game&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Thyroid+Disease"></a> </td><td><p class="normal"><b><a href="datasets/Thyroid+Disease">Thyroid Disease</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">10 separate databases from Garavan Institute&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Domain-Theory </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Real </p></td>
<td><p class="normal">7200 </p></td>
<td><p class="normal">21 </p></td>
<td><p class="normal">1987 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Trains"></a> </td><td><p class="normal"><b><a href="datasets/Trains">Trains</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">2 data formats (structured, one-instance-per-line)&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">10 </p></td>
<td><p class="normal">32 </p></td>
<td><p class="normal">1994 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/University"></a> </td><td><p class="normal"><b><a href="datasets/University">University</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Data in original (LISP-readable) form&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">285 </p></td>
<td><p class="normal">17 </p></td>
<td><p class="normal">1988 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Congressional+Voting+Records"></a> </td><td><p class="normal"><b><a href="datasets/Congressional+Voting+Records">Congressional Voting Records</a></b></p></td></tr></table></td>

```

<td> [Water Treatment Plant](#) </td> -->

<td> <p class="normal">Multivariate </p></td>

<td> <p class="normal">Classification </p></td>

<td> <p class="normal">Categorical </p></td>

<td> <p class="normal">435 </p></td>

<td> <p class="normal">16 </p></td>

<td> <p class="normal">1987 </p></td>

<!-- <td> <p class="normal">Social&nbsp;</p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td> <table><tr><td> </td><td> <p class="normal">Water Treatment Plant</p></td></tr></table></td>

<!-- <td> <p class="normal">Multiple classes predict plant state&nbsp;</p></td> -->

<td> <p class="normal">Multivariate </p></td>

<td> <p class="normal">Clustering </p></td>

<td> <p class="normal">Integer, Real </p></td>

<td> <p class="normal">527 </p></td>

<td> <p class="normal">38 </p></td>

<td> <p class="normal">1993 </p></td>

<!-- <td> <p class="normal">Physical&nbsp;</p></td> -->

</tr><tr>

<td> <table><tr><td> </td><td> <p class="normal">Waveform Database Generator (Version 1</p></td></tr></table></td>

<!-- <td> <p class="normal">CART book's waveform domains&nbsp;</p></td> -->

<td> <p class="normal">Multivariate, Data-Generator </p></td>

<td> <p class="normal">Classification </p></td>

<td> <p class="normal">Real </p></td>

<td> <p class="normal">5000 </p></td>

<td> <p class="normal">21 </p></td>

<td> <p class="normal">1988 </p></td>

<!-- <td> <p class="normal">Physical&nbsp;</p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td> <table><tr><td> </td><td> <p class="normal">Waveform Database Generator (Version 2</p></td></tr></table></td>

<!-- <td> <p class="normal">CART book's waveform domains&nbsp;</p></td> -->

<td> <p class="normal">Multivariate, Data-Generator </p></td>

<td> <p class="normal">Classification </p></td>

<td> <p class="normal">Real </p></td>

<td> <p class="normal">5000 </p></td>

<td> <p class="normal">40 </p></td>

<td> <p class="normal">1988 </p></td>

<!-- <td> <p class="normal">Physical&nbsp;</p></td> -->

</tr><tr>

<td> <table><tr><td> </td><td> <p class="normal">Wine</p></td></tr></table></td>

<!-- <td> <p class="normal">Using chemical analysis determine the origin of wines&nbsp;</p></td> -->

<td> <p class="normal">Multivariate </p></td>

<td> <p class="normal">Classification </p></td>

<td> <p class="normal">Integer, Real </p></td>

<td> <p class="normal">178 </p></td>

<td> <p class="normal">13 </p></td>

<td> <p class="normal">1991 </p></td>

<!-- <td> <p class="normal">Physical&nbsp;</p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td> <table><tr><td> </td><td> <p class="normal">Yeast</p></td></tr></table></td>

<!-- <td> <p class="normal">Predicting the Cellular Localization Sites of Proteins&nbsp;</p></td> -->

<td> <p class="normal">Multivariate </p></td>

<td> <p class="normal">Classification </p></td>

<td> <p class="normal">Real </p></td>

<td> <p class="normal">1484 </p></td>

<td> <p class="normal">8 </p></td>

<td> <p class="normal">1996 </p></td>

<!-- <td> <p class="normal">Life&nbsp;</p></td> -->

</tr><tr>

<td> <table><tr><td> </td><td> <n class="normal">Zoo</n></td></tr></table></td>

```

src="assets/MLimages/SmallLarge114.jpg"/></a> </td><td><p class="normal"><b><a href="datasets/Zoo">Zoo</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Artificial, 7 classes of animals<br/></p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">101 </p></td>
<td><p class="normal">17 </p></td>
<td><p class="normal">1990 </p></td>
<!-- <td><p class="normal">Life<br/></p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Undocumented"></a> </td><td><p class="normal"><b><a href="datasets/Undocumented">Undocumented</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Various datasets without documentation (feel free to explore!)<br/></p></td> -->
<td><p class="normal"> </p></td>
<td><p class="normal">Other<br/></p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Twenty+Newsgroups"></a> </td><td><p class="normal"><b><a href="datasets/Twenty+Newsgroups">Twenty Newsgroups</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This data set consists of 20000 messages taken from 20 newsgroups.<br/></p></td> -->
<td><p class="normal">Text </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">20000 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">1999 </p></td>
<td><p class="normal">Other<br/></p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Australian+Sign+Language+signs"></a> </td><td><p class="normal"><b><a href="datasets/Australian+Sign+Language+signs">Australian Sign Language signs</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This data consists of sample of Auslan (Australian Sign Language) signs . Examples of 95 signs were collected from five signers with a total of 6650 sign samples.<br/></p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Real </p></td>
<td><p class="normal">6650 </p></td>
<td><p class="normal">15 </p></td>
<td><p class="normal">1999 </p></td>
<td><p class="normal">Other<br/></p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Australian+Sign+Language+signs%28High+Quality%29"></a> </td><td><p class="normal"><b><a href="datasets/Australian+Sign+Language+signs%28High+Quality%29">Australian Sign Language signs ( High Quality)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This data consists of sample of Auslan (Australian Sign Language) signs . 27 examples of each of 95 Auslan signs were captured from a native signer using high-quality position trackers.<br/></p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">2565 </p></td>
<td><p class="normal">22 </p></td>
<td><p class="normal">2002 </p></td>
<td><p class="normal">Other<br/></p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/US+Census+Data%281990%29"></a> </td><td><p class="normal"><b><a href="datasets/US+Census+Data%281990%29">US Census Data (1990)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.<br/></p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Clustering </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">2458285 </p></td>
<td><n class="normal">68 </n></td>

```

```

<--><p class="normal"><br>/><br>
<td><p class="normal"> </p></td>
<!-- <td><p class="normal">Social&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Census-Income+%28KDD%29"></a> </td><td><p class="normal"><b><a
href="datasets/Census-Income+%28KDD%29">Census-Income (KDD)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This data set contains weighted census data extracted from the 1994 and
1995 current population surveys conducted by the U.S. Census Bureau.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">299285 </p></td>
<td><p class="normal">40 </p></td>
<td><p class="normal">2000 </p></td>
<!-- <td><p class="normal">Social&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Coil+1999+Competition+Data"></a> </td><td><p class="normal"><b><a
href="datasets/Coil+1999+Competition+Data">Coil 1999 Competition Data</a></b></p></td></tr></table>
</td>
<!-- <td><p class="normal">This data set is from the 1999 Computational Intelligence and Learning
(COIL) competition. The data contains measurements of river chemical concentrations and algae dens
ities.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">Categorical, Real </p></td>
<td><p class="normal">340 </p></td>
<td><p class="normal">17 </p></td>
<td><p class="normal">1999 </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Corel+Image+Features"></a> </td><td><p class="normal"><b><a
href="datasets/Corel+Image+Features">Corel Image Features</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This dataset contains image features extracted from a Corel image
collection. Four sets of features are available based on the color histogram, color histogram
layout, color moments, and co-occurrence&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">68040 </p></td>
<td><p class="normal">89 </p></td>
<td><p class="normal">1999 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/E.+Coli+Genes"></a> </td><td><p class="normal"><b><a
href="datasets/E.+Coli+Genes">E. Coli Genes</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Data giving characteristics of each ORF (potential gene) in the E. coli
genome. Sequence, homology (similarity to other genes) and structural information, and function (i
f known) are provided.&nbsp;</p></td> -->
<td><p class="normal">Relational </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2001 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/EEG+Database"></a> </td><td><p class="normal"><b><a
href="datasets/EEG+Database">EEG Database</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This data arises from a large study to examine EEG correlates of
genetic predisposition to alcoholism. It contains measurements from 64 electrodes placed on the sc
alp sampled at 256 Hz&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">Categorical, Integer, Real </p></td>
<td><p class="normal">122 </p></td>
<td><p class="normal">4 </p></td>
<td><p class="normal">1999 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/El+Nino"></a> </td><td><p class="normal"><b><a
href="datasets/El+Nino">El Nino</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The data set contains oceanographic and surface meteorological readings

```

-->-- <td> The data set contains oceanographic and surface meteorological readings taken from a series of buoys positioned throughout the equatorial Pacific. </td> -->

<td><p class="normal">Spatio-temporal </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">178080 </p></td>
<td><p class="normal">12 </p></td>
<td><p class="normal">1999 </p></td>
<!-- <td><p class="normal">Physical&nbsnbsp;</p></td> -->
</tr><tr>
<td><table><tr><td> </td><td><p class="normal">Entree Chicago Recommendation Data</p></td></tr></table></td>
<!-- <td><p class="normal">This data contains a record of user interactions with the Entree Chicago o restaurant recommendation system.&nbsnbsp;</p></td> -->
<td><p class="normal">Transactional, Sequential </p></td>
<td><p class="normal">Recommender-Systems </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">50672 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2000 </p></td>
<!-- <td><p class="normal">Other&nbsnbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td> </td><td><p class="normal">CMU Face Images</p></td></tr></table></td>
<!-- <td><p class="normal">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&nbsnbsp;</p></td> -->
<td><p class="normal">Image </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">640 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">1999 </p></td>
<!-- <td><p class="normal">Other&nbsnbsp;</p></td> -->
</tr><tr>
<td><table><tr><td> </td><td><p class="normal">Insurance Company Benchmark (COIL 2000)</p></td></tr></table></td>
<!-- <td><p class="normal">This data set used in the CoIL 2000 Challenge contains information on customers of an insurance company. The data consists of 86 variables and includes product usage data and socio-demographic data&nbsnbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression, Description </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">9000 </p></td>
<td><p class="normal">86 </p></td>
<td><p class="normal">2000 </p></td>
<!-- <td><p class="normal">Social&nbsnbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td> </td><td><p class="normal">Internet Usage Data</p></td></tr></table></td>
<!-- <td><p class="normal">This data contains general demographic information on internet users in 1997.&nbsnbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">10104 </p></td>
<td><p class="normal">72 </p></td>
<td><p class="normal">1999 </p></td>
<!-- <td><p class="normal">Computer&nbsnbsp;</p></td> -->
</tr><tr>
<td><table><tr><td> </td><td><p class="normal">IPUMS Census Database</p></td></tr></table></td>
<!-- <td><p class="normal">This data set contains unweighted PUMS census data from the Los Angeles and Long Beach areas for the years 1970, 1980, and 1990.&nbsnbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">256932 </p></td>
<td><p class="normal">61 </p></td>
<td><p class="normal">1999 </p></td>

```

<!-- <tr><p class="normal">SOCIAL&nbsp;-->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Japanese+Vowels"></a> </td><td><p class="normal"><b><a href="datasets/Japanese+Vowels">Japanese Vowels</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This dataset records 640 time series of 12 LPC cepstrum coefficients ta
ken from nine male speakers.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">640 </p></td>
<td><p class="normal">12 </p></td>
<td><p class="normal"> </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/KDD+Cup+1998+Data"></a> </td><td><p class="normal"><b><a href="datasets/KDD+Cup+1998+Data">KDD Cup 1998 Data</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This is the data set used for The Second International Knowledge Discov
ery and Data Mining Tools Competition, which was held in conjunction with KDD-98&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">191779 </p></td>
<td><p class="normal">481 </p></td>
<td><p class="normal">1998 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/KDD+Cup+1999+Data"></a> </td><td><p class="normal"><b><a href="datasets/KDD+Cup+1999+Data">KDD Cup 1999 Data</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">4000000 </p></td>
<td><p class="normal">42 </p></td>
<td><p class="normal">1999 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/M.+Tuberculosis+Genes"></a> </td><td><p class="normal"><b><a href="datasets/M.+Tuberculosis+Genes">M. Tuberculosis Genes</a></b></p></td></tr></table></td>
<!-- <td><p class="normal"> Data giving characteristics of each ORF (potential gene) in the M.
tuberculosis bacterium. Sequence, homology (similarity to other genes) and structural information,
and function (if known) are provided&nbsp;</p></td> -->
<td><p class="normal">Relational </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2001 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Movie"></a> </td><td><p class="normal"><b><a href="datasets/Movie">Movie</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This data set contains a list of over 10000 films including many older,
odd, and cult films. There is information on actors, casts, directors, producers, studios,
etc.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Relational </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">10000 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">1999 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/MSNBC.com+Anonymous+Web+Data"></a> </td><td><p class="normal"><b><a href="datasets/MSNBC.com+Anonymous+Web+Data">MSNBC.com Anonymous Web Data</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This data describes the page visits of users who visited msnbc.com on S
eptember 28, 1999. Visits are recorded at the level of URL category (see description) and are reco
-->
```

raea in time order. </p></td> -->
<td><p class="normal">Sequential </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">989818 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td> </td><td><p class="normal">NSF Research Award Abstracts 1990-2003</p></td></tr></table></td>
<!-- <td><p class="normal">This data set consists of (a) 129,000 abstracts describing NSF awards for basic research, (b) bag-of-word data files extracted from the abstracts, (c) a list of words used for indexing the bag-of-word </p></td> -->
<td><p class="normal">Text </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">129000 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2003 </p></td>
<!-- <td><p class="normal">Other </p></td> -->
</tr><tr>
<td><table><tr><td> </td><td><p class="normal">Pioneer-1 Mobile Robot Data</p></td></tr></table></td>
<!-- <td><p class="normal">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></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">Categorical, Real </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">1999 </p></td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td> </td><td><p class="normal">Pseudo Periodic Synthetic Time Series</p></td></tr></table></td>
<!-- <td><p class="normal">This data set is designed for testing indexing schemes in time series databases. The data appears highly periodic, but never exactly repeats itself. </p></td> -->
<td><p class="normal">Univariate, Time-Series </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">100000 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">1999 </p></td>
<!-- <td><p class="normal">Other </p></td> -->
</tr><tr>
<td><table><tr><td> </td><td><p class="normal">Reuters-21578 Text Categorization Collection</p></td></tr></table></td>
<!-- <td><p class="normal">This is a collection of documents that appeared on Reuters newswire in 1987. The documents were assembled and indexed with categories. </p></td> -->
<td><p class="normal">Text </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">21578 </p></td>
<td><p class="normal">5 </p></td>
<td><p class="normal">1997 </p></td>
<!-- <td><p class="normal">Other </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td> </td><td><p class="normal">Robot Execution Failures</p></td></tr></table></td>
<!-- <td><p class="normal">This dataset contains force and torque measurements on a robot after failure detection. Each failure is characterized by 15 force/torque samples collected at regular time intervals </p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
.....

```
<td><p class="normal">463 </p></td>
<td><p class="normal">90 </p></td>
<td><p class="normal">1999 </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Synthetic+Control+Chart+Time+Series"></a> </td><td><p class="normal"><b><a href="datasets/Synthetic+Control+Chart+Time+Series">Synthetic Control Chart Time Series</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This data consists of synthetically generated control charts.&nbsp;</p></td> -->
<td><p class="normal">Time-Series </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">600 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">1999 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Syskill+and+Webert+Web+Page+Ratings"></a> </td><td><p class="normal"><b><a href="datasets/Syskill+and+Webert+Web+Page+Ratings">Syskill and Webert Web Page Ratings</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This database contains HTML source of web pages plus the ratings of a single user on these web pages. Web pages are on four seperate subjects (Bands- recording artists; Goats; Sheep; and BioMedical)&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Text </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">332 </p></td>
<td><p class="normal">5 </p></td>
<td><p class="normal">1998 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/UNIX+User+Data"></a> </td><td><p class="normal"><b><a href="datasets/UNIX+User+Data">UNIX User Data</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbsp;</p></td> -->
<td><p class="normal">Text, Sequential </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Volcanoes+on+Venus+-+JARtool+experiment"></a> </td><td><p class="normal"><b><a href="datasets/Volcanoes+on+Venus+-+JARtool+experiment">Volcanoes on Venus - JARtool experiment</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbsp;</p></td> -->
<td><p class="normal">Image </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Statlog%28Australian+Credit+Approval%29"></a> </td><td><p class="normal"><b><a href="datasets/Statlog%28Australian+Credit+Approval%29">Statlog (Australian Credit Approval)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This file concerns credit card applications. This database exists elsewhere in the repository (Credit Screening Database) in a slightly different form&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer, Real </p></td>
<td><p class="normal">690 </p></td>
<td><p class="normal">14 </p></td>
<td><p class="normal"> </p></td>
<!-- <td><p class="normal">Financial&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Statlog%28German+Credit+Data%29"></a> </td><td><p class="normal"><b><a href="datasets/Statlog%28German+Credit+Data%29">Statlog (German Credit Data)</a></b></p></td></tr></table></td>
```

```

src="assets/MLimages/SmallLargedefault.jpg"/></a> </td><p class="normal"><b><a href="datasets/Statlog+28German+Credit+Data29">Statlog (German Credit Data)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This dataset classifies people described by a set of attributes as good or bad credit risks. Comes in two formats (one all numeric). Also comes with a cost matrix&nbsp;.</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">1000 </p></td>
<td><p class="normal">20 </p></td>
<td><p class="normal">1994 </p></td>
<!-- <td><p class="normal">Financial&nbsp;.</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Statlog+28Heart29"></a> </td><td><p class="normal"><b><a href="datasets/Statlog+28Heart29">Statlog (Heart)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This dataset is a heart disease database similar to a database already present in the repository (Heart Disease databases) but in a slightly different form&nbsp;.</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Real </p></td>
<td><p class="normal">270 </p></td>
<td><p class="normal">13 </p></td>
<td><p class="normal"> </p></td>
<!-- <td><p class="normal">Life&nbsp;.</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Statlog+28Landsat+Satellite29"></a> </td><td><p class="normal"><b><a href="datasets/Statlog+28Landsat+Satellite29">Statlog (Landsat Satellite)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Multi-spectral values of pixels in 3x3 neighbourhoods in a satellite image, and the classification associated with the central pixel in each neighbourhood&nbsp;.</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">6435 </p></td>
<td><p class="normal">36 </p></td>
<td><p class="normal">1993 </p></td>
<!-- <td><p class="normal">Physical&nbsp;.</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Statlog+28Image+Segmentation29"></a> </td><td><p class="normal"><b><a href="datasets/Statlog+28Image+Segmentation29">Statlog (Image Segmentation)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbsp;.</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">2310 </p></td>
<td><p class="normal">19 </p></td>
<td><p class="normal">1990 </p></td>
<!-- <td><p class="normal">Other&nbsp;.</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Statlog+28Shuttle29"></a> </td><td><p class="normal"><b><a href="datasets/Statlog+28Shuttle29">Statlog (Shuttle)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The shuttle dataset contains 9 attributes all of which are numerical. Approximately 80% of the data belongs to class 1&nbsp;.</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">58000 </p></td>
<td><p class="normal">9 </p></td>
<td><p class="normal"> </p></td>
<!-- <td><p class="normal">Physical&nbsp;.</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Statlog+28Vehicle+Silhouettes29"></a> </td><td><p class="normal"><b><a href="datasets/Statlog+28Vehicle+Silhouettes29">Statlog (Vehicle Silhouettes)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">3D objects within a 2D image by application of an ensemble of shape feature extractors to the 2D silhouettes of the objects.&nbsp;.</p></td> -->

```

```
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">946 </p></td>
<td><p class="normal">18 </p></td>
<td><p class="normal"> </p></td>
<!-- <td><p class="normal">Other &nbsp; </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Connectionist+Bench+%28Nettalk+Corpus%29"></a> </td><td><p class="normal"><b><a href="datasets/Connectionist+Bench+%28Nettalk+Corpus%29">Connectionist Bench (Nettalk Corpus)</a></b></a></p></td></tr></table></td>
<!-- <td><p class="normal">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 &nbsp; </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">20008 </p></td>
<td><p class="normal">4 </p></td>
<td><p class="normal"> </p></td>
<!-- <td><p class="normal">Other &nbsp; </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29"></a> </td><td><p class="normal"><b><a href="datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29">Connectionist Bench (Sonar, Mines vs. Rocks)</a></b></a></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbsp; </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">208 </p></td>
<td><p class="normal">60 </p></td>
<td><p class="normal"> </p></td>
<!-- <td><p class="normal">Physical &nbsp; </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Connectionist+Bench+%28Vowel+Recognition+-Deterding+Data%29"></a> </td><td><p class="normal"><b><a href="datasets/Connectionist+Bench+%28Vowel+Recognition+-Deterding+Data%29">Connectionist Bench (Vowel Recognition - Deterding Data)</a></b></a></p></td></tr></table></td>
<!-- <td><p class="normal">Speaker independent recognition of the eleven steady state vowels of British English using a specified training set of lpc derived log area ratios.&nbsp; </p></td> -->
<td><p class="normal"> </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">528 </p></td>
<td><p class="normal">10 </p></td>
<td><p class="normal"> </p></td>
<!-- <td><p class="normal">Other &nbsp; </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Economic+Sanctions"></a> </td><td><p class="normal"><b><a href="datasets/Economic+Sanctions">Economic Sanctions</a></b></a></p></td></tr></table></td>
<!-- <td><p class="normal">Domain Theory on Economic Sanctions; Undocumented&nbsp; </p></td> -->
<td><p class="normal">Domain-Theory </p></td>
<td><p class="normal"> </p></td>
<!-- <td><p class="normal">Financial &nbsp; </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Protein+Data"></a> </td><td><p class="normal"><b><a href="datasets/Protein+Data">Protein Data</a></b></a></p></td></tr></table></td>
<!-- <td><p class="normal">Undocumented&nbsp; </p></td> -->
<td><p class="normal"> </p></td>
<!-- <td><p class="normal">Life &nbsp; </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Cloud"></a> </td><td><p class="normal"><b><a href="datasets/Cloud">Cloud</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Little Documentation&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">1024 </p></td>
<td><p class="normal">10 </p></td>
<td><p class="normal">1989 </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/CalIt2+Building+People+Counts"></a> </td><td><p class="normal"><b><a href="datasets/CalIt2+Building+People+Counts">CalIt2 Building People Counts</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This data comes from the main door of the CalIt2 building at UCI.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">10080 </p></td>
<td><p class="normal">4 </p></td>
<td><p class="normal">2006 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Dodgers+Loop+Sensor"></a> </td><td><p class="normal"><b><a href="datasets/Dodgers+Loop+Sensor">Dodgers Loop Sensor</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Loop sensor data was collected for the Glendale on ramp for the 101 North freeway in Los Angeles&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">50400 </p></td>
<td><p class="normal">3 </p></td>
<td><p class="normal">2006 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Poker+Hand"></a> </td><td><p class="normal"><b><a href="datasets/Poker+Hand">Poker Hand</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Purpose is to predict poker hands&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">1025010 </p></td>
<td><p class="normal">11 </p></td>
<td><p class="normal">2007 </p></td>
<!-- <td><p class="normal">Game&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/MAGIC+Gamma+Telescope"></a> </td><td><p class="normal"><b><a href="datasets/MAGIC+Gamma+Telescope">MAGIC Gamma Telescope</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Data are MC generated to simulate registration of high energy gamma particles in an atmospheric Cherenkov telescope&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">19020 </p></td>
<td><p class="normal">11 </p></td>
<td><p class="normal">2007 </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/UJI+Pen+Characters"></a> </td><td><p class="normal"><b><a href="datasets/UJI+Pen+Characters">UJI Pen Characters</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Data consists of written characters in a UNIPEN-like format&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Sequential </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">1364 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2007 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Mammographic+Mass"></a> </td><td><p class="normal"><b><a href="datasets/Mammographic+Mass">Mammographic Mass</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Discrimination of benign and malignant mammographic masses based on BI-RADS attributes and the patient's age.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">961 </p></td>
<td><p class="normal">6 </p></td>
<td><p class="normal">2007 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Forest+Fires"></a> </td><td><p class="normal"><b><a href="datasets/Forest+Fires">Forest Fires</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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).&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">517 </p></td>
<td><p class="normal">13 </p></td>
<td><p class="normal">2008 </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Reuters+Transcribed+Subset"></a> </td><td><p class="normal"><b><a href="datasets/Reuters+Transcribed+Subset">Reuters Transcribed Subset</a></b></p></td></tr></table></td>
<!-- <p class="normal">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.&nbsp;</p></td> -->
<td><p class="normal">Text </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">200 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2008 </p></td>
<!-- <td><p class="normal">Business&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Bag+of+Words"></a> </td><td><p class="normal"><b><a href="datasets/Bag+of+Words">Bag of Words</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This data set contains five text collections in the form of bags-of-words.&nbsp;</p></td> -->
<td><p class="normal">Text </p></td>
<td><p class="normal">Clustering </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">8000000 </p></td>
<td><p class="normal">100000 </p></td>
<td><p class="normal">2008 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Concrete+Compressive+Strength"></a> </td><td><p class="normal"><b><a href="datasets/Concrete+Compressive+Strength">Concrete Compressive Strength</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Concrete is the most important material in civil engineering. The concrete compressive strength is a highly nonlinear function of age and ingredients. &nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">1030 </p></td>
<td><p class="normal">9 </p></td>
<td><p class="normal">2007 </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Hill-Valley"></a> </td><td><p class="normal"><b><a href="datasets/Hill-Valley">Hill-Valley</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Each record represents 100 points on a two-dimensional graph. When plotted in order (from 1 through 100) as the Y co-ordinate, the points will create either a Hill (a bump in the terrain) or a Valley (a dip in the terrain).&nbsp;</p></td> -->
<td><p class="normal">Sequential </p></td>

```

<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">606 </p></td>
<td><p class="normal">101 </p></td>
<td><p class="normal">2008 </p></td>
<!-- <td><p class="normal">Other </p></td> -->
</tr><tr>
<td><table><tr><td> </td><td><p class="normal">Arcene</p></td></tr></table></td>
<!-- <td><p class="normal">ARCENE's task is to distinguish cancer versus normal patterns from mass-spectrometric data. This is a two-class classification problem with continuous input variables. This dataset is one of 5 datasets of the NIPS 2003 feature selection challenge. </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">900 </p></td>
<td><p class="normal">10000 </p></td>
<td><p class="normal">2008 </p></td>
<!-- <td><p class="normal">Life </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td> </td><td><p class="normal">Dexter</p></td></tr></table></td>
<!-- <td><p class="normal">DEXTER is a text classification problem in a bag-of-word representation. This is a two-class classification problem with sparse continuous input variables. This dataset is one of five datasets of the NIPS 2003 feature selection challenge.
 </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">2600 </p></td>
<td><p class="normal">20000 </p></td>
<td><p class="normal">2008 </p></td>
<!-- <td><p class="normal">Other </p></td> -->
</tr><tr>
<td><table><tr><td> </td><td><p class="normal">Dorothea</p></td></tr></table></td>
<!-- <td><p class="normal">DOROTHEA is a drug discovery dataset. Chemical compounds represented by structural molecular features must be classified as active (binding to thrombin) or inactive. This is one of 5 datasets of the NIPS 2003 feature selection challenge. </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">1950 </p></td>
<td><p class="normal">100000 </p></td>
<td><p class="normal">2008 </p></td>
<!-- <td><p class="normal">Life </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td> </td><td><p class="normal">Gisette</p></td></tr></table></td>
<!-- <td><p class="normal">GISETTE is a handwritten digit recognition problem. The problem is to separate the highly confusable digits '4' and '9'. This dataset is one of five datasets of the NIPS 2003 feature selection challenge.
 </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">13500 </p></td>
<td><p class="normal">5000 </p></td>
<td><p class="normal">2008 </p></td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr><tr>
<td><table><tr><td> </td><td><p class="normal">Madelon</p></td></tr></table></td>
<!-- <td><p class="normal">MADELON is an artificial dataset, which was part of the NIPS 2003 feature selection challenge. This is a two-class classification problem with continuous input variables. The difficulty is that the problem is multivariate and highly non-linear. </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">4400 </p></td>
<td><p class="normal">500 </p></td>

```

<td><p class="normal">2008 </p></td>
<!-- <td><p class="normal">Other &nbsp; </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Ozone+Level+Detection"></a> </td><td><p class="normal"><b><a
href="datasets/Ozone+Level+Detection">Ozone Level Detection</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Two ground ozone level data sets are included in this collection. One i
s the eight hour peak set (eighthr.data), the other is the one hour peak set (onehr.data). Those d
ata were collected from 1998 to 2004 at the Houston, Galveston and Brazoria area.&nbsp;</p></td> --
>
<td><p class="normal">Multivariate, Sequential, Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">2536 </p></td>
<td><p class="normal">73 </p></td>
<td><p class="normal">2008 </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Abscisic+Acid+Signaling+Network"></a> </td><td><p class="normal"><b><a
href="datasets/Abscisic+Acid+Signaling+Network">Abscisic Acid Signaling Network</a></b></p></td></
tr></table></td>
<!-- <td><p class="normal">The objective is to determine the set of boolean rules that describe th
e interactions of the nodes within this plant signaling network. The dataset includes 300
separate boolean pseudodynamic simulations using an asynchronous update scheme. &nbsp;</p></td> --
>
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Causal-Discovery </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">300 </p></td>
<td><p class="normal">43 </p></td>
<td><p class="normal">2008 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Parkinsons"></a> </td><td><p class="normal"><b><a
href="datasets/Parkinsons">Parkinsons</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Oxford Parkinson's Disease Detection Dataset&nbsp;</p></td> --
>
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">197 </p></td>
<td><p class="normal">23 </p></td>
<td><p class="normal">2008 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Character+Trajectories"></a> </td><td><p class="normal"><b><a
href="datasets/Character+Trajectories">Character Trajectories</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Multiple, labelled samples of pen tip trajectories recorded whilst writ
ing individual characters. All samples are from the same writer, for the purposes of primitive ext
raction. Only characters with a single pen-down segment were considered.&nbsp;</p></td> --
>
<td><p class="normal">Time-Series </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">2858 </p></td>
<td><p class="normal">3 </p></td>
<td><p class="normal">2008 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Blood+Transfusion+Service+Center"></a> </td><td><p class="normal"><b><a
href="datasets/Blood+Transfusion+Service+Center">Blood Transfusion Service Center</a></b></p></td>
</tr></table></td>
<!-- <td><p class="normal">Data taken from the Blood Transfusion Service Center in Hsin-Chu City i
n Taiwan -- this is a classification problem. &nbsp;</p></td> --
>
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">748 </p></td>
<td><p class="normal">5 </p></td>
<td><p class="normal">2008 </p></td>
<!-- <td><p class="normal">Business&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/UJI+Pen+Characters+%28Version+2%29"></a> </td><td><p class="normal"><b><a
href="datasets/UJI+Pen+Characters+%28Version+2%29">UJI Pen Characters (Version 2)</a></b></p></td>

```

```

</tr></table></td>
<!-- <td><p class="normal">A pen-based database with more than 11k isolated handwritten
characters&nbsp;.</p></td> -->
<td><p class="normal">Multivariate, Sequential </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">11640 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2009 </p></td>
<!-- <td><p class="normal">Computer&nbsp;.</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Semeion+Handwritten+Digit"></a> </td><td><p class="normal"><b><a
href="datasets/Semeion+Handwritten+Digit">Semeion Handwritten Digit</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">1593 handwritten digits from around 80 persons were scanned, stretched
in a rectangular box 16x16 in a gray scale of 256 values.&nbsp;.</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">1593 </p></td>
<td><p class="normal">256 </p></td>
<td><p class="normal">2008 </p></td>
<!-- <td><p class="normal">Computer&nbsp;.</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/SECOM"></a> </td><td><p class="normal"><b><a
href="datasets/SECOM">SECOM</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Data from a semi-conductor manufacturing process&nbsp;.</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Causal-Discovery </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">1567 </p></td>
<td><p class="normal">591 </p></td>
<td><p class="normal">2008 </p></td>
<!-- <td><p class="normal">Computer&nbsp;.</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Plants"></a> </td><td><p class="normal"><b><a
href="datasets/Plants">Plants</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Data has been extracted from the USDA plants database. It contains all
plants (species and genera) in the database and the states of USA and Canada where they occur.&nbs
p;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Clustering </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">22632 </p></td>
<td><p class="normal">70 </p></td>
<td><p class="normal">2008 </p></td>
<!-- <td><p class="normal">Life&nbsp;.</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Libras+Movement"></a> </td><td><p class="normal"><b><a
href="datasets/Libras+Movement">Libras Movement</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The data set contains 15 classes of 24 instances each. Each class
references to a hand movement type in LIBRAS (Portuguese
name 'Língua Brasileira de Sinais', oficial brazilian signal language).&nbsp;.</p></td> -->
<td><p class="normal">Multivariate, Sequential </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">360 </p></td>
<td><p class="normal">91 </p></td>
<td><p class="normal">2009 </p></td>
<!-- <td><p class="normal">Other&nbsp;.</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Concrete+Slump+Test"></a> </td><td><p class="normal"><b><a
href="datasets/Concrete+Slump+Test">Concrete Slump Test</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Concrete is a highly complex material. The slump flow of concrete is no
t only determined by the water content, but that is also influenced by other concrete
ingredients.&nbsp;.</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">103 </p></td>
<td><p class="normal">10 </p></td>
<td><p class="normal">2009 </p></td>

```

```

<!-- <td><p class="normal">Computer </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Communities+and+Crime"></a> </td><p class="normal"><b><a
href="datasets/Communities+and+Crime">Communities and Crime</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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 f
rom the 1995 FBI UCR.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">1994 </p></td>
<td><p class="normal">128 </p></td>
<td><p class="normal">2009 </p></td>
<!-- <td><p class="normal">Social &nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Acute+Inflammations"></a> </td><p class="normal"><b><a
href="datasets/Acute+Inflammations">Acute Inflammations</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The data was created by a medical expert as a data set to test the expe
rt system,
which will perform the presumptive diagnosis of two diseases of the urinary system.
&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer </p></td>
<td><p class="normal">120 </p></td>
<td><p class="normal">6 </p></td>
<td><p class="normal">2009 </p></td>
<!-- <td><p class="normal">Life &nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Wine+Quality"></a> </td><td><p class="normal"><b><a
href="datasets/Wine+Quality">Wine Quality</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Two datasets are included, related to red and white vinhos verde wine sa
mples, 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/).&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">4898 </p></td>
<td><p class="normal">12 </p></td>
<td><p class="normal">2009 </p></td>
<!-- <td><p class="normal">Business &nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/URL+Reputation"></a> </td><td><p class="normal"><b><a
href="datasets/URL+Reputation">URL Reputation</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Anonymized 120-day subset of the ICML-09 URL data containing 2.4
million examples and 3.2 million features.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">2396130 </p></td>
<td><p class="normal">3231961 </p></td>
<td><p class="normal">2009 </p></td>
<!-- <td><p class="normal">Computer &nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/p53+Mutants"></a> </td><td><p class="normal"><b><a
href="datasets/p53+Mutants">p53 Mutants</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The goal is to model mutant p53 transcriptional activity (active vs
inactive) based on data extracted from biophysical simulations.
&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">16772 </p></td>
<td><p class="normal">5409 </p></td>
<td><p class="normal">2010 </p></td>
<!-- <td><p class="normal">Life &nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Parkinsons+Telemonitoring"></a> </td><td><p class="normal"><b><a
href="datasets/Parkinsons+Telemonitoring">Parkinsons Telemonitoring</a></b></p></td></tr></table><
/td>
<!-- <td><p class="normal">Oxford Parkinson's Disease Telemonitoring Dataset &nbsp;</p></td> -->

```

```

<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">5875 </p></td>
<td><p class="normal">26 </p></td>
<td><p class="normal">2009 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Demospongiae"></a> </td><td><p class="normal"><b><a href="datasets/Demospongiae">Demospongiae</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Marine sponges of the Demospongiae class classification domain.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">503 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2010 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Opinosis+Opinion+%26frasl%3B+Review"></a> </td><td><p class="normal"><b><a href="datasets/Opinosis+Opinion+%26frasl%3B+Review">Opinosis Opinion / Review</a></b></p></td></tr></table></td>
<!-- <td><p 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". &nbsp;</p></td> -->
<td><p class="normal">Text </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">51 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2010 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Breast+Tissue"></a> </td><td><p class="normal"><b><a href="datasets/Breast+Tissue">Breast Tissue</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Dataset with electrical impedance measurements of freshly excised tissue samples from the breast.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">106 </p></td>
<td><p class="normal">10 </p></td>
<td><p class="normal">2010 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Cardiotocography"></a> </td><td><p class="normal"><b><a href="datasets/Cardiotocography">Cardiotocography</a></b></p></td></tr></table></td>
<!-- <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.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">2126 </p></td>
<td><p class="normal">23 </p></td>
<td><p class="normal">2010 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Wall-Following+Robot+Navigation+Data"></a> </td><td><p class="normal"><b><a href="datasets/Wall-Following+Robot+Navigation+Data">Wall-Following Robot Navigation Data</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The data were collected as the SCITOS G5 robot navigates through the room following the wall in a clockwise direction, for 4 rounds, using 24 ultrasound sensors arranged circularly around its 'waist'.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Sequential </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">5456 </p></td>
<td><p class="normal">24 </p></td>
<td><p class="normal">2010 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->

```

```

</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Spoken+Arabic+Digit"></a> </td><td><p class="normal"><b><a href="datasets/Spoken+Arabic+Digit">Spoken Arabic Digit</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This dataset contains timeseries of mel-frequency cepstrum coefficients (MFCCs) corresponding to spoken Arabic digits. Includes data from 44 male and 44 female native Arabic speakers.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">8800 </p></td>
<td><p class="normal">13 </p></td>
<td><p class="normal">2010 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Localization+Data+for+Person+Activity"></a> </td><td><p class="normal"><b><a href="datasets/Localization+Data+for+Person+Activity">Localization Data for Person Activity</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Data contains recordings of five people performing different activities. Each person wore four sensors (tags) while performing the same scenario five times. &nbs;p;</p></td> -->
<td><p class="normal">Univariate, Sequential, Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">164860 </p></td>
<td><p class="normal">8 </p></td>
<td><p class="normal">2010 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/AutoUniv"></a> </td><td><p class="normal"><b><a href="datasets/AutoUniv">AutoUniv</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical, Integer, Real </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2010 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Steel+Plates+Faults"></a> </td><td><p class="normal"><b><a href="datasets/Steel+Plates+Faults">Steel Plates Faults</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">A dataset of steel plates' faults, classified into 7 different types. The goal was to train machine learning for automatic pattern recognition.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">1941 </p></td>
<td><p class="normal">27 </p></td>
<td><p class="normal">2010 </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/MiniBooNE+particle+identification"></a> </td><td><p class="normal"><b><a href="datasets/MiniBooNE+particle+identification">MiniBooNE particle identification</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This dataset is taken from the MiniBooNE experiment and is used to distinguish electron neutrinos (signal) from muon neutrinos (background).&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">130065 </p></td>
<td><p class="normal">50 </p></td>
<td><p class="normal">2010 </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/YearPredictionMSD"></a> </td><td><p class="normal"><b><a href="datasets/YearPredictionMSD">YearPredictionMSD</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbsp;</p></td>

```

model, ~~model~~, commercial flights ranging from 1922 to 2011, which is found in one year ~~every year~~, . /p></td> -->

<td><p class="normal">Multivariate </p></td>

<td><p class="normal">Regression </p></td>

<td><p class="normal">Real </p></td>

<td><p class="normal">515345 </p></td>

<td><p class="normal">90 </p></td>

<td><p class="normal">2011 </p></td>

<!-- <td><p class="normal">Other </p></td> -->

</tr><tr bgcolor="DDEEFF">

<td><table><tr><td> </td><td><p class="normal">PEMS-SF</p></td></tr></table></td>

<!-- <td><p class="normal">15 months worth of daily data (440 daily records) that describes the occupancy rate, between 0 and 1, of different car lanes of the San Francisco bay area freeways across time. </p></td> -->

<td><p class="normal">Multivariate, Time-Series </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Real </p></td>

<td><p class="normal">440 </p></td>

<td><p class="normal">138672 </p></td>

<td><p class="normal">2011 </p></td>

<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">OpinRank Review Dataset</p></td></tr></table></td>

<!-- <td><p class="normal">This data set contains user reviews of cars and hotels collected from Tripadvisor (~259,000 reviews) and Edmunds (~42,230 reviews). </p></td> -->

<td><p class="normal">Text </p></td>

<td><p class="normal"> </p></td>

<td><p class="normal"> </p></td>

<td><p class="normal"> </p></td>

<td><p class="normal"> </p></td>

<td><p class="normal">2011 </p></td>

<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr bgcolor="DDEEFF">

<td><table><tr><td> </td><td><p class="normal">Relative location of CT slices on axial axis</p></td></tr></table></td>

<!-- <td><p class="normal">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. </p></td> -->

<td><p class="normal">Domain-Theory </p></td>

<td><p class="normal">Regression </p></td>

<td><p class="normal">Real </p></td>

<td><p class="normal">53500 </p></td>

<td><p class="normal">386 </p></td>

<td><p class="normal">2011 </p></td>

<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">Online Handwritten Assamese Characters Dataset</p></td></tr></table></td>

<!-- <td><p class="normal">This is a dataset of 8235 online handwritten assamese characters. The "online" process involves capturing of data as text is written on a digitizing tablet with an electronic pen. </p></td> -->

<td><p class="normal">Multivariate, Sequential </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Integer </p></td>

<td><p class="normal">8235 </p></td>

<td><p class="normal"> </p></td>

<td><p class="normal">2011 </p></td>

<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr bgcolor="DDEEFF">

<td><table><tr><td> </td><td><p class="normal">PubChem Bioassay Data</p></td></tr></table></td>

<!-- <td><p class="normal">These highly imbalanced bioassay datasets are from the differing types of screening that can be performed using HTS technology. 21 datasets were created from 12 bioassays. </p></td> -->

<td><p class="normal">Multivariate </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Integer, Real </p></td>

<td><n class="normal"> </n></td>

```

<!-- <p class="normal" > -->
<td><p class="normal"> </p></td>
<td><p class="normal">2011 </p></td>
<!-- <td><p class="normal">Life&nbsp;:</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Record+Linkage+Comparison+Patterns"></a> </td><td><p class="normal"><b><a
href="datasets/Record+Linkage+Comparison+Patterns">Record Linkage Comparison Patterns</a></b></p><
/td></tr></table></td>
<!-- <td><p class="normal">Element-wise comparison of records with personal data from a record lin
kage setting. The task is to decide from a comparison pattern whether the underlying records belon
g to one person.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">5749132 </p></td>
<td><p class="normal">12 </p></td>
<td><p class="normal">2011 </p></td>
<!-- <td><p class="normal">Other&nbsp;:</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Communities+and+Crime+Unnormalized"></a> </td><td><p class="normal"><b><a
href="datasets/Communities+and+Crime+Unnormalized">Communities and Crime Unnormalized</a></b></p><
/td></tr></table></td>
<!-- <td><p class="normal">Communities in the US. Data combines socio-economic data from the '90 C
ensus, law enforcement data from the 1990 Law Enforcement Management and Admin Stats survey, and c
rime data from the 1995 FBI UCR&nbsp;:</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">2215 </p></td>
<td><p class="normal">147 </p></td>
<td><p class="normal">2011 </p></td>
<!-- <td><p class="normal">Social&nbsp;:</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Vertebral+Column"></a> </td><td><p class="normal"><b><a
href="datasets/Vertebral+Column">Vertebral Column</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Data set containing values for six biomechanical features used to class
ify orthopaedic patients into 3 classes (normal, disk hernia or spondilolisthesis) or 2 classes (n
ormal or abnormal).&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">310 </p></td>
<td><p class="normal">6 </p></td>
<td><p class="normal">2011 </p></td>
<!-- <td><p class="normal">&nbsp;:</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/EMG+Physical+Action+Data+Set"></a> </td><td><p class="normal"><b><a
href="datasets/EMG+Physical+Action+Data+Set">EMG Physical Action Data Set</a></b></p></td></tr></t
able></td>
<!-- <td><p class="normal">The Physical Action Data Set includes 10 normal and 10 aggressive physi
cal actions that measure the human activity. The data have been collected by 4 subjects using the
Delsys EMG wireless apparatus.&nbsp;</p></td> -->
<td><p class="normal">Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">10000 </p></td>
<td><p class="normal">8 </p></td>
<td><p class="normal">2011 </p></td>
<!-- <td><p class="normal">Physical&nbsp;:</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Vicon+Physical+Action+Data+Set"></a> </td><td><p class="normal"><b><a
href="datasets/Vicon+Physical+Action+Data+Set">Vicon Physical Action Data Set</a></b></p></td></tr
></table></td>
<!-- <td><p class="normal">The Physical Action Data Set includes 10 normal and 10 aggressive physi
cal actions that measure the human activity. The data have been collected by 10 subjects using the
Vicon 3D tracker.&nbsp;</p></td> -->
<td><p class="normal">Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">3000 </p></td>
<td><p class="normal">27 </p></td>
<td><p class="normal">2011 </p></td>
<!-- <td><p class="normal">Physical&nbsp;:</p></td> -->

```

```

.    <tr><td> Normal / Unspecified </td></tr>
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Amazon+Commerce+reviews+set"></a> </td><td><p class="normal"><b><a
href="datasets/Amazon+Commerce+reviews+set">Amazon Commerce reviews set</a></b></p></td></tr></tab
le></td>
<!-- <td><p class="normal">The dataset is used for authorship identification in online Writeprint
which is a new research field of pattern recognition. &nbsp;</p></td> -->
<td><p class="normal">Multivariate, Text, Domain-Theory </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">1500 </p></td>
<td><p class="normal">10000 </p></td>
<td><p class="normal">2011 </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Amazon+Access+Samples"></a> </td><td><p class="normal"><b><a
href="datasets/Amazon+Access+Samples">Amazon Access Samples</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Amazon's InfoSec is getting smarter about the way Access data is
leveraged. This is an anonymized sample of access provisioned within the company.&nbsp;</p></td> --
->
<td><p class="normal">Time-Series, Domain-Theory </p></td>
<td><p class="normal">Regression, Clustering, Causal-Discovery </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">30000 </p></td>
<td><p class="normal">20000 </p></td>
<td><p class="normal">2011 </p></td>
<!-- <td><p class="normal">Business&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Reuter_50_50"></a> </td><td><p class="normal"><b><a
href="datasets/Reuter_50_50">Reuter_50_50</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The dataset is used for authorship identification in online Writeprint
which is a new research field of pattern recognition. &nbsp;</p></td> -->
<td><p class="normal">Multivariate, Text, Domain-Theory </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">2500 </p></td>
<td><p class="normal">10000 </p></td>
<td><p class="normal">2011 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Farm+Ads"></a> </td><td><p class="normal"><b><a
href="datasets/Farm+Ads">Farm Ads</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This data was collected from text ads found on twelve websites that dea
l with various farm animal related topics. The binary labels are based on whether or not the cont
ent owner approves of the ad.&nbsp;</p></td> -->
<td><p class="normal">Text </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">4143 </p></td>
<td><p class="normal">54877 </p></td>
<td><p class="normal">2011 </p></td>
<!-- <td><p class="normal">Business&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/DBWorld+e-mails"></a> </td><td><p class="normal"><b><a
href="datasets/DBWorld+e-mails">DBWorld e-mails</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">It contains 64 e-mails which I have manually collected from DBWorld mai
ling list. They are classified in: 'announces of conferences' and 'everything else'.&nbsp;</p></td
> -->
<td><p class="normal">Text </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">64 </p></td>
<td><p class="normal">4702 </p></td>
<td><p class="normal">2011 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29"></a> </td><td><p class="normal"><b><a
href="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29">KEGG Metabolic Relation Network (Di
rected)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">KEGG Metabolic pathways modeled as directed relation network. Variety o
f graphical features presented.&nbsp;</p></td> -->
</tr><n class="normal">Multivariate Uniariate Text </n></td>

```

<td><p class="normal">Classification, Regression, Clustering </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">53414 </p></td>
<td><p class="normal">24 </p></td>
<td><p class="normal">2011 </p></td>
<!-- <td><p class="normal">Life -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td> </td><td><p class="normal">KEGG Metabolic Reaction Network (Undirected)</p></td></tr></table></td>
<!-- <td><p class="normal">KEGG Metabolic pathways modeled as un-directed reaction network. Variety of graphical features presented. -->
<td><p class="normal">Multivariate, Univariate, Text </p></td>
<td><p class="normal">Classification, Regression, Clustering </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">65554 </p></td>
<td><p class="normal">29 </p></td>
<td><p class="normal">2011 </p></td>
<!-- <td><p class="normal">Life -->
</tr><tr>
<td><table><tr><td> </td><td><p class="normal">Bank Marketing</p></td></tr></table></td>
<!-- <td><p class="normal">The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe a term deposit (variable y). -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">45211 </p></td>
<td><p class="normal">17 </p></td>
<td><p class="normal">2012 </p></td>
<!-- <td><p class="normal">Business -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td> </td><td><p class="normal">YouTube Comedy Slam Preference Data</p></td></tr></table></td>
<!-- <td><p class="normal">This dataset provides user vote data on which video from a pair of videos is funnier collected on YouTube Comedy Slam. The task is to automatically predict this preference based on video metadata. -->
<td><p class="normal">Text </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">1138562 </p></td>
<td><p class="normal">3 </p></td>
<td><p class="normal">2012 </p></td>
<!-- <td><p class="normal">Computer -->
</tr><tr>
<td><table><tr><td> </td><td><p class="normal">Gas Sensor Array Drift Dataset</p></td></tr></table></td>
<!-- <td><p class="normal">This archive contains 13910 measurements from 16 chemical sensors utilized in simulations for drift compensation in a discrimination task of 6 gases at various levels of concentrations. -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">13910 </p></td>
<td><p class="normal">128 </p></td>
<td><p class="normal">2012 </p></td>
<!-- <td><p class="normal">Computer -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td> </td><td><p class="normal">ILPD (Indian Liver Patient Dataset)</p></td></tr></table></td>
<!-- <td><p class="normal">This data set contains 10 variables that are age, gender, total Bilirubin, direct Bilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT and Alkphos. -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">583 </p></td>
<td><p class="normal">10 </p></td>

<td><p class="normal">10 </p></td>

<td><p class="normal">2012 </p></td>

<!-- <td><p class="normal">Life ;</p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">OPPORTUNITY Activity Recognition</p></td>

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

<!-- <td><p class="normal">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). </p></td> -->

<td><p class="normal">Multivariate, Time-Series </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Real </p></td>

<td><p class="normal">2551 </p></td>

<td><p class="normal">242 </p></td>

<td><p class="normal">2012 </p></td>

<!-- <td><p class="normal">Computer ;</p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">Nomao</p></td></tr></table></td>

<!-- <td><p class="normal">Nomao collects data about places (name, phone, localization...) from many sources.

Deduplication consists in detecting what data refer to the same place.

Instances in the dataset compare 2 spots. </p></td> -->

<td><p class="normal">Univariate </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Real </p></td>

<td><p class="normal">34465 </p></td>

<td><p class="normal">120 </p></td>

<td><p class="normal">2012 </p></td>

<!-- <td><p class="normal">Computer ;</p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">SMS Spam Collection</p></td></tr></table></td>

<!-- <td><p class="normal">The SMS Spam Collection is a public set of SMS labeled messages that have been collected for mobile phone spam research. </p></td> -->

<td><p class="normal">Multivariate, Text, Domain-Theory </p></td>

<td><p class="normal">Classification, Clustering </p></td>

<td><p class="normal">Real </p></td>

<td><p class="normal">5574 </p></td>

<td><p class="normal"> </p></td>

<td><p class="normal">2012 </p></td>

<!-- <td><p class="normal">Computer ;</p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">Skin Segmentation</p></td></tr></table></td>

<!-- <td><p class="normal">The Skin Segmentation dataset is constructed over B, G, R color space. Skin and Nonskin dataset is generated using skin textures from face images of diversity of age, gender, and race people. </p></td> -->

<td><p class="normal">Univariate </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Real </p></td>

<td><p class="normal">245057 </p></td>

<td><p class="normal">4 </p></td>

<td><p class="normal">2012 </p></td>

<!-- <td><p class="normal">Computer ;</p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">Planning Relax</p></td></tr></table></td>

<!-- <td><p class="normal">The dataset concerns with the classification of two mental stages from recorded EEG signals: Planning (during imagination of motor act) and Relax state. </p></td> -->

<td><p class="normal">Univariate </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Real </p></td>

<td><p class="normal">182 </p></td>

<td><p class="normal">13 </p></td>

<td><p class="normal">2012 </p></td>

<!-- <td><p class="normal">Computer ;</p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">PAMAP2+Physical+Activity+Monitoring</p></td></tr></table></td>

```

ets/MLimages/SmallLargeDefault.jpg"/></a> </td><td><p class="normal"><b><a href="datasets/PAMAP2+Physical+Activity+Monitoring">PAMAP2 Physical Activity Monitoring</a></b></p></td>
<!-- <td><p class="normal">The PAMAP2 Physical Activity Monitoring dataset contains data of 18 different physical activities, performed by 9 subjects wearing 3 inertial measurement units and a heart rate monitor.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">3850505 </p></td>
<td><p class="normal">52 </p></td>
<td><p class="normal">2012 </p></td>
<!-- <td><p class="normal">Computer&nbs;p;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Restaurant+%26+consumer+data"></a> </td><td><p class="normal"><b><a href="datasets/Restaurant+%26+consumer+data">Restaurant &amp; consumer data</a></b></p></td></tr></table></td>
<!-- <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. &nbs;p;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">138 </p></td>
<td><p class="normal">47 </p></td>
<td><p class="normal">2012 </p></td>
<!-- <td><p class="normal">Computer&nbs;p;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/CNAE-9"></a> </td><td><p class="normal"><b><a href="datasets/CNAE-9">CNAE-9</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This is a data set containing 1080 documents of free text business descriptions of Brazilian companies categorized into a subset of 9 categories&nbs;p;</p></td> -->
<td><p class="normal">Multivariate, Text </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">1080 </p></td>
<td><p class="normal">857 </p></td>
<td><p class="normal">2012 </p></td>
<!-- <td><p class="normal">Business&nbs;p;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Individual+household+electric+power+consumption"></a> </td><td><p class="normal"><b><a href="datasets/Individual+household+electric+power+consumption">Individual household electric power consumption</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbs;p;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Regression, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">2075259 </p></td>
<td><p class="normal">9 </p></td>
<td><p class="normal">2012 </p></td>
<!-- <td><p class="normal">Physical&nbs;p;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/seeds"></a> </td><td><p class="normal"><b><a href="datasets/seeds">seeds</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbs;p;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">210 </p></td>
<td><p class="normal">7 </p></td>
<td><p class="normal">2012 </p></td>
<!-- <td><p class="normal">Life&nbs;p;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Northix"></a> </td><td><p class="normal"><b><a href="datasets/Northix">Northix</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Northix is designed to be a schema matching benchmark problem for data integration of two entity relationship databases. &nbs;p;</p></td> -->

```

```

<td><p class="normal">Multivariate, Univariate, Text </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">115 </p></td>
<td><p class="normal">200 </p></td>
<td><p class="normal">2012 </p></td>
<!-- <td><p class="normal">Computer&nbsp; </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/QtyT40I10D100K"></a> </td><td><p class="normal"><b><a href="datasets/QtyT40I10D100K">QtyT40I10D100K</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Since there is no numerical sequential data stream available in standard data sets, this data set is generated from the original T40I10D100K data set&nbsp;</p></td>
d> -->
<td><p class="normal">Sequential </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">3960456 </p></td>
<td><p class="normal">4 </p></td>
<td><p class="normal">2012 </p></td>
<!-- <td><p class="normal">&nbsp; </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Legal+Case+Reports"></a> </td><td><p class="normal"><b><a href="datasets/Legal+Case+Reports">Legal Case Reports</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbsp;</p></td> -->
<td><p class="normal">Text </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2012 </p></td>
<!-- <td><p class="normal">Other&nbsp; </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Human+Activity+Recognition+Using+Smartphones"></a> </td><td><p class="normal"><b><a href="datasets/Human+Activity+Recognition+Using+Smartphones">Human Activity Recognition Using Smartphones</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Human Activity Recognition database built from the recordings of 30 subjects performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">10299 </p></td>
<td><p class="normal">561 </p></td>
<td><p class="normal">2012 </p></td>
<!-- <td><p class="normal">Computer&nbsp; </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/One-hundred+plant+species+leaves+data+set"></a> </td><td><p class="normal"><b><a href="datasets/One-hundred+plant+species+leaves+data+set">One-hundred plant species leaves data set</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Sixteen samples of leaf each of one-hundred plant species. For each sample, a shape descriptor, fine scale margin and texture histogram are given.&nbsp;</p></td> -->
<td><p class="normal"> </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">1600 </p></td>
<td><p class="normal">64 </p></td>
<td><p class="normal">2012 </p></td>
<!-- <td><p class="normal">Life&nbsp; </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Energy+efficiency"></a> </td><td><p class="normal"><b><a href="datasets/Energy+efficiency">Energy efficiency</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This study looked into assessing the heating load and cooling load requirements of buildings (that is, energy efficiency) as a function of building parameters.&nbsp;</p></td>
</td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">768 </p></td>
<td><p class="normal">8 </p></td>
<td><p class="normal">2012 </p></td>

```

```

<!-- <td><p class="normal">Computer </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Yacht+Hydrodynamics"></a> </td><td><p class="normal"><b><a href="datasets/Yacht+Hydrodynamics">Yacht Hydrodynamics</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Delft data set, used to predict the hydodynamic performance of sailing yachts from dimensions and velocity. </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">308 </p></td>
<td><p class="normal">7 </p></td>
<td><p class="normal">2013 </p></td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Fertility"></a> </td><td><p class="normal"><b><a href="datasets/Fertility">Fertility</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">100 volunteers provide a semen sample analyzed according to the WHO 2010 criteria. Sperm concentration are related to socio-demographic data, environmental factors, health status, and life habits </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">100 </p></td>
<td><p class="normal">10 </p></td>
<td><p class="normal">2013 </p></td>
<!-- <td><p class="normal">Life </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Daphnet+Freezing+of+Gait"></a> </td><td><p class="normal"><b><a href="datasets/Daphnet+Freezing+of+Gait">Daphnet Freezing of Gait</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This dataset contains the annotated readings of 3 acceleration sensors at the hip and leg of Parkinson's disease patients that experience freezing of gait (FoG) during walking tasks.
<br/></p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">237 </p></td>
<td><p class="normal">9 </p></td>
<td><p class="normal">2013 </p></td>
<!-- <td><p class="normal">Life </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29"></a> </td><td><p class="normal"><b><a href="datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29">3D Road Network (North Jutland, Denmark)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">3D road network with highly accurate elevation information (+-20cm) from Denmark used in eco-routing and fuel/Co2-estimation routing algorithms. </p></td> -->
<td><p class="normal">Sequential, Text </p></td>
<td><p class="normal">Regression, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">434874 </p></td>
<td><p class="normal">4 </p></td>
<td><p class="normal">2013 </p></td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/ISTANBUL+STOCK+EXCHANGE"></a> </td><td><p class="normal"><b><a href="datasets/ISTANBUL+STOCK+EXCHANGE">ISTANBUL STOCK EXCHANGE</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&ampnbsp</p></td> -->
<td><p class="normal">Multivariate, Univariate, Time-Series </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">536 </p></td>
<td><p class="normal">8 </p></td>
<td><p class="normal">2013 </p></td>
<!-- <td><p class="normal">Business </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Buzz+in+social+media+"></a> </td><td><p class="normal"><b><a href="datasets/Buzz+in+social+media+">Buzz in social media </a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This data-set contains examples of buzz events from two different socia

```

l networks: Twitter, and Tom's Hardware, a forum network focusing on new technology with more conservative dynamics. </p></td> -->

<td><p class="normal">Time-Series, Multivariate </p></td>

<td><p class="normal">Regression, Classification </p></td>

<td><p class="normal">Integer, Real </p></td>

<td><p class="normal">140000 </p></td>

<td><p class="normal">77 </p></td>

<td><p class="normal">2013 </p></td>

<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">First-order theorem proving</p></td></tr></table></td>

<!-- <td><p class="normal">Given a theorem, predict which of five heuristics will give the fastest proof when used by a first-order prover. A sixth prediction declines to attempt a proof, should the theorem be too difficult. </p></td> -->

<td><p class="normal">Multivariate </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Real </p></td>

<td><p class="normal">6118 </p></td>

<td><p class="normal">51 </p></td>

<td><p class="normal">2013 </p></td>

<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">Wearable Computing: Classification of Body Postures and Movements (PUC-Rio)</p></td></tr></table></td>

<!-- <td><p class="normal">A dataset with 5 classes (sitting-down, standing-up, standing, walking, and sitting) collected on 8 hours of activities of 4 healthy subjects. We also established a baseline performance index. </p></td> -->

<td><p class="normal">Sequential </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Integer, Real </p></td>

<td><p class="normal">165632 </p></td>

<td><p class="normal">18 </p></td>

<td><p class="normal">2013 </p></td>

<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">Gas sensor arrays in open sampling settings</p></td></tr></table></td>

<!-- <td><p class="normal">The dataset contains 18000 time-series recordings from a chemical detection platform at six different locations in a wind tunnel facility in response to ten high-priority chemical gaseous substances </p></td> -->

<td><p class="normal">Multivariate, Time-Series </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Real </p></td>

<td><p class="normal">18000 </p></td>

<td><p class="normal">1950000 </p></td>

<td><p class="normal">2013 </p></td>

<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">Climate Model Simulation Crashes</p></td></tr></table></td>

<!-- <td><p class="normal">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. </p></td> -->

<td><p class="normal">Multivariate </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Real </p></td>

<td><p class="normal">540 </p></td>

<td><p class="normal">18 </p></td>

<td><p class="normal">2013 </p></td>

<!-- <td><p class="normal">Physical </p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">MicroMass</p></td></tr></table></td>

<!-- <td><p class="normal">A dataset to explore machine learning approaches for the identification of microorganisms from mass-spectrometry data. </p></td> -->

```

<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">931 </p></td>
<td><p class="normal">1300 </p></td>
<td><p class="normal">2013 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/QSAR+biodegradation"></a> </td><td><p class="normal"><b><a href="datasets/QSAR+biodegradation">QSAR biodegradation</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Data set containing values for 41 attributes (molecular descriptors) used to classify 1055 chemicals into 2 classes (ready and not ready biodegradable).&nbsp;</p></td> --
->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">1055 </p></td>
<td><p class="normal">41 </p></td>
<td><p class="normal">2013 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/BLOGGER"></a> </td><td><p class="normal"><b><a href="datasets/BLOGGER">BLOGGER</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">In this paper, we look for to recognize the causes of users tend to cyber space in Kohkiloye and Boyer Ahmad Province in Iran&nbsp;</p></td> --
->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">100 </p></td>
<td><p class="normal">6 </p></td>
<td><p class="normal">2013 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Daily+and+Sports+Activities"></a> </td><td><p class="normal"><b><a href="datasets/Daily+and+Sports+Activities">Daily and Sports Activities</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The dataset comprises motion sensor data of 19 daily and sports activities each performed by 8 subjects in their own style for 5 minutes. Five Xsens MTx units are used on the torso, arms, and legs.
&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">9120 </p></td>
<td><p class="normal">5625 </p></td>
<td><p class="normal">2013 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/User+Knowledge+Modeling"></a> </td><td><p class="normal"><b><a href="datasets/User+Knowledge+Modeling">User Knowledge Modeling</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbsp;</p></td>
-->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">403 </p></td>
<td><p class="normal">5 </p></td>
<td><p class="normal">2013 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection"></a> </td><td><p class="normal"><b><a href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection">Reuters RCV1 RCV2 Multilingual, Multiview Text Categorization Test collection</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This test collection contains feature characteristics of documents originally written in five different languages and their translations, over a common set of 6 categories. &nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>

```

```

<td><p class="normal">Real </p></td>
<td><p class="normal">111740 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2013 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/NYSK"></a> </td><td><p class="normal"><b><a href="datasets/NYSK">NYSK</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">NYSK (New York v. Strauss-Kahn) is a collection of English news articles about the case relating to allegations of sexual assault against the former IMF director Dominique Strauss-Kahn (May 2011).&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Sequential, Text </p></td>
<td><p class="normal">Clustering </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">10421 </p></td>
<td><p class="normal">7 </p></td>
<td><p class="normal">2013 </p></td>
<!-- <td><p class="normal">Social&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Turkiye+Student+Evaluation"></a> </td><td><p class="normal"><b><a href="datasets/Turkiye+Student+Evaluation">Turkiye Student Evaluation</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">5820 </p></td>
<td><p class="normal">33 </p></td>
<td><p class="normal">2013 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machine"></a> </td><td><p class="normal"><b><a href="datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machine">ser Knowledge Modeling Data (Students' Knowledge Levels on DC Electrical Machines)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">403 </p></td>
<td><p class="normal">5 </p></td>
<td><p class="normal">2013 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/EEG+Eye+State"></a> </td><td><p class="normal"><b><a href="datasets/EEG+Eye+State">EEG Eye State</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The data set consists of 14 EEG values and a value indicating the eye state.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Sequential, Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">14980 </p></td>
<td><p class="normal">15 </p></td>
<td><p class="normal">2013 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure"></a> </td><td><p class="normal"><b><a href="datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure">Physicochemical Properties of Protein Tertiary Structure</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This is a data set of Physicochemical Properties of Protein Tertiary Structure. The data set is taken from CASP 5-9. There are 45730 decoys and size varying from 0 to 21 armstrong.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">45730 </p></td>

```

<p class="normal">9</p>	<p class="normal">2013</p>	<p class="normal"><!-- <td><p class="normal">Life &nbsp;</p></td> --></p>
<p class="normal"><td><table><tr><td> </td><td><p class="normal">seismic-bumps</p></td></tr></table></td></p>	<p class="normal"><!-- <td><p class="normal">The data describe the problem of high energy (higher than 10^4 J) seismic bumps forecasting in a coal mine. Data come from two of longwalls located in a Polish coal mine.&nbsp;</p></td> --></p>	<p class="normal"><td><p class="normal">Multivariate </p></td></p>
<p class="normal"><td><p class="normal">Classification </p></td></p>	<p class="normal"><td><p class="normal">Real </p></td></p>	<p class="normal"><td><p class="normal">2584 </p></td></p>
<p class="normal"><td><p class="normal">19 </p></td></p>	<p class="normal"><td><p class="normal">2013 </p></td></p>	<p class="normal"><!-- <td><p class="normal">Other&nbsp;</p></td> --></p>
<p class="normal"></tr><tr></p>	<p class="normal"><td><table><tr><td> </td><td><p class="normal">banknote authentication</p></td></tr></table></td></p>	<p class="normal"><!-- <td><p class="normal">Data were extracted from images that were taken for the evaluation of a n authentication procedure for banknotes.&nbsp;</p></td> --></p>
<p class="normal"><td><p class="normal">Multivariate </p></td></p>	<p class="normal"><td><p class="normal">Classification </p></td></p>	<p class="normal"><td><p class="normal">Real </p></td></p>
<p class="normal"><td><p class="normal">1372 </p></td></p>	<p class="normal"><td><p class="normal">5 </p></td></p>	<p class="normal"><td><p class="normal">2013 </p></td></p>
<p class="normal"><!-- <td><p class="normal">Computer&nbsp;</p></td> --></p>	<p class="normal"></tr><tr></p>	<p class="normal"><td><table><tr><td> </td><td><p class="normal">USPTO Algorithm Challenge, run by NASA-Harvard Tournament Lab and TopCoder Problem: Pat</p></td></tr></table></td></p>
<p class="normal"><!-- <td><p class="normal">Data used for USPTO Algorithm Competition. Contains drawing pages from US patents with manually labeled figure and part labels.&nbsp;</p></td> --></p>	<p class="normal"><td><p class="normal">Domain-Theory </p></td></p>	<p class="normal"><td><p class="normal">Classification </p></td></p>
<p class="normal"><td><p class="normal">Integer </p></td></p>	<p class="normal"><td><p class="normal">306 </p></td></p>	<p class="normal"><td><p class="normal">5 </p></td></p>
<p class="normal"><td><p class="normal">2013 </p></td></p>	<p class="normal"><!-- <td><p class="normal">Other&nbsp;</p></td> --></p>	<p class="normal"></tr><tr></p>
<p class="normal"><td><table><tr><td> </td><td><p class="normal">YouTube Multiview Video Games Dataset</p></td></tr></table></td></p>	<p class="normal"><!-- <td><p class="normal">This dataset contains about 120k instances, each described by 13 feature types, with class information, specially useful for exploring multiview topics (cotraining, ensembles, clustering,...).&nbsp;</p></td> --></p>	<p class="normal"><td><p class="normal">Multivariate, Text </p></td></p>
<p class="normal"><td><p class="normal">Classification, Clustering </p></td></p>	<p class="normal"><td><p class="normal">Integer, Real </p></td></p>	<p class="normal"><td><p class="normal">120000 </p></td></p>
<p class="normal"><td><p class="normal">1000000 </p></td></p>	<p class="normal"><td><p class="normal">2013 </p></td></p>	<p class="normal"><td><p class="normal">2013 </p></td></p>
<p class="normal"><!-- <td><p class="normal">Computer&nbsp;</p></td> --></p>	<p class="normal"></tr><tr></p>	<p class="normal"><td><table><tr><td> </td><td><p class="normal">Gas Sensor Array Drift Dataset at Different Concentrations</p></td></tr></table></td></p>
<p class="normal"><!-- <td><p class="normal">This archive contains 13910 measurements from 16 chemical sensors exposed to 6 different gases at various concentration levels.&nbsp;</p></td> --></p>	<p class="normal"><td><p class="normal">Multivariate, Time-Series </p></td></p>	<p class="normal"><td><p class="normal">Classification, Regression, Clustering, Causa </p></td></p>
<p class="normal"><td><p class="normal">Real </p></td></p>	<p class="normal"><td><p class="normal">13910 </p></td></p>	<p class="normal"><td><p class="normal">129 </p></td></p>
<p class="normal"><td><p class="normal">2013 </p></td></p>	<p class="normal"><!-- <td><p class="normal">Computer&nbsp;</p></td> --></p>	<p class="normal"></tr><tr></p>
<p class="normal"><td><table><tr><td> </td><td><p class="normal">...</p></td></tr></table></td></p>	<p class="normal"><!-- <td><p class="normal">...</p></td> --></p>	<p class="normal"><td><p class="normal">... </p></td></p>

href="datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors"> </td><td><p class="normal">Activities of Daily Living (ADLs) Recognition Using Binary Sensors</p></td></tr></table></td>
 <!-- <td><p class="normal">This dataset comprises information regarding the ADLs performed by two users on a daily basis in their own homes. </p></td> -->
 <td><p class="normal">Multivariate, Sequential, Time-Series </p></td>
 <td><p class="normal">Classification, Clustering </p></td>
 <td><p class="normal"> </p></td>
 <td><p class="normal">2747 </p></td>
 <td><p class="normal"> </p></td>
 <td><p class="normal">2013 </p></td>
 <!-- <td><p class="normal">Computer </p></td> -->
 </tr><tr bgcolor="DDEEFF">
 <td><table><tr><td> </td><td><p class="normal">SkillCraft1 Master Table Dataset</p></td>
 </tr></table></td>
 <!-- <td><p class="normal">This data was used in Thompson et al. (2013). A list of possible game actions is discussed in Thompson, Blair, Chen, & Henrey (2013). </p></td> -->
 <td><p class="normal">Multivariate </p></td>
 <td><p class="normal">Regression </p></td>
 <td><p class="normal">Integer, Real </p></td>
 <td><p class="normal">3395 </p></td>
 <td><p class="normal">20 </p></td>
 <td><p class="normal">2013 </p></td>
 <!-- <td><p class="normal">Game </p></td> -->
 </tr><tr>
 <td><table><tr><td> </td><td><p class="normal">Weight Lifting Exercises monitored with Inertial Measurement Units</p></td></tr></table></td>
 <!-- <td><p class="normal">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. </p></td> -->
 <td><p class="normal">Multivariate </p></td>
 <td><p class="normal">Classification </p></td>
 <td><p class="normal">Real </p></td>
 <td><p class="normal">39242 </p></td>
 <td><p class="normal">152 </p></td>
 <td><p class="normal">2013 </p></td>
 <!-- <td><p class="normal">Physical </p></td> -->
 </tr><tr bgcolor="DDEEFF">
 <td><table><tr><td> </td><td><p class="normal">SML2010</p></td></tr></table></td>
 <!-- <td><p class="normal">This dataset is collected from a monitor system mounted in a domotic house. It corresponds to approximately 40 days of monitoring data. </p></td> -->
 <td><p class="normal">Multivariate, Sequential, Time-Series, Text </p></td>
 <td><p class="normal">Regression </p></td>
 <td><p class="normal">Real </p></td>
 <td><p class="normal">4137 </p></td>
 <td><p class="normal">24 </p></td>
 <td><p class="normal">2014 </p></td>
 <!-- <td><p class="normal">Computer </p></td> -->
 </tr><tr>
 <td><table><tr><td> </td><td><p class="normal">Bike Sharing Dataset</p></td></tr></table></td>
 <!-- <td><p class="normal">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></td> -->
 <td><p class="normal">Univariate </p></td>
 <td><p class="normal">Regression </p></td>
 <td><p class="normal">Integer, Real </p></td>
 <td><p class="normal">17389 </p></td>
 <td><p class="normal">16 </p></td>
 <td><p class="normal">2013 </p></td>
 <!-- <td><p class="normal">Social </p></td> -->
 </tr><tr bgcolor="DDEEFF">
 <td><table><tr><td> </td><td><p class="normal">Predict keywords activities in a online social media</p></td></tr></table></td>
 <!-- <td><p class="normal">The data from Twitter was collected during 360 consecutive days. It was

done by querying 1497 English keywords sampled from Wikipedia. This dataset is proposed in a Learning to rank setting. </p></td> -->

<td><p class="normal">Multivariate, Sequential, Time-Series </p></td>

<td><p class="normal"> </p></td>

<td><p class="normal">Integer, Real </p></td>

<td><p class="normal">51 </p></td>

<td><p class="normal">35 </p></td>

<td><p class="normal">2013 </p></td>

<!-- <td><p class="normal">Computer&nbs;p;</p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">Thoracic Surgery Data</p></td></tr></table></td>

<!-- <td><p class="normal">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.&nbs;p;</p></td> -->
--

<td><p class="normal">Multivariate </p></td>
--

<td><p class="normal">Classification </p></td>
--

<td><p class="normal">Integer, Real </p></td>

<td><p class="normal">470 </p></td>

<td><p class="normal">17 </p></td>

<td><p class="normal">2013 </p></td>

<!-- <td><p class="normal">Life&nbs;p;</p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">EMG dataset in Lower Limb</p></td></tr></table></td>

<!-- <td><p class="normal">3 different exercises: sitting, standing and walking in the muscles: biceps femoris, vastus medialis, rectus femoris and semitendinosus addition to goniometry in the exercises.&nbs;p;</p></td> -->

<td><p class="normal">Multivariate, Time-Series </p></td>

<td><p class="normal"> </p></td>

<td><p class="normal">Real </p></td>

<td><p class="normal">132 </p></td>

<td><p class="normal">5 </p></td>

<td><p class="normal">2014 </p></td>

<!-- <td><p class="normal">Computer&nbs;p;</p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">SUSY</p></td></tr></table></td>
--

<!-- <td><p class="normal">This is a classification problem to distinguish between a signal process which produces supersymmetric particles and a background process which does not.&nbs;p;</p></td> -->
--

<td><p class="normal"> </p></td>

<td><p class="normal">Classification </p></td>
--

<td><p class="normal">Real </p></td>

<td><p class="normal">5000000 </p></td>

<td><p class="normal">18 </p></td>

<td><p class="normal">2014 </p></td>

<!-- <td><p class="normal">Physical&nbs;p;</p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">HIGGS</p></td></tr></table></td>

<!-- <td><p class="normal">This is a classification problem to distinguish between a signal process which produces Higgs bosons and a background process which does not. &nbs;p;</p></td> -->

<td><p class="normal"> </p></td>

<td><p class="normal">Classification </p></td>
--

<td><p class="normal">Real </p></td>

<td><p class="normal">11000000 </p></td>
--

<td><p class="normal">28 </p></td>

<td><p class="normal">2014 </p></td>

<!-- <td><p class="normal">Physical&nbs;p;</p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">Qualitative_Bankruptcy</p></td></tr></table></td>
--

<!-- <td><p class="normal">Predict the Bankruptcy from Qualitative parameters from experts.&nbs;p;</p></td> -->

<td><p class="normal">Multivariate </p></td>
--

<td><p class="normal">Classification </p></td>
--

<td><p class="normal"> </p></td>

<td><p class="normal">250 </p></td>

<td><p class="normal">7 </p></td>

<td><p class="normal">2014 </p></td>

```

<!-- <td><p class="normal">Computer </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/LSVT+Voice+Rehabilitation"></a> </td><td><p class="normal"><b><a href="datasets/LSVT+Voice+Rehabilitation">LSVT Voice Rehabilitation</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">126 samples from 14 participants, 309 features. Aim: assess whether voice rehabilitation treatment lead to phonations considered 'acceptable' or 'unacceptable' (binary class classification problem). </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">126 </p></td>
<td><p class="normal">309 </p></td>
<td><p class="normal">2014 </p></td>
<!-- <td><p class="normal">Life </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer"></a> </td><td><p class="normal"><b><a href="datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer">Dataset for ADL Recognition with Wrist-worn Accelerometer</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Recordings of 16 volunteers performing 14 Activities of Daily Living (ADL) while carrying a single wrist-worn tri-axial accelerometer. </p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">3 </p></td>
<td><p class="normal">2014 </p></td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Wilt"></a> </td><td><p class="normal"><b><a href="datasets/Wilt">Wilt</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">High-resolution Remote Sensing data set (Quickbird). Small number of training samples of diseased trees, large number for other land cover. Testing data set from stratified random sample of image.&ampnbsp</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">4889 </p></td>
<td><p class="normal">6 </p></td>
<td><p class="normal">2014 </p></td>
<!-- <td><p class="normal">Life </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/User+Identification+From+Walking+Activity"></a> </td><td><p class="normal"><b><a href="datasets/User+Identification+From+Walking+Activity">User Identification From Walking Activity</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The dataset collects data from an Android smartphone positioned in the chest pocket from 22 participants walking in the wild over a predefined path.&ampnbsp</p></td> -->
<td><p class="normal">Univariate, Sequential, Time-Series </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2014 </p></td>
<!-- <td><p class="normal">Other </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer"></a> </td><td><p class="normal"><b><a href="datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer">Activity Recognition from Single Chest-Mounted Accelerometer</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The dataset collects data from a wearable accelerometer mounted on the chest. The dataset is intended for Activity Recognition research purposes.&ampnbsp</p></td> -->
<td><p class="normal">Univariate, Sequential, Time-Series </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2014 </p></td>
<!-- <td><p class="normal">Other </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Leaf"></a> </td><td><p class="normal"><b><a href="datasets/Leaf">Leaf</a></b></p></td></tr></table></td>

```

href="datasets/Leaf">Leaf</p></td></tr></table></td>
 <!-- <td><p class="normal">This dataset consists in a collection of shape and texture features extracted from digital images of leaf specimens originating from a total of 40 different plant species. </p></td> -->
 <td><p class="normal">Multivariate </p></td>
 <td><p class="normal">Classification </p></td>
 <td><p class="normal">Real </p></td>
 <td><p class="normal">340 </p></td>
 <td><p class="normal">16 </p></td>
 <td><p class="normal">2014 </p></td>
 <!-- <td><p class="normal">Computer </p></td> -->
 </tr><tr bgcolor="#DDEEFF">
 <td><table><tr><td> </td><td><p class="normal">Dresses_Attribute_Sales</p></td></tr></table></td>
 <!-- <td><p class="normal">This dataset contain Attributes of dresses and their recommendations according to their sales.Sales are monitor on the basis of alternate days. </p></td> -->
 <td><p class="normal">Text </p></td>
 <td><p class="normal">Classification, Clustering </p></td>
 <td><p class="normal"> </p></td>
 <td><p class="normal">501 </p></td>
 <td><p class="normal">13 </p></td>
 <td><p class="normal">2014 </p></td>
 <!-- <td><p class="normal">Computer </p></td> -->
 </tr><tr>
 <td><table><tr><td> </td><td><p class="normal">Tamilnadu Electricity Board Hourly Readings</p></td></tr></table></td>
 <!-- <td><p class="normal">This data can be effectively produced the result to fewer parameter of the Load profile can be reduced in the Database </p></td> -->
 <td><p class="normal">Multivariate </p></td>
 <td><p class="normal">Classification, Regression, Clustering </p></td>
 <td><p class="normal">Real </p></td>
 <td><p class="normal">45781 </p></td>
 <td><p class="normal">5 </p></td>
 <td><p class="normal">2013 </p></td>
 <!-- <td><p class="normal">Life </p></td> -->
 </tr><tr bgcolor="#DDEEFF">
 <td><table><tr><td> </td><td><p class="normal">Airfoil Self-Noise</p></td></tr></table></td>
 <!-- <td><p class="normal">NASA data set, obtained from a series of aerodynamic and acoustic tests of two and three-dimensional airfoil blade sections conducted in an anechoic wind tunnel. </p></td> -->
 <td><p class="normal">Multivariate </p></td>
 <td><p class="normal">Regression </p></td>
 <td><p class="normal">Real </p></td>
 <td><p class="normal">1503 </p></td>
 <td><p class="normal">6 </p></td>
 <td><p class="normal">2014 </p></td>
 <!-- <td><p class="normal">Physical </p></td> -->
 </tr><tr>
 <td><table><tr><td> </td><td><p class="normal">Wholesale customers</p></td></tr></table></td>
 <!-- <td><p class="normal">The data set refers to clients of a wholesale distributor. It includes the annual spending in monetary units (m.u.) on diverse product categories </p></td> -->
 <td><p class="normal">Multivariate </p></td>
 <td><p class="normal">Classification, Clustering </p></td>
 <td><p class="normal">Integer </p></td>
 <td><p class="normal">440 </p></td>
 <td><p class="normal">8 </p></td>
 <td><p class="normal">2014 </p></td>
 <!-- <td><p class="normal">Business </p></td> -->
 </tr><tr bgcolor="#DDEEFF">
 <td><table><tr><td> </td><td><p class="normal">Twitter Data set for Arabic Sentiment Analysis</p></td></tr></table></td>
 <!-- <td><p class="normal">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. </p></td> -->
 <td><p class="normal">Text </p></td>
 <td><p class="normal">Classification </p></td>
 <td><p class="normal"> </p></td>
 <td><p class="normal">2000 </p></td>

```

<td><p class="normal">2 </p></td>
<td><p class="normal">2014 </p></td>
<!-- <td><p class="normal">Social &nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Combined+Cycle+Power+Plant"></a> </td><td><p class="normal"><b><a href="datasets/Combined+Cycle+Power+Plant">Combined Cycle Power Plant</a></b></p></td></tr>
</table></td>
<!-- <td><p class="normal">The dataset contains 9568 data points collected from a Combined Cycle P ower Plant over 6 years (2006-2011), when the plant was set to work with full load. &nbsp;</p></td>
> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">9568 </p></td>
<td><p class="normal">4 </p></td>
<td><p class="normal">2014 </p></td>
<!-- <td><p class="normal">Computer &nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Urban+Land+Cover"></a> </td><td><p class="normal"><b><a href="datasets/Urban+Land+Cover">Urban Land Cover</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Classification of urban land cover using high resolution aerial imagery. Intended to assist sustainable urban planning efforts.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">168 </p></td>
<td><p class="normal">148 </p></td>
<td><p class="normal">2014 </p></td>
<!-- <td><p class="normal">Physical &nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Diabetes+130-US+hospitals+for+years+1999-2008"></a> </td><td><p class="normal"><b><a href="datasets/Diabetes+130-US+hospitals+for+years+1999-2008">Diabetes 130-US hospitals for years 1999-2008</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This data has been prepared to analyze factors related to readmission a s well as other

outcomes pertaining to patients with diabetes.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">100000 </p></td>
<td><p class="normal">55 </p></td>
<td><p class="normal">2014 </p></td>
<!-- <td><p class="normal">Life &nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Bach+Choral+Harmony"></a> </td><td><p class="normal"><b><a href="datasets/Bach+Choral+Harmony">Bach Choral Harmony</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbsp;</p></td> -->
<td><p class="normal">Sequential </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">5665 </p></td>
<td><p class="normal">17 </p></td>
<td><p class="normal">2014 </p></td>
<!-- <td><p class="normal">Other &nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/StoneFlakes"></a> </td><td><p class="normal"><b><a href="datasets/StoneFlakes">StoneFlakes</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Clustering, Causal-Discovery </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">79 </p></td>
<td><p class="normal">8 </p></td>
<td><p class="normal">2014 </p></td>
<!-- <td><p class="normal">Other &nbsp;</p></td> -->
</tr><tr bacolor="#DDEEFF">
```

```

., etc. -- ~~~~~ .
<td><table><tr><td><a href="datasets/Tennis+Major+Tournament+Match+Statistics"><img border="1" src ="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td><p class="normal"><b><a href="datasets/Tennis+Major+Tournament+Match+Statistics">Tennis Major Tournament Match Statistics</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Regression, Clustering </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">127 </p></td>
<td><p class="normal">42 </p></td>
<td><p class="normal">2014 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings"><img border="1" src ="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td><p class="normal"><b><a href="datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings">Parkinson Speech Dataset with Multiple Types of Sound Recordings</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The training data belongs to 20 Parkinson's Disease (PD) patients and 2 0 healthy subjects. From all subjects, multiple types of sound recordings (26) are taken.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">1040 </p></td>
<td><p class="normal">26 </p></td>
<td><p class="normal">2014 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Gesture+Phase+Segmentation"><img border="1" src ="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td><p class="normal"><b><a href="datasets/Gesture+Phase+Segmentation">Gesture Phase Segmentation</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Sequential, Time-Series </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">9900 </p></td>
<td><p class="normal">50 </p></td>
<td><p class="normal">2014 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Perfume+Data"><img border="1" src ="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td><p class="normal"><b><a href="datasets/Perfume+Data">Perfume Data</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbsp;</p></td> -->
<td><p class="normal">Univariate, Domain-Theory </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">560 </p></td>
<td><p class="normal">2 </p></td>
<td><p class="normal">2014 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/BlogFeedback"><img border="1" src ="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td><p class="normal"><b><a href="datasets/BlogFeedback">BlogFeedback</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Instances in this dataset contain features extracted from blog posts. The task associated with the data is to predict how many comments the post will receive.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">60021 </p></td>
<td><p class="normal">281 </p></td>
<td><p class="normal">2014 </p></td>
<!-- <td><p class="normal">Social&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/REALDISP+Activity+Recognition+Dataset"><img border="1" src ="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td><p class="normal"><b><a href="datasets/REALDISP+Activity+Recognition+Dataset">REAL.DISP+Activity+Recognition+Dataset</a></b></p></td></tr></table>
```

<td> REALDISP Activity Recognition </td>

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

<!-- <td><p class="normal">The REALDISP dataset is devised to evaluate techniques dealing with the effects of sensor displacement in wearable activity recognition as well as to benchmark general activity recognition algorithms </p></td> -->

<td><p class="normal">Multivariate, Time-Series </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Real </p></td>

<td><p class="normal">1419 </p></td>

<td><p class="normal">120 </p></td>

<td><p class="normal">2014 </p></td>

<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">Newspaper and magazine images segmentation dataset</p></td></tr></table></td>

<!-- <td><p class="normal">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. </p></td> -->

<td><p class="normal"> </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal"> </p></td>

<td><p class="normal">101 </p></td>

<td><p class="normal"> </p></td>

<td><p class="normal">2014 </p></td>

<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">AAAI 2014 Accepted Papers</p></td></tr></table></td>

<!-- <td><p class="normal">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></td> -->

<td><p class="normal">Multivariate </p></td>

<td><p class="normal">Clustering </p></td>

<td><p class="normal"> </p></td>

<td><p class="normal">399 </p></td>

<td><p class="normal">6 </p></td>

<td><p class="normal">2014 </p></td>

<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">Gas sensor array under flow modulation</p></td></tr></table></td>

<!-- <td><p class="normal">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. </p></td> -->

<td><p class="normal">Multivariate, Time-Series </p></td>

<td><p class="normal">Classification, Regression </p></td>

<td><p class="normal">Real </p></td>

<td><p class="normal">58 </p></td>

<td><p class="normal">120432 </p></td>

<td><p class="normal">2014 </p></td>

<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">Gas sensor array exposed to turbulent gas mixtures</p></td></tr></table></td>

<!-- <td><p class="normal">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. </p></td> -->

<td><p class="normal">Multivariate, Time-Series </p></td>

<td><p class="normal">Classification, Regression </p></td>

<td><p class="normal">Real </p></td>

<td><p class="normal">180 </p></td>

<td><p class="normal">150000 </p></td>

<td><p class="normal">2014 </p></td>

<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">UJIIndoorLoc</p></td></tr></table></td>

<!-- <td><p class="normal">The UJIIndoorLoc is a Multi-Building Multi-Floor indoor localization database to test Indoor Positioning System that rely on WLAN/WiFi fingerprint .</p></td> -->

case to test through positioning system that rely on many with longitude & latitude, ~, ~, ~, ~

<p class="normal">Multivariate</p>	<p class="normal">Classification, Regression</p>	<p class="normal">Integer, Real</p>
<p class="normal">21048</p>	<p class="normal">529</p>	<p class="normal">2014</p>
<p>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.</p>		
<p>Text</p>		
<p>Classification</p>		
<p>Integer</p>		
<p>2014</p>		
<p>Other</p>		

Dow+Jones+Index

This dataset contains weekly data for the Dow Jones Industrial Index. It has been used in computational investing research.

<p>Time-Series</p>	<p>Classification, Clustering</p>	<p>Integer, Real</p>
<p>750</p>	<p>16</p>	<p>2014</p>
<p>Business</p>		

The "sEMG for Basic Hand movements" includes 2 databases of surface electromyographic signals of 6 hand movements using Delsys' EMG System. Healthy subjects conducted six daily life grasps.

<p>Time-Series</p>	<p>Classification</p>	<p>Real</p>
<p>3000</p>	<p>2500</p>	<p>2014</p>
<p>Life</p>		

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.

<p>Multivariate</p>	<p>Clustering</p>	<p>150</p>
<p>5</p>	<p>2014</p>	<p>Computer</p>

Instances in this dataset contain audio features extracted from 1059 wave files. The task associated with the data is to predict the geographical origin of music.

<p>Multivariate</p>	<p>Classification, Regression</p>	<p>Real</p>
<p>1059</p>		

```

<!--><p class="normal" /></p>
<td><p class="normal">2014 </p></td>
<!-- <td><p class="normal">Other &nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants"></a> </td><td><p class="normal"><b><a href="datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants">Condition Based Maintenance of Naval Propulsion Plants</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">11934 </p></td>
<td><p class="normal">16 </p></td>
<td><p class="normal">2014 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Grammatical+Facial+Expressions"></a> </td><td><p class="normal"><b><a href="datasets/Grammatical+Facial+Expressions">Grammatical Facial Expressions</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This dataset supports the development of models that make possible to interpret Grammatical Facial Expressions from Brazilian Sign Language (Libras).&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Sequential </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">27965 </p></td>
<td><p class="normal">100 </p></td>
<td><p class="normal">2014 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/NoisyOffice"></a> </td><td><p class="normal"><b><a href="datasets/NoisyOffice">NoisyOffice</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Corpus intended to do cleaning (or binarization) and enhancement of noisy grayscale printed text images using supervised learning methods. Noisy images and their corresponding ground truth provided.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">216 </p></td>
<td><p class="normal">216 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/MHEALTH+Dataset"></a> </td><td><p class="normal"><b><a href="datasets/MHEALTH+Dataset">MHEALTH Dataset</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The MHEALTH (Mobile Health) dataset is devised to benchmark techniques dealing with human behavior analysis based on multimodal body sensing.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">120 </p></td>
<td><p class="normal">23 </p></td>
<td><p class="normal">2014 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Student+Performance"></a> </td><td><p class="normal"><b><a href="datasets/Student+Performance">Student Performance</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Predict student performance in secondary education (high school).&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">649 </p></td>
<td><p class="normal">33 </p></td>
<td><p class="normal">2014 </p></td>
<!-- <td><p class="normal">Social&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/ElectricityLoadDiagrams20112014"></a> </td><td><p class="normal"><b><a href="datasets/ElectricityLoadDiagrams20112014">ElectricityLoadDiagrams20112014</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This dataset contains electricity consumption of 270 houses/ollients

```

```

<!-- <tr><p class="normal">This data set contains electricity consumption of 570 points/clients.  

&nbsp;</p></td> -->
<td><p class="normal">Time-Series </p></td>
<td><p class="normal">Regression, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">370 </p></td>
<td><p class="normal">140256 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Gas+sensor+array+under+dynamic+gas+mixtures"></a> </td><td><p class="normal"><b><a href="datasets/Gas+sensor+array+under+dynamic+gas+mixtures">Gas sensor array under dynamic gas mixtures</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The data set contains the recordings of 16 chemical sensors exposed to two dynamic gas mixtures at varying concentrations. For each mixture, signals were acquired continuously during 12 hours.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">4178504 </p></td>
<td><p class="normal">19 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/microblogPCU"></a> </td><td><p class="normal"><b><a href="datasets/microblogPCU">microblogPCU</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">MicroblogPCU data is crawled from sina weibo microblog[http://weibo.com/]. This data can be used to study machine learning methods as well as do some social network research. &nbsp;</p></td> -->
<td><p class="normal">Multivariate, Univariate, Sequential, Text </p></td>
<td><p class="normal">Classification, Causal-Discovery </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">221579 </p></td>
<td><p class="normal">20 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Firm-Teacher_Clave-Direction_Classification"></a> </td><td><p class="normal"><b><a href="datasets/Firm-Teacher_Clave-Direction_Classification">Firm-Teacher_Clave-Direction_Classification</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The data are binary attack-point vectors and their clave-direction classes(es) according to the partido-alto-based paradigm.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">10800 </p></td>
<td><p class="normal">20 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Dataset+for+Sensorless+Drive+Diagnosis"></a> </td><td><p class="normal"><b><a href="datasets/Dataset+for+Sensorless+Drive+Diagnosis">Dataset for Sensorless Drive Diagnosis</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Features are extracted from motor current. The motor has intact and defective components. This results in 11 different classes with different conditions. &nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">58509 </p></td>
<td><p class="normal">49 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/TV+News+Channel+Commercial+Detection+Dataset"></a> </td><td><p class="normal"><b><a href="datasets/TV+News+Channel+Commercial+Detection+Dataset">TV News Channel Commercial Detection Dataset</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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). &nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Clustering </p></td>

```

```

<ta><p class="normal">Real </p></ta>
<td><p class="normal">129685 </p></td>
<td><p class="normal">12 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Phishing+Websites"></a> </td><td><p class="normal"><b><a href="datasets/Phishing+Websites">Phishing Websites</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This dataset collected mainly from: PhishTank archive, MillerSmiles archive, Google's searching operators.&nbsp;</p></td> -->
<td><p class="normal"> </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">2456 </p></td>
<td><p class="normal">30 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Computer Security&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Greenhouse+Gas+Observing+Network"></a> </td><td><p class="normal"><b><a href="datasets/Greenhouse+Gas+Observing+Network">Greenhouse Gas Observing Network</a></b></p></td>
</tr></table></td>
<!-- <td><p class="normal">Design an observing network to monitor emissions of a greenhouse gas (GHG) in California given time series of synthetic observations and tracers from weather model simulations.
&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">2921 </p></td>
<td><p class="normal">5232 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Diabetic+Retinopathy+Debrecen+Data+Set"></a> </td><td><p class="normal"><b><a href="datasets/Diabetic+Retinopathy+Debrecen+Data+Set">Diabetic Retinopathy Debrecen Data Set</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This dataset contains features extracted from the Messidor image set to predict whether an image contains signs of diabetic retinopathy or not. &nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">1151 </p></td>
<td><p class="normal">20 </p></td>
<td><p class="normal">2014 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/HIV-1+protease+cleavage"></a> </td><td><p class="normal"><b><a href="datasets/HIV-1+protease+cleavage">HIV-1 protease cleavage</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The data contains lists of octamers (8 amino acids) and a flag (-1 or 1) depending on whether HIV-1 protease will cleave in the central position (between amino acids 4 and 5).&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal">6590 </p></td>
<td><p class="normal">1 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Sentiment+Labelled+Sentences"></a> </td><td><p class="normal"><b><a href="datasets/Sentiment+Labelled+Sentences">Sentiment Labelled Sentences</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The dataset contains sentences labelled with positive or negative sentiment.&nbsp;</p></td> -->
<td><p class="normal">Text </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">3000 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">

```

```

<td><table><tr><td><a href="datasets/Online+News+Popularity"></a> </td><td><p class="normal"><b><a href="datasets/Online+News+Popularity">Online News Popularity</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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).&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">39797 </p></td>
<td><p class="normal">61 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Business&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Forest+type+mapping"></a> </td><td><p class="normal"><b><a href="datasets/Forest+type+mapping">Forest type mapping</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Multi-temporal remote sensing data of a forested area in Japan. The goal is to map different forest types using spectral data.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">326 </p></td>
<td><p class="normal">27 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/wiki4HE"></a> </td><td><p class="normal"><b><a href="datasets/wiki4HE">wiki4HE</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Survey of faculty members from two Spanish universities on teaching uses of Wikipedia&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression, Clustering, Causal-Discovery </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">913 </p></td>
<td><p class="normal">53 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Social&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset"></a> </td><td><p class="normal"><b><a href="datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset">Online Video Characteristics and Transcoding Time Dataset</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The dataset contains a million randomly sampled video instances listing 10 fundamental video characteristics along with the YouTube video ID. &nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">168286 </p></td>
<td><p class="normal">11 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Chronic_Kidney_Disease"></a> </td><td><p class="normal"><b><a href="datasets/Chronic_Kidney_Disease">Chronic_Kidney_Disease</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This dataset can be used to predict the chronic kidney disease and it can be collected from the hospital nearly 2 months of period.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">400 </p></td>
<td><p class="normal">25 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014"></a> </td><td><p class="normal"><b><a href="datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014">Machine Learning based ZZAlpha Ltd. Stock Recommendations 2012-2014</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The data here are the ZZAlpha® machine learning recommendations made for various US traded stock portfolios the morning of each day during the 3 year period Jan 1, 2012 - Dec 31, 2014. &nbsp;</p></td> -->
<td><p class="normal">Sequential, Time-Series </p></td>
<td><p class="normal">Classification </p></td>

```

```

<td><p class="normal">Real </p></td>
<td><p class="normal">314080 </p></td>
<td><p class="normal">0 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Business&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Folio"></a> </td><td><p class="normal"><b><a href="datasets/Folio">Folio</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">20 photos of leaves for each of 32 different species.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">637 </p></td>
<td><p class="normal">20 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015"></a> </td><td><p class="normal"><b><a href="datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015">Taxi Service Trajectory - Prediction Challenge, ECML PKDD 2015</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">An accurate dataset describing trajectories performed by all the 442 taxis running in the city of Porto, in Portugal.
&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Sequential, Time-Series, Domain-Theory </p></td>
<td><p class="normal">Clustering, Causal-Discovery </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">1710671 </p></td>
<td><p class="normal">9 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Cuff-Less+Blood+Pressure+Estimation"></a> </td><td><p class="normal"><b><a href="datasets/Cuff-Less+Blood+Pressure+Estimation">Cuff-Less Blood Pressure Estimation</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">12000 </p></td>
<td><p class="normal">3 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions"></a> </td><td><p class="normal"><b><a href="datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions">Smartphone-Based Recognition of Human Activities and Postural Transitions</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Activity recognition data set built from the recordings of 30 subjects performing basic activities and postural transitions while carrying a waist-mounted smartphone with embedded inertial sensors.
&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">10929 </p></td>
<td><p class="normal">561 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Mice+Protein+Expression"></a> </td><td><p class="normal"><b><a href="datasets/Mice+Protein+Expression">Mice Protein Expression</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Expression levels of 77 proteins measured in the cerebral cortex of 8 classes of control and Down syndrome mice exposed to context fear conditioning, a task used to assess associative learning.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">1080 </p></td>

```

```

<td><p class="normal">82 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Life &nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/UJIIndoorLoc-Mag"></a> </td><td><p class="normal"><b><a href="datasets/UJIIndoorLoc-Mag">UJIIndoorLoc-Mag</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The UJIIndoorLoc-Mag is an indoor localization database to test Indoor Positioning System that rely on Earth's magnetic field variations.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Sequential, Time-Series </p></td>
<td><p class="normal">Classification, Regression, Clustering </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">40000 </p></td>
<td><p class="normal">13 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Heterogeneity+Activity+Recognition"></a> </td><td><p class="normal"><b><a href="datasets/Heterogeneity+Activity+Recognition">Heterogeneity Activity Recognition</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The Heterogeneity Human Activity Recognition (HHAR) dataset from Smartphones and Smartwatches is a dataset devised to benchmark human activity recognition algorithms (classification, automatic data segmentation, sensor fusion, feature extraction, etc.) in real-world contexts; specifically, the dataset is gathered with a variety of different device models and use-scenarios, in order to reflect sensing heterogeneities to be expected in real deployments.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">43930257 </p></td>
<td><p class="normal">16 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Educational+Process+Mining%28EPM%29%3A+A+Learning+Analytics+Data+Set"></a> </td><td><p class="normal"><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></p></td></tr></table></td>
<!-- <td><p class="normal">Educational Process Mining data set is built from the recordings of 115 subjects' activities through a logging application while learning with an educational simulator.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Sequential, Time-Series </p></td>
<td><p class="normal">Classification, Regression, Clustering </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">230318 </p></td>
<td><p class="normal">13 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/HEPMASS"></a> </td><td><p class="normal"><b><a href="datasets/HEPMASS">HEPMASS</a></b></p></td></tr></table></td>
<!-- <td><p 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.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">10500000 </p></td>
<td><p class="normal">28 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Indoor+User+Movement+Prediction+from+RSS+data"></a> </td><td><p class="normal"><b><a href="datasets/Indoor+User+Movement+Prediction+from+RSS+data">Indoor User Movement Prediction from RSS data</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Sequential, Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">13197 </p></td>
<td><p class="normal">4 </p></td>

```

```

<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Open+University+Learning+Analytics+dataset"><img border="1" s
rc="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td><p class="normal"><b><a href="datasets/Open+University+Learning+Analytics+dataset">Open University Learning Analytics data
set</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Open University Learning Analytics Dataset contains data about courses,
students and their interactions with Virtual Learning Environment for seven selected courses and m
ore than 30000 students.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Sequential, Time-Series </p></td>
<td><p class="normal">Classification, Regression, Clustering </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/default+of+credit+card+clients"><img border="1" s
rc="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td><p class="normal"><b><a href="datasets/default+of+credit+card+clients">default of credit card clients</a></b></p></td></tr>
</td>
<!-- <td><p class="normal">This research aimed at the case of customers' default payments in
Taiwan and compares the predictive accuracy of probability of default among six data mining method
s.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">30000 </p></td>
<td><p class="normal">24 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Business&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Mesothelioma%E2%80%99s+disease+data+set+"><img border="1" s
rc="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td><p class="normal"><b><a href="datasets/Mesothelioma%E2%80%99s+disease+data+set+">Mesothelioma's disease data set </a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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 featu
res.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">324 </p></td>
<td><p class="normal">34 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Online+Retail"></a> </td><td><p class="normal"><b><a href="datasets/Online+Retail">Online Retail</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This is a transnational data set which contains all the transactions oc
curring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online
retail.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Sequential, Time-Series </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">541909 </p></td>
<td><p class="normal">8 </p></td>
<td><p class="normal">2015 </p></td>
<!-- <td><p class="normal">Business&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/SIFT10M"></a> </td><td><p class="normal"><b><a href="datasets/SIFT10M">SIFT10M</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">In SIFT10M, each data point is a SIFT feature which is extracted from C
altech-256 by the open source VLFeat library. The corresponding patches of the SIFT features are p
rovided.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Causal-Discovery </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">11164866 </p></td>
<td><p class="normal">128 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr><tr>

```

```

<td><table><tr><td><a href="datasets/GPS+Trajectories"></a> </td><td><p class="normal"><b><a href="datasets/GPS+Trajectories">GPS Trajectories</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The dataset has been feed by Android app called Go!Track. It is available at Goolge Play Store(https://play.google.com/store/apps/details?id=com.go.router). &nbsp ;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">163 </p></td>
<td><p class="normal">15 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Computer&nbs;p;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Detect+Malacious+Executable%28AntiVirus%29"></a> </td><td><p class="normal"><b><a href="datasets/Detect+Malacious+Executable%28AntiVirus%29">Detect Malacious Executable(AntiVirus)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">I extract features from malacious and non-malacious and create and training dataset to teach svm classifier.Dataset made of unknown executable to detect if it is virus or normal safe executable.&nbs;p;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">373 </p></td>
<td><p class="normal">513 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Computer&nbs;p;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Occupancy+Detection+"></a> </td><td><p class="normal"><b><a href="datasets/Occupancy+Detection+">Occupancy Detection </a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Experimental data used for binary classification (room occupancy) from Temperature,Humidity,Light and CO2. Ground-truth occupancy was obtained from time stamped pictures that were taken every minute.&nbs;p;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">20560 </p></td>
<td><p class="normal">7 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Computer&nbs;p;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Parkinson%E2%80%Disease"></a> </td><td><p class="normal"><b><a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Parkinson%E2%80%Disease">Improved Spiral Test Using Digitized Graphics Tablet for Monitoring Parkinson's Disease</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbs;p;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Regression, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">40 </p></td>
<td><p class="normal">7 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Computer&nbs;p;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/News+Aggregator"></a> </td><td><p class="normal"><b><a href="datasets/News+Aggregator">News Aggregator</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">References to news pages collected from an web aggregator in the period from 10-March-2014 to 10-August-2014. The resources are grouped into clusters that represent pages discussing the same story.&nbs;p;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">422937 </p></td>
<td><p class="normal">5 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Other&nbs;p;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Air+Quality"></a> </td><td><p class="normal"><b><a href="datasets/Air+Quality">Air Quality</a></b></p></td></tr></table></td>

```

[Air Quality](datasets/Air+Quality)</p></td></tr></table></td>
 <!-- <td><p class="normal">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. </p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">9358 </p></td>
<td><p class="normal">15 </p></td>
<td><p class="normal">2016 </p></td>
 <!-- <td><p class="normal">Computer </p></td> -->
</tr><tr>
<td><table><tr><td> </td><td><p class="normal">Twin gas sensor arrays</p></td></tr></table></td>
 <!-- <td><p class="normal">5 replicates of an 8-MOX gas sensor array were exposed to different gas conditions (4 volatiles at 10 concentration levels each). </p></td> -->
<td><p class="normal">Multivariate, Time-Series, Domain-Theory </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">640 </p></td>
<td><p class="normal">480000 </p></td>
<td><p class="normal">2016 </p></td>
 <!-- <td><p class="normal">Computer </p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td><p class="normal">Gas sensors for home activity monitoring</p></td></tr></table></td>
 <!-- <td><p class="normal">100 recordings of a sensor array under different conditions in a home setting: background, wine and banana presentations. The array includes 8 MOX gas sensors, and humidity and temperature sensors.
 </p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">919438 </p></td>
<td><p class="normal">11 </p></td>
<td><p class="normal">2016 </p></td>
 <!-- <td><p class="normal">Computer </p></td> -->
</tr><tr>
<td><table><tr><td> </td><td><p class="normal">Facebook Comment Volume Dataset</p></td></tr></table></td>
 <!-- <td><p class="normal">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. </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">40949 </p></td>
<td><p class="normal">54 </p></td>
<td><p class="normal">2016 </p></td>
 <!-- <td><p class="normal">Other </p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td><p class="normal">Smartphone Dataset for Human Activity Recognition (HAR) in Ambient Assisted Living (AAL)</p></td></tr></table></td>
 <!-- <td><p class="normal">This data is an addition to an existing dataset on UCI. We collected more data to improve the accuracy of our human activity recognition algorithms applied in the domain of Ambient Assisted Living. </p></td> -->
<td><p class="normal">Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">5744 </p></td>
<td><p class="normal">561 </p></td>
<td><p class="normal">2016 </p></td>
 <!-- <td><p class="normal">Computer </p></td> -->
</tr><tr>
<td><table><tr><td> </td><td><p class="normal">Polish companies bankruptcy data</p></td></tr></table></td>

```

</tr></table></td>
<!-- <td><p class="normal">The dataset is about bankruptcy prediction of Polish companies. The bankrupt companies were analyzed in the period 2000-2012, while the still operating companies were evaluated from 2007 to 2013.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">10503 </p></td>
<td><p class="normal">64 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Business&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29"></a> </td><td><p class="normal"><b><a href="datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29">Activity Recognition system based on Multisensor data fusion (AReM)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This dataset contains temporal data from a Wireless Sensor Network worn by an actor performing the activities: bending, cycling, lying down, sitting, standing, walking.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Sequential, Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">42240 </p></td>
<td><p class="normal">6 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Dota2+Games+Results"></a> </td><td><p class="normal"><b><a href="datasets/Dota2+Games+Results">Dota2 Games Results</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">102944 </p></td>
<td><p class="normal">116 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Game&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Facebook+metrics"></a> </td><td><p class="normal"><b><a href="datasets/Facebook+metrics">Facebook metrics</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Facebook performance metrics of a renowned cosmetic's brand Facebook page.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">500 </p></td>
<td><p class="normal">19 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Business&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/UbiqLog+%28smartphone+lifelogging%29"></a> </td><td><p class="normal"><b><a href="datasets/UbiqLog+%28smartphone+lifelogging%29">UbiqLog (smartphone lifelogging)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">UbiqLog is the smartphone lifelogging tool that runs on the smartphone of 35 users for about 2 months.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Causal-Discovery </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">9782222 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/NIPS+Conference+Papers+1987-2015"></a> </td><td><p class="normal"><b><a href="datasets/NIPS+Conference+Papers+1987-2015">NIPS Conference Papers 1987-2015</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This data set contains the distribution of words in the full text of the NIPS conference papers published from 1987 to 2015.&nbsp;</p></td> -->
<td><p class="normal">Text </p></td>

```

```

<td><p class="normal">Clustering </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">11463 </p></td>
<td><p class="normal">5812 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/HTRU2"></a> </td><td><p class="normal"><b><a href="datasets/HTRU2">HTRU2</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">17898 </p></td>
<td><p class="normal">9 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Drug+consumption+%28quantified%29"></a> </td><td><p class="normal"><b><a href="datasets/Drug+consumption+%28quantified%29">Drug consumption (quantified)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Classify type of drug consumer by personality data&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">1885 </p></td>
<td><p class="normal">32 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Social&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Appliances+energy+prediction"></a> </td><td><p class="normal"><b><a href="datasets/Appliances+energy+prediction">Appliances energy prediction</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Experimental data used to create regression models of appliances energy use in a low energy building.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">19735 </p></td>
<td><p class="normal">29 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Miskolc+IIS+Hybrid+IPS"></a> </td><td><p class="normal"><b><a href="datasets/Miskolc+IIS+Hybrid+IPS">Miskolc IIS Hybrid IPS</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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. &nbsp;</p></td> -->
<td><p class="normal">Text </p></td>
<td><p class="normal">Classification, Clustering, Causal-Discovery </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">1540 </p></td>
<td><p class="normal">67 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/KDC-4007+dataset+Collection"></a> </td><td><p class="normal"><b><a href="datasets/KDC-4007+dataset+Collection">KDC-4007 dataset Collection</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">KDC-4007 dataset Collection is the Kurdish Documents Classification text used in categories regarding Kurdish Sorani news and articles.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Text </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">4007 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Geo-

```

Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone"> </td><td><p class="normal"><a href="datasets/Geo-

Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone">Geo-

Magnetic field and WLAN dataset for indoor localisation from wristband and smartphone</p></td></tr></table></td>

<!-- <td><p class="normal">A multisource and multivariate dataset for indoor localisation methods based on WLAN and Geo-Magnetic field fingerprinting </p></td> -->

<td><p class="normal">Multivariate, Sequential, Time-Series </p></td>

<td><p class="normal">Classification, Regression, Clustering </p></td>

<td><p class="normal">Integer, Real </p></td>

<td><p class="normal">153540 </p></td>

<td><p class="normal">25 </p></td>

<td><p class="normal">2017 </p></td>

<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">DrivFace</p></td></tr></table></td>

<!-- <td><p class="normal">The DrivFace contains images sequences of subjects while driving in real scenarios. It is composed of 606 samples of 640×480, acquired over different days from 4 drivers with several facial features. </p></td> -->

<td><p class="normal">Multivariate </p></td>

<td><p class="normal">Classification, Regression, Clustering </p></td>

<td><p class="normal">Real </p></td>

<td><p class="normal">606 </p></td>

<td><p class="normal">6400 </p></td>

<td><p class="normal">2016 </p></td>

<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">Website Phishing</p></td></tr></table></td>

<!-- <td><p class="normal">

 </p></td> -->

<td><p class="normal">Multivariate </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Integer </p></td>

<td><p class="normal">1353 </p></td>

<td><p class="normal">10 </p></td>

<td><p class="normal">2016 </p></td>

<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">YouTube Spam Collection</p></td></tr></table></td>

<!-- <td><p class="normal">It is a public set of comments collected for spam research. It has five datasets composed by 1,956 real messages extracted from five videos that were among the 10 most viewed on the collection period. </p></td> -->

<td><p class="normal">Text </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal"> </p></td>

<td><p class="normal">1956 </p></td>

<td><p class="normal">5 </p></td>

<td><p class="normal">2017 </p></td>

<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">Beijing PM2.5 Data</p></td></tr></table></td>

<!-- <td><p class="normal">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.

 </p></td> -->

<td><p class="normal">Multivariate, Time-Series </p></td>

<td><p class="normal">Regression </p></td>

<td><p class="normal">Integer, Real </p></td>

<td><p class="normal">43824 </p></td>

<td><p class="normal">13 </p></td>

<td><p class="normal">2017 </p></td>

<!-- <td><p class="normal">Physical </p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">Cargo 2000 Freight Tracking and Tracing</p></td></tr></table></td>

<!-- <td><p class="normal">Sanitized and anonymized Cargo 2000 (C2K) airfreight tracking and

tracing events, covering five months of business execution (3,942 process instances, 7,932 transport legs, 56,082 activities). -->

<td><p class="normal">Multivariate, Sequential </p></td>

<td><p class="normal">Classification, Regression </p></td>

<td><p class="normal">Integer </p></td>

<td><p class="normal">3942 </p></td>

<td><p class="normal">98 </p></td>

<td><p class="normal">2016 </p></td>

<!-- <td><p class="normal">Business </p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">Cervical cancer (Risk Factors)</p></td>

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

<!-- <td><p class="normal">This dataset focuses on the prediction of indicators/diagnosis of cervical cancer. The features cover demographic information, habits, and historic medical records. </p></td> -->

<td><p class="normal">Multivariate </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Integer, Real </p></td>

<td><p class="normal">858 </p></td>

<td><p class="normal">36 </p></td>

<td><p class="normal">2017 </p></td>

<!-- <td><p class="normal">Life </p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">Quality Assessment of Digital Colposcopies</p></td></tr></table></td>

<!-- <td><p class="normal">This dataset explores the subjective quality assessment of digital colposcopies. </p></td> -->

<td><p class="normal">Multivariate </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Real </p></td>

<td><p class="normal">287 </p></td>

<td><p class="normal">69 </p></td>

<td><p class="normal">2017 </p></td>

<!-- <td><p class="normal">Life </p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">KASANDR</p></td></tr></table></td>

<!-- <td><p class="normal">KASANDR is a novel, publicly available collection for recommendation systems that records the behavior of customers of the European leader in e-Commerce advertising, Kelkoo. </p></td> -->

<td><p class="normal">Multivariate </p></td>

<td><p class="normal">Causal-Discovery </p></td>

<td><p class="normal">Integer </p></td>

<td><p class="normal">17764280 </p></td>

<td><p class="normal">2158859 </p></td>

<td><p class="normal">2017 </p></td>

<!-- <td><p class="normal">Life </p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">FMA: A Dataset For Music Analysis</p></td></tr></table></td>

<!-- <td><p class="normal">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. </p></td> -->

<td><p class="normal">Multivariate, Time-Series </p></td>

<td><p class="normal">Classification, Clustering </p></td>

<td><p class="normal">Real </p></td>

<td><p class="normal">106574 </p></td>

<td><p class="normal">518 </p></td>

<td><p class="normal">2017 </p></td>

<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">Air quality</p></td></tr></table></td>

<!-- <td><p class="normal"> Contains the responses of a gas multisensor device deployed on the field in an Italian city. </p></td> -->

<td><p class="normal">Multivariate, Time-Series </p></td>

<td><p class="normal">Regression </p></td>

<td><p class="normal">Real </p></td>

<td><p class="normal">9358 </p></td>

```

<td><p class="normal">15 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Other &nbsp; ;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Epileptic+Seizure+Recognition"></a> </td><td><p class="normal"><b><a href="datasets/Epileptic+Seizure+Recognition">Epileptic Seizure Recognition</a></b></p></td></tr></td>
<!-- <td><p class="normal">This dataset is a pre-processed and re-structured/reshaped version of a very commonly used dataset featuring epileptic seizure detection. &nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">11500 </p></td>
<td><p class="normal">179 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Life&nbsp; ;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Devanagari+Handwritten+Character+Dataset"></a> </td><td><p class="normal"><b><a href="datasets/Devanagari+Handwritten+Character+Dataset">Devanagari Handwritten Character Dataset</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This is an image database of Handwritten Devanagari characters. There are 46 classes of characters with 2000 examples each. The dataset is split into training set(85%) and testing set(15%). &nbsp;</p></td> -->
<td><p class="normal"> </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">92000 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Computer&nbsp; ;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Stock+portfolio+performance"></a> </td><td><p class="normal"><b><a href="datasets/Stock+portfolio+performance">Stock portfolio performance</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The data set of performances of weighted scoring stock portfolios are obtained with mixture design from the US stock market historical database.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">315 </p></td>
<td><p class="normal">12 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Business&nbsp; ;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/MoCap+Hand+Postures"></a> </td><td><p class="normal"><b><a href="datasets/MoCap+Hand+Postures">MoCap Hand Postures</a></b></p></td></tr></table></td>
<!-- <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.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">78095 </p></td>
<td><p class="normal">38 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Computer&nbsp; ;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech"></a> </td><td><p class="normal"><b><a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech">Early biomarkers of Parkinson's disease based on natural connected speech</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Predict a pattern of neurodegeneration in the dataset of speech features obtained from patients with early untreated Parkinson's disease.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">130 </p></td>
<td><p class="normal">65 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Life&nbsp; ;</p></td> -->

```

```

.. . . . . <tr><td>  </td><td>
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting"></a> </td><td><p class="normal"><b><a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting">Data for Software Engineering Teamwork Assessment in Education Setting</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Data include over 100 Team Activity Measures and outcomes (ML classes) obtained from activities of 74 student teams during the creation of final class project in SW Eng. classes at SFSU, Fulda, FAU&nbsp;</p></td> -->
<td><p class="normal">Sequential, Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">74 </p></td>
<td><p class="normal">102 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/PM2.5+Data+of+Five+Chinese+Cities"></a> </td><td><p class="normal"><b><a href="datasets/PM2.5+Data+of+Five+Chinese+Cities">PM2.5 Data of Five Chinese Cities</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This hourly data set contains the PM2.5 data in Beijing, Shanghai, Guan gzhou, Chengdu and Shenyang. Meanwhile, meteorological data for each city are also included.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">52854 </p></td>
<td><p class="normal">86 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet"></a> </td><td><p class="normal"><b><a href="datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet">Parkinson Disease Spiral Drawings Using Digitized Graphics Tablet</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Handwriting database consists of 62 PWP(People with Parkinson) and 15 h ealthy individuals. Three types of recordings (Static Spiral Test, Dynamic Spiral Test and Stability Test) are taken.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Regression, Clustering </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">77 </p></td>
<td><p class="normal">7 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Sales_Transactions_Dataset_Weekly"></a> </td><td><p class="normal"><b><a href="datasets/Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Contains weekly purchased quantities of 800 over products over 52 weeks. Normalised values are provided too.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Clustering </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">811 </p></td>
<td><p class="normal">53 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Las+Vegas+Strip"></a> </td><td><p class="normal"><b><a href="datasets/Las+Vegas+Strip">Las Vegas Strip</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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).&nbsp;</p></td> -->
<td><p class="normal"> </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">504 </p></td>
<td><p class="normal">20 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Business&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Eco-hotel"><img border="1"

```

```

<td><a href="datasets/Eco-hotel"></a></td><td><p class="normal"><b><a href="datasets/Eco-hotel">Eco-hotel</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This dataset includes Online Textual Reviews from both online (e.g., TripAdvisor) and offline (e.g., Guests' book) sources from the Areias do Seixo Eco-Resort.&nbsp;</p>
</td> -->
<td><p class="normal">Text </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">401 </p></td>
<td><p class="normal">1 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Business&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/MEU-Mobile+KSD"></a></td><td><p class="normal"><b><a href="datasets/MEU-Mobile+KSD">MEU-Mobile KSD</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This dataset contains keystroke dynamics data collected on a touch mobile device (Nexus 7). The dataset contains 2856 records, 51 records per subject for 56 subjects. &nbs;p;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">2856 </p></td>
<td><p class="normal">71 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Crowdsourced+Mapping"></a></td><td><p class="normal"><b><a href="datasets/Crowdsourced+Mapping">Crowdsourced Mapping</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Crowdsourced data from OpenStreetMap is used to automate the classification of satellite images into different land cover classes (impervious, farm, forest, grass, orchard, water). &nbs;p;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">10546 </p></td>
<td><p class="normal">29 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/gene+expression+cancer+RNA-Seq"></a></td><td><p class="normal"><b><a href="datasets/gene+expression+cancer+RNA-Seq">gene expression cancer RNA-Seq</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This collection of data is part of the RNA-Seq (HiSeq) PANCAN data set, it is a random extraction of gene expressions of patients having different types of tumor: BRCA, KIRC, COAD, LUAD and PRAD.&nbs;p;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">801 </p></td>
<td><p class="normal">20531 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer"></a></td><td><p class="normal"><b><a href="datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer">Hybrid Indoor Positioning Dataset from WiFi RSSI, Bluetooth and magnetometer</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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;</p></td> -->
<td><p class="normal">Multivariate, Sequential, Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">1540 </p></td>
<td><p class="normal">65 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/chestnut+%E2%80%93+LARVIC"></a></td><td><p class="normal"><b><a href="datasets/chestnut+%E2%80%93+LARVIC">chestnut - LARVIC</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The research project presents this database shows the images of

```

<!-- This research project presents this database, shows the images of chestnuts that will be processed to determine the presence or absence of defects-->

<p class="normal"></p></td></p>	<p class="normal">Classification, Clustering </p></td></p>
<p class="normal"></p></td></p>	<p class="normal">1451 </p></td></p>
<p class="normal"></p></td></p>	<p class="normal">3 </p></td></p>
<p class="normal"></p></td></p>	<p class="normal">2017 </p></td></p>
<p class="normal"></td><tr></p>	<p class="normal"><p class="normal">Computer&nbsp;</p></td> --></p>
<p class="normal"><td><table><tr><td> </td><td><p class="normal">b>Burst Header Packet (BHP) flooding attack on Optical Burst Switching (OBS) Network</p></td></tr></table></td></p>	<p class="normal"><!-- <td><p class="normal">One of the primary challenges in identifying the risks of the Burst Header Packet (BHP) flood attacks in Optical Burst Switching networks (OBS) is the scarcity of reliable historical data. &nbsp;</p></td> --></p>
<p class="normal"><td><p class="normal">Text </p></td></p>	<p class="normal"><td><p class="normal">Classification </p></td></p>
<p class="normal"><td><p class="normal">Integer </p></td></p>	<p class="normal"><td><p class="normal">1075 </p></td></p>
<p class="normal"><td><p class="normal">22 </p></td></p>	<p class="normal"><td><p class="normal">2017 </p></td></p>
<p class="normal"><!-- <td><p class="normal">Computer&nbsp;</p></td> --></p>	<p class="normal"></td><tr bgcolor="#DDEEFF"></p>
<p class="normal"><td><table><tr><td> </td><td><p class="normal">b>Motion Capture Hand Postures</p></td></tr></table></td></p>	<p class="normal"><!-- <td><p class="normal">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.&nbsp;</p></td> --></p>
<p class="normal"><td><p class="normal">Multivariate </p></td></p>	<p class="normal"><td><p class="normal">Classification, Clustering </p></td></p>
<p class="normal"><td><p class="normal">Real </p></td></p>	<p class="normal"><td><p class="normal">78095 </p></td></p>
<p class="normal"><td><p class="normal">38 </p></td></p>	<p class="normal"><td><p class="normal">38 </p></td></p>
<p class="normal"><td><p class="normal">2017 </p></td></p>	<p class="normal"><!-- <td><p class="normal">Computer&nbsp;</p></td> --></p>
<p class="normal"></td><tr></p>	<p class="normal"><td><table><tr><td> </td><td><p class="normal">b>Anuran Calls (MFCCs)</p></td></tr></table></td></p>
<p class="normal"><!-- <td><p class="normal">Acoustic features extracted from syllables of anuran (frogs) calls, including the family, the genus, and the species labels (multilabel). &nbsp;</p></td> --></p>	<p class="normal"><td><p class="normal">Multivariate </p></td></p>
<p class="normal"><td><p class="normal">Classification, Clustering </p></td></p>	<p class="normal"><td><p class="normal">Real </p></td></p>
<p class="normal"><td><p class="normal">7195 </p></td></p>	<p class="normal"><td><p class="normal">22 </p></td></p>
<p class="normal"><td><p class="normal">2017 </p></td></p>	<p class="normal"><td><p class="normal">2017 </p></td></p>
<p class="normal"><!-- <td><p class="normal">Life&nbsp;</p></td> --></p>	<p class="normal"><!-- <td><p class="normal">Computer&nbsp;</p></td> --></p>
<p class="normal"></td><tr bgcolor="#DDEEFF"></p>	<p class="normal"></td><tr><td> </td><td><p class="normal">b>TTC-3600: Benchmark dataset for Turkish text categorization</p></td></tr></table></td></p>
<p class="normal"><!-- <td><p class="normal">The TTC-3600 data set is a collection of Turkish news and articles including categorized 3,600 documents from 6 well-known portals in Turkey. It has 4 different forms in ARFF Weka format.&nbsp;</p></td> --></p>	<p class="normal"><td><p class="normal">Text </p></td></p>
<p class="normal"><td><p class="normal">Classification, Clustering </p></td></p>	<p class="normal"><td><p class="normal">Integer </p></td></p>
<p class="normal"><td><p class="normal">3600 </p></td></p>	<p class="normal"><td><p class="normal">4814 </p></td></p>
<p class="normal"><td><p class="normal">2017 </p></td></p>	<p class="normal"><td><p class="normal">2017 </p></td></p>
<p class="normal"><!-- <td><p class="normal">Computer&nbsp;</p></td> --></p>	<p class="normal"></td><tr></p>
<p class="normal"><td><table><tr><td> </td><td><p class="normal">b>Gastrointestinal Lesions in Regular Colonoscopy</p></td></tr></table></td></p>	<p class="normal"><!-- <td><p class="normal">This dataset contains features extracted from colonoscopy videos used to --></p>

```

<!-- <tr><td> Normal / This dataset contains features extracted from colonoscopy videos used to detect gastrointestinal lesions. It contains 76 lesions: 15 serrated adenomas, 21 hyperplastic lesions and 40 adenoma. &nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">76 </p></td>
<td><p class="normal">698 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Daily+Demand+Forecasting+Orders"></a> </td><td><p class="normal"><b><a href="datasets/Daily+Demand+Forecasting+Orders">Daily Demand Forecasting Orders</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The dataset was collected during 60 days, this is a real database of a brazilian logistics company.&nbsp;</p></td> -->
<td><p class="normal">Time-Series </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">60 </p></td>
<td><p class="normal">13 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Business&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Paper+Reviews"></a> </td><td><p class="normal"><b><a href="datasets/Paper+Reviews">Paper Reviews</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This sentiment analysis data set contains scientific paper reviews from an international conference on computing and informatics. The task is to predict the orientation or the evaluation of a review.&nbsp;</p></td> -->
<td><p class="normal">Text </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">405 </p></td>
<td><p class="normal">10 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/extention+of+Z-Alizadeh+sani+dataset"></a> </td><td><p class="normal"><b><a href="datasets/extention+of+Z-Alizadeh+sani+dataset">extention of Z-Alizadeh sani dataset</a></b><a href="datasets/extention+of+Z-Alizadeh+sani+dataset">extention of Z-Alizadeh sani dataset</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">It was collected for CAD diagnosis.&nbsp;</p></td> -->
<td><p class="normal"> </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">303 </p></td>
<td><p class="normal">59 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Z-Alizadeh+Sani"></a> </td><td><p class="normal"><b><a href="datasets/Z-Alizadeh+Sani">Z-Alizadeh Sani</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">It was collected for CAD diagnosis.&nbsp;</p></td> -->
<td><p class="normal"> </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">303 </p></td>
<td><p class="normal">56 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Dynamic+Features+of+VirusShare+Executables"></a> </td><td><p class="normal"><b><a href="datasets/Dynamic+Features+of+VirusShare+Executables">Dynamic Features of VirusShare Executables</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This dataset contains the dynamic features of 107,888 executables, collected by VirusShare from Nov/2010 to Jul/2014.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">107888 </p></td>
<td><p class="normal">482 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>

```

```

</tr><tr>
<td><table><tr><td><a href="datasets/IDA2016Challenge"></a> </td><td><p class="normal"><b><a href="datasets/IDA2016Challenge">IDA2016Challenge</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The dataset consists of data collected from heavy Scania trucks in everyday usage. &nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">76000 </p></td>
<td><p class="normal">171 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/DSRC+Vehicle+Communications"></a> </td><td><p class="normal"><b><a href="datasets/DSRC+Vehicle+Communications">DSRC Vehicle Communications</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This set Provides data regarding wireless communications between vehicles and road side units. two separate data sets are provided (normal scenario) and in the presence of attacker (jammer).&nbsp;</p></td> -->
<td><p class="normal">Sequential, Text </p></td>
<td><p class="normal">Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">10000 </p></td>
<td><p class="normal">5 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Mturk+User-Perceived+Clusters+over+Images"></a> </td><td><p class="normal"><b><a href="datasets/Mturk+User-Perceived+Clusters+over+Images">Mturk User-Perceived Clusters over Images</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Text </p></td>
<td><p class="normal">Clustering </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">180 </p></td>
<td><p class="normal">500 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Character+Font+Images"></a> </td><td><p class="normal"><b><a href="datasets/Character+Font+Images">Character Font Images</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Character images from scanned and computer generated fonts.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">745000 </p></td>
<td><p class="normal">411 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Instance+Labels"></a> </td><td><p class="normal"><b><a href="datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Instance+Labels">DeliciousMIL: A Data Set for Multi-Label Multi-Instance Learning with Instance Labels</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This dataset includes 1) 12234 documents (8251 training, 3983 test) extracted from DeliciousT140 dataset, 2) class labels for all documents, 3) labels for a subset of sentences of the test documents.&nbsp;</p></td> -->
<td><p class="normal">Text </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">12234 </p></td>
<td><p class="normal">8519 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++"></a> </td><td><p class="normal"><b><a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++">Autistic Spectrum Disorder Screening Data for Children ++</a></b></p></td></tr></table></td>

```

Screening data for Children </p></td></tr></table></td>

<!-- <td><p class="normal">Children screening data for autism suitable for classification and predictive tasks </p></td> -->

<td><p class="normal">Multivariate </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Integer </p></td>

<td><p class="normal">292 </p></td>

<td><p class="normal">21 </p></td>

<td><p class="normal">2017 </p></td>

<!-- <td><p class="normal">Life </p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">Autistic Spectrum Disorder Screening Data for Adolescent </p></td></tr></table></td>

<!-- <td><p class="normal">Autistic Spectrum Disorder Screening Data for Adolescent. This dataset is related to classification and predictive tasks. </p></td> -->

<td><p class="normal">Multivariate </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Integer </p></td>

<td><p class="normal">104 </p></td>

<td><p class="normal">21 </p></td>

<td><p class="normal">2017 </p></td>

<!-- <td><p class="normal">Life </p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">APS Failure at Scania Trucks</p></td></tr></table></td>

<!-- <td><p class="normal">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. </p></td> -->

<td><p class="normal">Multivariate </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Integer, Real </p></td>

<td><p class="normal">60000 </p></td>

<td><p class="normal">171 </p></td>

<td><p class="normal">2017 </p></td>

<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">Wireless Indoor Localization</p></td></tr></table></td>

<!-- <td><p class="normal">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. </p></td> -->

<td><p class="normal">Multivariate </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Real </p></td>

<td><p class="normal">2000 </p></td>

<td><p class="normal">7 </p></td>

<td><p class="normal">2017 </p></td>

<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">HCC Survival</p></td></tr></table></td>

<!-- <td><p class="normal">Hepatocellular Carcinoma dataset (HCC dataset) was collected at a University Hospital in Portugal. It contains real clinical data of 165 patients diagnosed with HCC.&nbs p;</p></td> -->

<td><p class="normal">Multivariate </p></td>

<td><p class="normal">Classification </p></td>

<td><p class="normal">Integer, Real </p></td>

<td><p class="normal">165 </p></td>

<td><p class="normal">49 </p></td>

<td><p class="normal">2017 </p></td>

<!-- <td><p class="normal">Life </p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">CSM (Conventional and Social Media Movies) Dataset 2014 and 2015</p></td></tr></table></td>

<!-- <td><p class="normal">12 features categorized as conventional and social media features. Both conventional features, collected from movies databases on Web as well as social media features (YouTube, Twitter). </p></td> -->

```

<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">217 </p></td>
<td><p class="normal">12 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29"></a> </td><td><p class="normal"><b><a href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29">University of Tehran Question Dataset 2016 (UTQD.2016)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Persian questions gathered from a jeopardy game broadcasted on Iranian national television. &nbsp;</p></td> -->
<td><p class="normal">Text </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">1175 </p></td>
<td><p class="normal">3 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Autism+Screening+Adult"></a> </td><td><p class="normal"><b><a href="datasets/Autism+Screening+Adult">Autism Screening Adult</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Autistic Spectrum Disorder Screening Data for Adult. This dataset is related to classification and predictive tasks.&nbsp;</p></td> -->
<td><p class="normal"> </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">704 </p></td>
<td><p class="normal">21 </p></td>
<td><p class="normal">2017 </p></td>
<!-- <td><p class="normal">Social&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor"></a> </td><td><p class="normal"><b><a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor">vity recognition with healthy older people using a batteryless wearable sensor</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Sequential motion data from 14 healthy older people aged 66 to 86 years old using a batteryless, wearable sensor on top of their clothing for the recognition of activities in clinical environments.&nbsp;</p></td> -->
<td><p class="normal">Sequential </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">75128 </p></td>
<td><p class="normal">9 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Immunotherapy+Dataset"></a> </td><td><p class="normal"><b><a href="datasets/Immunotherapy+Dataset">Immunotherapy Dataset</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This dataset contains information about wart treatment results of 90 patients using immunotherapy.&nbsp;</p></td> -->
<td><p class="normal">Univariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">90 </p></td>
<td><p class="normal">8 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Cryotherapy+Dataset+"></a> </td><td><p class="normal"><b><a href="datasets/Cryotherapy+Dataset+">Cryotherapy Dataset </a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This dataset contains information about wart treatment results of 90 patients using cryotherapy.&nbsp;</p></td> -->
<td><p class="normal">Univariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">90 </p></td>
<td><p class="normal">7 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->

```

```

</tr><tr>
<td><table><tr><td><a href="datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes"></a> </td><td><p class="normal"><b><a href="datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes">OCT data &amp; Color Fundus Images of Left &amp; Right Eyes</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This dataset contains OCT data (in mat format) and color fundus data (in jpg format) of left & right eyes of 50 healthy persons.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">50 </p></td>
<td><p class="normal">2 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Discrete+Tone+Image+Dataset"></a> </td><td><p class="normal"><b><a href="datasets/Discrete+Tone+Image+Dataset">Discrete Tone Image Dataset</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">71 </p></td>
<td><p class="normal">11 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/News+Popularity+in+Multiple+Social+Media+Platforms"></a> </td><td><p class="normal"><b><a href="datasets/News+Popularity+in+Multiple+Social+Media+Platforms">News Popularity in Multiple Social Media Platforms</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Large data set of news items and their respective social feedback on multiple platforms: Facebook, Google+ and LinkedIn.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series, Text </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">93239 </p></td>
<td><p class="normal">11 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Ultrasonic+flowmeter+diagnostics"></a> </td><td><p class="normal"><b><a href="datasets/Ultrasonic+flowmeter+diagnostics">Ultrasonic flowmeter diagnostics</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Fault diagnosis of four liquid ultrasonic flowmeters&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">540 </p></td>
<td><p class="normal">173 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/ICMLA+2014+Accepted+Papers+Data+Set"></a> </td><td><p class="normal"><b><a href="datasets/ICMLA+2014+Accepted+Papers+Data+Set">ICMLA 2014 Accepted Papers Data Set</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This data set compromises the metadata for the 2014 ICMLA conference's accepted papers, including ID, paper titles, author's keywords, abstracts and sessions in which they were exposed.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">105 </p></td>
<td><p class="normal">5 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation"></a> </td><td><p class="normal"><b><a href="datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation">BLE RSSI Dataset for Indoor localization and Navigation</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">This dataset contains RSSI readings gathered from an array of Bluetooth

```

Low Energy (BLE) iBeacons in a real-world and operational indoor environment for localization and navigation purposes. </td> -->

<td><p class="normal">Multivariate, Sequential, Time-Series </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">6611 </p></td>
<td><p class="normal">15 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">Container Crane Controller Data Set</p></td></tr></table></td>
<!-- <td><p class="normal">A container crane has the function of transporting containers from one point to another point. </p></td> -->

<td><p class="normal">Univariate, Domain-Theory </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">15 </p></td>
<td><p class="normal">3 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">Residential Building Data Set</p></td></tr></table></td>
<!-- <td><p class="normal">Data set includes construction cost, sale prices, project variables, and economic variables corresponding to real estate single-family residential apartments in Tehran, Iran. </p></td> -->

<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">372 </p></td>
<td><p class="normal">105 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">Health News in Twitter</p></td></tr></table></td>
<!-- <td><p class="normal">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></td> -->

<td><p class="normal">Text </p></td>
<td><p class="normal">Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">58000 </p></td>
<td><p class="normal">25000 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Computer </p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">chipseq</p></td></tr></table></td>
<!-- <td><p class="normal">ChIP-seq experiments characterize protein modifications or binding at specific genomic locations in specific samples. The machine learning problem in these data is structured binary classification. </p></td> -->

<td><p class="normal">Sequential </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">4960 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Life </p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">SGEMM GPU kernel performance</p></td></tr></table></td>
<!-- <td><p class="normal">Running times for multiplying two 2048 x 2048 matrices using a GPU OpenCL SGEMM kernel with varying parameters (using the library 'CLTune'). </p></td> -->

<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">241600 </p></td>

<p class="normal">18</p>	<p class="normal">2018</p>	<p class="normal"><!-- <td><p class="normal">Computer </p></td> --></p>
<p class="normal"></tr><tr bgcolor="#DDEEFF"></p>	<p class="normal"><td><table><tr><td> </td><td><p class="normal">Repeat Consumption Matrices</p></td></tr></table></p>	<p class="normal"><!-- <td><p class="normal">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.&nbsp;</p></td> --></p>
<p class="normal"><td><p class="normal">Multivariate </p></td></p>	<p class="normal"><td><p class="normal">Clustering </p></td></p>	<p class="normal"><td><p class="normal">Real </p></td></p>
<p class="normal"><td><p class="normal">130000 </p></td></p>	<p class="normal"><td><p class="normal">21000 </p></td></p>	<p class="normal"><td><p class="normal">2018 </p></td></p>
<p class="normal"><!-- <td><p class="normal">Computer </p></td> --></p>	<p class="normal"></tr><tr></p>	<p class="normal"><td><table><tr><td> </td><td><p class="normal">detection_of_IoT_botnet_attacks_N_BaIoT</p></td></tr></table></p>
<p class="normal"><!-- <td><p class="normal">This dataset addresses the lack of public botnet datasets, especially f or the IoT. It suggests *real* traffic data, gathered from 9 commercial IoT devices authentically infected by Mirai and BASHLITE.&nbsp;</p></td> --></p>	<p class="normal"><td><p class="normal">Multivariate, Sequential </p></td></p>	<p class="normal"><td><p class="normal">Classification, Clustering </p></td></p>
<p class="normal"><td><p class="normal">Real </p></td></p>	<p class="normal"><td><p class="normal">7062606 </p></td></p>	<p class="normal"><td><p class="normal">115 </p></td></p>
<p class="normal"><td><p class="normal">2018 </p></td></p>	<p class="normal"><!-- <td><p class="normal">Computer </p></td> --></p>	<p class="normal"></tr><tr bgcolor="#DDEEFF"></p>
<p class="normal"><td><table><tr><td> </td><td><p class="normal">Absenteeism at work</p></td></tr></table></p>	<p class="normal"><!-- <td><p class="normal">The database was created with records of absenteeism at work from July 2007 to July 2010 at a courier company in Brazil.&nbsp;</p></td> --></p>	<p class="normal"><td><p class="normal">Multivariate, Time-Series </p></td></p>
<p class="normal"><td><p class="normal">Classification, Clustering </p></td></p>	<p class="normal"><td><p class="normal">Integer, Real </p></td></p>	<p class="normal"><td><p class="normal">740 </p></td></p>
<p class="normal"><td><p class="normal">21 </p></td></p>	<p class="normal"><td><p class="normal">2018 </p></td></p>	<p class="normal"><!-- <td><p class="normal">Business&nbsp;</p></td> --></p>
<p class="normal"></tr><tr></p>	<p class="normal"><td><table><tr><td> </td><td><p class="normal">SCADI</p></td></tr></table></p>	<p class="normal"><!-- <td><p class="normal">First self-care activities dataset based on ICF-CY.&nbsp;</p></td> --></p>
<p class="normal"><td><p class="normal">Multivariate </p></td></p>	<p class="normal"><td><p class="normal">Classification, Clustering </p></td></p>	<p class="normal"><td><p class="normal"></p></td></p>
<p class="normal"><td><p class="normal">70 </p></td></p>	<p class="normal"><td><p class="normal">206 </p></td></p>	<p class="normal"><td><p class="normal">2018 </p></td></p>
<p class="normal"><td><p class="normal">2018 </p></td></p>	<p class="normal"><!-- <td><p class="normal">Life&nbsp;</p></td> --></p>	<p class="normal"></tr><tr bgcolor="#DDEEFF"></p>
<p class="normal"><td><table><tr><td> </td><td><p class="normal">Condition monitoring of hydraulic systems</p></td></tr></table></p>	<p class="normal"><!-- <td><p class="normal">The data set addresses the condition assessment of a hydraulic test rig based on multi sensor data. Four fault types are superimposed with several severity grades impeding selective quantification.&nbsp;</p></td> --></p>	<p class="normal"><td><p class="normal">Multivariate, Time-Series </p></td></p>
<p class="normal"><td><p class="normal">Classification, Regression </p></td></p>	<p class="normal"><td><p class="normal">Real </p></td></p>	<p class="normal"><td><p class="normal">2205 </p></td></p>
<p class="normal"><td><p class="normal">43680 </p></td></p>	<p class="normal"><td><p class="normal">2018 </p></td></p>	<p class="normal"><td><p class="normal">2018 </p></td></p>
<p class="normal"><td><p class="normal">Computer&nbsp;</p></td> --></p>	<p class="normal"></tr><tr></p>	<p class="normal"><td><table><tr><td> </td><td><p class="normal">Carbon Nanotubes</p></td></tr></table></p>

```

<!-- <td><p class="normal">This dataset contains 10721 initial and calculated atomic coordinates o
f carbon nanotubes.&nbsp;</p></td> -->
<td><p class="normal">Univariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">10721 </p></td>
<td><p class="normal">8 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Optical+Interconnection+Network+"></a> </td><td><p class="normal"><b><a
href="datasets/Optical+Interconnection+Network+">Optical Interconnection Network </a></b></p></td>
</tr></table></td>
<!-- <td><p class="normal">This dataset contains 640 performance measurements from a simulation of
2-Dimensional Multiprocessor Optical Interconnection Network. &nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">640 </p></td>
<td><p class="normal">10 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Sports+articles+for+objectivity+analysis"><img border="1" src
="assets/MLimages/SmallLargedefault.jpg"/></a> </td><td><p class="normal"><b><a
href="datasets/Sports+articles+for+objectivity+analysis">Sports articles for objectivity
analysis </a></b></p></td></tr></table></td>
<!-- <td><p class="normal">1000 sports articles were labeled using Amazon Mechanical Turk as
objective or subjective. The raw texts, extracted features, and the URLs from which the articles w
ere retrieved are provided.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Text </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">1000 </p></td>
<td><p class="normal">59 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Social&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Breast+Cancer+Coimbra"></a> </td><td><p class="normal"><b><a
href="datasets/Breast+Cancer+Coimbra">Breast Cancer Coimbra </a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Clinical features were observed or measured for 64 patients with breast
cancer and 52 healthy controls. &nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">116 </p></td>
<td><p class="normal">10 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data"></a> </td><td><p class="normal"><b><a
href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data">GNFUV Unmanned Surface Vehicles Sensor
Data </a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The data-set contains four (4) sets of mobile sensor readings data (hum
idity, temperature) corresponding to a swarm of four (4) Unmanned Surface Vehicles (USVs) in a tes
t-bed in Athens (Greece). &nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">1672 </p></td>
<td><p class="normal">5 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Dishonest+Internet+users+Dataset"></a> </td><td><p class="normal"><b><a
href="datasets/Dishonest+Internet+users+Dataset">Dishonest Internet users Dataset </a></b></p></td>
</tr></table></td>
<!-- <td><p class="normal">The dataset was used to test an architecture based on a trust model
capable to cope with the evaluation of the trustworthiness of users interacting in pervasive
environments.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal"> </p></td>

```

<td><p class="normal">322 </p></td>
<td><p class="normal">5 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr><tr>
<td><table><tr><td> </td><td><p class="normal">Victorian Era Authorship Attribution </td></tr></table></td>
<!-- <td><p class="normal">To create the largest authorship attribution dataset, we extracted works of 50 well-known authors. To have a non-exhaustive learning, in training there are 45 authors whereas, in the testing, it's 50 </p></td> -->
<td><p class="normal">Text </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">93600 </p></td>
<td><p class="normal">1000 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td> </td><td><p class="normal">Simulated Falls and Daily Living Activities Data Set</p></td></tr></table></td>
<!-- <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. </p></td> -->
<td><p class="normal">Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">3060 </p></td>
<td><p class="normal">138 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Life </p></td> -->
</tr><tr>
<td><table><tr><td> </td><td><p class="normal">Multimodal Damage Identification for Humanitarian Computing</p></td></tr></table></td>
<!-- <td><p class="normal">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, Infrastructure, Human, Non-damage. </p></td> -->
<td><p class="normal">Multivariate, Text </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">5879 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Social </p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td> </td><td><p class="normal">EEG Steady-State Visual Evoked Potential Signals</p></td></tr></table></td>
<!-- <td><p class="normal">This database consists on 30 subjects performing Brain Computer Interface for Steady State Visual Evoked Potentials (BCI-SSVEP). </p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">9200 </p></td>
<td><p class="normal">16 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Life </p></td> -->
</tr><tr>
<td><table><tr><td> </td><td><p class="normal">Roman Urdu Data Set</p></td></tr></table></td>
<!-- <td><p class="normal">Roman Urdu (the scripting style for Urdu language) is one of the limited resource languages.A data corpus comprising of more than 20000 records was collected. </p></td> -->
<td><p class="normal">Text </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">20000 </p></td>
<td><p class="normal">2 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Computer </p></td> -->

```

</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Avila"></a> </td><td><p class="normal"><b><a href="datasets/Avila">Avila</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The Avila data set has been extracted from 800 images of the 'Avila Bib
le', an XII century giant Latin copy of the Bible. The prediction task consists in associating ea
ch pattern to a copyist.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">20867 </p></td>
<td><p class="normal">10 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/PANDOR"></a> </td><td><p class="normal"><b><a href="datasets/PANDOR">PANDOR</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">PANDOR is a novel and publicly available dataset for online
recommendation provided by Purch (http://www.purch.com/). &nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Recommendation </p></td>
<td><p class="normal">Categorical </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Drug+Review+Dataset+%28Druglib.com%29"></a> </td><td><p class="normal"><b><a href="datasets/Drug+Review+Dataset+%28Druglib.com%29">Drug Review Dataset (Druglib.com)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The dataset provides patient reviews on specific drugs along with
related conditions. Reviews and ratings are grouped into reports on the three aspects benefits, si
de effects and overall comment.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Text </p></td>
<td><p class="normal">Classification, Regression, Clustering </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">4143 </p></td>
<td><p class="normal">8 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Drug+Review+Dataset+%28Drugs.com%29"></a> </td><td><p class="normal"><b><a href="datasets/Drug+Review+Dataset+%28Drugs.com%29">Drug Review Dataset (Drugs.com)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The dataset provides patient reviews on specific drugs along with
related conditions and a 10 star patient rating reflecting overall patient satisfaction.&nbsp;</p>
</td> -->
<td><p class="normal">Multivariate, Text </p></td>
<td><p class="normal">Classification, Regression, Clustering </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">215063 </p></td>
<td><p class="normal">6 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Physical+Unclonable+Functions"></a> </td><td><p class="normal"><b><a href="datasets/Physical+Unclonable+Functions">Physical Unclonable Functions</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The dataset is generated from Physical Unclonable Functions (PUFs)
simulation, specifically XOR Arbiter PUFs. PUFs are used for authentication purposes. For more inf
o, refer to our paper below.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">6000000 </p></td>
<td><p class="normal">129 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Superconductivty+Data"></a> </td><td><p class="normal"><b><a href="datasets/Superconductivty+Data">Superconductivty Data</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Two file s contain data on 21263 superconductors and their relevant

```

```

features.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">21263 </p></td>
<td><p class="normal">81 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <p class="normal">Physical&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/WESAD%28Wearable+Stress+and+Affect+Detection%29"></a> </td><td><p class="normal"><b><a href="datasets/WESAD%28Wearable+Stress+and+Affect+Detection%29">WESAD (Wearable Stress and Affect Detection)</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">WESAD (Wearable Stress and Affect Detection) contains data of 15 subjects during a stress-affect lab study, while wearing physiological and motion sensors.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">63000000 </p></td>
<td><p class="normal">12 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2"></a> </td><td><p class="normal"><b><a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2">GNFUV Unmanned Surface Vehicles Sensor Data Set 2</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The data-set contains eight (2x4) data-sets of mobile sensor readings data (humidity, temperature) corresponding to a swarm of four Unmanned Surface Vehicles (USVs) in a test-bed, Athens, Greece.&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Sequential, Time-Series </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">10190 </p></td>
<td><p class="normal">6 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Student+Academics+Performance"></a> </td><td><p class="normal"><b><a href="datasets/Student+Academics+Performance">Student Academics Performance</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The dataset tried to find the end semester percentage prediction based on different social, economic and academic attributes. &nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">300 </p></td>
<td><p class="normal">22 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Online+Shoppers+Purchasing+Intention+Dataset"></a> </td><td><p class="normal"><b><a href="datasets/Online+Shoppers+Purchasing+Intention+Dataset">Online Shoppers Purchasing Intention Dataset</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Of the 12,330 sessions in the dataset, 84.5% (10,422) were negative class samples that did not end with shopping, and the rest (1908) were positive class samples ending with shopping.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">12330 </p></td>
<td><p class="normal">18 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <p class="normal">Business&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/PMU-UD"></a> </td><td><p class="normal"><b><a href="datasets/PMU-UD">PMU-UD</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">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). &nbsp;</p></td> -->
<td><p class="normal">Univariate </p></td>
<td><p class="normal">Classification </p></td>

```

```

<td><p class="normal"> </p></td>
<td><p class="normal">5180 </p></td>
<td><p class="normal">9 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Parkinson%27s+Disease+Classification"></a> </td><td><p class="normal"><b><a href="datasets/Parkinson%27s+Disease+Classification">Parkinson's Disease Classification</a></b></a></td></tr></table></td>
<!-- <td><p class="normal">The data used in this study were gathered from 188 patients with PD (10 7 men and 81 women) with ages ranging from 33 to 87 (65.1±10.9).&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">756 </p></td>
<td><p class="normal">754 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Electrical+Grid+Stability+Simulated+Data+"></a> </td><td><p class="normal"><b><a href="datasets/Electrical+Grid+Stability+Simulated+Data+">Electrical Grid Stability Simulated Data</a></b></a></td></tr></table></td>
<!-- <td><p class="normal">The local stability analysis of the 4-node star system (electricity producer is in the center) implementing Decentral Smart Grid Control concept. &nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">10000 </p></td>
<td><p class="normal">14 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Caesarian+Section+Classification+Dataset"></a> </td><td><p class="normal"><b><a href="datasets/Caesarian+Section+Classification+Dataset">Caesarian Section Classification Dataset</a></b></a></td></tr></table></td>
<!-- <td><p class="normal">This dataset contains information about caesarian section results of 8 0 pregnant women with the most important characteristics of delivery problems in the medical field .&nbsp;</p></td> -->
<td><p class="normal">Univariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">80 </p></td>
<td><p class="normal">5 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/BAUM-1"></a> </td><td><p class="normal"><b><a href="datasets/BAUM-1">BAUM-1</a></b></a></td></tr></table></td>
<!-- <td><p class="normal">BAUM-1 dataset contains 1184 multimodal facial video clips collected from 31 subjects. The 1184 video clips contain spontaneous facial expressions and speech of 13 emotional and mental states.&nbsp;</p></td> -->
<td><p class="normal">Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">1184 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/BAUM-2"></a> </td><td><p class="normal"><b><a href="datasets/BAUM-2">BAUM-2</a></b></a></td></tr></table></td>
<!-- <td><p class="normal">A multilingual audio-visual affective face database consisting of 1047 video clips of 286 subjects. &nbsp;</p></td> -->
<td><p class="normal">Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">1047 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Audit+Data"></a> </td><p class="normal"><b><a href="datasets/Audit+Data">Audit Data</a></b></p></td></tr></table></td>
<!-- <td><p class="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.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">777 </p></td>
<td><p class="normal">18 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/BuddyMove+Data+Set"></a> </td><td><p class="normal"><b><a href="datasets/BuddyMove+Data+Set">BuddyMove Data Set</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">User interest information extracted from user reviews published in holi dayiq.com about various types of point of interests in South India&nbsp;</p></td> -->
<td><p class="normal">Multivariate, Text </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">249 </p></td>
<td><p class="normal">7 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Real+estate+valuation+data+set"></a> </td><td><p class="normal"><b><a href="datasets/Real+estate+valuation+data+set">Real estate valuation data set</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">The "real estate valuation" is a regression problem. The market historical data set of real estate valuation are collected from Sidian Dist., New Taipei City, Taiwan. &nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Regression </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">414 </p></td>
<td><p class="normal">7 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Business&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Early+biomarkers+of+Parkinson%20%99s+disease+based+on+natural+connected+speech+Data+et+></a> </td><td><p class="normal"><b><a href="datasets/Early+biomarkers+of+Parkinson%20%99s+disease+based+on+natural+connected+speech+Data+et+>Early biomarkers of Parkinson's disease based on natural connected speech Data Set </a></b></p></td></tr></table></td>
<!-- <td><p class="normal">.&nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr bgcolor="#DDEEFF">
<td><table><tr><td><a href="datasets/Somerville+Happiness+Survey"></a> </td><td><p class="normal"><b><a href="datasets/Somerville+Happiness+Survey">Somerville Happiness Survey</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">A data extract of a non-federal dataset posted here https://catalog.data.gov/dataset/somerville-happiness-survey-responses-2011-2013-2015&nbsp;</p></td> -->
<td><p class="normal"> </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">143 </p></td>
<td><p class="normal">7 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/2.4+GHZ+Indoor+Channel+Measurements"></a> </td><td><p class="normal"><b><a href="datasets/2.4+GHZ+Indoor+Channel+Measurements">2.4 GHZ Indoor Channel Measurements</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Measurement of the S21, consists of 10 sweeps, each sweep contains 601 f

```

reency points with spacing of 0.167MHz to cover a 100MHz band centered at 2.4GHz. </p></td>

-->

<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">7840 </p></td>
<td><p class="normal">5 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Computer&nbs;p;</p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">EMG data for gestures</p></td></tr></table></td>
<!-- <td><p class="normal">These are files of raw EMG data recorded by MYO Thalmic bracelet&nbs;p;</p></td> -->

<td><p class="normal">Time-Series </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">30000 </p></td>
<td><p class="normal">6 </p></td>
<td><p class="normal">2019 </p></td>
<!-- <td><p class="normal">Life&nbs;p;</p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">Parking Birmingham</p></td></tr></table></td>
<!-- <td><p class="normal">Data collected from car parks in Birmingham that are operated by NCP fr om

Birmingham City Council. UK Open Government Licence (OGL).
<https://data.birmingham.gov.uk/dataset/birmingham-parking&nbs;p;> </td> -->

<td><p class="normal">Multivariate, Univariate, Sequential, Time-Series </p></td>
<td><p class="normal">Classification, Regression, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">35717 </p></td>
<td><p class="normal">4 </p></td>
<td><p class="normal">2019 </p></td>
<!-- <td><p class="normal">Computer&nbs;p;</p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">Behavior of the u rban traffic of the city of Sao Paulo in Brazil</p></td></tr></table></td>
<!-- <td><p class="normal">The database was created with records of behavior of the urban traffic of the city of Sao Paulo in Brazil.&nbs;p;</p></td> -->

<td><p class="normal">Multivariate, Time-Series </p></td>
<td><p class="normal">Classification, Regression </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">135 </p></td>
<td><p class="normal">18 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Computer&nbs;p;</p></td> -->

</tr><tr>

<td><table><tr><td> </td><td><p class="normal">Travel Reviews</p></td></tr></table></td>
<!-- <td><p class="normal">Reviews on destinations in 10 categories mentioned across East Asia.

Each traveler rating is mapped as Excellent(4), Very Good(3), Average(2), Poor(1), and Terrible(0) and average rating is used.&nbs;p;</p></td> -->

<td><p class="normal">Multivariate, Text </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">980 </p></td>
<td><p class="normal">11 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Other&nbs;p;</p></td> -->

</tr><tr bgcolor="#DDEEFF">

<td><table><tr><td> </td><td><p class="normal">Tarvel Review Ratings</p></td></tr></table></td>
<!-- <td><p class="normal">Google reviews on attractions from 24 categories across Europe are cons idered. Google user rating ranges from 1 to 5 and average user rating per category is calculated.&nbs;p;</p></td> -->

<td><p class="normal">Multivariate, Text </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><p class="normal">5456 </p></td>

```

<td><p class="normal">25 </p></td>
<td><p class="normal">2018 </p></td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Rice+Leaf+Diseases"></a> </td><td><p class="normal"><b><a href="datasets/Rice+Leaf+Diseases">Rice Leaf Diseases</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">There are three classes/diseases: Bacterial leaf blight, Brown spot, and Leaf smut, each having 40 images. The format of all images is jpg. &nbsp;</p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer </p></td>
<td><p class="normal">120 </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">2019 </p></td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr></table>
</td>
</tr>
</table>
<table align="center" cellpadding="5"><tr valign="center">
<td><p class="normal">Supported By:</p></td>
<td> </td>
<td><p class="normal"> In Collaboration With:</p></td>
<td></td>
</tr></table>
<center>
<span class="normal">
<a href="about.html">About</a> ||
<a href="citation_policy.html">Citation Policy</a> ||
<a href="donation_policy.html">Donation Policy</a> ||
<a href="contact.html">Contact</a> ||
<a href="http://cml.ics.uci.edu">CML</a>
</span>
</center>
</body></html>

```

In [12]:

```

print(database.prettify())

```

```

<!DOCTYPE HTML>
<html>
<body>
<p>
  "-//W3C//DTD HTML 4.01 Transitional//EN\"&gt;;
</p>
<title>
  UCI Machine Learning Repository: Data Sets
</title>
<!-- Stylesheet link -->
<link href="assets/ml.css" rel="stylesheet" type="text/css"/>
<script language="JavaScript" type="text/javascript">
  <!--
function checkform ( form )
{
  // see http://www.thesitewizard.com/archive/validation.shtml
  // for an explanation of this script and how to use it on your
  // own website

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

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

  // ** END **

```

```

        return true ;
    }

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

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

</a>
<br/>
<a href="http://cml.ics.uci.edu">
<font color="FFDD33">
    Center for Machine Learning and Intelligent Systems
</font>
</a>
</span>
</td>
<td align="right" valign="top" width="100%">
<span class="whitetext">
<a href="about.html">
    About
</a>
<a href="citation_policy.html">
    Citation Policy
</a>
<a href="donation_policy.html">
    Donate a Data Set
</a>
<a href="contact.html">
    Contact
</a>
</span>
<br/>
<br/>
<!-- Search Google -->
<form action="http://www.google.com/custom" method="GET" onsubmit="return checkform(this);">
    <input maxlength="255" name="q" size="30" type="text" value="" />
    <input name="sa" type="submit" value="Search" />
    <input name="cof" type="hidden" value="AH:center;LH:130;L:http://archive.ics.uci.edu/assets/logo.gif;LW:384;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>
</form>

```

```

</span>
<input name="sitesearch" type="radio" value="" />
<span class="whitetext">
<font size="1">
  Web
</font>
</span>
<a href="http://www.google.com/search">
  
</a>
<br/>
</form>
<!-- Search Google -->
<span class="whitetext">
<a href="datasets.php">
  <font color="#FFDD33" size="3">
    <b>
      View ALL Data Sets
    </b>
  </font>
</a>
</span>
<br/>
</td>
</tr>
</table>
<br/>
<table cellpadding="3">
<tr>
  <td valign="top">
    <table>
      <tr>
        <td>
          <p align="center" class="big">
            Browse Through:
          </p>
        </td>
      </tr>
    </table>
    <table border="2" cellpadding="2" width="150">
      <tr>
        <td bgcolor="#003366">
          <p class="whitetext">
            <b>
              Default Task
            </b>
          </p>
        </td>
      </tr>
      <tr>
        <td valign="top">
          <p class="normal">
            <a href="datasets.php?
format=&task=cla&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=
e">
              Classification
            </a>
            <font color="red">
              (350)
            </font>
            <br/>
            <a href="datasets.php?
format=&task=reg&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=
e">
              Regression
            </a>
            <font color="red">
              (96)
            </font>
            <br/>
            <a href="datasets.php?
format=&task=clu&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=
e">
              Clustering
            </a>
            <font color="red">

```

```

(84)
</font>
<br/>
<a href="datasets.php?
format=&task=other&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=
ble">
    Other
</a>
<font color="red">
    (55)
</font>
</p>
</td>
</tr>
<tr>
<td bgcolor="#003366">
    <p class="whitetext">
        <b>
            Attribute Type
        </b>
    </p>
</td>
</tr>
<tr>
<td valign="top">
    <p class="normal">
        <a href="datasets.php?
format=&task=&att=cat&area=&numAtt=&numIns=&type=&sort=nameUp&view=
e">
            Categorical
        </a>
        <font color="red">
            (38)
        </font>
        <br/>
        <a href="datasets.php?
format=&task=&att=num&area=&numAtt=&numIns=&type=&sort=nameUp&view=
e">
            Numerical
        </a>
        <font color="red">
            (307)
        </font>
        <br/>
        <a href="datasets.php?
format=&task=&att=mix&area=&numAtt=&numIns=&type=&sort=nameUp&view=
e">
            Mixed
        </a>
        <font color="red">
            (55)
        </font>
        </p>
    </td>
</tr>
<tr>
<td bgcolor="#003366">
    <p class="whitetext">
        <b>
            Data Type
        </b>
    </p>
</td>
</tr>
<tr>
<td valign="top">
    <p class="normal">
        <a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=mvar&sort=nameUp&view=
le">
            Multivariate
        </a>
        <font color="red">
            (357)
        </font>
        <br/>
        <a href="datasets.php?

```

```
format=&task=&att=&area=&numAtt=&numIns=&type=uvar&sort=nameUp&view=le">
    Univariate
  </a>
  <font color="red">
    (23)
  </font>
  <br/>
  <a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=seq&sort=nameUp&view=e">
    Sequential
  </a>
  <font color="red">
    (47)
  </font>
  <br/>
  <a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=ts&sort=nameUp&view=t">
    Time-Series
  </a>
  <font color="red">
    (91)
  </font>
  <br/>
  <a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=text&sort=nameUp&view=le">
    Text
  </a>
  <font color="red">
    (53)
  </font>
  <br/>
  <a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=dt&sort=nameUp&view=t">
    Domain-Theory
  </a>
  <font color="red">
    (23)
  </font>
  <br/>
  <a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=other&sort=nameUp&view=ble">
    Other
  </a>
  <font color="red">
    (21)
  </font>
  <br/>
</p>
</td>
</tr>
<tr>
<td bgcolor="#003366">
<p class="whitetext">
<b>
    Area
  </b>
</p>
</td>
</tr>
<tr>
<td valign="top">
<p class="normal">
  <a href="datasets.php?
format=&task=&att=&area=life&numAtt=&numIns=&type=&sort=nameUp&view=le">
    Life Sciences
  </a>
  <font color="red">
    (107)
  </font>
  <br/>
```

```
<a href="datasets.php?
format=&task=&att=&area=phys&numAtt=&numIns=&type=&sort=nameUp&viewle">
    Physical Sciences
</a>
<font color="red">
    (49)
</font>
<br/>
<a href="datasets.php?
format=&task=&att=&area=comp&numAtt=&numIns=&type=&sort=nameUp&viewle">
    CS / Engineering
</a>
<font color="red">
    (170)
</font>
<br/>
<a href="datasets.php?
format=&task=&att=&area=soc&numAtt=&numIns=&type=&sort=nameUp&viewle">
    Social Sciences
</a>
<font color="red">
    (26)
</font>
<br/>
<a href="datasets.php?
format=&task=&att=&area=bus&numAtt=&numIns=&type=&sort=nameUp&viewle">
    Business
</a>
<font color="red">
    (29)
</font>
<br/>
<a href="datasets.php?
format=&task=&att=&area=game&numAtt=&numIns=&type=&sort=nameUp&viewle">
    Game
</a>
<font color="red">
    (10)
</font>
<br/>
<a href="datasets.php?
format=&task=&att=&area=other&numAtt=&numIns=&type=&sort=nameUp&viewble">
    Other
</a>
<font color="red">
    (73)
</font>
</p>
</td>
</tr>
<tr>
<td bgcolor="#003366">
    <p class="whitetext">
        <b>
            # Attributes
        </b>
    </p>
</td>
</tr>
<tr>
    <td valign="top">
        <p class="normal">
            <a href="datasets.php?
format=&task=&att=&area=&numAtt=less10&numIns=&type=&sort=nameUp&viewle">
                Less than 10
            </a>
            <font color="red">
                (113)
            </font>
            <br/>
```

```

~~~
<a href="datasets.php?
format=&task=&att=&area=&numAtt=10to100&numIns=&type=&sort=nameUp&view=table">
    10 to 100
</a>
<font color="red">
    (210)
</font>
<br/>
<a href="datasets.php?
format=&task=&att=&area=&numAtt=greater100&numIns=&type=&sort=nameUp&view=table">
    Greater than 100
</a>
<font color="red">
    (84)
</font>
</p>
</td>
</tr>
<tr>
    <td bgcolor="#003366">
        <p class="whitetext">
            <b>
                # Instances
            </b>
            </p>
        </td>
    </tr>
    <tr>
        <td valign="top">
            <p class="normal">
                <a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=less100&type=&sort=nameUp&view=table">
                    Less than 100
                </a>
                <font color="red">
                    (27)
                </font>
                <br/>
                <a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=100to1000&type=&sort=nameUp&view=table">
                    100 to 1000
                </a>
                <font color="red">
                    (162)
                </font>
                <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>
                </p>
            </td>
        </tr>
        <tr>
            <td bgcolor="#003366">
                <p class="whitetext">
                    <b>
                        Format Type
                    </b>
                    </p>
                </td>
            </tr>
            <tr>
                <td valign="top">
                    <p class="normal">
                        <a href="datasets.php?
format=mat&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=matrix">
                            Matrix
                        </a>
                    </p>
                </td>
            </tr>
        </table>
    </div>

```

```

        ...
    </a>
    <font color="red">
        (324)
    </font>
    <br/>
    <a href="datasets.php?
format=nonmat&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&visible">
        Non-Matrix
    </a>
    <font color="red">
        (145)
    </font>
    </p>
    </td>
    </tr>
</table>
</td>
<td valign="top">
<table width="100%">
    <tr>
        <td>
            <p class="big">
                <b>
                    469
                </b>
                Data Sets
            </p>
        </td>
        <td align="right">
            <p class="normal">
                <font color="gray">
                    Table View
                </font>
            <a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list">
                List View
            </a>
            </p>
        </td>
    </tr>
</table>
<table border="1" cellpadding="5">
    <tr bgcolor="#003366">
        <td class="normal, whitetext">
            <p class="normal, whitetext">
                <a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameDown&view=t">
                    <b>
                        Name
                    </b>
                </a>
            </p>
        </td>
        <!-- <td><p class="normal, whitetext"><b>Abstract</b></p></td> -->
        <td>
            <p class="normal, whitetext">
                <a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=&sort=typeUp&view=tak">
                    <b>
                        Data Types
                    </b>
                </a>
            </p>
        </td>
        <td>
            <p class="normal, whitetext">
                <a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=&sort=taskUp&view=tak">
                    <b>
                        Default Task
                    </b>
                </a>
            </p>
        </td>
    </tr>
</table>

```

```

</td>
<td>
<p class="normal, whitetext">
<a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=&sort=attTypeUp&view=e">
<b>
Attribute Types
</b>
</a>
</p>
</td>
<td>
<p class="normal, whitetext">
<a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=&sort=instUp&view=tak
<b>
# Instances
</b>
</a>
</p>
</td>
<td>
<p class="normal, whitetext">
<a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=&sort=attUp&view=tabl
<b>
# Attributes
</b>
</a>
</p>
</td>
<td>
<p class="normal, whitetext">
<a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=&sort=dateUp&view=tak
<b>
Year
</b>
</a>
</p>
</td>
<!-- <td><p class="normal, whitetext"><b>Area</b></p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Abalone">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Abalone">
Abalone
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Predict the age of abalone from physical measurements&nbsp;</p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>

```

```

<!-->
<p class="normal">
  Classification
</p>
</td>
<td>
<p class="normal">
  Categorical, Integer, Real
</p>
</td>
<td>
<p class="normal">
  4177
</p>
</td>
<td>
<p class="normal">
  8
</p>
</td>
<td>
<p class="normal">
  1995
</p>
</td>
<!-- <td><p class="normal">Life &nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Adult">
  
</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Adult">
  Adult
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Predict whether income exceeds $50K/yr based on census data. Also known as "Census Income" dataset.&nbsp;</p></td> -->
<td>
<p class="normal">
  Multivariate
</p>
</td>
<td>
<p class="normal">
  Classification
</p>
</td>
<td>
<p class="normal">
  Categorical, Integer
</p>
</td>
<td>
<p class="normal">
  48842
</p>
</td>
<td>
<p class="normal">
  14
</p>
</td>
<td>
<p class="normal">
  100%
</p>

```

```
    1990
  </p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
  <a href="datasets/Annealing">
    
  </a>
</td>
<td>
  <p class="normal">
    <b>
      <a href="datasets/Annealing">
        Annealing
      </a>
    </b>
  </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Steel annealing data </p></td> -->
<td>
  <p class="normal">
    Multivariate
  </p>
</td>
<td>
<p class="normal">
  Classification
</p>
</td>
<td>
<p class="normal">
  Categorical, Integer, Real
</p>
</td>
<td>
<p class="normal">
  798
</p>
</td>
<td>
<p class="normal">
  38
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
  <a href="datasets/Anonymous+Microsoft+Web+Data">
    
  </a>
</td>
<td>
  <p class="normal">
    <b>
      <a href="datasets/Anonymous+Microsoft+Web+Data">
        Anonymous Microsoft Web Data
      </a>
    </b>
  </p>
</td>
</tr>
</table>
</td>

```

```
</table>
</td>
<!-- <td><p class="normal">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.&nbsp;</p></td> -->
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
    Recommender-Systems
</p>
</td>
<td>
<p class="normal">
    Categorical
</p>
</td>
<td>
<p class="normal">
    37711
</p>
</td>
<td>
<p class="normal">
    294
</p>
</td>
<td>
<p class="normal">
    1998
</p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Arrhythmia">
    
</a>
</td>
<td>
<p class="normal">
<b>
    <a href="datasets/Arrhythmia">
        Arrhythmia
    </a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Distinguish between the presence and absence of cardiac
arrhythmia and classify it in one of the 16 groups.&nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">
    Categorical, Integer, Real
</p>
</td>
<td>
<p class="normal">
    452
</p>
...<br/>
```

```

</td>
<td>
<p class="normal">
279
</p>
</td>
<td>
<p class="normal">
1998
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Artificial+Characters">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Artificial+Characters">
    Artificial Characters
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Dataset artificially generated by using first order theory which
describes structure of ten capital letters of English alphabet </p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Categorical, Integer, Real
</p>
</td>
<td>
<p class="normal">
6000
</p>
</td>
<td>
<p class="normal">
7
</p>
</td>
<td>
<p class="normal">
1992
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Audiology%28Original%29">

</a>
</td>

```

```

<td>
  <p class="normal">
    <b>
      <a href="datasets/Audiology%28Original%29">
        Audiology (Original)
      </a>
    </b>
  </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Nominal audiology dataset from Baylor&nbsp;</p></td> -->
<td>
  <p class="normal">
    Multivariate
  </p>
</td>
<td>
  <p class="normal">
    Classification
  </p>
</td>
<td>
  <p class="normal">
    Categorical
  </p>
</td>
<td>
  <p class="normal">
    226
  </p>
</td>
<td>
  <p class="normal">
    1987
  </p>
</td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
  <table>
    <tr>
      <td>
        <a href="datasets/Audiology%28Standardized%29">
          
        </a>
      </td>
      <td>
        <p class="normal">
          <b>
            <a href="datasets/Audiology%28Standardized%29">
              Audiology (Standardized)
            </a>
          </b>
        </p>
      </td>
    </tr>
  </table>
</td>
<!-- <td><p class="normal">Standardized version of the original audiology database&nbsp;</p></td> -->
</p></td>
<td>
  <p class="normal">
    Multivariate
  </p>
</td>
<td>
  <p class="normal">
    Classification
  </p>
</td>

```

```
<td>
<p class="normal">
  Categorical
</p>
</td>
<td>
<p class="normal">
  226
</p>
</td>
<td>
<p class="normal">
  69
</p>
</td>
<td>
<p class="normal">
  1992
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Auto+MPG">
  
</a>
</td>
<td>
<p class="normal">
  <b>
    <a href="datasets/Auto+MPG">
      Auto MPG
    </a>
  </b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Revised from CMU StatLib library, data concerns city-cycle fuel
consumption </p></td> -->
<td>
<p class="normal">
  Multivariate
</p>
</td>
<td>
<p class="normal">
  Regression
</p>
</td>
<td>
<p class="normal">
  Categorical, Real
</p>
</td>
<td>
<p class="normal">
  398
</p>
</td>
<td>
<p class="normal">
  8
</p>
</td>
<td>
<p class="normal">
  1993
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
```

```

<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Automobile">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Automobile">
    Automobile
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">From 1985 Ward's Automotive Yearbook </p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Regression
</p>
</td>
<td>
<p class="normal">
    Categorical, Integer, Real
</p>
</td>
<td>
<p class="normal">
    205
</p>
</td>
<td>
<p class="normal">
    26
</p>
</td>
<td>
<p class="normal">
    1987
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Badges">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Badges">
    Badges
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Badges labeled with a "+" or "-" as a function of a person's
name </p></td> -->

```

```
<td>
<p class="normal">
    Univariate, Text
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">
    294
</p>
</td>
<td>
<p class="normal">
    1
</p>
</td>
<td>
<p class="normal">
    1994
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Balance+Scale">
    
</a>
</td>
<td>
<p class="normal">
    <b>
        <a href="datasets/Balance+Scale">
            Balance Scale
        </a>
    </b>
    </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Balance scale weight & distance database </p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">
    Categorical
</p>
</td>
<td>
<p class="normal">
    625
</p>
</td>
<td>
<p class="normal">
    4
</p>
```

```

</td>
<td>
<p class="normal">
  1994
</p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Balloons">
  
</a>
</td>
<td>
<p class="normal">
<b>
  <a href="datasets/Balloons">
    Balloons
  </a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Data previously used in cognitive psychology experiment; 4 data sets represent different conditions of an experiment </p></td> -->
<td>
<p class="normal">
  Multivariate
</p>
</td>
<td>
<p class="normal">
  Classification
</p>
</td>
<td>
<p class="normal">
  Categorical
</p>
</td>
<td>
<p class="normal">
  16
</p>
</td>
<td>
<p class="normal">
  4
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Breast+Cancer">
  
</a>
</td>
<td>
<p class="normal">
<b>
  <a href="datasets/Breast+Cancer">
    Breast Cancer
  </a>
</b>
</p>
</td>

```

```

        </b>
        </p>
        </td>
        </tr>
        </table>
        <td>
        <!-- <td><p class="normal">Breast Cancer Data (Restricted Access) &nbsp;</p></td> -->
        <p class="normal">
          Multivariate
        </p>
        </td>
        <td>
          <p class="normal">
            Classification
          </p>
        </td>
        <td>
          <p class="normal">
            Categorical
          </p>
        </td>
        <td>
          <p class="normal">
            286
          </p>
        </td>
        <td>
          <p class="normal">
            9
          </p>
        </td>
        <td>
          <p class="normal">
            1988
          </p>
        </td>
        <!-- <td><p class="normal">Life&nbsp;</p></td> -->
      </tr>
      <tr>
        <td>
          <table>
            <tr>
              <td>
                <a href="datasets/Breast+Cancer+Wisconsin%28Original%29">
                  
                </a>
              </td>
              <td>
                <p class="normal">
                  <b>
                    <a href="datasets/Breast+Cancer+Wisconsin%28Original%29">
                      Breast Cancer Wisconsin (Original)
                    </a>
                  </b>
                </p>
              </td>
            </tr>
          </table>
        </td>
        <!-- <td><p class="normal">Original Wisconsin Breast Cancer Database&nbsp;</p></td> -->
      <td>
        <p class="normal">
          Multivariate
        </p>
      </td>
      <td>
        <p class="normal">
          Classification
        </p>
      </td>
      <td>
        <p class="normal">
          Integer
        </p>
      </td>
    </td>

```

```
<p class="normal">
  699
</p>
</td>
<td>
<p class="normal">
  10
</p>
</td>
<td>
<p class="normal">
  1992
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Breast+Cancer+Wisconsin+%28Prognostic%29">
  
</a>
</td>
<td>
<p class="normal">
<b>
  <a href="datasets/Breast+Cancer+Wisconsin+%28Prognostic%29">
    Breast Cancer Wisconsin (Prognostic)
  </a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Prognostic Wisconsin Breast Cancer Database </p></td> -->
<td>
<p class="normal">
  Multivariate
</p>
</td>
<td>
<p class="normal">
  Classification, Regression
</p>
</td>
<td>
<p class="normal">
  Real
</p>
</td>
<td>
<p class="normal">
  198
</p>
</td>
<td>
<p class="normal">
  34
</p>
</td>
<td>
<p class="normal">
  1995
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29">
  
```

```

        </a>
    </td>
    <td>
        <p class="normal">
            <b>
                <a href="datasets/Breast+Cancer+Wisconsin%28Diagnostic%29">
                    Breast Cancer Wisconsin (Diagnostic)
                </a>
            </b>
        </p>
    </td>
    </tr>
</table>
</td>
<!-- <td><p class="normal">Diagnostic Wisconsin Breast Cancer Database </p></td> -->
<td>
    <p class="normal">
        Multivariate
    </p>
</td>
<td>
    <p class="normal">
        Classification
    </p>
</td>
<td>
    <p class="normal">
        Real
    </p>
</td>
<td>
    <p class="normal">
        569
    </p>
</td>
<td>
    <p class="normal">
        32
    </p>
</td>
<td>
    <p class="normal">
        1995
    </p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
    <td>
        <table>
            <tr>
                <td>
                    <a href="datasets/Pittsburgh+Bridges">
                        
                    </a>
                </td>
                <td>
                    <p class="normal">
                        <b>
                            <a href="datasets/Pittsburgh+Bridges">
                                Pittsburgh Bridges
                            </a>
                        </b>
                    </p>
                </td>
            </tr>
        </table>
    </td>
    <!-- <td><p class="normal">Bridges database that has original and numeric-discretized
datasets </p></td> -->
    <td>
        <p class="normal">
            Multivariate
        </p>
    </td>
    <td>
        <p class="normal">

```

```

`Classification
</p>
</td>
<td>
<p class="normal">
  Categorical, Integer
</p>
</td>
<td>
<p class="normal">
  108
</p>
</td>
<td>
<p class="normal">
  13
</p>
</td>
<td>
<p class="normal">
  1990
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Car+Evaluation">
  
</a>
</td>
<td>
<p class="normal">
<b>
  <a href="datasets/Car+Evaluation">
    Car Evaluation
  </a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Derived from simple hierarchical decision model, this database m
ay be useful for testing constructive induction and structure discovery methods. </p></td> -->
<td>
<p class="normal">
  Multivariate
</p>
</td>
<td>
<p class="normal">
  Classification
</p>
</td>
<td>
<p class="normal">
  Categorical
</p>
</td>
<td>
<p class="normal">
  1728
</p>
</td>
<td>
<p class="normal">
  6
</p>
</td>
<td>
<p class="normal">
  1997
</p>

```

```

</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Census+Income">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Census+Income">
    Census Income
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Predict whether income exceeds $50K/yr based on census data. Also known as "Adult" dataset.&ampnbsp</p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">
    Categorical, Integer
</p>
</td>
<td>
<p class="normal">
    48842
</p>
</td>
<td>
<p class="normal">
    14
</p>
</td>
<td>
<p class="normal">
    1996
</p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Chess%28King-Rook+vs.+King-Knight%29">

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

```

```

        .. ~~
    </tr>
</table>
</td>
<!-- <td><p class="normal">Knight Pin Chess End-Game Database Creator&nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate, Data-Generator
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">
    Categorical, Integer
</p>
</td>
<td>
<p class="normal">
    22
</p>
</td>
<td>
<p class="normal">
    1988
</p>
</td>
<!-- <td><p class="normal">Game&nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Chess%28King-Rook+vs.+King-Pawn%29">
    
</a>
</td>
<td>
<p class="normal">
    <b>
        <a href="datasets/Chess%28King-Rook+vs.+King-Pawn%29">
            Chess (King-Rook vs. King-Pawn)
        </a>
    </b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">King+Rook versus King+Pawn on a7 (usually abbreviated
KRKPA7).&nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">
    Categorical
</p>
</td>
<td>
<p class="normal">
    3196
</p>
</td>

```

```

        `````  

</td>

<td>

<p class="normal">

 36

</p>

</td>

<td>

<p class="normal">

 1989

</p>

</td>

<!-- <td><p class="normal">Game </p></td> -->

</tr>

<tr>

<td>

<table>

<tr>

<td>

</td>

<td>

<p class="normal">

 Chess (King-Rook vs. King)

</p>

</td>

</tr>

</table>

</td>

<!-- <td><p class="normal">Chess Endgame Database for White King and Rook against Black

King (KRK). </p></td> -->

<td>

<p class="normal">

 Multivariate

</p>

</td>

<td>

<p class="normal">

 Classification

</p>

</td>

<td>

<p class="normal">

 Categorical, Integer

</p>

</td>

<td>

<p class="normal">

 28056

</p>

</td>

<td>

<p class="normal">

 6

</p>

</td>

<td>

<p class="normal">

 1994

</p>

</td>

<!-- <td><p class="normal">Game </p></td> -->

</tr>

<tr bgcolor="DDEEFF">

<td>

<table>

<tr>

<td>

</td>

```



```
~\p>
</td>
<td>
<p class="normal">
 100
</p>
</td>
<td>
<p class="normal">
 6
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Connect-4

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Contains connect-4 positions </p></td> -->
<td>
<p class="normal">
 Multivariate, Spatial
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Categorical
</p>
</td>
<td>
<p class="normal">
 67557
</p>
</td>
<td>
<p class="normal">
 42
</p>
</td>
<td>
<p class="normal">
 1995
</p>
</td>
<!-- <td><p class="normal">Game </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
~\p>
```

```


</td>
<td>
 <p class="normal">

 Credit Approval

 </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This data concerns credit card applications; good mix of
attributes&nbsp;</p></td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Classification
 </p>
</td>
<td>
 <p class="normal">
 Categorical, Integer, Real
 </p>
</td>
<td>
 <p class="normal">
 690
 </p>
</td>
<td>
 <p class="normal">
 15
 </p>
</td>
<td>
 <p class="normal">
 Financial&nbsp;
 </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Japanese Credit Screening

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">Includes domain theory (generated by talking to Japanese domain
experts); data in Lisp&nbsp;</p></td> -->
<td>
 <p class="normal">
 Multivariate, Domain-Theory
 </p>
</td>
<td>

```

```

<ta>
 <p class="normal">
 Classification
 </p>
</td>
<td>
 <p class="normal">
 Categorical, Real, Integer
 </p>
</td>
<td>
 <p class="normal">
 125
 </p>
</td>
<td>
 <p class="normal">
 1992
 </p>
</td>
<!-- <td><p class="normal">Financial </p></td> -->
</tr>
<tr>
<td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Computer Hardware

 </p>
 </td>
 </tr>
 </table>
</td>
<!-- <td><p class="normal">Relative CPU Performance Data, described in terms of its cycle time, memory size, etc. </p></td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Regression
 </p>
</td>
<td>
 <p class="normal">
 Integer
 </p>
</td>
<td>
 <p class="normal">
 209
 </p>
</td>
<td>
 <p class="normal">
 9
 </p>
</td>
<td>
 <p class="normal">
 1987
 </p>
</td>

```

```
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Contraceptive Method Choice

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Dataset is a subset of the 1987 National Indonesia Contraceptive Prevalence Survey. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Categorical, Integer
</p>
</td>
<td>
<p class="normal">
 1473
</p>
</td>
<td>
<p class="normal">
 9
</p>
</td>
<td>
<p class="normal">
 1997
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Covertype

</p>
</td>
<td>
..
```

```

 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">Forest CoverType dataset
</p></td> -->
 <td>
 <p class="normal">
 Multivariate
 </p>
 </td>
 <td>
 <p class="normal">
 Classification
 </p>
 </td>
 <td>
 <p class="normal">
 Categorical, Integer
 </p>
 </td>
 <td>
 <p class="normal">
 581012
 </p>
 </td>
 <td>
 <p class="normal">
 54
 </p>
 </td>
 <td>
 <p class="normal">
 1998
 </p>
 </td>
 <!-- <td><p class="normal">Life
</p></td> -->
 </tr>
 <tr bgcolor="DDEEFF">
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Cylinder Bands

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">Used in decision tree induction for mitigating process delays
known as "cylinder bands" in rotogravure printing
</p></td> -->
 <td>
 <p class="normal">
 Multivariate
 </p>
 </td>
 <td>
 <p class="normal">
 Classification
 </p>
 </td>
 <td>
 <p class="normal">
 Categorical, Integer, Real
 </p>
 </td>
 <td>
 <p class="normal">
 512
 </p>
 </td>
 </tr>
 </table>

```

```

 </p>
 </td>
 <td>
 <p class="normal">
 39
 </p>
 </td>
 <td>
 <p class="normal">
 1995
 </p>
 </td>
 <!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Dermatology

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">Aim for this dataset is to determine the type of Eryhemato-Squamous Disease. </p></td> -->
 <td>
 <p class="normal">
 Multivariate
 </p>
 </td>
 <td>
 <p class="normal">
 Classification
 </p>
 </td>
 <td>
 <p class="normal">
 Categorical, Integer
 </p>
 </td>
 <td>
 <p class="normal">
 366
 </p>
 </td>
 <td>
 <p class="normal">
 33
 </p>
 </td>
 <td>
 <p class="normal">
 1998
 </p>
 </td>
 <!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
 <table>
 <tr>
 <td>

 </td>

```

```

</td>
<td>
<p class="normal">

 Diabetes

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This diabetes dataset is from AIM '94 </p></td> -->
<td>
<p class="normal">
 Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 Categorical, Integer
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 20
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 DGP2 - The Second Data Generation Program

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Generates application domains based on specific parameters,
number of features, and proportion of positive to negative examples </p></td> -->
<td>
<p class="normal">
 Data-Generator
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">

```

```

 Real
 </p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Document Understanding

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Five concepts, expressed as predicates, to be learned </p>
</td> -->
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
1994
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
</td>
<td>
```



```
<p class="normal">
 132
</p>
</td>
<td>
<p class="normal">
 12
</p>
</td>
<td>
<p class="normal">
 1989
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Ecoli

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This data contains protein localization sites </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 336
</p>
</td>
<td>
<p class="normal">
 8
</p>
</td>
<td>
<p class="normal">
 1996
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>


```

```


 </td>
 <td>
 <p class="normal">

 Flags

 </p>
 </td>
 </tr>
</table>
<td>
<!-- <td><p class="normal">From Collins Gem Guide to Flags, 1986 </p></td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Classification
 </p>
</td>
<td>
 <p class="normal">
 Categorical, Integer
 </p>
</td>
<td>
 <p class="normal">
 194
 </p>
</td>
<td>
 <p class="normal">
 30
 </p>
</td>
<td>
 <p class="normal">
 1990
 </p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Function Finding

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">Cases collected mostly from investigations in physical science; intention is to evaluate function-finding algorithms </p></td> -->
 <td>
 <p class="normal">
 </p>
 </td>
 <td>
 <p class="normal">
 Function-Learning
 </p>
 </td>

```

```
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 352
</p>
</td>
<td>
<p class="normal">
 1990
</p>
</td>
<td>
<p class="normal">
Physical </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Glass Identification

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">From USA Forensic Science Service; 6 types of glass; defined in
terms of their oxide content (i.e. Na, Fe, K, etc) </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 214
</p>
</td>
<td>
<p class="normal">
 10
</p>
</td>
<td>
<p class="normal">
 1987
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
```

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

</td>
<td>
<p class="normal">

Haberman's Survival

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Dataset contains cases from study conducted on the survival of patients who had undergone surgery for breast cancer </p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Integer
</p>
</td>
<td>
<p class="normal">
306
</p>
</td>
<td>
<p class="normal">
3
</p>
</td>
<td>
<p class="normal">
1999
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Hayes-Roth

</p>
</td>
</tr>
</table>
</td>
```

```
<!-- <td><p class="normal">Topic: human subjects study </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Categorical
</p>
</td>
<td>
<p class="normal">
 160
</p>
</td>
<td>
<p class="normal">
 5
</p>
</td>
<td>
<p class="normal">
 1989
</p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Heart Disease

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">4 databases: Cleveland, Hungary, Switzerland, and the VA Long
Beach </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Categorical, Integer, Real
</p>
</td>
<td>
<p class="normal">
 303
</p>
</td>
</td>
```

```

 ...
<p class="normal">
 75
</p>
</td>
<td>
<p class="normal">
 1988
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
 <p class="normal">

 Hepatitis

 </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">From G.Gong: CMU; Mostly Boolean or numeric-valued attribute
types; Includes cost data (donated by Peter Turney) </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Categorical, Integer, Real
</p>
</td>
<td>
<p class="normal">
 155
</p>
</td>
<td>
<p class="normal">
 19
</p>
</td>
<td>
<p class="normal">
 1988
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
 <p class="normal">

```

```

 <td>

 Horse Colic

 </p>
 </td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Well documented attributes; 368 instances with 28 attributes
(continuous, discrete, and nominal); 30% missing values
</p></td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Classification
 </p>
</td>
<td>
 <p class="normal">
 Categorical, Integer, Real
 </p>
</td>
<td>
 <p class="normal">
 368
 </p>
</td>
<td>
 <p class="normal">
 27
 </p>
</td>
<td>
 <p class="normal">
 1989
 </p>
</td>
<!-- <td><p class="normal">Life
</p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 ICU

 </p>
 </td>
 </tr>
 </table>
</td>
<!-- <td><p class="normal">Data set prepared for the use of participants for the 1994 AAAI
Spring Symposium on Artificial Intelligence in Medicine.
</p></td> -->
<td>
 <p class="normal">
 Multivariate, Time-Series
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <+>

```

```

<!--
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Image Segmentation

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Image data described by high-level numeric-valued attributes, 7
classes </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 2310
</p>
</td>
<td>
<p class="normal">
 19
</p>
</td>
<td>
<p class="normal">
 1990
</p>
</td>
<td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>

```

```

<!--
<td>

</td>
<td>
 <p class="normal">

 Internet Advertisements

 </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset represents a set of possible advertisements on
Internet pages. </p></td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Classification
 </p>
</td>
<td>
 <p class="normal">
 Categorical, Integer, Real
 </p>
</td>
<td>
 <p class="normal">
 3279
 </p>
</td>
<td>
 <p class="normal">
 1558
 </p>
</td>
<td>
 <p class="normal">
 1998
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Ionosphere

 </p>
 </td>
 </tr>
 </table>
</td>
<!-- <td><p class="normal">Classification of radar returns from the ionosphere </p>
-->
<td>
 <p class="normal">
 M...1+....+....+...

```

```
 multivariate
 </p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Integer, Real
</p>
</td>
<td>
<p class="normal">
 351
</p>
</td>
<td>
<p class="normal">
 34
</p>
</td>
<td>
<p class="normal">
 1989
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Iris

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Famous database; from Fisher, 1936 </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 150
</p>
</td>
<td>
<p class="normal">
 4
</p>
</td>
<!-- -->
```

```

<ta>
 <p class="normal">
 1988
 </p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 ISOLET

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal"> Goal: Predict which letter-name was spoken--a simple classification task.&nbsp</p></td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Classification
 </p>
</td>
<td>
 <p class="normal">
 Real
 </p>
</td>
<td>
 <p class="normal">
 7797
 </p>
</td>
<td>
 <p class="normal">
 617
 </p>
</td>
<td>
 <p class="normal">
 1994
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Kinship

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

```

```


 </p>
 </td>
</tr>
</table>
<td>
<!-- <td><p class="normal">Relational dataset </p></td> -->
<td>
<p class="normal">
 Relational
</p>
</td>
<td>
<p class="normal">
 Relational-Learning
</p>
</td>
<td>
<p class="normal">
 Categorical
</p>
</td>
<td>
<p class="normal">
 104
</p>
</td>
<td>
<p class="normal">
 12
</p>
</td>
<td>
<p class="normal">
 1990
</p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr>
<td>
<table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Labor Relations

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">From Collective Bargaining Review </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Categorical, Integer, Real
</p>
</td>
<td>
<p class="normal">
 --
</p>

```

```

57
</p>
</td>
<td>
<p class="normal">
 16
</p>
</td>
<td>
<p class="normal">
 1988
</p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 LED Display Domain

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">From Classification and Regression Trees book; We provide here 2
C programs for generating sample databases </p></td> -->
<td>
<p class="normal">
 Multivariate, Data-Generator
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Categorical
</p>
</td>
<td>
<p class="normal">
 7
</p>
</td>
<td>
<p class="normal">
 1988
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>


```

```

</td>
<td>
<p class="normal">

 Lenses

</p>
</td>
</tr>
</table>
</td>
<!!-- <td><p class="normal">Database for fitting contact lenses </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Categorical
</p>
</td>
<td>
<p class="normal">
 24
</p>
</td>
<td>
<p class="normal">
 4
</p>
</td>
<td>
<p class="normal">
 1990
</p>
</td>
<!!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Letter Recognition

</p>
</td>
</tr>
</table>
</td>
<!!-- <td><p class="normal">Database of character image features; try to identify the
letter </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification

```

```
</p>
</td>
<td>
<p class="normal">
 Integer
</p>
</td>
<td>
<p class="normal">
 20000
</p>
</td>
<td>
<p class="normal">
 16
</p>
</td>
<td>
<p class="normal">
 1991
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Liver Disorders

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">BUPA Medical Research Ltd. database donated by Richard S.
Forsyth </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 Categorical, Integer, Real
</p>
</td>
<td>
<p class="normal">
 345
</p>
</td>
<td>
<p class="normal">
 7
</p>
</td>
<td>
<p class="normal">
 1990
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
```

```

</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Logic Theorist

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">All code for Logic Theorist </p></td> -->
<td>
<p class="normal">
 Domain-Theory
</p>
</td>
<td>
<p class="normal">
Computer
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Lung Cancer

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Lung cancer data; no attribute definitions </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>

```

```

<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Integer
</p>
</td>
<td>
<p class="normal">
32
</p>
</td>
<td>
<p class="normal">
56
</p>
</td>
<td>
<p class="normal">
1992
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Lymphography

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This lymphography domain was obtained from the University
Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia. (Restricted access) </p></td>
-->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Categorical
</p>
</td>
<td>
<p class="normal">
148
</p>
</td>
<td>
<p class="normal">
18
</p>
</td>
<td>

```

```

<p class="normal">
 1988
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Mechanical Analysis

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">Fault diagnosis problem of electromechanical devices; also PUMPS
DATA SET is newer version with domain theory and results </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Categorical, Integer, Real
</p>
</td>
<td>
<p class="normal">
 209
</p>
</td>
<td>
<p class="normal">
 8
</p>
</td>
<td>
<p class="normal">
 1990
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Meta-data

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

```

```

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">Meta-Data was used in order to give advice about which classification method is appropriate for a particular dataset (taken from results of Statlog project). -->
</p></td> -->
 <td>
 <p class="normal">
 Multivariate
 </p>
 </td>
 <td>
 <p class="normal">
 Classification
 </p>
 </td>
 <td>
 <p class="normal">
 Categorical, Integer, Real
 </p>
 </td>
 <td>
 <p class="normal">
 528
 </p>
 </td>
 <td>
 <p class="normal">
 22
 </p>
 </td>
 <td>
 <p class="normal">
 1996
 </p>
 </td>
 <!-- <td><p class="normal">Other -->
</tr>
<tr>
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Mobile Robots

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">Learning concepts from sensor data of a mobile robot; set of data sets-->
</td>
 <p class="normal">
 Domain-Theory
 </p>
 </td>
 <td>
 <p class="normal">
 </p>
 </td>
 <td>
 <p class="normal">
 Categorical, Integer, Real
 </p>
 </td>

```

```

<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
1995
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Molecular Biology (Promoter Gene Sequences)

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">E. Coli promoter gene sequences (DNA) with partial domain
theory </p></td> -->
<td>
<p class="normal">
Sequential, Domain-Theory
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Categorical
</p>
</td>
<td>
<p class="normal">
106
</p>
</td>
<td>
<p class="normal">
58
</p>
</td>
<td>
<p class="normal">
1990
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>


```

```


 </td>
 <td>
 <p class="normal">

 Molecular Biology (Protein Secondary Structure)

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">From CMU connectionist bench repository; Classifies secondary
structure of certain globular proteins&nbsp;</p></td> -->
<td>
 <p class="normal">
 Sequential
 </p>
</td>
<td>
 <p class="normal">
 Classification
 </p>
</td>
<td>
 <p class="normal">
 Categorical
 </p>
</td>
<td>
 <p class="normal">
 128
 </p>
</td>
<td>
 <p class="normal">
 </p>
 </td>
 <td>
 <p class="normal">
 </p>
 </td>
 <td>
 <p class="normal">
 </p>
 </td>
 <td>
 <p class="normal">
 </p>
 </td>
 <td>
 <p class="normal">
 </p>
 </td>
 <td>
 <p class="normal">
 </p>
 </td>
 <td>
 <p class="normal">

 </p>
 </td>
 </td>
 </td>
 </td>
 </td>
 </td>
</tr>
<tr bgcolor="DDEEFF">
 <td>
 <table>
 <tr>
 <td>

 Molecular Biology (Splice-junction Gene Sequences)

 </td>
 <td>
 <p class="normal">
 Primate splice-junction gene sequences (DNA) with associated impre
fect domain theory&nbsp;</p></td> -->
 <td>
 <p class="normal">
 Sequential, Domain-Theory
 </p>
 </td>
 <td>
 <p class="normal">
 Classification
 </p>
 </td>
 </tr>
 </table>
 </td>
</pre>

```

```

</p>
</td>
<td>
<p class="normal">
 Categorical
</p>
</td>
<td>
<p class="normal">
 3190
</p>
</td>
<td>
<p class="normal">
 61
</p>
</td>
<td>
<p class="normal">
 1992
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 MONK's Problems

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">A set of three artificial domains over the same attribute space;
Used to test a wide range of induction algorithms </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Categorical
</p>
</td>
<td>
<p class="normal">
 432
</p>
</td>
<td>
<p class="normal">
 7
</p>
</td>
<td>
<p class="normal">
 1992
</p>
</td>
</tr>

```

```

 ., -->
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Moral Reasoner

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Horn-clause model that qualitatively simulates moral reasoning; Theory includes negated literals </p></td> -->
<td>
<p class="normal">
 Domain-Theory
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 202
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 1994
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Multiple Features

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset consists of features of handwritten numerals ('0'-->'9') extracted from a collection of Dutch utility maps&nbsn:</p></td> -->

```

, extracted from a collection of Bacon writing manuscripts, 1570-1616

<td>	<p class="normal"> Multivariate		
</td>	</p>		
<td>	<p class="normal"> Classification		
</td>	</p>		
<td>	<p class="normal"> Integer, Real		
</td>	</p>		
<td>	<p class="normal"> 2000		
</td>	</p>		
<td>	<p class="normal"> 649		
</td>	</p>		
<td>	<p class="normal"> </p>		
</td>	</p>		
<!-- <td><p class="normal">Computer </p></td> -->			
</tr>			
<tr bgcolor="DDEEFF">			
<td>			
<table>			
<tr>			
<td>	<a href="datasets/Mushroom">  <td></td> <td></td>		
</a>			
</td>			
<td>	<p class="normal">		
<b>			
<a href="datasets/Mushroom">			
Mushroom			
</a>			
</b>			
</p>			
</td>			
</tr>			
</table>			
</td>			
<!-- <td><p class="normal">From Audobon Society Field Guide; mushrooms described in terms of physical characteristics; classification: poisonous or edible </p></td> -->			
<td>			
<p class="normal">			
Multivariate			
</p>			
</td>			
<td>	<p class="normal">		
Classification			
</p>			
</td>			
<td>	<p class="normal">		
Categorical			
</p>			
</td>			
<td>	<p class="normal">		
8124			
</p>			
</td>			
<td>	<p class="normal">		
??			

```

<-->
</p>
</td>
<td>
<p class="normal">
 1987
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Musk (Version 1)

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The goal is to learn to predict whether new molecules will be musks or non-musks </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Integer
</p>
</td>
<td>
<p class="normal">
 476
</p>
</td>
<td>
<p class="normal">
 168
</p>
</td>
<td>
<p class="normal">
 1994
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">


```

```


 Musk (Version 2)

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The goal is to learn to predict whether new molecules will be musks or non-musks&nbsp</p></td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Classification
 </p>
</td>
<td>
 <p class="normal">
 Integer
 </p>
</td>
<td>
 <p class="normal">
 6598
 </p>
</td>
<td>
 <p class="normal">
 168
 </p>
</td>
<td>
 <p class="normal">
 1994
 </p>
</td>
<!-- <td><p class="normal">Physical&nbsp</p></td> -->
</tr>
<tr>
<td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Nursery

 </p>
 </td>
 </tr>
 </table>
</td>
<!-- <td><p class="normal"> Nursery Database was derived from a hierarchical decision model originally developed to rank applications for nursery schools.&nbsp</p></td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Classification
 </p>
</td>
<td>
 <p class="normal">

```

```

<p class="normal">
 Categorical
</p>
</td>
<td>
<p class="normal">
 12960
</p>
</td>
<td>
<p class="normal">
 8
</p>
</td>
<td>
<p class="normal">
 1997
</p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
 <p class="normal">

 Othello Domain Theory

 </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Used in research to generate features for an inductive learning
system </p></td> -->
<td>
 <p class="normal">
 Domain-Theory
 </p>
</td>
<td>
 <p class="normal">
 1991
 </p>
</td>
<!-- <td><p class="normal">Game </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

```

```


</td>
<td>
 <p class="normal">

 Page Blocks Classification

 </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The problem consists of classifying all the blocks of the page l
ayout of a document that has been detected by a segmentation process. </p></td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Classification
 </p>
</td>
<td>
 <p class="normal">
 Integer, Real
 </p>
</td>
<td>
 <p class="normal">
 5473
 </p>
</td>
<td>
 <p class="normal">
 10
 </p>
</td>
<td>
 <p class="normal">
 1995
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Optical Recognition of Handwritten Digits

 </p>
 </td>
 </tr>
 </table>
</td>
<!-- <td><p class="normal">Two versions of this database available; see folder </p>
</td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
 ...

```

```

</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Integer
</p>
</td>
<td>
<p class="normal">
 5620
</p>
</td>
<td>
<p class="normal">
 64
</p>
</td>
<td>
<p class="normal">
 1998
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Pen-Based Recognition of Handwritten Digits

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Digit database of 250 samples from 44 writers </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Integer
</p>
</td>
<td>
<p class="normal">
 10992
</p>
</td>
<td>
<p class="normal">
 16
</p>
</td>
<td>
<p class="normal">
 ...

```

```
1998
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Post-Operative Patient

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Dataset of patient features </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Categorical, Integer
</p>
</td>
<td>
<p class="normal">
 90
</p>
</td>
<td>
<p class="normal">
 8
</p>
</td>
<td>
<p class="normal">
 1993
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Primary Tumor

</p>
</td>
`
```



```

<p class="normal">
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Qualitative Structure Activity Relationships

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Two sets of datasets are given: pyrimidines and triazines </p></td> -->
</p></td>
<td>
<p class="normal">
Domain-Theory
</p>
</td>
<td>
<p class="normal">
Physical
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>

</td>
<td>
<p class="normal">

Quadruped Mammals

</p>
</td>

```

```

 </tr>
 </table>
 </td>
 <!-- <td><p class="normal"> The file animals.c is a data generator of structured instances
representing quadruped animals&nbsp.</p></td> -->
 <td>
 <p class="normal">
 Multivariate, Data-Generator
 </p>
 </td>
 <td>
 <p class="normal">
 Classification
 </p>
 </td>
 <td>
 <p class="normal">
 Real
 </p>
 </td>
 <td>
 <p class="normal">
 72
 </p>
 </td>
 <td>
 <p class="normal">
 1992
 </p>
 </td>
 <!-- <td><p class="normal">Life&nbsp.</p></td> -->
</tr>
<tr>
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Servo

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">Data was from a simulation of a servo system&nbsp.</p></td> -->
 <td>
 <p class="normal">
 Multivariate
 </p>
 </td>
 <td>
 <p class="normal">
 Regression
 </p>
 </td>
 <td>
 <p class="normal">
 Categorical, Integer
 </p>
 </td>
 <td>
 <p class="normal">
 167
 </p>
 </td>

```

```
</td>
<td>
<p class="normal">
 4
</p>
</td>
<td>
<p class="normal">
 1993
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Shuttle Landing Control

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Tiny database; all nominal values </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Categorical
</p>
</td>
<td>
<p class="normal">
 15
</p>
</td>
<td>
<p class="normal">
 6
</p>
</td>
<td>
<p class="normal">
 1988
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
```

```

<p class="normal">

 Solar Flare

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Each class attribute counts the number of solar flares of a
certain class that occur in a 24 hour period </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Regression
</p>
</td>
<td>
<p class="normal">
 Categorical
</p>
</td>
<td>
<p class="normal">
 1389
</p>
</td>
<td>
<p class="normal">
 10
</p>
</td>
<td>
<p class="normal">
 1989
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Soybean (Large)

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Michalski's famous soybean disease database </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>

```

```
<td>
<p class="normal">
 Categorical
</p>
</td>
<td>
<p class="normal">
 307
</p>
</td>
<td>
<p class="normal">
 35
</p>
</td>
<td>
<p class="normal">
 1988
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Soybean (Small)

 </p>
</td>
</tr>
</tr>
</table>
</td>
<!-- <td><p class="normal">Michalski's famous soybean disease database </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Categorical
</p>
</td>
<td>
<p class="normal">
 47
</p>
</td>
<td>
<p class="normal">
 35
</p>
</td>
<td>
<p class="normal">
 1987
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
```

```

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

</td>
<td>
<p class="normal">

Challenger USA Space Shuttle O-Ring

</p>
</td>
</tr>
</table>
</td>
<!!-- <td><p class="normal">Task: predict the number of O-rings that experience thermal distress on a flight at 31 degrees F given data on the previous 23 shuttle flights </p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Regression
</p>
</td>
<td>
<p class="normal">
Integer
</p>
</td>
<td>
<p class="normal">
23
</p>
</td>
<td>
<p class="normal">
4
</p>
</td>
<td>
<p class="normal">
1993
</p>
</td>
<!!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Low Resolution Spectrometer

</p>
</td>
</tr>
</table>
</td>
<!!-- <td><p class="normal">From IRAS data -- NASA Ames Research Center </p></td> -->
<td>

```

```
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Integer, Real
</p>
</td>
<td>
<p class="normal">
 531
</p>
</td>
<td>
<p class="normal">
 102
</p>
</td>
<td>
<p class="normal">
 1988
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Spambase

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Classifying Email as Spam or Non-Spam </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Integer, Real
</p>
</td>
<td>
<p class="normal">
 4601
</p>
</td>
<td>
<p class="normal">
 57
</p>
```

```

</td>
<td>
<p class="normal">
 1999
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 SPECT Heart

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Data on cardiac Single Proton Emission Computed Tomography
(SPECT) images. Each patient classified into two categories: normal and abnormal.&nbsp</p></td> --
->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Categorical
</p>
</td>
<td>
<p class="normal">
 267
</p>
</td>
<td>
<p class="normal">
 22
</p>
</td>
<td>
<p class="normal">
 2001
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">


```

```

--> MICE datasets, SPECT Heart
 SPECTF Heart

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Data on cardiac Single Proton Emission Computed Tomography
(SPECT) images. Each patient classified into two categories: normal and abnormal. </p></td> --
->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Integer
</p>
</td>
<td>
<p class="normal">
 267
</p>
</td>
<td>
<p class="normal">
 44
</p>
</td>
<td>
<p class="normal">
 2001
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
 <p class="normal">

 Sponge

 </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Data on sponges; Attributes in spanish </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Clustering
</p>
</td>
<td>
<p class="normal">

```

```

 <p class="normal">
 Categorical, Integer
 </p>
</td>
<td>
<p class="normal">
 76
</p>
</td>
<td>
<p class="normal">
 45
</p>
</td>
<td>
<p class="normal">
 45
</p>
</td>
<td>
<p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Statlog Project

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Various Databases: Vehicle silhouettes, Landsat Sattelite, Shuttle
, Australian Credit Approval, Heart Disease, Image Segmentation, German Credit </p></td> -->
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 1992
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>


```

```

 ^,^
 </td>
 <td>
 <p class="normal">

 Student Loan Relational

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">Student Loan Relational Domain </p></td> -->
<td>
 <p class="normal">
 Domain-Theory
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
 1000
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
 1993
 </p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Teaching Assistant Evaluation

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">The data consist of evaluations of teaching performance; scores
are "low", "medium", or "high" </p></td> -->
 <td>
 <p class="normal">
 Multivariate
 </p>
 </td>
 <td>
 <p class="normal">
 Classification
 </p>
 </td>
</+^>

```

```

<!-->
<p class="normal">
 Categorical, Integer
</p>
</td>
<td>
<p class="normal">
 151
</p>
</td>
<td>
<p class="normal">
 5
</p>
</td>
<td>
<p class="normal">
 1997
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Tic-Tac-Toe Endgame

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">Binary classification task on possible configurations of tic-tac-toe game </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Categorical
</p>
</td>
<td>
<p class="normal">
 958
</p>
</td>
<td>
<p class="normal">
 9
</p>
</td>
<td>
<p class="normal">
 1991
</p>
</td>
<!-- <td><p class="normal">Game </p></td> -->
</tr>
<!-->

```

```

<tr style="border-top: 1px solid black; border-bottom: 1px solid black;">
<td>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Thyroid Disease

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">10 separate databases from Garavan Institute </p></td> -->
<td>
<p class="normal">
 Multivariate, Domain-Theory
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Categorical, Real
</p>
</td>
<td>
<p class="normal">
 7200
</p>
</td>
<td>
<p class="normal">
 21
</p>
</td>
<td>
<p class="normal">
 1987
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Trains

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">2 data formats (structured, one-instance-per-line) </p>
</td> -->
<td>

```

```

<ta>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Classification
 </p>
</td>
<td>
 <p class="normal">
 Categorical
 </p>
</td>
<td>
 <p class="normal">
 10
 </p>
</td>
<td>
 <p class="normal">
 32
 </p>
</td>
<td>
 <p class="normal">
 1994
 </p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 University

 </p>
 </td>
 </tr>
 </table>
</td>
<!-- <td><p class="normal">Data in original (LISP-readable) form </p></td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Classification
 </p>
</td>
<td>
 <p class="normal">
 Categorical, Integer
 </p>
</td>
<td>
 <p class="normal">
 285
 </p>
</td>
<td>
 <p class="normal">
 17
 </p>
<!-- <td><p class="normal">..</p></td> -->

```

```

</p>
</td>
<td>
<p class="normal">
 1988
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Congressional Voting Records

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">1984 United Stated Congressional Voting Records; Classify as
Republican or Democrat </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Categorical
</p>
</td>
<td>
<p class="normal">
 435
</p>
</td>
<td>
<p class="normal">
 16
</p>
</td>
<td>
<p class="normal">
 1987
</p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 ...

</p>

```

```

 Water Treatment Plant

 </p>
 </td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Multiple classes predict plant state </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Clustering
</p>
</td>
<td>
<p class="normal">
 Integer, Real
</p>
</td>
<td>
<p class="normal">
 527
</p>
</td>
<td>
<p class="normal">
 38
</p>
</td>
<td>
<p class="normal">
 1993
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Waveform Database Generator (Version 1)

 </p>
 </td>
</tr>
</table>
</td>
<!-- <td><p class="normal">CART book's waveform domains </p></td> -->
<td>
<p class="normal">
 Multivariate, Data-Generator
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
`.
```

```
</td>
<td>
<p class="normal">
 5000
</p>
</td>
<td>
<p class="normal">
 21
</p>
</td>
<td>
<p class="normal">
 1988
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Waveform Database Generator (Version 2)

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">CART book's waveform domains </p></td> -->
<td>
<p class="normal">
 Multivariate, Data-Generator
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 5000
</p>
</td>
<td>
<p class="normal">
 40
</p>
</td>
<td>
<p class="normal">
 1988
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
```

```


</td>
<td>
 <p class="normal">

 Wine

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

<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Classification
 </p>
</td>
<td>
 <p class="normal">
 Integer, Real
 </p>
</td>
<td>
 <p class="normal">
 178
 </p>
</td>
<td>
 <p class="normal">
 13
 </p>
</td>
<td>
 <p class="normal">
 1991
 </p>
</td>

</tr>
<tr bgcolor="DDEEFF">
<td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Yeast

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

<td>
 <p class="normal">
 Multivariate
 </p>

```

```
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Real
</p>
</td>
<td>
<p class="normal">
1484
</p>
</td>
<td>
<p class="normal">
8
</p>
</td>
<td>
<p class="normal">
1996
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Zoo

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Artificial, 7 classes of animals </p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Categorical, Integer
</p>
</td>
<td>
<p class="normal">
101
</p>
</td>
<td>
<p class="normal">
17
</p>
</td>
<td>
<p class="normal">
```



```

<td>
<p class="normal">
Text
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
20000
</p>
</td>
<td>
<p class="normal">
1999
</p>
</td>
<td>
<p class="normal">
Other
</p>
-->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Australian Sign Language signs

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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 s
amples. </p></td> -->
<td>
<p class="normal">
Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Categorical, Real
</p>
</td>
<td>
<p class="normal">
6650
</p>
</td>
<td>
<p class="normal">
15
</p>

```

```

</td>
<td>
<p class="normal">
 1999
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Australian Sign Language signs (High Quality)

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This data consists of sample of Auslan (Australian Sign
Language) signs. 27 examples of each of 95 Auslan signs were captured from a native signer using h
igh-quality position trackers </p></td> -->
<td>
<p class="normal">
 Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 2565
</p>
</td>
<td>
<p class="normal">
 22
</p>
</td>
<td>
<p class="normal">
 2002
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">


```

```

 US Census Data (1990)

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">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. </p></td> -->
 <td>
 <p class="normal">
 Multivariate
 </p>
 </td>
 <td>
 <p class="normal">
 Clustering
 </p>
 </td>
 <td>
 <p class="normal">
 Categorical
 </p>
 </td>
 <td>
 <p class="normal">
 2458285
 </p>
 </td>
 <td>
 <p class="normal">
 68
 </p>
 </td>
 <td>
 <p class="normal">
 </p>
 </td>
 <!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr>
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Census-Income (KDD)

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">This data set contains weighted census data extracted from the 1
994 and 1995 current population surveys conducted by the U.S. Census Bureau. </p></td> -->
 <td>
 <p class="normal">
 Multivariate
 </p>
 </td>
 <td>
 <p class="normal">
 Classification
 </p>
 </td>
 <td>
 <p class="normal">

```

```

 Categorical, Integer
 </p>
</td>
<td>
 <p class="normal">
 299285
 </p>
</td>
<td>
 <p class="normal">
 40
 </p>
</td>
<td>
 <p class="normal">
 2000
 </p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Coil 1999 Competition Data

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">This data set is from the 1999 Computational Intelligence and
Learning (COIL) competition. The data contains measurements of river chemical concentrations and a
lgae densities. </p></td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Categorical, Real
 </p>
</td>
<td>
 <p class="normal">
 340
 </p>
</td>
<td>
 <p class="normal">
 17
 </p>
</td>
<td>
 <p class="normal">
 1999
 </p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
<td>

```

```

<!-->
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Corel Image Features

</p>
</td>
</tr>
</table>
</td>
<!--<td><p class="normal">This dataset contains image features extracted from a Corel
image collection. Four sets of features are available based on the color histogram, color
histogram layout, color moments, and co-occurrence </p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
Real
</p>
</td>
<td>
<p class="normal">
68040
</p>
</td>
<td>
<p class="normal">
89
</p>
</td>
<td>
<p class="normal">
1999
</p>
</td>
<!--<td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

E. Coli Genes

</p>
</td>
</tr>
</table>
</td>
<!--<td><p class="normal">Data giving characteristics of each ORF (potential gene) in the
E. coli genome. Sequence, homology (similarity to other genes) and structural information, and fun
ction (if known) are provided. </p></td> -->

```

```
<td>
<p class="normal">
 Relational
</p>
</td>
<td>
<p class="normal">
 2001
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 EEG Database

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">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 sc
alp sampled at 256 Hz </p></td> -->
<td>
<p class="normal">
 Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 Categorical, Integer, Real
</p>
</td>
<td>
<p class="normal">
 122
</p>
</td>
<td>
<p class="normal">
 4
</p>
</td>
<td>
```

```

<!--
<p class="normal">
 1999
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 El Niño

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The data set contains oceanographic and surface meteorological
readings taken from a series of buoys positioned throughout the equatorial Pacific. </p></td>
-->
<td>
<p class="normal">
 Spatio-temporal
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 Integer, Real
</p>
</td>
<td>
<p class="normal">
 178080
</p>
</td>
<td>
<p class="normal">
 12
</p>
</td>
<td>
<p class="normal">
 1999
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Entree Chicago Recommendation Data

</p>
</td>

```

```

</td>
<td>
<p class="normal">This data contains a record of user interactions with the Entree
Chicago restaurant recommendation system. </p></td> -->
<td>
<p class="normal">
 Transactional, Sequential
</p>
</td>
<td>
<p class="normal">
 Recommender-Systems
</p>
</td>
<td>
<p class="normal">
 Categorical
</p>
</td>
<td>
<p class="normal">
 50672
</p>
</td>
<td>
<p class="normal">
 2000
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 CMU Face Images

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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 </p></td> -->
<td>
<p class="normal">
 Image
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Integer
</p>
</td>

```

```

~,~

<td>

<p class="normal">

 640

</p>

</td>

<td>

<p class="normal">

</p>

</td>

<td>

<p class="normal">

 1999

</p>

</td>

<!-- <td><p class="normal">Other </p></td> -->

</tr>

<tr>

<td>

<table>

<tr>

<td>

</td>

<td>

<p class="normal">

 Insurance Company Benchmark (COIL 2000)

</p>

</td>

</tr>

</table>

</td>

<!-- <td><p class="normal">This data set used in the CoIL 2000 Challenge contains

information on customers of an insurance company. The data consists of 86 variables and includes p

roduct usage data and socio-demographic data </p></td> -->

<td>

<p class="normal">

 Multivariate

</p>

</td>

<td>

<p class="normal">

 Regression, Description

</p>

</td>

<td>

<p class="normal">

 Categorical, Integer

</p>

</td>

<td>

<p class="normal">

 9000

</p>

</td>

<td>

<p class="normal">

 86

</p>

</td>

<td>

<p class="normal">

 2000

</p>

</td>

<!-- <td><p class="normal">Social </p></td> -->

</tr>

<tr bgcolor="DDEEFF">

<td>

<table>

<tr>

~,~
```

```

<!-->

</td>
<td>
 <p class="normal">

 Internet Usage Data

 </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This data contains general demographic information on internet u
users in 1997. </p></td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Categorical, Integer
 </p>
</td>
<td>
 <p class="normal">
 10104
 </p>
</td>
<td>
 <p class="normal">
 72
 </p>
</td>
<td>
 <p class="normal">
 1999
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 IPUMS Census Database

 </p>
 </td>
 </tr>
 </table>
</td>
<!-- <td><p class="normal">This data set contains unweighted PUMS census data from the Los
Angeles and Long Beach areas for the years 1970, 1980, and 1990. </p></td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>

```

```

</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 Categorical, Integer
</p>
</td>
<td>
<p class="normal">
 256932
</p>
</td>
<td>
<p class="normal">
 61
</p>
</td>
<td>
<p class="normal">
 1999
</p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Japanese Vowels

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset records 640 time series of 12 LPC cepstrum
coefficients taken from nine male speakers. </p></td> -->
<td>
<p class="normal">
 Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 640
</p>
</td>
<td>
<p class="normal">
 12
</p>
</td>
<td>
<p class="normal">
 ..
</p>

```

```

</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 KDD Cup 1998 Data

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This is the data set used for The Second International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-98 </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Regression
</p>
</td>
<td>
<p class="normal">
 Categorical, Integer
</p>
</td>
<td>
<p class="normal">
 191779
</p>
</td>
<td>
<p class="normal">
 481
</p>
</td>
<td>
<p class="normal">
 1998
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 KDD Cup 1999 Data

</p>
</td>

```



```
<p class="normal">
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
2001
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Movie

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This data set contains a list of over 10000 films including many
older, odd, and cult films. There is information on actors, casts, directors, producers, studios,
etc. </p></td> -->
<td>
<p class="normal">
Multivariate, Relational
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
10000
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
1999
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
</td>
```

```

<p class="normal">

 MSNBC.com Anonymous Web Data

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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. </p></td> -->
<td>
<p class="normal">
 Sequential
</p>
</td>
<td>
<p class="normal">
 Categorical
</p>
</td>
<td>
<p class="normal">
 989818
</p>
</td>
<td>
<p class="normal">
 Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 NSF Research Award Abstracts 1990-2003

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This data set consists of (a) 129,000 abstracts describing NSF a
wards for basic research, (b) bag-of-word data files extracted from the abstracts, (c) a list of w
ords used for indexing the bag-of-word </p></td> -->
<td>
<p class="normal">
 Text
</p>
</td>
<td>
<p class="normal">
 </p>
</td>
<td>

```

```

<p class="normal">
</p>
</td>
<td>
<p class="normal">
129000
</p>
</td>
<td>
<p class="normal">
2003
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Pioneer-1 Mobile Robot Data

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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></td> -->
<td>
<p class="normal">
Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
Categorical, Real
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
1999
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

```

```


</td>
<td>
 <p class="normal">

 Pseudo Periodic Synthetic Time Series

 </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This data set is designed for testing indexing schemes in time s
eries databases. The data appears highly periodic, but never exactly repeats itself. </p></td>
> -->
<td>
 <p class="normal">
 Univariate, Time-Series
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
 100000
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
 1999
 </p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Reuters-21578 Text Categorization Collection

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

```

```
<p class="normal">
 Classification
</p>
</td>
<td>
 <p class="normal">
 Categorical
 </p>
</td>
<td>
 <p class="normal">
 21578
 </p>
</td>
<td>
 <p class="normal">
 5
 </p>
</td>
<td>
 <p class="normal">
 1997
 </p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Robot Execution Failures

 </p>
 </td>
 </tr>
 </table>
</td>
<!-- <td><p class="normal">This dataset contains force and torque measurements on a robot
after failure detection. Each failure is characterized by 15 force/torque samples collected at
regular time intervals </p></td> -->
<td>
 <p class="normal">
 Multivariate, Time-Series
 </p>
</td>
<td>
 <p class="normal">
 Classification
 </p>
</td>
<td>
 <p class="normal">
 Integer
 </p>
</td>
<td>
 <p class="normal">
 463
 </p>
</td>
<td>
 <p class="normal">
 90
 </p>
</td>
<td>
 <p class="normal">
```

```

1999
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Synthetic Control Chart Time Series

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This data consists of synthetically generated control
charts. </p></td> -->
<td>
<p class="normal">
 Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification, Clustering
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 600
</p>
</td>
<td>
<p class="normal">
 1999
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Syskill and Webert Web Page Ratings

</p>
</td>

```

```

 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">This database contains HTML source of web pages plus the ratings
of a single user on these web pages. Web pages are on four seperate subjects (Bands- recording art
ists; Goats; Sheep; and BioMedical) </p></td> -->
 <td>
 <p class="normal">
 Multivariate, Text
 </p>
 </td>
 <td>
 <p class="normal">
 Classification
 </p>
 </td>
 <td>
 <p class="normal">
 Categorical
 </p>
 </td>
 <td>
 <p class="normal">
 332
 </p>
 </td>
 <td>
 <p class="normal">
 5
 </p>
 </td>
 <td>
 <p class="normal">
 1998
 </p>
 </td>
 <!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 UNIX User Data

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">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. </p></
td> -->
 <td>
 <p class="normal">
 Text, Sequential
 </p>
 </td>
 <td>
 <p class="normal">
 </p>
 </td>
 <td>
 <p class="normal">
 </p>
 </td>
 <td>
 <p class="normal">
 </p>
 </td>

```

```

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

</td>
<td>
<p class="normal">

Volcanoes on Venus - JARtool experiment

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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. </p></td> -->
<td>
<p class="normal">
Image
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Physical
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">


```

```

 Statlog (Australian Credit Approval)

 </p>
 </td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This file concerns credit card applications. This database
exists elsewhere in the repository (Credit Screening Database) in a slightly different form </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Categorical, Integer, Real
</p>
</td>
<td>
<p class="normal">
 690
</p>
</td>
<td>
<p class="normal">
 14
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<!-- <td><p class="normal">Financial </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Statlog (German Credit Data)

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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
 </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
</td>

```

```

 ...
<p class="normal">
 Categorical, Integer
</p>
</td>
<td>
<p class="normal">
 1000
</p>
</td>
<td>
<p class="normal">
 20
</p>
</td>
<td>
<p class="normal">
 1994
</p>
</td>
<!-- <td><p class="normal">Financial </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Statlog (Heart)

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset is a heart disease database similar to a database
already present in the repository (Heart Disease databases) but in a slightly different form -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Categorical, Real
</p>
</td>
<td>
<p class="normal">
 270
</p>
</td>
<td>
<p class="normal">
 13
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">

```

```

<-- ~~~~~
 -->
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Statlog (Landsat Satellite)

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Multi-spectral values of pixels in 3x3 neighbourhoods in a
satellite image, and the classification associated with the central pixel in each
neighbourhood
</p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Integer
</p>
</td>
<td>
<p class="normal">
 6435
</p>
</td>
<td>
<p class="normal">
 36
</p>
</td>
<td>
<p class="normal">
 1993
</p>
</td>
<!-- <td><p class="normal">Physical
</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Statlog (Image Segmentation)

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset is an image segmentation database similar to a dataset

```

.. ~~<td> class="normal"></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.  
`rm. </p></td> -->  
<td>  
<p class="normal">  
Multivariate  
</p>  
</td>  
<td>  
<p class="normal">  
Classification  
</p>  
</td>  
<td>  
<p class="normal">  
Real  
</p>  
</td>  
<td>  
<p class="normal">  
2310  
</p>  
</td>  
<td>  
<p class="normal">  
19  
</p>  
</td>  
<td>  
<p class="normal">  
1990  
</p>  
</td>  
<!-- <td><p class="normal">Other </p></td> -->  
</tr>  
<tr bgcolor="DDEEFF">  
<td>  
<table>  
<tr>  
<td>  
<a href="datasets/Statlog%28Shuttle%29">  
  
</a>  
</td>  
<td>  
<p class="normal">  
<b>  
<a href="datasets/Statlog%28Shuttle%29">  
Statlog (Shuttle)  
</a>  
</b>  
</p>  
</td>  
</tr>  
</table>  
</td>  
<!-- <td><p class="normal">The shuttle dataset contains 9 attributes all of which are numerical. Approximately 80% of the data belongs to class 1 </p></td> -->  
<td>  
<p class="normal">  
Multivariate  
</p>  
</td>  
<td>  
<p class="normal">  
Classification  
</p>  
</td>  
<td>  
<p class="normal">  
Integer  
</p>  
</td>  
<td>  
<p class="normal">  
58000  
</p>  
</td>

```

~,~

<td>

<p class="normal">

 9

</p>

</td>

<td>

<p class="normal">

</p>

</td>

<!-- <td><p class="normal">Physical </p></td> -->

</tr>

<tr>

<td>

<table>

<tr>

<td>

</td>

<td>

<p class="normal">

 Statlog (Vehicle Silhouettes)

</p>

</td>

</tr>

</table>

</td>

<!-- <td><p class="normal">3D objects within a 2D image by application of an ensemble of

shape feature extractors to the 2D silhouettes of the objects. </p></td> -->

<td>

<p class="normal">

 Multivariate

</p>

</td>

<td>

<p class="normal">

 Classification

</p>

</td>

<td>

<p class="normal">

 Integer

</p>

</td>

<td>

<p class="normal">

 946

</p>

</td>

<td>

<p class="normal">

 18

</p>

</td>

<td>

<p class="normal">

</p>

</td>

<!-- <td><p class="normal">Other </p></td> -->

</tr>

<tr bgcolor="DDEEFF">

<td>

<table>

<tr>

<td>

</td>

<td>

<p class="normal">

 ~

```

```

<tr>

 Connectionist Bench (Nettalk Corpus)

</td>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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 p
roper phonemes&nbsp.</p></td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
 Categorical
 </p>
</td>
<td>
 <p class="normal">
 20008
 </p>
</td>
<td>
 <p class="normal">
 4
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">Other&nbsp.</p></td> -->
</tr>
<tr>
<td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Connectionist Bench (Sonar, Mines vs. Rocks)

 </p>
 </td>
 </tr>
 </table>
</td>
<td>
 <p class="normal">The task is to train a network to discriminate between sonar sig
als bounced off a metal cylinder and those bounced off a roughly cylindrical rock.&nbsp.</p></td>
-->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Classification
 </p>
</td>
<td>

```

```

<ta>
 <p class="normal">
 Real
 </p>
</td>
<td>
 <p class="normal">
 208
 </p>
</td>
<td>
 <p class="normal">
 60
 </p>
</td>
<td>
 <p class="normal">
 Physical
 </p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Connectionist Bench (Vowel Recognition - Deterding Data)

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">Speaker independent recognition of the eleven steady state
vowels of British English using a specified training set of lpc derived log area ratios.&nbsp</p>
</td> -->
 <td>
 <p class="normal">
 </p>
 </td>
 <td>
 <p class="normal">
 Classification
 </p>
 </td>
 <td>
 <p class="normal">
 Real
 </p>
 </td>
 <td>
 <p class="normal">
 528
 </p>
 </td>
 <td>
 <p class="normal">
 10
 </p>
 </td>
 <td>
 <p class="normal">
 Other
 </p>
 </td>
 <!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
 <td>
 ...
 </td>

```



```
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">>Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Cloud

</p>
</td>
</tr>
</table>
</td>
<td>
<p class="normal">>Little Documentation </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 1024
</p>
</td>
<td>
<p class="normal">
 10
</p>
</td>
<td>
<p class="normal">
 1989
</p>
</td>
<td>
<p class="normal">>Physical </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
```

```


</td>
<td>
 <p class="normal">

 CalIt2 Building People Counts

 </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This data comes from the main door of the CalIt2 building at
UCI. </p></td> -->
<td>
 <p class="normal">
 Multivariate, Time-Series
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
 Categorical, Integer
 </p>
</td>
<td>
 <p class="normal">
 10080
 </p>
</td>
<td>
 <p class="normal">
 4
 </p>
</td>
<td>
 <p class="normal">
 2006
 </p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Dodgers Loop Sensor

 </p>
 </td>
 </tr>
 </table>
</td>
<!-- <td><p class="normal">Loop sensor data was collected for the Glendale on ramp for the
101 North freeway in Los Angeles </p></td> -->
<td>
 <p class="normal">
 Multivariate, Time-Series
 </p>
</td>

```

```
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 Categorical, Integer
</p>
</td>
<td>
<p class="normal">
 50400
</p>
</td>
<td>
<p class="normal">
 3
</p>
</td>
<td>
<p class="normal">
 2006
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Poker Hand

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Purpose is to predict poker hands </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Categorical, Integer
</p>
</td>
<td>
<p class="normal">
 1025010
</p>
</td>
<td>
<p class="normal">
 11
</p>
</td>
<td>
<p class="normal">
 2007
</p>
```

```

</td>
<!-- <td><p class="normal">Game </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

MAGIC Gamma Telescope

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Data are MC generated to simulate registration of high energy
gamma particles in an atmospheric Cherenkov telescope </p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Real
</p>
</td>
<td>
<p class="normal">
19020
</p>
</td>
<td>
<p class="normal">
11
</p>
</td>
<td>
<p class="normal">
2007
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

UJI Pen Characters

</p>
</td>
</tr>

```

```

 </table>
 </td>
 <!-- <td><p class="normal">Data consists of written characters in a UNIPEN-like
format </p></td> -->
 <td>
 <p class="normal">
 Multivariate, Sequential
 </p>
 </td>
 <td>
 <p class="normal">
 Classification
 </p>
 </td>
 <td>
 <p class="normal">
 Integer
 </p>
 </td>
 <td>
 <p class="normal">
 1364
 </p>
 </td>
 <td>
 <p class="normal">
 2007
 </p>
 </td>
 <!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Mammographic Mass

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">Discrimination of benign and malignant mammographic masses based
on BI-RADS attributes and the patient's age. </p></td> -->
 <td>
 <p class="normal">
 Multivariate
 </p>
 </td>
 <td>
 <p class="normal">
 Classification
 </p>
 </td>
 <td>
 <p class="normal">
 Integer
 </p>
 </td>
 <td>
 <p class="normal">
 961
 </p>
 </td>

```

```

</td>
<td>
<p class="normal">
 6
</p>
</td>
<td>
<p class="normal">
 2007
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Forest Fires

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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). </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Regression
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 517
</p>
</td>
<td>
<p class="normal">
 13
</p>
</td>
<td>
<p class="normal">
 2008
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>


```

```

</td>
<td>
<p class="normal">

 Reuters Transcribed Subset

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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.&nbsp</p></td> -->
<td>
<p class="normal">
 Text
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 200
</p>
</td>
<td>
<p class="normal">
 2008
</p>
</td>
<!-- <td><p class="normal">Business&nbsp</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Bag of Words

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This data set contains five text collections in the form of
bags-of-words.&nbsp</p></td> -->
<td>
<p class="normal">
 Text
</p>
</td>
<td>
<p class="normal">

```

```

 - Clustering
 </p>
</td>
<td>
<p class="normal">
 Integer
</p>
</td>
<td>
<p class="normal">
 8000000
</p>
</td>
<td>
<p class="normal">
 100000
</p>
</td>
<td>
<p class="normal">
 2008
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Concrete Compressive Strength

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">Concrete is the most important material in civil engineering.
The concrete compressive strength is a highly nonlinear function of age and ingredients. </p>
-->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Regression
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 1030
</p>
</td>
<td>
<p class="normal">
 9
</p>
</td>
<td>
<p class="normal">
 2007
</p>

```

```

</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Hill-Valley

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Each record represents 100 points on a two-dimensional graph.
When plotted in order (from 1 through 100) as the Y co-ordinate, the points will create either a H
ill (a bump in the terrain) or a Valley (a dip in the terrain). </p></td> -->
<td>
<p class="normal">
 Sequential
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 606
</p>
</td>
<td>
<p class="normal">
 101
</p>
</td>
<td>
<p class="normal">
 2008
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Arcene

</p>
</td>

```

```

 . .
 </td>
</tr>
</table>
</td>
<!-- <td><p class="normal">ARCENE's task is to distinguish cancer versus normal patterns
from mass-spectrometric data. This is a two-class classification problem with continuous input
variables. This dataset is one of 5 datasets of the NIPS 2003 feature selection challenge. </p></td> -->
 <td>
 <p class="normal">
 Multivariate
 </p>
 </td>
 <td>
 <p class="normal">
 Classification
 </p>
 </td>
 <td>
 <p class="normal">
 Real
 </p>
 </td>
 <td>
 <p class="normal">
 900
 </p>
 </td>
 <td>
 <p class="normal">
 10000
 </p>
 </td>
 <td>
 <p class="normal">
 2008
 </p>
 </td>
 <!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Dexter

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">DEXTER is a text classification problem in a bag-of-word
representation. This is a two-class classification problem with sparse continuous input variables.
This dataset is one of five datasets of the NIPS 2003 feature selection challenge.
 </p></td> -->
 <td>
 <p class="normal">
 Multivariate
 </p>
 </td>
 <td>
 <p class="normal">
 Classification
 </p>
 </td>
 <td>
 <p class="normal">

```

```

<td>
 Integer
</p>
</td>
<td>
<p class="normal">
 2600
</p>
</td>
<td>
<p class="normal">
 20000
</p>
</td>
<td>
<p class="normal">
 2008
</p>
</td>
<td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Dorothea

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">DOROTHEA is a drug discovery dataset. Chemical compounds
represented by structural molecular features must be classified as active (binding to thrombin) or
inactive. This is one of 5 datasets of the NIPS 2003 feature selection challenge. </p></td> -
-->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Integer
</p>
</td>
<td>
<p class="normal">
 1950
</p>
</td>
<td>
<p class="normal">
 100000
</p>
</td>
<td>
<p class="normal">
 2008
</p>
</td>
<td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>

```

```

<!-- -->
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Gisette

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">GISETTE is a handwritten digit recognition problem. The problem
is to separate the highly confusable digits '4' and '9'. This dataset is one of five datasets of t
he NIPS 2003 feature selection challenge.
 </p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Integer
</p>
</td>
<td>
<p class="normal">
13500
</p>
</td>
<td>
<p class="normal">
5000
</p>
</td>
<td>
<p class="normal">
2008
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Madelon

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

```

```

 >/> <td>
</td>
<!-- <td><p class="normal">MADELON is an artificial dataset, which was part of the NIPS
2003 feature selection challenge. This is a two-class classification problem with continuous input
variables. The difficulty is that the problem is multivariate and highly non-linear. </p></t
d> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 4400
</p>
</td>
<td>
<p class="normal">
 500
</p>
</td>
<td>
<p class="normal">
 2008
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Ozone Level Detection

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Two ground ozone level data sets are included in this
collection. One is the eight hour peak set (eighthr.data), the other is the one hour peak set (one
hr.data). Those data were collected from 1998 to 2004 at the Houston, Galveston and Brazoria area.
&nbs;p;</p></td> -->
<td>
<p class="normal">
 Multivariate, Sequential, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
</tr>

```

```

~,~

<td>

<p class="normal">

 2536

</p>

</td>

<td>

<p class="normal">

 73

</p>

</td>

<td>

<p class="normal">

 2008

</p>

</td>

<!-- <td><p class="normal">Physical </p></td> -->

</tr>

<tr>

<td>

<table>

<tr>

<td>

</td>

<td>

<p class="normal">

 Abscisic Acid Signaling Network

</p>

</td>

</tr>

</table>

</td>

<!-- <td><p class="normal">The objective is to determine the set of boolean rules that descr

ibe the interactions of the nodes within this plant signaling network. The dataset includes 300 s

eparate boolean pseudodynamic simulations using an asynchronous update scheme. </p></td> -->

<td>

<p class="normal">

 Multivariate

</p>

</td>

<td>

<p class="normal">

 Causal-Discovery

</p>

</td>

<td>

<p class="normal">

 Integer

</p>

</td>

<td>

<p class="normal">

 300

</p>

</td>

<td>

<p class="normal">

 43

</p>

</td>

<td>

<p class="normal">

 2008

</p>

</td>

<!-- <td><p class="normal">Life </p></td> -->

</tr>

<tr bgcolor="DDEEFF">

<td>

<table>

~,~
```

```

<tr>
<td>

</td>
<td>
<p class="normal">

Parkinsons

</p>
</td>
</tr>
</table>
</td>
<!!-- <td><p class="normal">Oxford Parkinson's Disease Detection Dataset </p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Real
</p>
</td>
<td>
<p class="normal">
197
</p>
</td>
<td>
<p class="normal">
23
</p>
</td>
<td>
<p class="normal">
2008
</p>
</td>
<!!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Character Trajectories

</p>
</td>
</tr>
</table>
</td>
<!!-- <td><p class="normal">Multiple, labelled samples of pen tip trajectories recorded
whilst writing individual characters. All samples are from the same writer, for the purposes of pr
imitive extraction. Only characters with a single pen-down segment were considered. </p></td>
-->
<td>


```

```
<p class="normal">
 Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification, Clustering
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 2858
</p>
</td>
<td>
<p class="normal">
 3
</p>
</td>
<td>
<p class="normal">
 2008
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
 <p class="normal">

 Blood Transfusion Service Center

 </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Data taken from the Blood Transfusion Service Center in Hsin-Chu
City in Taiwan -- this is a classification problem. </p></td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 748
</p>
</td>
<td>
<p class="normal">
 5
</p>
</td>
<td>
 ..
 ..

```

```

</p>
</td>
<td>
<p class="normal">
 2008
</p>
</td>
<!-- <td><p class="normal">Business </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 UJI Pen Characters (Version 2)

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">A pen-based database with more than 11k isolated handwritten cha-
racters </p></td> -->
<td>
<p class="normal">
 Multivariate, Sequential
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Integer
</p>
</td>
<td>
<p class="normal">
 11640
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 2009
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Semeion Handwritten Digit

</p>

```

```


 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">1593 handwritten digits from around 80 persons were scanned,
stretched in a rectangular box 16x16 in a gray scale of 256 values. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Integer
</p>
</td>
<td>
<p class="normal">
 1593
</p>
</td>
<td>
<p class="normal">
 256
</p>
</td>
<td>
<p class="normal">
 2008
</p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
 <td>

 </td>
 <td>
 <p class="normal">

 SECOM

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">Data from a semi-conductor manufacturing process&nbsp;</p></td>
-->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification, Causal-Discovery
</p>
</td>
<td>
<p class="normal">
 Real
 .
</pre>

```

```
</p>
</td>
<td>
<p class="normal">
 1567
</p>
</td>
<td>
<p class="normal">
 591
</p>
</td>
<td>
<p class="normal">
 2008
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Plants

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Data has been extracted from the USDA plants database. It
contains all plants (species and genera) in the database and the states of USA and Canada where th
ey occur. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Clustering
</p>
</td>
<td>
<p class="normal">
 Categorical
</p>
</td>
<td>
<p class="normal">
 22632
</p>
</td>
<td>
<p class="normal">
 70
</p>
</td>
<td>
<p class="normal">
 2008
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
```

```

<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Libras Movement

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The data set contains 15 classes of 24 instances each. Each class references to a hand movement type in LIBRAS (Portuguese name 'Língua Brasileira de Sinais', oficial brazilian signal language). </p></td> -->
<td>
<p class="normal">
 Multivariate, Sequential
</p>
</td>
<td>
<p class="normal">
 Classification, Clustering
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 360
</p>
</td>
<td>
<p class="normal">
 91
</p>
</td>
<td>
<p class="normal">
 2009
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Concrete Slump Test

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Concrete is a highly complex material. The slump flow of concrete is not only determined by the water content, but that is also influenced by other

```

```
concrete ingredients. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Regression
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 103
</p>
</td>
<td>
<p class="normal">
 10
</p>
</td>
<td>
<p class="normal">
 2009
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Communities and Crime

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Communities within the United States. The data combines socio-economic data from the 1990 US Census, law enforcement data from the 1990 US LEMAS survey, and crime data from the 1995 FBI UCR. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Regression
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 1994
</p>
</td>
```

```

<td>
<p class="normal">
128
</p>
</td>
<td>
<p class="normal">
2009
</p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Acute Inflammations

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The data was created by a medical expert as a data set to test
the expert system,
which will perform the presumptive diagnosis of two diseases of the urinary system.
 </p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Categorical, Integer
</p>
</td>
<td>
<p class="normal">
120
</p>
</td>
<td>
<p class="normal">
6
</p>
</td>
<td>
<p class="normal">
2009
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>


```

```

</td>
<td>
<p class="normal">

 Wine Quality

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Two datasets are included, related to red and white vinho verde
wine samples, from the north of Portugal. The goal is to model wine quality based on
physicochemical tests (see [Cortez et al., 2009], http://www3.dsi.uminho.pt/pcortez/wine/). <
/p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification, Regression
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 4898
</p>
</td>
<td>
<p class="normal">
 12
</p>
</td>
<td>
<p class="normal">
 2009
</p>
</td>
<!-- <td><p class="normal">Business </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 URL Reputation

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Anonymized 120-day subset of the ICML-09 URL data containing 2.4
million examples and 3.2 million features. </p></td> -->
<td>
<p class="normal">
 Multivariate, Time-Series
</p>
</td>

```

```
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Integer, Real
</p>
</td>
<td>
<p class="normal">
2396130
</p>
</td>
<td>
<p class="normal">
3231961
</p>
</td>
<td>
<p class="normal">
2009
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

p53 Mutants

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The goal is to model mutant p53 transcriptional activity (active vs inactive) based on data extracted from biophysical simulations.
 </p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Real
</p>
</td>
<td>
<p class="normal">
16772
</p>
</td>
<td>
<p class="normal">
5409
</p>
</td>
<td>
```

```

<p class="normal">
 2010
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
 <p class="normal">

 Parkinsons Telemonitoring

 </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Oxford Parkinson's Disease Telemonitoring Dataset </p></td>
-->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Regression
</p>
</td>
<td>
<p class="normal">
 Integer, Real
</p>
</td>
<td>
<p class="normal">
 5875
</p>
</td>
<td>
<p class="normal">
 26
</p>
</td>
<td>
<p class="normal">
 2009
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
 <p class="normal">

 Demospongiae

 </p>
</td>

```

```

 </p>
 </td>
 </tr>
 </table>
<td>
 <!-- <td><p class="normal">Marine sponges of the Demospongiae class classification
domain. </p></td> -->
 <td>
 <p class="normal">
 Multivariate
 </p>
 </td>
 <td>
 <p class="normal">
 Classification
 </p>
 </td>
 <td>
 <p class="normal">
 Integer
 </p>
 </td>
 <td>
 <p class="normal">
 503
 </p>
 </td>
 <td>
 <p class="normal">
 2010
 </p>
 </td>
 <!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

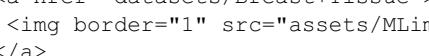
 Opinosis Opinion ✓ Review

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p 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
bsp;</p></td> -->
 <td>
 <p class="normal">
 Text
 </p>
 </td>
 <td>
 <p class="normal">
 </p>
 </td>

```

```

51
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
2010
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>



</td>
<td>
<p class="normal">

Breast Tissue

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Dataset with electrical impedance measurements of freshly excised tissue samples from the breast. </p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Real
</p>
</td>
<td>
<p class="normal">
106
</p>
</td>
<td>
<p class="normal">
10
</p>
</td>
<td>
<p class="normal">
2010
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>




```

```

 .. ~.
 </td>
 <td>
 <p class="normal">

 Cardiotocography

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <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. </p></td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Classification
 </p>
</td>
<td>
 <p class="normal">
 Real
 </p>
</td>
<td>
 <p class="normal">
 2126
 </p>
</td>
<td>
 <p class="normal">
 23
 </p>
</td>
<td>
 <p class="normal">
 2010
 </p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Wall-Following Robot Navigation Data

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">The data were collected as the SCITOS G5 robot navigates through the room following the wall in a clockwise direction, for 4 rounds, using 24 ultrasound sensors arranged circularly around its 'waist'. </p></td> -->
 <td>
 <p class="normal">
 Multivariate, Sequential
 </p>
 </td>

```

```

., ..
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 5456
</p>
</td>
<td>
<p class="normal">
 24
</p>
</td>
<td>
<p class="normal">
 2010
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Spoken Arabic Digit

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset contains timeseries of mel-frequency cepstrum
coefficients (MFCCs) corresponding to spoken Arabic digits. Includes data from 44 male and 44 fem
ale native Arabic speakers. </p></td> -->
<td>
<p class="normal">
 Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 8800
</p>
</td>
<td>
<p class="normal">
 13
</p>
</td>
<!-->

```

```

<!--
<p class="normal">
 2010
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Localization Data for Person Activity

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Data contains recordings of five people performing different
activities. Each person wore four sensors (tags) while performing the same scenario five times. &n
bsp;</p></td> -->
<td>
<p class="normal">
 Univariate, Sequential, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 164860
</p>
</td>
<td>
<p class="normal">
 8
</p>
</td>
<td>
<p class="normal">
 2010
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 AutoUniv

</p>

```

```

 ~ / a ~

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Categorical, Integer, Real
</p>
</td>
<td>
<p class="normal">
 2010
</p>
</td>
<td>
<p class="normal">

</td>
<td>
<p class="normal">

 Steel Plates Faults

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">A dataset of steel plates' faults, classified into 7 different ty
pes.
The goal was to train machine learning for automatic pattern recognition.
 </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Tnt+~~~~ Poo~1

```

```

 integer, real
 </p>
</td>
<td>
<p class="normal">
 1941
</p>
</td>
<td>
<p class="normal">
 27
</p>
</td>
<td>
<p class="normal">
 2010
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 MiniBooNE particle identification

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset is taken from the MiniBooNE experiment and is used
to distinguish electron neutrinos (signal) from muon neutrinos (background). </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 130065
</p>
</td>
<td>
<p class="normal">
 50
</p>
</td>
<td>
<p class="normal">
 2010
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
<td>

```

```

<table>
<tr>
<td>

</td>
<td>
<p class="normal">

YearPredictionMSD

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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 200
0s. </p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Regression
</p>
</td>
<td>
<p class="normal">
Real
</p>
</td>
<td>
<p class="normal">
515345
</p>
</td>
<td>
<p class="normal">
90
</p>
</td>
<td>
<p class="normal">
2011
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

PEMS-SF

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">15 months worth of daily data (440 daily records) that describes
the occupancy rate, between 0 and 1, of different car lanes of the San Francisco bay area freeways
-->

```



```

</td>
<td>
<p class="normal">
2011
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Relative location of CT slices on axial axis

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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. </p></td> -->
<td>
<p class="normal">
Domain-Theory
</p>
</td>
<td>
<p class="normal">
Regression
</p>
</td>
<td>
<p class="normal">
Real
</p>
</td>
<td>
<p class="normal">
53500
</p>
</td>
<td>
<p class="normal">
386
</p>
</td>
<td>
<p class="normal">
2011
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">


```

Online Handwritten Assamese Characters Dataset

</a>

</b>

</p>

</td>

</tr>

</table>

</td>

<!-- <td><p class="normal">This is a dataset of 8235 online handwritten assamese characters. The "online" process involves capturing of data as text is written on a digitizing tab let with an electronic pen.&nbsp;</p></td> -->

<td>

<p class="normal">

Multivariate, Sequential

</p>

</td>

<td>

<p class="normal">

Classification

</p>

</td>

<td>

<p class="normal">

Integer

</p>

</td>

<td>

<p class="normal">

8235

</p>

</td>

<td>

<p class="normal">

2011

</p>

</td>

<!-- <td><p class="normal">Computer&nbsp;</p></td> -->

</tr>

<tr bgcolor="#DDEEFF">

<td>

<table>

<tr>

<td>

<a href="datasets/PubChem+Bioassay+Data">



</a>

</td>

<td>

<p class="normal">

<b>

<a href="datasets/PubChem+Bioassay+Data">

PubChem Bioassay Data

</a>

</b>

</p>

</td>

</tr>

</table>

</td>

<!-- <td><p class="normal">These highly imbalanced bioassay datasets are from the differing types of screening that can be performed using HTS technology. 21 datasets were created from 12 bi oassays.&nbsp;</p></td> -->

<td>

<p class="normal">

Multivariate

</p>

</td>

<td>

<p class="normal">

Classification

</p>

</td>

<td>

```

<p class="normal">
 Integer, Real
</p>
</td>
<td>
 <p class="normal">
 </p>
 </td>
 <td>
 <p class="normal">
 </p>
 </td>
 <td>
 <p class="normal">
 2011
 </p>
 </td>
 <!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Record Linkage Comparison Patterns

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">Element-wise comparison of records with personal data from a
record linkage setting. The task is to decide from a comparison pattern whether the underlying rec
ords belong to one person. </p></td> -->
 <td>
 <p class="normal">
 Multivariate
 </p>
 </td>
 <td>
 <p class="normal">
 Classification
 </p>
 </td>
 <td>
 <p class="normal">
 Real
 </p>
 </td>
 <td>
 <p class="normal">
 5749132
 </p>
 </td>
 <td>
 <p class="normal">
 12
 </p>
 </td>
 <td>
 <p class="normal">
 2011
 </p>
 </td>
 <!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>

```

```

<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Communities and Crime Unnormalized

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Communities in the US. Data combines socio-economic data from th
e '90 Census, law enforcement data from the 1990 Law Enforcement Management and Admin Stats survey
, and crime data from the 1995 FBI UCR </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Regression
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 2215
</p>
</td>
<td>
<p class="normal">
 147
</p>
</td>
<td>
<p class="normal">
 2011
</p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Vertebral Column

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Data set containing values for six biomechanical features used t
o classify orthopaedic patients into 3 classes (normal, disk hernia or spondilolisthesis) or 2 cla

```

```

sses (normal or abnormal). </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 310
</p>
</td>
<td>
<p class="normal">
 6
</p>
</td>
<td>
<p class="normal">
 2011
</p>
</td>
<!-- <td><p class="normal"> </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 EMG Physical Action Data Set

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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 sub
jects using the Delsys EMG wireless apparatus. </p></td> -->
<td>
<p class="normal">
 Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 10000
</p>
</td>

```

```

<td>
<p class="normal">
 8
</p>
</td>
<td>
<p class="normal">
 2011
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Vicon Physical Action Data Set

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The Physical Action Data Set includes 10 normal and 10
aggressive physical actions that measure the human activity. The data have been collected by 10 su
bjects using the Vicon 3D tracker. </p></td> -->
<td>
<p class="normal">
 Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 3000
</p>
</td>
<td>
<p class="normal">
 27
</p>
</td>
<td>
<p class="normal">
 2011
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>

```

```

<td>
<p class="normal">

 Amazon Commerce reviews set

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The dataset is used for authorship identification in online Write
print which is a new research field of pattern recognition. </p></td> -->
<td>
<p class="normal">
 Multivariate, Text, Domain-Theory
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 1500
</p>
</td>
<td>
<p class="normal">
 10000
</p>
</td>
<td>
<p class="normal">
 2011
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Amazon Access Samples

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Amazon's InfoSec is getting smarter about the way Access data is
leveraged. This is an anonymized sample of access provisioned within the company. </p></td> --
->
<td>
<p class="normal">
 Time-Series, Domain-Theory
</p>
</td>
<td>
<p class="normal">

```

```

 Regression, Clustering, Causal-Discovery
 </p>
</td>
<td>
 <p class="normal">
 </p>
</td>
<td>
 <p class="normal">
 30000
 </p>
</td>
<td>
 <p class="normal">
 20000
 </p>
</td>
<td>
 <p class="normal">
 2011
 </p>
</td>
<!-- <td><p class="normal">Business </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
 <p class="normal">

 Reuter_50_50

 </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The dataset is used for authorship identification in online Write
print which is a new research field of pattern recognition. </p></td> -->
<td>
 <p class="normal">
 Multivariate, Text, Domain-Theory
 </p>
</td>
<td>
 <p class="normal">
 Classification, Clustering
 </p>
</td>
<td>
 <p class="normal">
 Real
 </p>
</td>
<td>
 <p class="normal">
 2500
 </p>
</td>
<td>
 <p class="normal">
 10000
 </p>
</td>
<td>
 <p class="normal">
 2011
 </p>
</td>

```

```

<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Farm Ads

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This data was collected from text ads found on twelve websites t
hat deal with various farm animal related topics. The binary labels are based on whether or not t
he content owner approves of the ad. </p></td> -->
<td>
<p class="normal">
 Text
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 4143
</p>
</td>
<td>
<p class="normal">
 54877
</p>
</td>
<td>
<p class="normal">
 2011
</p>
</td>
<!-- <td><p class="normal">Business </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 DBWorld e-mails

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

```

```
 . . .
 </td>
 <!-- <td><p class="normal">It contains 64 e-mails which I have manually collected from
DBWorld mailing list. They are classified in: 'announces of conferences' and 'everything
else'. </p></td> -->
 <td>
 <p class="normal">
 Text
 </p>
 </td>
 <td>
 <p class="normal">
 Classification
 </p>
 </td>
 <td>
 <p class="normal">
 64
 </p>
 </td>
 <td>
 <p class="normal">
 4702
 </p>
 </td>
 <td>
 <p class="normal">
 2011
 </p>
 </td>
 <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 KEGG Metabolic Relation Network (Directed)

 </p>
 </td>
 </tr>
 </table>
 </td>
 <td><p class="normal">KEGG Metabolic pathways modeled as directed relation network. Va
riety of graphical features presented. </p></td> -->
 <td>
 <p class="normal">
 Multivariate, Univariate, Text
 </p>
 </td>
 <td>
 <p class="normal">
 Classification, Regression, Clustering
 </p>
 </td>
 <td>
 <p class="normal">
 Integer, Real
 </p>
 </td>
 <td>
 <p class="normal">
 53414
 </p>
 </td>
```

```

 '' r'
</td>
<td>
<p class="normal">
 24
</p>
</td>
<td>
<p class="normal">
 2011
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 KEGG Metabolic Reaction Network (Undirected)

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">KEGG Metabolic pathways modeled as un-directed reaction network.
Variety of graphical features presented. </p></td> -->
<td>
<p class="normal">
 Multivariate, Univariate, Text
</p>
</td>
<td>
<p class="normal">
 Classification, Regression, Clustering
</p>
</td>
<td>
<p class="normal">
 Integer, Real
</p>
</td>
<td>
<p class="normal">
 65554
</p>
</td>
<td>
<p class="normal">
 29
</p>
</td>
<td>
<p class="normal">
 2011
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>

```

```

.. ..
<td>
<p class="normal">

 Bank Marketing

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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). </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 45211
</p>
</td>
<td>
<p class="normal">
 17
</p>
</td>
<td>
<p class="normal">
 2012
</p>
</td>
<!-- <td><p class="normal">Business </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 YouTube Comedy Slam Preference Data

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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. </p></td> -->
<td>
<p class="normal">
 Text
</p>
</td>
<+^>

```

```

<!--
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 1138562
</p>
</td>
<td>
<p class="normal">
 3
</p>
</td>
<td>
<p class="normal">
 2012
</p>
</td>
<td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Gas Sensor Array Drift Dataset

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This archive contains 13910 measurements from 16 chemical
sensors utilized in simulations for drift compensation in a discrimination task of 6 gases at vari
ous levels of concentrations. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 13910
</p>
</td>
<td>
<p class="normal">
 128
</p>
</td>
<td>
<p class="normal">
 2012
</p>

```

```

<!--
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 ILPD (Indian Liver Patient Dataset)

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This data set contains 10 variables that are age, gender, total
Bilirubin, direct Bilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT and Alkphos. </p>
-->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Integer, Real
</p>
</td>
<td>
<p class="normal">
 583
</p>
</td>
<td>
<p class="normal">
 10
</p>
</td>
<td>
<p class="normal">
 2012
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 OPPORTUNITY Activity Recognition

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

```

```

 </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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). </p></td> -->
<td>
<p class="normal">
 Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 2551
</p>
</td>
<td>
<p class="normal">
 242
</p>
</td>
<td>
<p class="normal">
 2012
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Nomao

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Nomao collects data about places (name, phone, localization...) from many sources.
Deduplication consists in detecting what data refer to the same place.
Instances in the dataset compare 2 spots. </p></td> -->
<td>
<p class="normal">
 Univariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>

```

```

<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 34465
</p>
</td>
<td>
<p class="normal">
 120
</p>
</td>
<td>
<p class="normal">
 2012
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 SMS Spam Collection

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">The SMS Spam Collection is a public set of SMS labeled messages
that have been collected for mobile phone spam research. </p></td> -->
<td>
 <p class="normal">
 Multivariate, Text, Domain-Theory
 </p>
</td>
<td>
 <p class="normal">
 Classification, Clustering
 </p>
</td>
<td>
 <p class="normal">
 Real
 </p>
</td>
<td>
 <p class="normal">
 5574
 </p>
</td>
<td>
 <p class="normal">
 2012
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
 ...

```

```

<table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Skin Segmentation

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">The Skin Segmentation dataset is constructed over B, G, R color space. Skin and Nonskin dataset is generated using skin textures from face images of diversity of age, gender, and race people. </p></td> -->
<td>
 <p class="normal">
 Univariate
 </p>
</td>
<td>
 <p class="normal">
 Classification
 </p>
</td>
<td>
 <p class="normal">
 Real
 </p>
</td>
<td>
 <p class="normal">
 245057
 </p>
</td>
<td>
 <p class="normal">
 4
 </p>
</td>
<td>
 <p class="normal">
 2012
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Planning Relax

 </p>
 </td>
 </tr>
 </table>
 </td>
<!-- <td><p class="normal">The dataset concerns with the classification of two mental stages from recorded EEG signals: Planning (during imagination of motor act) and Relax state. &nbs

```

```

p;</p></td> -->
<td>
<p class="normal">
 Univariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 182
</p>
</td>
<td>
<p class="normal">
 13
</p>
</td>
<td>
<p class="normal">
 2012
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 PAMAP2 Physical Activity Monitoring

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The PAMAP2 Physical Activity Monitoring dataset contains data of
18 different physical activities, performed by 9 subjects wearing 3 inertial measurement units and
a heart rate monitor. </p></td> -->
<td>
<p class="normal">
 Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 3850505
</p>
</td>

```

```

<td>
<p class="normal">
52
</p>
</td>
<td>
<p class="normal">
2012
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Restaurant & consumer data

</p>
</td>
</tr>
</table>
</td>
<!-- <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. ;
</p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
138
</p>
</td>
<td>
<p class="normal">
47
</p>
</td>
<td>
<p class="normal">
2012
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

```

```


CNAE-9

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This is a data set containing 1080 documents of free text
business descriptions of Brazilian companies categorized into a
subset of 9 categories&nbsp</p></td> -->
<td>
<p class="normal">
Multivariate, Text
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Integer
</p>
</td>
<td>
<p class="normal">
1080
</p>
</td>
<td>
<p class="normal">
857
</p>
</td>
<td>
<p class="normal">
2012
</p>
</td>
<!-- <td><p class="normal">Business&nbsp</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Individual household electric power consumption

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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 so
me sub-metering values are available.&nbsp</p></td> -->
<td>
<p class="normal">
Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
Regression, Clustering

```

```

 </p>
 </td>
 <td>
 <p class="normal">
 Real
 </p>
 </td>
 <td>
 <p class="normal">
 2075259
 </p>
 </td>
 <td>
 <p class="normal">
 9
 </p>
 </td>
 <td>
 <p class="normal">
 2012
 </p>
 </td>
 <!-- <td><p class="normal">Physical </p></td> -->
 </tr>
 <tr bgcolor="#DDEEFF">
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 seeds

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">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. </p></td> -->
 <td>
 <p class="normal">
 Multivariate
 </p>
 </td>
 <td>
 <p class="normal">
 Classification, Clustering
 </p>
 </td>
 <td>
 <p class="normal">
 Real
 </p>
 </td>
 <td>
 <p class="normal">
 210
 </p>
 </td>
 <td>
 <p class="normal">
 7
 </p>
 </td>
 <td>
 <p class="normal">
 2012
 </p>
 </td>
 </tr>

```

```

</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Northix

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Northix is designed to be a schema matching benchmark problem
for data integration of two entity relationship databases. </p></td> -->
<td>
<p class="normal">
Multivariate, Univariate, Text
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Integer, Real
</p>
</td>
<td>
<p class="normal">
115
</p>
</td>
<td>
<p class="normal">
200
</p>
</td>
<td>
<p class="normal">
2012
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

QtyT40I10D100K

</p>
</td>
</tr>

```

```

 </table>
 </td>
 <!-- <td><p class="normal">Since there is no numerical sequential data stream available in standard data sets, this data set is generated from the original T40I10D100K data set
</p></td>
-->
<td>
<p class="normal">
 Sequential
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 Integer
</p>
</td>
<td>
<p class="normal">
 3960456
</p>
</td>
<td>
<p class="normal">
 4
</p>
</td>
<td>
<p class="normal">
 2012
</p>
</td>
<!-- <td><p class="normal">&br/></p></td>
-->
</tr>
<tr>
<td>
<table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Legal Case Reports

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">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.&br/></p></td>
-->
<td>
<p class="normal">
 Text
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
</p>
</td>

```

```

</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
2012
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Human Activity Recognition Using Smartphones

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Human Activity Recognition database built from the recordings of
30 subjects performing activities of daily living (ADL) while carrying a waist-mounted smartphone
with embedded inertial sensors. </p></td> -->
<td>
<p class="normal">
Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
Classification, Clustering
</p>
</td>
<td>
<p class="normal">
10299
</p>
</td>
<td>
<p class="normal">
561
</p>
</td>
<td>
<p class="normal">
2012
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
</td>
<td>

```

```

<p class="normal">

 One-hundred plant species leaves data set

</p>
</td>
</tr>
</table>
</td>
!--<td><p class="normal">Sixteen samples of leaf each of one-hundred plant species. For
each sample, a shape descriptor, fine scale margin and texture histogram are given. </p></td>
-->
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 1600
</p>
</td>
<td>
<p class="normal">
 64
</p>
</td>
<td>
<p class="normal">
 2012
</p>
</td>
!--<td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Energy efficiency

</p>
</td>
</tr>
</table>
</td>
!--<td><p class="normal">This study looked into assessing the heating load and cooling
load requirements of buildings (that is, energy efficiency) as a function of building
parameters. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification, Regression

```

```
</p>
</td>
<td>
<p class="normal">
 Integer, Real
</p>
</td>
<td>
<p class="normal">
 768
</p>
</td>
<td>
<p class="normal">
 8
</p>
</td>
<td>
<p class="normal">
 2012
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Yacht Hydrodynamics

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Delft data set, used to predict the hydodynamic performance of
sailing yachts from dimensions and velocity. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Regression
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 308
</p>
</td>
<td>
<p class="normal">
 7
</p>
</td>
<td>
<p class="normal">
 2013
</p>
</td>
```

```

<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Fertility

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">100 volunteers provide a semen sample analyzed according to the WHO 2010 criteria. Sperm concentration are related to socio-demographic data, environmental factors, health status, and life habits&nbsp</p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification, Regression
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 100
</p>
</td>
<td>
<p class="normal">
 10
</p>
</td>
<td>
<p class="normal">
 2013
</p>
</td>
<!-- <td><p class="normal">Life &nbsp</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Daphnet Freezing of Gait

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

```

```

 ...
 </table>
</td>
<!-- <td><p class="normal">This dataset contains the annotated readings of 3 acceleration sensors at the hip and leg of Parkinson's disease patients that experience freezing of gait (FoG) during walking tasks.
 </p></td> -->
<td>
<p class="normal">
 Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 237
</p>
</td>
<td>
<p class="normal">
 9
</p>
</td>
<td>
<p class="normal">
 2013
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 3D Road Network (North Jutland, Denmark)

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">3D road network with highly accurate elevation information (+-20 cm) from Denmark used in eco-routing and fuel/Co2-estimation routing algorithms. </p></td> -->
<td>
<p class="normal">
 Sequential, Text
</p>
</td>
<td>
<p class="normal">
 Regression, Clustering
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>

```

```

., ..
<td>
<p class="normal">
434874
</p>
</td>
<td>
<p class="normal">
4
</p>
</td>
<td>
<p class="normal">
2013
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

ISTANBUL STOCK EXCHANGE

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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. </p></td> -->
<td>
<p class="normal">
Multivariate, Univariate, Time-Series
</p>
</td>
<td>
<p class="normal">
Classification, Regression
</p>
</td>
<td>
<p class="normal">
Real
</p>
</td>
<td>
<p class="normal">
536
</p>
</td>
<td>
<p class="normal">
8
</p>
</td>
<td>
<p class="normal">
2013
</p>
</td>
<!-- <td><p class="normal">Business </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>

```

```

<!--
<td>

</td>
<td>
 <p class="normal">

 Buzz in social media

 </p>
</td>
</tr>
</table>
</td>
<!-- <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. </p></td> -->
<td>
 <p class="normal">
 Time-Series, Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Regression, Classification
 </p>
</td>
<td>
 <p class="normal">
 Integer, Real
 </p>
</td>
<td>
 <p class="normal">
 140000
 </p>
</td>
<td>
 <p class="normal">
 77
 </p>
</td>
<td>
 <p class="normal">
 2013
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 First-order theorem proving

 </p>
 </td>
 </tr>
 </table>
</td>
<!-- <td><p class="normal">Given a theorem, predict which of five heuristics will give the fastest proof when used by a first-order prover. A sixth prediction declines to attempt a proof, should the theorem be too difficult. </p></td> -->
<!-->

```

```

<!-->
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 6118
</p>
</td>
<td>
<p class="normal">
 51
</p>
</td>
<td>
<p class="normal">
 2013
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Wearable Computing: Classification of Body Postures and Movements (PUC-Rio)

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">A dataset with 5 classes (sitting-down, standing-up, standing, walking, and sitting) collected on 8 hours of activities of 4 healthy subjects. We also established a baseline performance index. </p></td> -->
<td>
<p class="normal">
 Sequential
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Integer, Real
</p>
</td>
<td>
<p class="normal">
 165632
</p>
<!-->

```

```

</td>
<td>
<p class="normal">
 18
</p>
</td>
<td>
<p class="normal">
 2013
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Gas sensor arrays in open sampling settings

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The dataset contains 18000 time-series recordings from a chemical
detection platform at six different locations in a wind tunnel facility in response to ten high-pr
iority chemical gaseous substances</p></td> -->
<td>
<p class="normal">
 Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 18000
</p>
</td>
<td>
<p class="normal">
 1950000
</p>
</td>
<td>
<p class="normal">
 2013
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>

```

```

</td>
<td>
<p class="normal">

 Climate Model Simulation Crashes

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 540
</p>
</td>
<td>
<p class="normal">
 18
</p>
</td>
<td>
<p class="normal">
 2013
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 MicroMass

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">A dataset to explore machine learning approaches for the identification of microorganisms from mass-spectrometry data. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>

```

```
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Real
</p>
</td>
<td>
<p class="normal">
931
</p>
</td>
<td>
<p class="normal">
1300
</p>
</td>
<td>
<p class="normal">
2013
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>


```

```
2013
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 BLOGGER

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">In this paper, we look for to recognize the causes of users tend
to cyber space in Kohkiloye and Boyer Ahmad Province in
Iran </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 100
</p>
</td>
<td>
<p class="normal">
 6
</p>
</td>
<td>
<p class="normal">
 2013
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Daily and Sports Activities

</p>

```

```

 </td>
 </tr>
</table>
<td>
<!-- <td><p class="normal">The dataset comprises motion sensor data of 19 daily and sports activities each performed by 8 subjects in their own style for 5 minutes. Five Xsens MTx units are used on the torso, arms, and legs.
 </p></td> -->
<td>
<p class="normal">
 Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification, Clustering
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 9120
</p>
</td>
<td>
<p class="normal">
 5625
</p>
</td>
<td>
<p class="normal">
 2013
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 User Knowledge Modeling

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification, Clustering
</p>
</td>
<td>
<p class="normal">
 Integer

```

```

</p>
</td>
<td>
<p class="normal">
 403
</p>
</td>
<td>
<p class="normal">
 5
</p>
</td>
<td>
<p class="normal">
 2013
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Reuters RCV1 RCV2 Multilingual, Multiview Text Categorization Test collection

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This test collection contains feature characteristics of
documents originally written in five different languages and their translations, over a common set
of 6 categories. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 111740
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 2013
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>

```

```

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

</td>
<td>
<p class="normal">

NYSK

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">NYSK (New York v. Strauss-Kahn) is a collection of English news
articles about the case relating to allegations of sexual assault against the former IMF director
Dominique Strauss-Kahn (May 2011). </p></td> -->
<td>
<p class="normal">
Multivariate, Sequential, Text
</p>
</td>
<td>
<p class="normal">
Clustering
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
10421
</p>
</td>
<td>
<p class="normal">
7
</p>
</td>
<td>
<p class="normal">
2013
</p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Turkiye Student Evaluation

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This data set contains a total 5820 evaluation scores provided b
y students from Gazi University in Ankara (Turkey). There is a total of 28 course specific

```

```

questions and additional 5 attributes. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification, Clustering
</p>
</td>
<td>
<p class="normal">
 5820
</p>
</td>
<td>
<p class="normal">
 33
</p>
</td>
<td>
<p class="normal">
 2013
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

 </td>
 <td>
 <p class="normal">

 ser Knowledge Modeling Data (Students' Knowledge Levels on DC Electrical Machines)

 </p>
 </td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">

```

```

 403
 </p>
</td>
<td>
<p class="normal">
 5
</p>
</td>
<td>
<p class="normal">
 2013
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 EEG Eye State

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The data set consists of 14 EEG values and a value indicating
the eye state. </p></td> -->
<td>
<p class="normal">
 Multivariate, Sequential, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Integer, Real
</p>
</td>
<td>
<p class="normal">
 14980
</p>
</td>
<td>
<p class="normal">
 15
</p>
</td>
<td>
<p class="normal">
 2013
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>


```

```


 </td>
 <td>
 <p class="normal">

 Physicochemical Properties of Protein Tertiary Structure

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">This is a data set of Physicochemical Properties of Protein Tertiary Structure. The data set is taken from CASP 5-9. There are 45730 decoys and size varying from 0 to 21 armstrong. </p></td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Regression
 </p>
</td>
<td>
 <p class="normal">
 Real
 </p>
</td>
<td>
 <p class="normal">
 45730
 </p>
</td>
<td>
 <p class="normal">
 9
 </p>
</td>
<td>
 <p class="normal">
 2013
 </p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 seismic-bumps

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">The data describe the problem of high energy (higher than 10^4 J) seismic bumps forecasting in a coal mine. Data come from two of longwalls located in a Polish coal mine. </p></td> -->
 <td>
 <p class="normal">
 Multivariate
 </p>

```

```
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 2584
</p>
</td>
<td>
<p class="normal">
 19
</p>
</td>
<td>
<p class="normal">
 2013
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 banknote authentication

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Data were extracted from images that were taken for the
evaluation of an authentication procedure for banknotes. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 1372
</p>
</td>
<td>
<p class="normal">
 5
</p>
</td>
<td>
```

```

<p class="normal">
 2013
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-
Harvard+Tournament+Lab+and+TopCoder++++Problem%3A+Pat">

</td>
<td>
<p class="normal">

<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

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Data used for USPTO Algorithm Competition. Contains drawing
pages from US patents with manually labeled figure and part labels. </p></td> -->
<td>
<p class="normal">
 Domain-Theory
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Integer
</p>
</td>
<td>
<p class="normal">
 306
</p>
</td>
<td>
<p class="normal">
 5
</p>
</td>
<td>
<p class="normal">
 2013
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">


```

```

 YouTube Multiview Video Games Dataset

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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></td> -->
<td>
<p class="normal">
 Multivariate, Text
</p>
</td>
<td>
<p class="normal">
 Classification, Clustering
</p>
</td>
<td>
<p class="normal">
 Integer, Real
</p>
</td>
<td>
<p class="normal">
 120000
</p>
</td>
<td>
<p class="normal">
 1000000
</p>
</td>
<td>
<p class="normal">
 2013
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Gas Sensor Array Drift Dataset at Different Concentrations

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This archive contains 13910 measurements from 16 chemical
sensors exposed to 6 different gases at various concentration levels. </p></td> -->
<td>
<p class="normal">
 Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification, Regression, Clustering, Causa
</p>
</td>
<td>
```

```

 ...
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 13910
</p>
</td>
<td>
<p class="normal">
 129
</p>
</td>
<td>
<p class="normal">
 2013
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Activities of Daily Living (ADLs) Recognition Using Binary Sensors

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">This dataset comprises information regarding the ADLs performed by two users on a daily basis in their own homes. </p></td> -->
<td>
<p class="normal">
 Multivariate, Sequential, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification, Clustering
</p>
</td>
<td>
<p class="normal">
 2747
</p>
</td>
<td>
<p class="normal">
 2013
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>

```

```

<!--
<tr bgcolor="#DDEEFF">
<td>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

SkillCraft1 Master Table Dataset

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This data was used in Thompson et al. (2013). A list of possible game actions is discussed in Thompson, Blair, Chen, & Henrey (2013). </p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Regression
</p>
</td>
<td>
<p class="normal">
Integer, Real
</p>
</td>
<td>
<p class="normal">
3395
</p>
</td>
<td>
<p class="normal">
20
</p>
</td>
<td>
<p class="normal">
2013
</p>
</td>
<!-- <td><p class="normal">Game </p></td> -->
</tr>
<tr>
<td>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Weight Lifting Exercises monitored with Inertial Measurement Units

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Six young health subjects were asked to perform 5 variations of +

```

.. --> Class-- normal /six young healthy subjects were asked to perform 3 variations of the biceps curl weight lifting exercise. One of the variations is the one predicted by the health professional.&nbsp;</p></td> -->

<td>	<p class="normal">
	Multivariate
	</p>
</td>	
<td>	<p class="normal">
	Classification
	</p>
</td>	
<td>	<p class="normal">
	Real
	</p>
</td>	
<td>	<p class="normal">
	39242
	</p>
</td>	
<td>	<p class="normal">
	152
	</p>
</td>	
<td>	<p class="normal">
	2013
	</p>
</td>	
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->	
</tr>	
<tr bgcolor="DDEEFF">	
<td>	
<table>	
<tr>	
<td>	
<a href="datasets/SML2010">	
</a>	
</td>	
<td>	
<p class="normal">	
<b>	
<a href="datasets/SML2010">	
SML2010	
</a>	
</b>	
</p>	
</td>	
</tr>	
</table>	
</td>	
<!-- <td><p class="normal">This dataset is collected from a monitor system mounted in a domotic house. It corresponds to approximately 40 days of monitoring data.&nbsp;</p></td> -->	
<td>	
<p class="normal">	
Multivariate, Sequential, Time-Series, Text	
</p>	
</td>	
<td>	
<p class="normal">	
Regression	
</p>	
</td>	
<td>	
<p class="normal">	
Real	
</p>	
</td>	
<td>	
<p class="normal">	
4137	
</p>	
</td>	

```

~,~

<td>

<p class="normal">

 24

</p>

</td>

<td>

<p class="normal">

 2014

</p>

</td>

<!-- <td><p class="normal">Computer </p></td> -->

</tr>

<tr>

<td>

<table>

<tr>

<td>

</td>

<td>

<p class="normal">

 Bike Sharing Dataset

</p>

</td>

</tr>

</table>

</td>

<!-- <td><p class="normal">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></td> -->

<td>

<p class="normal">

 Univariate

</p>

</td>

<td>

<p class="normal">

 Regression

</p>

</td>

<td>

<p class="normal">

 Integer, Real

</p>

</td>

<td>

<p class="normal">

 17389

</p>

</td>

<td>

<p class="normal">

 16

</p>

</td>

<td>

<p class="normal">

 2013

</p>

</td>

<!-- <td><p class="normal">Social </p></td> -->

</tr>

<tr bgcolor="DDEEFF">

<td>

<table>

<tr>

<td>

~,~
```

```

</tr>
</table>
</td>
<!-- <td><p class="normal">The data from Twitter was collected during 360 consecutive days.
It was done by querying 1497 English keywords sampled from Wikipedia. This dataset is proposed in
a Learning to rank setting. </p></td> -->
<td>
<p class="normal">
 Multivariate, Sequential, Time-Series
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 Integer, Real
</p>
</td>
<td>
<p class="normal">
 51
</p>
</td>
<td>
<p class="normal">
 35
</p>
</td>
<td>
<p class="normal">
 2013
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Thoracic Surgery Data

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The data is dedicated to classification problem related to the pc
st-operative life expectancy in the lung cancer patients: class 1 - death within one year after su
rgery, class 2 - survival. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>

```

```

<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Integer, Real
</p>
</td>
<td>
<p class="normal">
470
</p>
</td>
<td>
<p class="normal">
17
</p>
</td>
<td>
<p class="normal">
2013
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>


```

```

</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

SUSY

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This is a classification problem to distinguish between a signal process which produces supersymmetric particles and a background process which does not. </p>
</td> -->
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Real
</p>
</td>
<td>
<p class="normal">
5000000
</p>
</td>
<td>
<p class="normal">
18
</p>
</td>
<td>
<p class="normal">
2014
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

HIGGS

</p>
</td>
<td>
<td>

```

```

</tr>
</table>
</td>
<!-- <td><p class="normal">This is a classification problem to distinguish between a signal
process which produces Higgs bosons and a background process which does not. </p></td> -->
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 11000000
</p>
</td>
<td>
<p class="normal">
 28
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Qualitative_Bankruptcy

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Predict the Bankruptcy from Qualitative parameters from
experts. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 250
</p>
</td>
<td>
<p class="normal">
 .
</p>

```

```

</td>
<td>
<p class="normal">
 7
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 LSVT Voice Rehabilitation

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">126 samples from 14 participants, 309 features. Aim: assess whether voice rehabilitation treatment lead to phonations considered 'acceptable' or 'unacceptable' (binary class classification problem). </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 126
</p>
</td>
<td>
<p class="normal">
 309
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>


```

```

</td>
<td>
<p class="normal">

 Dataset for ADL Recognition with Wrist-worn Accelerometer

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Recordings of 16 volunteers performing 14 Activities of Daily Living (ADL) while carrying a single wrist-worn tri-axial accelerometer. </p></td> -->
<td>
<p class="normal">
 Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification, Clustering
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 3
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Wilt

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">High-resolution Remote Sensing data set (Quickbird). Small number of training samples of diseased trees, large number for other land cover. Testing data set from stratified random sample of image. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification

```

```
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 4889
</p>
</td>
<td>
<p class="normal">
 6
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 User Identification From Walking Activity

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The dataset collects data from an Android smartphone positioned i
n the chest pocket from 22 participants walking in the wild over a predefined path.
 </p></td> -->
<td>
<p class="normal">
 Univariate, Sequential, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification, Clustering
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
```

```

<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Activity Recognition from Single Chest-Mounted Accelerometer

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The dataset collects data from a wearable accelerometer mounted on the chest. The dataset is intended for Activity Recognition research purposes. </p></td> -->
<td>
<p class="normal">
Univariate, Sequential, Time-Series
</p>
</td>
<td>
<p class="normal">
Classification, Clustering
</p>
</td>
<td>
<p class="normal">
Real
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
2014
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Leaf

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset consists in a collection of shape and texture features extracted from digital images of leaf specimens originating from a total of 40 different

```

```

plant species. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 340
</p>
</td>
<td>
<p class="normal">
 16
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Dresses_Attribute_Sales

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset contain Attributes of dresses and their recommendations according to their sales.Sales are monitor on the basis of alternate days. </p></td> -->
<td>
<p class="normal">
 Text
</p>
</td>
<td>
<p class="normal">
 Classification, Clustering
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 501
</p>
</td>
<td>
<p class="normal">

```

```

 13
 </p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Tamilnadu Electricity Board Hourly Readings

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This data can be effectively produced the result to fewer
parameter of the Load profile can be reduced in the Database </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification, Regression, Clustering
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 45781
</p>
</td>
<td>
<p class="normal">
 5
</p>
</td>
<td>
<p class="normal">
 2013
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">


```

```


 Airfoil Self-Noise

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">NASA data set, obtained from a series of aerodynamic and
acoustic tests of two and three-dimensional airfoil blade sections conducted in an anechoic wind t
unnel. </p></td> -->
 <td>
 <p class="normal">
 Multivariate
 </p>
 </td>
 <td>
 <p class="normal">
 Regression
 </p>
 </td>
 <td>
 <p class="normal">
 Real
 </p>
 </td>
 <td>
 <p class="normal">
 1503
 </p>
 </td>
 <td>
 <p class="normal">
 6
 </p>
 </td>
 <td>
 <p class="normal">
 2014
 </p>
 </td>
 <!-- <td><p class="normal">Physical </p></td> -->
 </tr>
 <tr>
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Wholesale customers

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">The data set refers to clients of a wholesale distributor. It
includes the annual spending in monetary units (m.u.) on diverse product categories </p></td>
-->
 <td>
 <p class="normal">
 Multivariate
 </p>
 </td>
 <td>
 <p class="normal">
 Classification, Clustering
 </p>

```

```

</td>
<td>
<p class="normal">
 Integer
</p>
</td>
<td>
<p class="normal">
 440
</p>
</td>
<td>
<p class="normal">
 8
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Business </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Twitter Data set for Arabic Sentiment Analysis

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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. </p></td> -->
<td>
<p class="normal">
 Text
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 </p>
</td>
<td>
<p class="normal">
 2000
</p>
</td>
<td>
<p class="normal">
 2
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Social </p></td> -->

```

```

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

</td>
<td>
<p class="normal">

 Combined Cycle Power Plant

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The dataset contains 9568 data points collected from a Combined
Cycle Power Plant over 6 years (2006-2011), when the plant was set to work with full load. <
/p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Regression
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 9568
</p>
</td>
<td>
<p class="normal">
 4
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Urban Land Cover

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

```

```

 .. ~~~~
 </td>
 <!-- <td><p class="normal">Classification of urban land cover using high resolution aerial imagery. Intended to assist sustainable urban planning efforts. </p></td> -->
 <td>
 <p class="normal">
 Multivariate
 </p>
 </td>
 <td>
 <p class="normal">
 Classification
 </p>
 </td>
 <td>
 <p class="normal">
 </p>
 </td>
 <td>
 <p class="normal">
 168
 </p>
 </td>
 <td>
 <p class="normal">
 148
 </p>
 </td>
 <td>
 <p class="normal">
 2014
 </p>
 </td>
 <!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Diabetes 130-US hospitals for years 1999-2008

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">This data has been prepared to analyze factors related to readmission as well as other
outcomes pertaining to patients with diabetes. </p></td> -->
 <td>
 <p class="normal">
 Multivariate
 </p>
 </td>
 <td>
 <p class="normal">
 Classification, Clustering
 </p>
 </td>
 <td>
 <p class="normal">
 Integer
 </p>
 </td>
 <td>
 <p class="normal">
 100000
 </p>
 </td>

```

```

+-----+
</p>
</td>
<td>
<p class="normal">
 55
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Bach Choral Harmony

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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.&nbsp</p></td> -->
<td>
<p class="normal">
 Sequential
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 5665
</p>
</td>
<td>
<p class="normal">
 17
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>


```

```

 <td>
 <p>StoneFlakes</p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">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. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification, Clustering, Causal-Discovery
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 79
</p>
</td>
<td>
<p class="normal">
 8
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
 <p>
 Tennis Major Tournament Match Statistics

 </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This is a collection of 8 files containing the match statistics f
or both women and men at the four major tennis tournaments of the year 2013. Each file has 42 colu
mns and a minimum of 76 rows. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>

```

```

~ / ~
</td>
<td>
<p class="normal">
Classification, Regression, Clustering
</p>
</td>
<td>
<p class="normal">
Integer, Real
</p>
</td>
<td>
<p class="normal">
127
</p>
</td>
<td>
<p class="normal">
42
</p>
</td>
<td>
<p class="normal">
2014
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Parkinson Speech Dataset with Multiple Types of Sound Recordings

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The training data belongs to 20 Parkinson's Disease (PD)
patients and 20 healthy subjects. From all subjects, multiple types of sound recordings (26) are t
aken. </p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Classification, Regression
</p>
</td>
<td>
<p class="normal">
Integer, Real
</p>
</td>
<td>
<p class="normal">
1040
</p>
</td>
<td>
<p class="normal">
26
</p>
</td>
~ / ~

```

```

</td>
<td>
<p class="normal">
2014
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Gesture Phase Segmentation

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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. </p></td> -->
<td>
<p class="normal">
Multivariate, Sequential, Time-Series
</p>
</td>
<td>
<p class="normal">
Classification, Clustering
</p>
</td>
<td>
<p class="normal">
Real
</p>
</td>
<td>
<p class="normal">
9900
</p>
</td>
<td>
<p class="normal">
50
</p>
</td>
<td>
<p class="normal">
2014
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Perfume Data

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

```

```

 Perfume Data

 </p>
 </td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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>
-->
<td>
<p class="normal">
 Univariate, Domain-Theory
</p>
</td>
<td>
<p class="normal">
 Classification, Clustering
</p>
</td>
<td>
<p class="normal">
 Integer
</p>
</td>
<td>
<p class="normal">
 560
</p>
</td>
<td>
<p class="normal">
 2
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
 <p class="normal">

 BlogFeedback

 </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Instances in this dataset contain features extracted from blog
posts. The task associated with the data is to predict how many comments the post will
receive. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Regression
</p>
</td>
.. ..

```

```

<td>
<p class="normal">
 Integer, Real
</p>
</td>
<td>
<p class="normal">
 60021
</p>
</td>
<td>
<p class="normal">
 281
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 REALDISP Activity Recognition Dataset

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The REALDISP dataset is devised to evaluate techniques dealing wi
th the effects of sensor displacement in wearable activity recognition as well as to benchmark gen
eral activity recognition algorithms </p></td> -->
<td>
<p class="normal">
 Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 1419
</p>
</td>
<td>
<p class="normal">
 120
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
``

```

```

</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Newspaper and magazine images segmentation dataset

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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. </p></td> -->
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
101
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
2014
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

AAAI 2014 Accepted Papers

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Clustering
</p>
</td>
<td>
<p class="normal">
 399
</p>
</td>
<td>
<p class="normal">
 6
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Gas sensor array under flow modulation

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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. </p></td> -->
<td>
<p class="normal">
 Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification, Regression
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 58
</p>
</td>
<td>

```

```

<p class="normal">
 120432
</p>
</td>
<td>
 <p class="normal">
 2014
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Gas sensor array exposed to turbulent gas mixtures

 </p>
 </td>
 </tr>
 </table>
</td>
<!-- <td><p class="normal">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. </p></td> -->
<td>
 <p class="normal">
 Multivariate, Time-Series
 </p>
</td>
<td>
 <p class="normal">
 Classification, Regression
 </p>
</td>
<td>
 <p class="normal">
 Real
 </p>
</td>
<td>
 <p class="normal">
 180
 </p>
</td>
<td>
 <p class="normal">
 150000
 </p>
</td>
<td>
 <p class="normal">
 2014
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
 <table>
 <tr>
 <td>

 </td>
 <td>

```

```

<p class="normal">

UJIIndoorLoc

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The UJIIndoorLoc is a Multi-Building Multi-Floor indoor
localization database to test Indoor Positioning System that rely on WLAN/WiFi fingerprint. <
/p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Classification, Regression
</p>
</td>
<td>
<p class="normal">
Integer, Real
</p>
</td>
<td>
<p class="normal">
21048
</p>
</td>
<td>
<p class="normal">
529
</p>
</td>
<td>
<p class="normal">
2014
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

Sentence Classification

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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. </p></td> -->
<td>
<p class="normal">
Text
</p>
</td>
<td>
<p class="normal">

```

```

Classification
</p>
</td>
<td>
<p class="normal">
 Integer
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Dow Jones Index

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset contains weekly data for the Dow Jones Industrial I
ndex. It has been used in computational investing research. </p></td> -->
<td>
<p class="normal">
 Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification, Clustering
</p>
</td>
<td>
<p class="normal">
 Integer, Real
</p>
</td>
<td>
<p class="normal">
 750
</p>
</td>
<td>
<p class="normal">
 16
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Business </p></td> -->

```

```

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

</td>
<td>
<p class="normal">

 sEMG for Basic Hand movements

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The "sEMG for Basic Hand movements" includes 2 databases of surface electromyographic signals of 6 hand movements using Delsys' EMG System. Healthy subjects conducted six daily life grasps. </p></td> -->
<td>
<p class="normal">
 Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 3000
</p>
</td>
<td>
<p class="normal">
 2500
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 AAAI 2013 Accepted Papers

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

```

```

</td>
<!-- <td><p class="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. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Clustering
</p>
</td>
<td>
<p class="normal">
 150
</p>
</td>
<td>
<p class="normal">
 5
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
 <p class="normal">

 Geographical Original of Music

 </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Instances in this dataset contain audio features extracted from
1059 wave files. The task associated with the data is to predict the geographical origin of music.
 </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification, Regression
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 1059

```

```

</p>
</td>
<td>
<p class="normal">
 68
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Condition Based Maintenance of Naval Propulsion Plants

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Data have been generated from a sophisticated simulator of a Gas Turbines (GT), mounted on a Frigate characterized by a COnbined Diesel eLectric And Gas (CODLAG) propulsion plant type. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Regression
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 11934
</p>
</td>
<td>
<p class="normal">
 16
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>


```

```


 </td>
 <td>
 <p class="normal">

 Grammatical Facial Expressions

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">This dataset supports the development of models that make
possible to interpret Grammatical Facial Expressions from Brazilian Sign Language (Libras). </p></td> -->
<td>
 <p class="normal">
 Multivariate, Sequential
 </p>
</td>
<td>
 <p class="normal">
 Classification, Clustering
 </p>
</td>
<td>
 <p class="normal">
 Real
 </p>
</td>
<td>
 <p class="normal">
 27965
 </p>
</td>
<td>
 <p class="normal">
 100
 </p>
</td>
<td>
 <p class="normal">
 2014
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 NoisyOffice

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">Corpus intended to do cleaning (or binarization) and enhancement
of noisy grayscale printed text images using supervised learning methods. Noisy images and their c
orresponding ground truth provided. </p></td> -->
 <td>
 <p class="normal">
 Multivariate
 </p>

```

```

</td>
<td>
<p class="normal">
Classification, Regression
</p>
</td>
<td>
<p class="normal">
Real
</p>
</td>
<td>
<p class="normal">
216
</p>
</td>
<td>
<p class="normal">
216
</p>
</td>
<td>
<p class="normal">
2015
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

MHEALTH Dataset

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The MHEALTH (Mobile Health) dataset is devised to benchmark techniques dealing with human behavior analysis based on multimodal body sensing. </p></td> -->
<td>
<p class="normal">
Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Real
</p>
</td>
<td>
<p class="normal">
120
</p>
</td>
<td>
<p class="normal">
23
</p>
</td>
<td>

```

```
...
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
 <p class="normal">

 Student Performance

 </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Predict student performance in secondary education (high school). </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification, Regression
</p>
</td>
<td>
<p class="normal">
 Integer
</p>
</td>
<td>
<p class="normal">
 649
</p>
</td>
<td>
<p class="normal">
 33
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
 <p class="normal">

 ElectricityLoadDiagrams20112014

 </p>
</td>

```

```

 `~~~
 </p>
 </td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This data set contains electricity consumption of 370
points/clients.
 </p></td> -->
<td>
<p class="normal">
 Time-Series
</p>
</td>
<td>
<p class="normal">
 Regression, Clustering
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 370
</p>
</td>
<td>
<p class="normal">
 140256
</p>
</td>
<td>
<p class="normal">
 2015
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Gas sensor array under dynamic gas mixtures

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The data set contains the recordings of 16 chemical sensors
exposed to two dynamic gas mixtures at varying concentrations. For each mixture, signals were acqui
red continuously during 12 hours. </p></td> -->
<td>
<p class="normal">
 Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification, Regression
</p>
</td>
<td>
<p class="normal">
 Real
</p>

```

```
 <td>
 <p class="normal">
 4178504
 </p>
 </td>
 <td>
 <p class="normal">
 19
 </p>
 </td>
 <td>
 <p class="normal">
 2015
 </p>
 </td>
 <!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 microblogPCU

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">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. </p></td> -->
 <td>
 <p class="normal">
 Multivariate, Univariate, Sequential, Text
 </p>
 </td>
 <td>
 <p class="normal">
 Classification, Causal-Discovery
 </p>
 </td>
 <td>
 <p class="normal">
 Integer, Real
 </p>
 </td>
 <td>
 <p class="normal">
 221579
 </p>
 </td>
 <td>
 <p class="normal">
 20
 </p>
 </td>
 <td>
 <p class="normal">
 2015
 </p>
 </td>
 <!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
```

```

<!--


```

```

<!-->
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 58509
</p>
</td>
<td>
<p class="normal">
 49
</p>
</td>
<td>
<p class="normal">
 2015
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 TV News Channel Commercial Detection Dataset

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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). </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification, Clustering
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 129685
</p>
</td>
<td>
<!-->

```

```
<p class="normal">
 12
</p>
</td>
<td>
<p class="normal">
 2015
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Phishing Websites

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset collected mainly from: PhishTank archive,
MillerSmiles archive, Google's searching operators. </p></td> -->
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Integer
</p>
</td>
<td>
<p class="normal">
 2456
</p>
</td>
<td>
<p class="normal">
 30
</p>
</td>
<td>
<p class="normal">
 2015
</p>
</td>
<!-- <td><p class="normal">Computer Security </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

```

```


 Greenhouse Gas Observing Network

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Design an observing network to monitor emissions of a greenhouse
gas (GHG) in California given time series of synthetic observations and tracers from weather model
simulations.
 </p></td> -->
<td>
<p class="normal">
 Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
 Regression
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 2921
</p>
</td>
<td>
<p class="normal">
 5232
</p>
</td>
<td>
<p class="normal">
 2015
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
 <p class="normal">

 Diabetic Retinopathy Debrecen Data Set

 </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset contains features extracted from the Messidor image
set to predict whether an image contains signs of diabetic retinopathy or not. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
<..>

```

```
</td>
<td>
<p class="normal">
 Integer, Real
</p>
</td>
<td>
<p class="normal">
 1151
</p>
</td>
<td>
<p class="normal">
 20
</p>
</td>
<td>
<p class="normal">
 2014
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 HIV-1 protease cleavage

 </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Categorical
</p>
</td>
<td>
<p class="normal">
 6590
</p>
</td>
<td>
<p class="normal">
 1
</p>
</td>
<td>
<p class="normal">
 2015
</p>
</td>
```

```

<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Sentiment Labelled Sentences

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The dataset contains sentences labelled with positive or
negative sentiment.&nbsp</p></td> -->
<td>
<p class="normal">
 Text
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 3000
</p>
</td>
<td>
<p class="normal">
 2015
</p>
</td>
<!-- <td><p class="normal">Other&nbsp</p></td> -->
</tr>
|
| |

```

articles published by Mashable in a period of two years. The goal is to predict the number of shares in social networks (popularity).&nbsp;</p></td> -->

<td>	<p class="normal"> Multivariate
</td>	</td>
<td>	<p class="normal"> Classification, Regression
</td>	</td>
<td>	<p class="normal"> Integer, Real
</td>	</td>
<td>	<p class="normal"> 39797
</td>	</td>
<td>	<p class="normal"> 61
</td>	</td>
<td>	<p class="normal"> 2015
</td>	</td>
<td>	<!-- <td><p class="normal">Business&nbsp;</p></td> -->
</tr>	
<tr>	
<td>	
<table>	
<tr>	
<td>	<a href="datasets/Forest+type+mapping"> 
</td>	
<td>	<p class="normal"> <b>
</td>	<a href="datasets/Forest+type+mapping"> Forest type mapping
</td>	</b>
</td>	</p>
</tr>	
</table>	
</td>	
<!-- <td><p class="normal">Multi-temporal remote sensing data of a forested area in Japan.	
The goal is to map different forest types using spectral data.&nbsp;</p></td> -->	
<td>	
<p class="normal"> Multivariate	
</p>	
</td>	
<td>	
<p class="normal"> Classification	
</p>	
</td>	
<td>	
<p class="normal"> 326	
</p>	
</td>	
<td>	

```

<p class="normal">
 27
</p>
</td>
<td>
 <p class="normal">
 2015
 </p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
 <p class="normal">

 wiki4HE

 </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Survey of faculty members from two Spanish universities on
teaching uses of Wikipedia </p></td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Regression, Clustering, Causal-Discovery
 </p>
</td>
<td>
 <p class="normal">
 913
 </p>
</td>
<td>
 <p class="normal">
 53
 </p>
</td>
<td>
 <p class="normal">
 2015
 </p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
 <p class="normal">


```

```


 Online Video Characteristics and Transcoding Time Dataset

 </p>
 </td>
 </tr>
 </table>
</td>
<!-- <td><p class="normal">The dataset contains a million randomly sampled video instances listing 10 fundamental video characteristics along with the YouTube video ID. </p></td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Regression
 </p>
</td>
<td>
 <p class="normal">
 Integer, Real
 </p>
</td>
<td>
 <p class="normal">
 168286
 </p>
</td>
<td>
 <p class="normal">
 11
 </p>
</td>
<td>
 <p class="normal">
 2015
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Chronic_Kidney_Disease

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">This dataset can be used to predict the chronic kidney disease and it can be collected from the hospital nearly 2 months of period. </p></td> -->
 <td>
 <p class="normal">
 Multivariate
 </p>
 </td>
 <td>
 <p class="normal">
 Classification
 </p>
 </td>
 <td>

```

```

<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 400
</p>
</td>
<td>
<p class="normal">
 25
</p>
</td>
<td>
<p class="normal">
 2015
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 <a href="datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014"
 Machine Learning based ZZAlpha Ltd. Stock Recommendations 2012-2014

 </p>
 </td>
 </tr>
 </table>
</td>
<!-- <td><p class="normal">The data here are the ZZAlpha® machine learning recommendations
made for various US traded stock portfolios the morning of each day during the 3 year period Jan 1
, 2012 - Dec 31, 2014. </p></td> -->
<td>
<p class="normal">
 Sequential, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 314080
</p>
</td>
<td>
<p class="normal">
 0
</p>
</td>
<td>
<p class="normal">
 2015
</p>
</td>
<!-- <td><p class="normal">Business </p></td> -->

```

```

</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Folio

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">20 photos of leaves for each of 32 different species. </p>-->
</td>
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification, Clustering
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 637
</p>
</td>
<td>
<p class="normal">
 20
</p>
</td>
<td>
<p class="normal">
 2015
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Taxi Service Trajectory - Prediction Challenge, ECML PKDD 2015

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">An accurate dataset describing trajectories performed by all the

```

442 taxis running in the city of Porto, in Portugal.

&nbsp;</p></td> -->

<td>

<p class="normal">  
    Multivariate, Sequential, Time-Series, Domain-Theory  
</p>

</td>

<td>

<p class="normal">  
    Clustering, Causal-Discovery  
</p>

</td>

<td>

<p class="normal">  
    Real  
</p>

</td>

<td>

<p class="normal">  
    1710671  
</p>

</td>

<td>

<p class="normal">  
    9  
</p>

</td>

<td>

<p class="normal">  
    2015  
</p>

</td>

<!-- <td><p class="normal">Computer&nbsp;</p></td> -->

```

</td>
<td>
<p class="normal">
 3
</p>
</td>
<td>
<p class="normal">
 2015
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
 <tr>
 <td>
 <a href="datasets/Smartphone-
Based+Recognition+of+Human+Activities+and+Postural+Transitions">

 </td>
 <td>
 <p class="normal">

 <a href="datasets/Smartphone-
Based+Recognition+of+Human+Activities+and+Postural+Transitions">
 Smartphone-Based Recognition of Human Activities and Postural Transitions

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">Activity recognition data set built from the recordings of 30
subjects performing basic activities and postural transitions while carrying a waist-mounted smart
phone with embedded inertial sensors.
 </p></td> -->
<td>
<p class="normal">
 Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 10929
</p>
</td>
<td>
<p class="normal">
 561
</p>
</td>
<td>
<p class="normal">
 2015
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
 <tr>
 <td>

```

```


</td>
<td>
 <p class="normal">

 Mice Protein Expression

 </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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 t
o assess associative learning. </p></td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Classification, Clustering
 </p>
</td>
<td>
 <p class="normal">
 Real
 </p>
</td>
<td>
 <p class="normal">
 1080
 </p>
</td>
<td>
 <p class="normal">
 82
 </p>
</td>
<td>
 <p class="normal">
 2015
 </p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 UJIIndoorLoc-Mag

 </p>
 </td>
 </tr>
 </table>
</td>
<!-- <td><p class="normal">The UJIIndoorLoc-Mag is an indoor localization database to test I
ndoor Positioning System that rely on Earth's magnetic field variations. </p></td> -->
<td>
 <p class="normal">
 Multivariate. Sequential. Time-Series

```

```

 Multivariate, Sequential, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification, Regression, Clustering
</p>
</td>
<td>
<p class="normal">
 Integer, Real
</p>
</td>
<td>
<p class="normal">
 40000
</p>
</td>
<td>
<p class="normal">
 13
</p>
</td>
<td>
<p class="normal">
 2015
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Heterogeneity Activity Recognition

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The Heterogeneity Human Activity Recognition (HHAR) dataset from
Smartphones and Smartwatches is a dataset devised to benchmark human activity recognition
algorithms (classification, automatic data segmentation, sensor fusion, feature extraction, etc.)
in real-world contexts; specifically, the dataset is gathered with a variety of different device m
odels and use-scenarios, in order to reflect sensing heterogeneities to be expected in real
deployments. </p></td> -->
<td>
<p class="normal">
 Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification, Clustering
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 43930257
</p>
</td>
<td>

```

```

<-->
<p class="normal">
 16
</p>
</td>
<td>
<p class="normal">
 2015
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Educational Process Mining (EPM) : A Learning Analytics Data Set

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">Educational Process Mining data set is built from the recordings of 115 subjects' activities through a logging application while learning with an educational simulator. </p></td> -->
<td>
 <p class="normal">
 Multivariate, Sequential, Time-Series
 </p>
</td>
<td>
 <p class="normal">
 Classification, Regression, Clustering
 </p>
</td>
<td>
 <p class="normal">
 Integer
 </p>
</td>
<td>
 <p class="normal">
 230318
 </p>
</td>
<td>
 <p class="normal">
 13
 </p>
</td>
<td>
 <p class="normal">
 2015
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
 <tr>
 <td>

 </td>
 </tr>

```

```

~ / ~
</td>
<td>
<p class="normal">

 HEPMASS

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p 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. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 10500000
</p>
</td>
<td>
<p class="normal">
 28
</p>
</td>
<td>
<p class="normal">
 2016
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Indoor User Movement Prediction from RSS data

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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></td> -->
<td>
<p class="normal">
 Multivariate, Sequential, Time-Series
</p>
</td>

```

```

~,~,
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 13197
</p>
</td>
<td>
<p class="normal">
 4
</p>
</td>
<td>
<p class="normal">
 2016
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Open University Learning Analytics dataset

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Open University Learning Analytics Dataset contains data about
courses, students and their interactions with Virtual Learning Environment for seven selected cour
ses and more than 30000 students. </p></td> -->
<td>
<p class="normal">
 Multivariate, Sequential, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification, Regression, Clustering
</p>
</td>
<td>
<p class="normal">
 Integer
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 2015
</p>

```

```

<td>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 default of credit card clients

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This research aimed at the case of customers' default payments
in Taiwan and compares the predictive accuracy of probability of default among six data mining met
hods.&nbsp</p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Integer, Real
</p>
</td>
<td>
<p class="normal">
 30000
</p>
</td>
<td>
<p class="normal">
 24
</p>
</td>
<td>
<p class="normal">
 2016
</p>
</td>
<!-- <td><p class="normal">Business </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Mesothelioma's disease data set

</p>
</td>
</tr>

```

```

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">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 featu
res. </p></td> -->
 <td>
 <p class="normal">
 Multivariate
 </p>
 </td>
 <td>
 <p class="normal">
 Classification
 </p>
 </td>
 <td>
 <p class="normal">
 Real
 </p>
 </td>
 <td>
 <p class="normal">
 324
 </p>
 </td>
 <td>
 <p class="normal">
 34
 </p>
 </td>
 <td>
 <p class="normal">
 2016
 </p>
 </td>
 <!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Online Retail

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">This is a transnational data set which contains all the
transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store o
nline retail. </p></td> -->
 <td>
 <p class="normal">
 Multivariate, Sequential, Time-Series
 </p>
 </td>
 <td>
 <p class="normal">
 Classification, Clustering
 </p>
 </td>
 <td>
 <p class="normal">
 ...
 </p>
 </td>

```

```
 integer, real
 </p>
</td>
<td>
<p class="normal">
 541909
</p>
</td>
<td>
<p class="normal">
 8
</p>
</td>
<td>
<p class="normal">
 2015
</p>
</td>
<!-- <td><p class="normal">Business </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 SIFT10M

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">In SIFT10M, each data point is a SIFT feature which is extracted
from Caltech-256 by the open source VLFeat library. The corresponding patches of the SIFT features
are provided. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Causal-Discovery
</p>
</td>
<td>
<p class="normal">
 Integer
</p>
</td>
<td>
<p class="normal">
 11164866
</p>
</td>
<td>
<p class="normal">
 128
</p>
</td>
<td>
<p class="normal">
 2016
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
`.
```

```

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

</td>
<td>
<p class="normal">

 GPS Trajectories

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification, Regression
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 163
</p>
</td>
<td>
<p class="normal">
 15
</p>
</td>
<td>
<p class="normal">
 2016
</p>
</td>
<!-- <td><p class="normal">Computer&nbs;p;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Detect Malacious Executable(AntiVirus)

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">I extract features from malacious and non-malacious and create a

```

nd training dataset to teach svm classifier.Dataset made of unknown executable to detect if it is virus or normal safe executable.&nbsp;</p></td> -->

<td>	<p class="normal">
	Multivariate
	</p>
</td>	
<td>	<p class="normal">
	Classification
	</p>
</td>	
<td>	<p class="normal">
	Real
	</p>
</td>	
<td>	<p class="normal">
	373
	</p>
</td>	
<td>	<p class="normal">
	513
	</p>
</td>	
<td>	<p class="normal">
	2016
	</p>
</td>	
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->	
</tr>	
<tr>	
<td>	
<table>	
<tr>	
<td>	<a href="datasets/Occupancy+Detection+">
	</a>
</td>	
<td>	<p class="normal">
	<b>
	<a href="datasets/Occupancy+Detection+">
	Occupancy Detection
	</a>
	</b>
	</p>
</td>	
</tr>	
</table>	
</td>	
<!-- <td><p class="normal">Experimental data used for binary classification (room occupancy) from Temperature, Humidity, Light and CO2. Ground-truth occupancy was obtained from time stamped pictures that were taken every minute.&nbsp;</p></td> -->	
<td>	
<p class="normal">	
Multivariate, Time-Series	
</p>	
</td>	
<td>	
<p class="normal">	
Classification	
</p>	
</td>	
<td>	
<p class="normal">	
Real	
</p>	
</td>	
<td>	
<p class="normal">	
20560	
</p>	

```

</td>
<td>
<p class="normal">
 7
</p>
</td>
<td>
<p class="normal">
 2016
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">
 Improved Spiral Test Using Digitized Graphics Tablet for Monitoring Parkinson's Disease
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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.&nbsp</p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification, Regression, Clustering
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 40
</p>
</td>
<td>
<p class="normal">
 7
</p>
</td>
<td>
<p class="normal">
 2016
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>

```

```

<tr>
<td>

</td>
<td>
<p class="normal">

 News Aggregator

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">References to news pages collected from an web aggregator in the
period from 10-March-2014 to 10-August-2014. The resources are grouped into clusters that represent
pages discussing the same story. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification, Clustering
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 422937
</p>
</td>
<td>
<p class="normal">
 5
</p>
</td>
<td>
<p class="normal">
 2016
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Air Quality

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Contains the responses of a gas multisensor device deployed on t
he field in an Italian city. Hourly responses averages are recorded along with gas concentrations
references from a certified analyzer. </p></td> -->
<td>

```

```

<p class="normal">
 Multivariate, Time-Series
</p>
</td>
<td>
 <p class="normal">
 Regression
 </p>
</td>
<td>
 <p class="normal">
 Real
 </p>
</td>
<td>
 <p class="normal">
 9358
 </p>
</td>
<td>
 <p class="normal">
 15
 </p>
</td>
<td>
 <p class="normal">
 2016
 </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Twin gas sensor arrays

 </p>
 </td>
 </tr>
 </table>
</td>
<!-- <td><p class="normal">5 replicates of an 8-MOX gas sensor array were exposed to different gas conditions (4 volatiles at 10 concentration levels each). </p></td> -->
<td>
 <p class="normal">
 Multivariate, Time-Series, Domain-Theory
 </p>
</td>
<td>
 <p class="normal">
 Classification, Regression
 </p>
</td>
<td>
 <p class="normal">
 Real
 </p>
</td>
<td>
 <p class="normal">
 640
 </p>
</td>
<td>
 <p class="normal">
 480000
 </p>

```

```

</p>
</td>
<td>
<p class="normal">
 2016
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Gas sensors for home activity monitoring

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">100 recordings of a sensor array under different conditions in a
home setting: background, wine and banana presentations. The array includes 8 MOX gas sensors, and
humidity and temperature sensors.
 </p></td> -->
<td>
<p class="normal">
 Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 919438
</p>
</td>
<td>
<p class="normal">
 11
</p>
</td>
<td>
<p class="normal">
 2016
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

```

```


 Facebook Comment Volume Dataset

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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 re
ceive.&nbsp;</p></td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Regression
 </p>
</td>
<td>
 <p class="normal">
 Integer, Real
 </p>
</td>
<td>
 <p class="normal">
 40949
 </p>
</td>
<td>
 <p class="normal">
 54
 </p>
</td>
<td>
 <p class="normal">
 2016
 </p>
</td>
<!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
 <td>
 <table>
 <tr>
 <td>
 <a
 href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Livi
28AAL%29">

 </td>
 <td>
 <p class="normal">

 <a
 href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Livi
28AAL%29">
 Smartphone Dataset for Human Activity Recognition (HAR) in Ambient Assisted Living (A
AL)

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">This data is an addition to an existing dataset on UCI. We
collected more data to improve the accuracy of our human activity recognition algorithms applied i
n the domain of Ambient Assisted Living. &nbsp;</p></td> -->
 <td>
 <p class="normal">
 Time-Series
</pre>

```

```
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 5744
</p>
</td>
<td>
<p class="normal">
 561
</p>
</td>
<td>
<p class="normal">
 2016
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Polish companies bankruptcy data

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The dataset is about bankruptcy prediction of Polish companies. The bankrupt companies were analyzed in the period 2000-2012, while the still operating companies were evaluated from 2007 to 2013. </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 10503
</p>
</td>
<td>
<p class="normal">
 64
</p>
```

```

</td>
<td>
<p class="normal">
2016
</p>
</td>
<!-- <td><p class="normal">Business </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>


```

```


 Dota2 Games Results

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">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. </p></td> -->
<td>
 <p class="normal">
 Multivariate
 </p>
</td>
<td>
 <p class="normal">
 Classification
 </p>
</td>
<td>
 <p class="normal">
 102944
 </p>
</td>
<td>
 <p class="normal">
 116
 </p>
</td>
<td>
 <p class="normal">
 2016
 </p>
</td>
<!-- <td><p class="normal">Game </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
 <td>
 <table>
 <tr>
 <td>

 </td>
 <td>
 <p class="normal">

 Facebook metrics

 </p>
 </td>
 </tr>
 </table>
 </td>
 <!-- <td><p class="normal">Facebook performance metrics of a renowned cosmetic's brand Facek
ook page. </p></td> -->
 <td>
 <p class="normal">
 Multivariate
 </p>
 </td>
 <td>
 <p class="normal">
 Regression
 </p>
 </td>

```

```

 .. .
<td>
<p class="normal">
 Integer
</p>
</td>
<td>
<p class="normal">
 500
</p>
</td>
<td>
<p class="normal">
 19
</p>
</td>
<td>
<p class="normal">
 2016
</p>
</td>
<!-- <td><p class="normal">Business </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 UbiqLog (smartphone lifelogging)

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">UbiqLog is the smartphone lifelogging tool that runs on the
smartphone of 35 users for about 2 months.
 </p></td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Causal-Discovery
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 9782222
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
 2016
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bacolor="#DDEEFF">

```

```

<-- >>>>> <td>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

NIPS Conference Papers 1987-2015

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This data set contains the distribution of words in the full
text of the NIPS conference papers published from 1987 to 2015. </p></td> -->
<td>
<p class="normal">
Text
</p>
</td>
<td>
<p class="normal">
Clustering
</p>
</td>
<td>
<p class="normal">
Integer
</p>
</td>
<td>
<p class="normal">
11463
</p>
</td>
<td>
<p class="normal">
5812
</p>
</td>
<td>
<p class="normal">
2016
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

HTRU2

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Pulsar candidates collected during the HTRU survey. Pulsars are
a type of star of considerable scientific interest. Candidates must be classified in to pulsar

```

a type of star, of considerable scientific interest. Candidates must be classified in to pulsar and non-pulsar classes to aid discovery.&nbsp;</p></td> -->

```
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification, Clustering
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 17898
</p>
</td>
<td>
<p class="normal">
 9
</p>
</td>
<td>
<p class="normal">
 2017
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Drug consumption (quantified)

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Classify type of drug consumer by personality data </p>
</td> -->
<td>
<p class="normal">
 Multivariate
</p>
</td>
<td>
<p class="normal">
 Classification
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 1885
</p>
</td>
<!-->
```

```

<!-->
<p class="normal">
 32
</p>
</td>
<td>
<p class="normal">
 2016
</p>
</td>
<!-- <td><p class="normal">Social </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 Appliances energy prediction

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Experimental data used to create regression models of appliances
energy use in a low energy building. </p></td> -->
<td>
<p class="normal">
 Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
 Regression
</p>
</td>
<td>
<p class="normal">
 Real
</p>
</td>
<td>
<p class="normal">
 19735
</p>
</td>
<td>
<p class="normal">
 29
</p>
</td>
<td>
<p class="normal">
 2017
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

```

```

<p class="normal">

 Miskolc IIS Hybrid IPS

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The dataset was created for the comparison and evaluation of
hybrid indoor positioning methods. The dataset presented contains data from W-LAN and Bluetooth in
terfaces, and Magnetometer. </p></td> -->
<td>
<p class="normal">
 Text
</p>
</td>
<td>
<p class="normal">
 Classification, Clustering, Causal-Discovery
</p>
</td>
<td>
<p class="normal">
 Integer
</p>
</td>
<td>
<p class="normal">
 1540
</p>
</td>
<td>
<p class="normal">
 67
</p>
</td>
<td>
<p class="normal">
 2016
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>

</td>
<td>
<p class="normal">

 KDC-4007 dataset Collection

</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">KDC-4007 dataset Collection is the Kurdish Documents
Classification text used in categories regarding Kurdish Sorani news and articles. </p></td>
-->
<td>
<p class="normal">
 Multivariate, Text
</p>
</td>
<td>
<p class="normal">
 Classification, Clustering, Causal-Discovery
</p>
</td>

```

```

Classification, Regression
</p>
</td>
<td>
<p class="normal">
 Integer
</p>
</td>
<td>
<p class="normal">
 4007
</p>
</td>
<td>
<p class="normal">
 2017
</p>
</td>
<td>
<p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
 <tr>
 <td>
 <a href="datasets/Geo-
Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone">

 </td>
 <td>
 <p class="normal">

 <a href="datasets/Geo-
Magnetic field and WLAN dataset for indoor localisation from wristband and smartp
hone

 </p>
 </td>
 </tr>
</table>
</td>
<!-- <td><p class="normal">A multisource and multivariate dataset for indoor localisation m
ethods based on WLAN and Geo-Magnetic field fingerprinting </p></td> -->
<td>
<p class="normal">
 Multivariate, Sequential, Time-Series
</p>
</td>
<td>
<p class="normal">
 Classification, Regression, Clustering
</p>
</td>
<td>
<p class="normal">
 Integer, Real
</p>
</td>
<td>
<p class="normal">
 153540
</p>
</td>
<td>
<p class="normal">
 25
</p>
</td>
<td>
<p class="normal">
  ~~~~
```

```

    201 /
</p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/DrivFace">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/DrivFace">
    DrivFace
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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.&nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Classification, Regression, Clustering
</p>
</td>
<td>
<p class="normal">
    Real
</p>
</td>
<td>
<p class="normal">
    606
</p>
</td>
<td>
<p class="normal">
    6400
</p>
</td>
<td>
<p class="normal">
    2016
</p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Website+Phishing">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Website+Phishing">
    Website Phishing
</a>
</b>
</p>
</td>

```

```

        </p>
    </td>
</tr>
</table>
</td>
<!-- <td><p class="normal">

&nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">
    Integer
</p>
</td>
<td>
<p class="normal">
    1353
</p>
</td>
<td>
<p class="normal">
    10
</p>
</td>
<td>
<p class="normal">
    2016
</p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/YouTube+Spam+Collection">
    
</a>
</td>
<td>
<p class="normal">
<b>
    <a href="datasets/YouTube+Spam+Collection">
        YouTube Spam Collection
    </a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">It is a public set of comments collected for spam research. It has five datasets composed by 1,956 real messages extracted from five videos that were among the 10 most viewed on the collection period.&nbsp;</p></td> -->
<td>
<p class="normal">
    Text
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">
    .
</p>

```

```

</td>
<td>
<p class="normal">
  1956
</p>
</td>
<td>
<p class="normal">
  5
</p>
</td>
<td>
<p class="normal">
  2017
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Beijing+PM2.5+Data">
  
</a>
</td>
<td>
<p class="normal">
<b>
  <a href="datasets/Beijing+PM2.5+Data">
    Beijing PM2.5 Data
  </a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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.  </p></td> -->
<td>
<p class="normal">
  Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
  Regression
</p>
</td>
<td>
<p class="normal">
  Integer, Real
</p>
</td>
<td>
<p class="normal">
  43824
</p>
</td>
<td>
<p class="normal">
  13
</p>
</td>
<td>
<p class="normal">
  2017
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr>
<td>
<table>

```

```

<tr>
<td>
<a href="datasets/Cargo+2000+Freight+Tracking+and+Tracing">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Cargo+2000+Freight+Tracking+and+Tracing">
Cargo 2000 Freight Tracking and Tracing
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Sanitized and anonymized Cargo 2000 (C2K) airfreight tracking and tracing events, covering five months of business execution (3,942 process instances, 7,932 transport legs, 56,082 activities). &nbsp;</p></td> -->
<td>
<p class="normal">
Multivariate, Sequential
</p>
</td>
<td>
<p class="normal">
Classification, Regression
</p>
</td>
<td>
<p class="normal">
Integer
</p>
</td>
<td>
<p class="normal">
3942
</p>
</td>
<td>
<p class="normal">
98
</p>
</td>
<td>
<p class="normal">
2016
</p>
</td>
<!-- <td><p class="normal">Business&nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Cervical+cancer+%28Risk+Factors%29">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Cervical+cancer+%28Risk+Factors%29">
Cervical cancer (Risk Factors)
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset focuses on the prediction of indicators/diagnosis of cervical cancer. The features cover demographic information, habits, and historic medical records.&nbsp;</p></td> -->

```

```
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">
    Integer, Real
</p>
</td>
<td>
<p class="normal">
    858
</p>
</td>
<td>
<p class="normal">
    36
</p>
</td>
<td>
<p class="normal">
    2017
</p>
</td>
<!-- <td><p class="normal">Life &nbsp;</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
    <a href="datasets/Quality+Assessment+of+Digital+Colposcopies">
        
    </a>
</td>
<td>
    <p class="normal">
        <b>
            <a href="datasets/Quality+Assessment+of+Digital+Colposcopies">
                Quality Assessment of Digital Colposcopies
            </a>
        </b>
    </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset explores the subjective quality assessment of
digital colposcopies.&nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">
    Real
</p>
</td>
<td>
<p class="normal">
    287
</p>
</td>
<td>
<p class="normal">
```

```
    69
  </p>
</td>
<td>
<p class="normal">
  2017
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/KASANDR">
  
</a>
</td>
<td>
<p class="normal">
<b>
  <a href="datasets/KASANDR">
    KASANDR
  </a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">KASANDR is a novel, publicly available collection for
recommendation systems that records the behavior of customers of the European leader in e-Commerce
advertising, Kelkoo. &nbsp;</p></td> -->
<td>
<p class="normal">
  Multivariate
</p>
</td>
<td>
<p class="normal">
  Causal-Discovery
</p>
</td>
<td>
<p class="normal">
  Integer
</p>
</td>
<td>
<p class="normal">
  17764280
</p>
</td>
<td>
<p class="normal">
  2158859
</p>
</td>
<td>
<p class="normal">
  2017
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/FMA%3A+A+Dataset+For+Music+Analysis">
  
</a>
</td>
<td>
<p class="normal">
```

```

<b>
  <a href="datasets/FMA%3A+A+Dataset+For+Music+Analysis">
    FMA: A Dataset For Music Analysis
  </a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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.&nbsp;</p></td> -->
<td>
  <p class="normal">
    Multivariate, Time-Series
  </p>
</td>
<td>
  <p class="normal">
    Classification, Clustering
  </p>
</td>
<td>
  <p class="normal">
    Real
  </p>
</td>
<td>
  <p class="normal">
    106574
  </p>
</td>
<td>
  <p class="normal">
    518
  </p>
</td>
<td>
  <p class="normal">
    2017
  </p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
  <table>
    <tr>
      <td>
        <a href="datasets/Air+quality">
          
        </a>
      </td>
      <td>
        <p class="normal">
          <b>
            <a href="datasets/Air+quality">
              Air quality
            </a>
          </b>
        </p>
      </td>
    </tr>
  </table>
</td>
<!-- <td><p class="normal"> Contains the responses of a gas multisensor device deployed on
the field in an Italian city. &nbsp;</p></td> -->
<td>
  <p class="normal">
    Multivariate, Time-Series
  </p>
</td>
<td>
  <p class="normal">
    Regression
  </p>
</td>

```

```
</td>
<td>
<p class="normal">
  Real
</p>
</td>
<td>
<p class="normal">
  9358
</p>
</td>
<td>
<p class="normal">
  15
</p>
</td>
<td>
<p class="normal">
  2016
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Epileptic+Seizure+Recognition">
  
</a>
</td>
<td>
<p class="normal">
<b>
  <a href="datasets/Epileptic+Seizure+Recognition">
    Epileptic Seizure Recognition
  </a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset is a pre-processed and re-structured/reshaped
version of a very commonly used dataset featuring epileptic seizure detection.  </p></td> -->
<td>
<p class="normal">
  Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
  Classification, Clustering
</p>
</td>
<td>
<p class="normal">
  Integer, Real
</p>
</td>
<td>
<p class="normal">
  11500
</p>
</td>
<td>
<p class="normal">
  179
</p>
</td>
<td>
<p class="normal">
  2017
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
```

```

</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Devanagari+Handwritten+Character+Dataset">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Devanagari+Handwritten+Character+Dataset">
    Devanagari Handwritten Character Dataset
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This is an image database of Handwritten Devanagari characters.
There are 46 classes of characters with 2000 examples each. The dataset is split into training set
(85%) and testing set(15%). &nbsp;</p></td> -->
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">
    Integer
</p>
</td>
<td>
<p class="normal">
    92000
</p>
</td>
<td>
<p class="normal">
    2016
</p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Stock+portfolio+performance">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Stock+portfolio+performance">
    Stock portfolio performance
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The data set of performances of weighted scoring stock portfolios

```

are obtained with mixture design from the US stock market historical database.&nbsp;</p></td> -->

<td>	<p class="normal"> Multivariate
</td>	</p>
<td>	<p class="normal"> Regression
</td>	</p>
<td>	<p class="normal"> Real
</td>	</p>
<td>	<p class="normal"> 315
</td>	</p>
<td>	<p class="normal"> 12
</td>	</p>
<td>	<p class="normal"> 2016
</td>	</p>
<!-- <td><p class="normal">Business&nbsp;</p></td> -->	
</tr>	
<tr bgcolor="#DDEEFF">	
<td>	
<table>	
<tr>	
<td>	
<a href="datasets/MoCap+Hand+Postures">	
</a>	
</td>	
<td>	
<p class="normal">	
<b>	
<a href="datasets/MoCap+Hand+Postures">	
MoCap Hand Postures	
</a>	
</b>	
</p>	
</td>	
</tr>	
</table>	
</td>	
<!-- <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.&nbsp;</p></td> -->	
<td>	
<p class="normal">	
Multivariate	
</p>	
</td>	
<td>	
<p class="normal">	
Classification, Clustering	
</p>	
</td>	
<td>	
<p class="normal">	
Integer, Real	
</p>	
</td>	
<td>	
<p class="normal">	
78095	
</p>	
</td>	

```

<td>
<p class="normal">
38
</p>
</td>
<td>
<p class="normal">
2016
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech">


```

```

<td> <a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting">
    
  </a>
</td>
<td>
  <p class="normal">
    <b>
      <a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting">
        Data for Software Engineering Teamwork Assessment in Education Setting
      </a>
    </b>
  </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Data include over 100 Team Activity Measures and outcomes (ML classes) obtained from activities of 74 student teams during the creation of final class project in SW Eng. classes at SFSU, Fulda, FAU&ampnbsp;</p></td> -->
<td>
  <p class="normal">
    Sequential, Time-Series
  </p>
</td>
<td>
  <p class="normal">
    Classification
  </p>
</td>
<td>
  <p class="normal">
    Integer, Real
  </p>
</td>
<td>
  <p class="normal">
    74
  </p>
</td>
<td>
  <p class="normal">
    102
  </p>
</td>
<td>
  <p class="normal">
    2017
  </p>
</td>
<!-- <td><p class="normal">Computer&ampnbsp;</p></td> -->
</tr>
<tr>
<td>
  <table>
    <tr>
      <td>
        <a href="datasets/PM2.5+Data+of+Five+Chinese+Cities">
          
        </a>
      </td>
      <td>
        <p class="normal">
          <b>
            <a href="datasets/PM2.5+Data+of+Five+Chinese+Cities">
              PM2.5 Data of Five Chinese Cities
            </a>
          </b>
        </p>
      </td>
    </tr>
  </table>
</td>
<!-- <td><p class="normal">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.&ampnbsp;</p></td> -->
<td>
  <p class="normal">

```

```

<p class="normal">
    Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
    Regression
</p>
</td>
<td>
<p class="normal">
    Integer, Real
</p>
</td>
<td>
<p class="normal">
    52854
</p>
</td>
<td>
<p class="normal">
    86
</p>
</td>
<td>
<p class="normal">
    2017
</p>
</td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
    <a href="datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet">
        
    </a>
</td>
<td>
    <p class="normal">
        <b>
            <a href="datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet">
                Parkinson Disease Spiral Drawings Using Digitized Graphics Tablet
            </a>
        </b>
    </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Handwriting database consists of 62 PWP(People with Parkinson) and 15 healthy individuals. Three types of recordings (Static Spiral Test, Dynamic Spiral Test and Stability Test) are taken.&nbsp;</p></td> -->
<td>
    <p class="normal">
        Multivariate
    </p>
</td>
<td>
<p class="normal">
    Classification, Regression, Clustering
</p>
</td>
<td>
<p class="normal">
    Integer
</p>
</td>
<td>
<p class="normal">
    77
</p>
</td>
<td>
<p class="normal">
    7

```

```

        '
    </p>
</td>
<td>
<p class="normal">
    2017
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Sales_Transactions_Dataset_Weekly">
    
</a>
</td>
<td>
<p class="normal">
    <b>
        <a href="datasets/Sales_Transactions_Dataset_Weekly">
            Sales_Transactions_Dataset_Weekly
        </a>
    </b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Contains weekly purchased quantities of 800 over products over
52 weeks. Normalised values are provided too.&nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
    Clustering
</p>
</td>
<td>
<p class="normal">
    Integer, Real
</p>
</td>
<td>
<p class="normal">
    811
</p>
</td>
<td>
<p class="normal">
    53
</p>
</td>
<td>
<p class="normal">
    2017
</p>
</td>
<!-- <td><p class="normal">&nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/LasVegasStrip">
    
</a>
</td>
<td>
<p class="normal">
    <b>
        <a href="datasets/LasVegasStrip">
            Las+Vegas+Strip
        </a>
    </b>
</p>

```



```
<p class="normal">
</p>
</td>
<td>
<p class="normal">
401
</p>
</td>
<td>
<p class="normal">
1
</p>
</td>
<td>
<p class="normal">
2017
</p>
</td>
<!-- <td><p class="normal">Business&nbsp;</p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/MEU-Mobile+KSD">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/MEU-Mobile+KSD">
MEU-Mobile KSD
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset contains keystroke dynamics data collected on a
touch mobile device (Nexus 7). The dataset contains 2856 records, 51 records per subject for 56 su
bjects. &nbsp;</p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Integer, Real
</p>
</td>
<td>
<p class="normal">
2856
</p>
</td>
<td>
<p class="normal">
71
</p>
</td>
<td>
<p class="normal">
2016
</p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr>
<td>
```

```

<ta>
<table>
<tr>
<td>
<a href="datasets/Crowdsourced+Mapping">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Crowdsourced+Mapping">
    Crowdsourced Mapping
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Crowdsourced data from OpenStreetMap is used to automate the classification of satellite images into different land cover classes (impervious, farm, forest, grass, orchard, water). &nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">
    10546
</p>
</td>
<td>
<p class="normal">
    29
</p>
</td>
<td>
<p class="normal">
    2016
</p>
</td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/gene+expression+cancer+RNA-Seq">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/gene+expression+cancer+RNA-Seq">
    gene expression cancer RNA-Seq
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This collection of data is part of the RNA-Seq (HiSeq) PANCAN data set, it is a random extraction of gene expressions of patients having different types of tumo-->

```

```

r: BRCA, KIRC, COAD, LUAD and PRAD.&nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Classification, Clustering
</p>
</td>
<td>
<p class="normal">
    Real
</p>
</td>
<td>
<p class="normal">
    801
</p>
</td>
<td>
<p class="normal">
    20531
</p>
</td>
<td>
<p class="normal">
    2016
</p>
</td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer">
    
</a>
</td>
<td>
<p class="normal">
    <b>
        <a href="datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer">
            Hybrid Indoor Positioning Dataset from WiFi RSSI, Bluetooth and magnetometer
        </a>
    </b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The dataset was created for the comparison and evaluation of
hybrid indoor positioning methods. The dataset presented contains data from W-LAN and Bluetooth in
terfaces, and Magnetometer. &nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate, Sequential, Time-Series
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">
    Real
</p>
</td>
<td>
<p class="normal">
    1540
</p>

```

```

</p>
</td>
<td>
<p class="normal">
  65
</p>
</td>
<td>
<p class="normal">
  2016
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/chestnut+E2%80%93+LARVIC">
  
</a>
</td>
<td>
<p class="normal">
<b>
  <a href="datasets/chestnut+E2%80%93+LARVIC">
    chestnut - LARVIC
  </a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The research project presents this database, shows the images of
chestnuts that will be processed to determine the presence or absence of defects </p></td> --
>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
  Classification, Clustering
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
  1451
</p>
</td>
<td>
<p class="normal">
  3
</p>
</td>
<td>
<p class="normal">
  2017
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a
href="datasets/Burst+Header+Packet+28BHP%29+flooding+attack+on+Optical+Burst+Switching+28OBS%29+N
rk">
  
</td>

```

```

        </a>
    </td>
    <td>
        <p class="normal">
            <b>
                <a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%28OBS%29+Nrk">
                    Burst Header Packet (BHP) flooding attack on Optical Burst Switching (OBS) Network
                </a>
            </b>
        </p>
    </td>
    </tr>
</table>
</td>
<!-- &lt;td&gt;&lt;p class="normal"&gt;One of the primary challenges in identifying the risks of the Burst Header Packet (BHP) flood attacks in Optical Burst Switching networks (OBS) is the scarcity of reliable historical data. &amp;nbsp;&lt;/p&gt;&lt;/td&gt; --&gt;
&lt;td&gt;
    &lt;p class="normal"&gt;
        Text
    &lt;/p&gt;
&lt;/td&gt;
&lt;td&gt;
    &lt;p class="normal"&gt;
        Classification
    &lt;/p&gt;
&lt;/td&gt;
&lt;td&gt;
    &lt;p class="normal"&gt;
        Integer
    &lt;/p&gt;
&lt;/td&gt;
&lt;td&gt;
    &lt;p class="normal"&gt;
        1075
    &lt;/p&gt;
&lt;/td&gt;
&lt;td&gt;
    &lt;p class="normal"&gt;
        22
    &lt;/p&gt;
&lt;/td&gt;
&lt;td&gt;
    &lt;p class="normal"&gt;
        2017
    &lt;/p&gt;
&lt;/td&gt;
<!-- &lt;td&gt;&lt;p class="normal"&gt;Computer&amp;nbsp;&lt;/p&gt;&lt;/td&gt; --&gt;
&lt;/tr&gt;
&lt;tr bgcolor="DDEEFF"&gt;
    &lt;td&gt;
        &lt;table&gt;
            &lt;tr&gt;
                &lt;td&gt;
                    &lt;a href="datasets/Motion+Capture+Hand+Postures"&gt;
                        &lt;img border="1" src="assets/MLimages/SmallLargedefault.jpg"/&gt;
                    &lt;/a&gt;
                &lt;/td&gt;
                &lt;td&gt;
                    &lt;p class="normal"&gt;
                        &lt;b&gt;
                            &lt;a href="datasets/Motion+Capture+Hand+Postures"&gt;
                                Motion Capture Hand Postures
                            &lt;/a&gt;
                        &lt;/b&gt;
                    &lt;/p&gt;
                &lt;/td&gt;
            &lt;/tr&gt;
        &lt;/table&gt;
    &lt;/td&gt;
    &lt;!-- &lt;td&gt;&lt;p class="normal"&gt;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.&amp;nbsp;&lt;/p&gt;&lt;/td&gt; --&gt;
    &lt;td&gt;
        &lt;p class="normal"&gt;
</pre>

```

```

        Multivariate
    </p>
</td>
<td>
<p class="normal">
    Classification, Clustering
</p>
</td>
<td>
<p class="normal">
    Real
</p>
</td>
<td>
<p class="normal">
    78095
</p>
</td>
<td>
<p class="normal">
    38
</p>
</td>
<td>
<p class="normal">
    2017
</p>
</td>
<!-- <td><p class="normal">Computer &nbsp;</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Anuran+Calls+%28MFCCs%29">
    
</a>
</td>
<td>
<p class="normal">
    <b>
        <a href="datasets/Anuran+Calls+%28MFCCs%29">
            Anuran Calls (MFCCs)
        </a>
    </b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Acoustic features extracted from syllables of anuran (frogs)
calls, including the family, the genus, and the species labels (multilabel). &nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Classification, Clustering
</p>
</td>
<td>
<p class="normal">
    Real
</p>
</td>
<td>
<p class="normal">
    7195
</p>
</td>
<td>
<p class="normal">
    22
</p>

```

```

</td>
<td>
<p class="normal">
  2017
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization">
  
</a>
</td>
<td>
<p class="normal">
<b>
  <a href="datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization">
    TTC-3600: Benchmark dataset for Turkish text categorization
  </a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The TTC-3600 data set is a collection of Turkish news and
articles including categorized 3,600 documents from 6 well-known portals in Turkey. It has 4
different forms in ARFF Weka format.&nbsp;</p></td> -->
<td>
<p class="normal">
  Text
</p>
</td>
<td>
<p class="normal">
  Classification, Clustering
</p>
</td>
<td>
<p class="normal">
  Integer
</p>
</td>
<td>
<p class="normal">
  3600
</p>
</td>
<td>
<p class="normal">
  4814
</p>
</td>
<td>
<p class="normal">
  2017
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy">
  
</a>
</td>
<td>
<p class="normal">
<b>
  <a href="datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy">

```

```

        Gastrointestinal Lesions in Regular Colonoscopy
    </a>
    </b>
    </p>
    </td>
    </tr>
    </table>
</td>
<!-- <td><p class="normal">This dataset contains features extracted from colonoscopy videos used to detect gastrointestinal lesions. It contains 76 lesions: 15 serrated adenomas, 21 hyperplastic lesions and 40 adenoma. &nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">
    Real
</p>
</td>
<td>
<p class="normal">
    76
</p>
</td>
<td>
<p class="normal">
    698
</p>
</td>
<td>
<p class="normal">
    2016
</p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
    <td>
        <table>
            <tr>
                <td>
                    <a href="datasets/Daily+Demand+Forecasting+Orders">
                        
                    </a>
                </td>
                <td>
                    <p class="normal">
                        <b>
                            <a href="datasets/Daily+Demand+Forecasting+Orders">
                                Daily Demand Forecasting Orders
                            </a>
                        </b>
                    </p>
                </td>
            </tr>
        </table>
    </td>
    <!-- <td><p class="normal">The dataset was collected during 60 days, this is a real database of a brazilian logistics company.&nbsp;</p></td> -->
    <td>
        <p class="normal">
            Time-Series
        </p>
    </td>
    <td>
        <p class="normal">
            Regression
        </p>
    </td>
    <td>

```

```

<p class="normal">
  Integer
</p>
</td>
<td>
  <p class="normal">
    60
  </p>
</td>
<td>
  <p class="normal">
    13
  </p>
</td>
<td>
  <p class="normal">
    2017
  </p>
</td>
<!-- <td><p class="normal">Business </p></td> -->
</tr>
<tr>
<td>
<table>
  <tr>
    <td>
      <a href="datasets/Paper+Reviews">
        
      </a>
    </td>
    <td>
      <p class="normal">
        <b>
          <a href="datasets/Paper+Reviews">
            Paper Reviews
          </a>
        </b>
      </p>
    </td>
  </tr>
</table>
</td>
<!-- <td><p class="normal">This sentiment analysis data set contains scientific paper
reviews from an international conference on computing and informatics. The task is to predict the
orientation or the evaluation of a review. </p></td> -->
<td>
  <p class="normal">
    Text
  </p>
</td>
<td>
  <p class="normal">
    Classification, Regression
  </p>
</td>
<td>
  <p class="normal">
    Integer
  </p>
</td>
<td>
  <p class="normal">
    405
  </p>
</td>
<td>
  <p class="normal">
    10
  </p>
</td>
<td>
  <p class="normal">
    2017
  </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>

```

```
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/extention+of+Z-Alizadeh+sani+dataset">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/extention+of+Z-Alizadeh+sani+dataset">
extention of Z-Alizadeh sani dataset
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!!-- <td><p class="normal">It was collected for CAD diagnosis.&nbsp;</p></td> -->
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Integer, Real
</p>
</td>
<td>
<p class="normal">
303
</p>
</td>
<td>
<p class="normal">
59
</p>
</td>
<td>
<p class="normal">
2017
</p>
</td>
<!!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Z-Alizadeh+Sani">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Z-Alizadeh+Sani">
Z-Alizadeh Sani
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!!-- <td><p class="normal">It was collected for CAD diagnosis.&nbsp;</p></td> -->
<td>
<p class="normal">
```

```
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Integer, Real
</p>
</td>
<td>
<p class="normal">
303
</p>
</td>
<td>
<p class="normal">
56
</p>
</td>
<td>
<p class="normal">
2017
</p>
</td>
<!-- <td><p class="normal">Life &nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Dynamic+Features+of+VirusShare+Executables">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Dynamic+Features+of+VirusShare+Executables">
Dynamic Features of VirusShare Executables
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset contains the dynamic features of 107,888
executables, collected by VirusShare from Nov/2010 to Jul/2014.&nbsp;</p></td> -->
<td>
<p class="normal">
Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
Classification, Regression
</p>
</td>
<td>
<p class="normal">
Integer
</p>
</td>
<td>
<p class="normal">
107888
</p>
</td>
<td>
<p class="normal">
482
</p>
</td>
```

```

<td>
<p class="normal">
2017
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/IDA2016Challenge">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/IDA2016Challenge">
IDA2016Challenge
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The dataset consists of data collected from heavy Scania trucks
in everyday usage. &nbsp;</p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Integer
</p>
</td>
<td>
<p class="normal">
76000
</p>
</td>
<td>
<p class="normal">
171
</p>
</td>
<td>
<p class="normal">
2017
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/DSRC+Vehicle+Communications">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/DSRC+Vehicle+Communications">
DSRC Vehicle Communications
</a>

```

```

        </b>
    </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This set Provides data regarding wireless communications between vehicles and road side units. two separate data sets are provided (normal scenario) and in the presence of attacker (jammer).&nbsp;</p></td> -->
<td>
<p class="normal">
    Sequential, Text
</p>
</td>
<td>
<p class="normal">
    Clustering
</p>
</td>
<td>
<p class="normal">
    Real
</p>
</td>
<td>
<p class="normal">
    10000
</p>
</td>
<td>
<p class="normal">
    5
</p>
</td>
<td>
<p class="normal">
    2017
</p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Mturk+User-Perceived+Clusters+over+Images">
    
</a>
</td>
<td>
<p class="normal">
    <b>
        <a href="datasets/Mturk+User-Perceived+Clusters+over+Images">
            Mturk User-Perceived Clusters over Images
        </a>
    </b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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.&nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate, Text
</p>
</td>
<td>
<p class="normal">
    Clustering
</p>
</td>
<td>
<p class="normal">

```

```

    <p class="normal">
        Integer
    </p>
</td>
<td>
<p class="normal">
    180
</p>
</td>
<td>
<p class="normal">
    500
</p>
</td>
<td>
<p class="normal">
    2016
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Character+Font+Images">
    
</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Character+Font+Images">
    Character Font Images
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Character images from scanned and computer generated
fonts. </p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">
    Integer, Real
</p>
</td>
<td>
<p class="normal">
    745000
</p>
</td>
<td>
<p class="normal">
    411
</p>
</td>
<td>
<p class="normal">
    2016
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>

```

```

<-->
<table>
  <tr>
    <td>
      <a href="datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-
Instance+Learning+with+Instance+Labels">
        
      </a>
    </td>
    <td>
      <p class="normal">
        <b>
          <a href="datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-
Instance+Learning+with+Instance+Labels">
            DeliciousMIL: A Data Set for Multi-Label Multi-Instance Learning with Instance Labels
          </a>
        </b>
      </p>
    </td>
  </tr>
</table>
</td>
<!-- <td><p class="normal">This dataset includes 1) 12234 documents (8251 training, 3983
test) extracted from DeliciousT140 dataset, 2) class labels for all documents, 3) labels for a sub
set of sentences of the test documents.&nbsp;</p></td> -->
<td>
  <p class="normal">
    Text
  </p>
</td>
<td>
  <p class="normal">
    Classification
  </p>
</td>
<td>
  <p class="normal">
    Integer
  </p>
</td>
<td>
  <p class="normal">
    12234
  </p>
</td>
<td>
  <p class="normal">
    8519
  </p>
</td>
<td>
  <p class="normal">
    2016
  </p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
  <td>
    <table>
      <tr>
        <td>
          <a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++">
            
          </a>
        </td>
        <td>
          <p class="normal">
            <b>
              <a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++">
                Autistic Spectrum Disorder Screening Data for Children
              </a>
            </b>
          </p>
        </td>
      </tr>
    </table>
  </td>
</tr>
</+>

```

```
~,~  
<!-- <td><p class="normal">Children screening data for autism suitable for classification a  
nd predictive tasks &nbsp;</p></td> -->  
<td>  
  <p class="normal">  
    Multivariate  
  </p>  
</td>  
<td>  
  <p class="normal">  
    Classification  
  </p>  
</td>  
<td>  
  <p class="normal">  
    Integer  
  </p>  
</td>  
<td>  
  <p class="normal">  
    292  
  </p>  
</td>  
<td>  
  <p class="normal">  
    21  
  </p>  
</td>  
<td>  
  <p class="normal">  
    2017  
  </p>  
</td>  
<!-- <td><p class="normal">Life&nbsp;</p></td> -->  
</tr>  
<tr>  
<td>  
  <table>  
    <tr>  
      <td>  
        <a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent++">  
            
        </a>  
      </td>  
      <td>  
        <p class="normal">  
          <b>  
            <a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent++">  
              Autistic Spectrum Disorder Screening Data for Adolescent  
            </a>  
          </b>  
        </p>  
      </td>  
    </tr>  
  </table>  
</td>  
<!-- <td><p class="normal">Autistic Spectrum Disorder Screening Data for Adolescent. This d  
ataset is related to classification and predictive tasks.&nbsp;</p></td> -->  
<td>  
  <p class="normal">  
    Multivariate  
  </p>  
</td>  
<td>  
  <p class="normal">  
    Classification  
  </p>  
</td>  
<td>  
  <p class="normal">  
    Integer  
  </p>  
</td>  
<td>  
  <p class="normal">  
    104  
  </p>  
</td>
```

```

~,~  

<td>  

<p class="normal">  

  21  

</p>  

</td>  

<td>  

<p class="normal">  

  2017  

</p>  

</td>  

<!-- <td><p class="normal">Life </p></td> -->  

</tr>  

<tr bgcolor="#DDEEFF">  

<td>  

<table>  

<tr>  

<td>  

<a href="datasets/APS+Failure+at+Scania+Trucks">  

</a>  

</td>  

<td>  

<p class="normal">  

<b>  

<a href="datasets/APS+Failure+at+Scania+Trucks">  

  APS Failure at Scania Trucks  

</a>  

</b>  

</p>  

</td>  

</tr>  

</table>  

</td>  

<!-- <td><p class="normal">The datasets' positive class consists of component failures for  

a specific component of the APS system. The negative class consists of trucks with failures for co  

mponents not related to the APS. </p></td> -->  

<td>  

<p class="normal">  

  Multivariate  

</p>  

</td>  

<td>  

<p class="normal">  

  Classification  

</p>  

</td>  

<td>  

<p class="normal">  

  Integer, Real  

</p>  

</td>  

<td>  

<p class="normal">  

  60000  

</p>  

</td>  

<td>  

<p class="normal">  

  171  

</p>  

</td>  

<td>  

<p class="normal">  

  2017  

</p>  

</td>  

<!-- <td><p class="normal">Computer </p></td> -->  

</tr>  

<tr>  

<td>  

<table>  

<tr>  

<td>  

<a href="datasets/Wireless+Indoor+Localization">  

</a>  

~,~
```

```

</ul>
<td>
<p class="normal">
<b>
<a href="datasets/Wireless+Indoor+Localization">
    Wireless Indoor Localization
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Collected in indoor space by observing signal strengths of seven WiFi signals visible on a smartphone. The decision variable is one of the four rooms. &nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">
    Real
</p>
</td>
<td>
<p class="normal">
    2000
</p>
</td>
<td>
<p class="normal">
    7
</p>
</td>
<td>
<p class="normal">
    2017
</p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/HCC+Survival">
    
</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/HCC+Survival">
    HCC Survival
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Hepatocellular Carcinoma dataset (HCC dataset) was collected at a University Hospital in Portugal. It contains real clinical data of 165 patients diagnosed with HCC.&nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
</tr>

```

```

<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Integer, Real
</p>
</td>
<td>
<p class="normal">
165
</p>
</td>
<td>
<p class="normal">
49
</p>
</td>
<td>
<p class="normal">
2017
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/CSM%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/CSM%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015">
CSM (Conventional and Social Media Movies) Dataset 2014 and 2015
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">12 features categorized as conventional and social media
features. Both conventional features, collected from movies databases on Web as well as social med
ia features(YouTube, Twitter).&nbsp;</p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Classification, Regression
</p>
</td>
<td>
<p class="normal">
Integer
</p>
</td>
<td>
<p class="normal">
217
</p>
</td>
<td>
<p class="normal">
12
</p>
</td>
<!-- <td><p class="normal">..</p></td> -->

```

```

<td>
<p class="normal">
2017
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29">
University of Tehran Question Dataset 2016 (UTQD.2016)
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Persian questions gathered from a jeopardy game broadcasted on Iranian national television.  </p></td> -->
<td>
<p class="normal">
Text
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
1175
</p>
</td>
<td>
<p class="normal">
3
</p>
</td>
<td>
<p class="normal">
2017
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Autism+Screening+Adult">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Autism+Screening+Adult">
Autism Screening Adult
</a>
</b>
</p>
</td>

```

```

        </p>
        </td>
        </tr>
        </table>
        <td>
        <!-- <td><p class="normal">Autistic Spectrum Disorder Screening Data for Adult. This
dataset is related to classification and predictive tasks.&nbsp;</p></td> -->
        <td>
        <p class="normal">
        Classification
        </p>
        </td>
        <td>
        <p class="normal">
        Integer
        </p>
        </td>
        <td>
        <p class="normal">
        704
        </p>
        </td>
        <td>
        <p class="normal">
        21
        </p>
        </td>
        <td>
        <p class="normal">
        2017
        </p>
        </td>
        <!-- <td><p class="normal">Social&nbsp;</p></td> -->
        </tr>
        <tr bgcolor="DDEEFF">
        <td>
        <table>
        <tr>
        <td>
        <a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor">
            
        </a>
        </td>
        <td>
        <p class="normal">
            <b>
                <a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor">
                    Activity recognition with healthy older people using a batteryless wearable sensor
                </a>
            </b>
            </p>
        </td>
        </tr>
        </table>
        </td>
        <!-- <td><p class="normal">Sequential motion data from 14 healthy older people aged 66 to 86
years old using a batteryless, wearable sensor on top of their clothing for the recognition of act
ivities in clinical environments.&nbsp;</p></td> -->
        <td>
        <p class="normal">
            Sequential
        </p>
        </td>
        <td>
        <p class="normal">
            Classification
        </p>
        </td>
        <td>
        .. ...
    
```

```
<p class="normal">
  Real
</p>
</td>
<td>
  <p class="normal">
    75128
  </p>
</td>
<td>
  <p class="normal">
    9
  </p>
</td>
<td>
  <p class="normal">
    2016
  </p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
  <a href="datasets/Immunotherapy+Dataset">
    
  </a>
</td>
<td>
  <p class="normal">
    <b>
      <a href="datasets/Immunotherapy+Dataset">
        Immunotherapy Dataset
      </a>
    </b>
  </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset contains information about wart treatment results
of 90 patients using immunotherapy. </p></td> -->
<td>
  <p class="normal">
    Univariate
  </p>
</td>
<td>
  <p class="normal">
    Classification
  </p>
</td>
<td>
  <p class="normal">
    Integer, Real
  </p>
</td>
<td>
  <p class="normal">
    90
  </p>
</td>
<td>
  <p class="normal">
    8
  </p>
</td>
<td>
  <p class="normal">
    2018
  </p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
```

```

<td>
<table>
<tr>
<td>
<a href="datasets/Cryotherapy+Dataset+">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Cryotherapy+Dataset+">
Cryotherapy Dataset
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset contains information about wart treatment results
of 90 patients using cryotherapy.&nbsp;</p></td> -->
<td>
<p class="normal">
Univariate
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Integer, Real
</p>
</td>
<td>
<p class="normal">
90
</p>
</td>
<td>
<p class="normal">
7
</p>
</td>
<td>
<p class="normal">
2018
</p>
</td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes">
OCT data &amp; Color Fundus Images of Left &amp; Right Eyes
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset contains OCT data (in mat format) and color fundus
data (in jpg format) of left & right eyes of 50 healthy persons.&nbsp;</p></td> -->

```

```

<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">
    Real
</p>
</td>
<td>
<p class="normal">
    50
</p>
</td>
<td>
<p class="normal">
    2
</p>
</td>
<td>
<p class="normal">
    2016
</p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Discrete+Tone+Image+Dataset">
    
</a>
</td>
<td>
<p class="normal">
<b>
    <a href="datasets/Discrete+Tone+Image+Dataset">
        Discrete Tone Image Dataset
    </a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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.
&nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">
    71
</p>
</td>
<td>
<p class="normal">

```

```

    11
  </p>
</td>
<td>
<p class="normal">
  2018
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/News+Popularity+in+Multiple+Social+Media+Platforms">
  
</a>
</td>
<td>
<p class="normal">
<b>
  <a href="datasets/News+Popularity+in+Multiple+Social+Media+Platforms">
    News Popularity in Multiple Social Media Platforms
  </a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Large data set of news items and their respective social
feedback on multiple platforms: Facebook, Google+ and LinkedIn.&ampnbsp</p></td> -->
<td>
<p class="normal">
  Multivariate, Time-Series, Text
</p>
</td>
<td>
<p class="normal">
  Regression
</p>
</td>
<td>
<p class="normal">
  Integer, Real
</p>
</td>
<td>
<p class="normal">
  93239
</p>
</td>
<td>
<p class="normal">
  11
</p>
</td>
<td>
<p class="normal">
  2018
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Ultrasonic+flowmeter+diagnostics">
  
</a>
</td>
<td>
<p class="normal">
<b>

```

```

        <a href="datasets/Ultrasonic+flowmeter+diagnostics">
            Ultrasonic flowmeter diagnostics
        </a>
        </b>
        </p>
        </td>
    </tr>
</table>
</td>
<!-- <td><p class="normal">Fault diagnosis of four liquid ultrasonic flowmeters </p>
</td> -->
<td>
    <p class="normal">
        Multivariate
    </p>
</td>
<td>
    <p class="normal">
        Classification
    </p>
</td>
<td>
    <p class="normal">
        Real
    </p>
</td>
<td>
    <p class="normal">
        540
    </p>
</td>
<td>
    <p class="normal">
        173
    </p>
</td>
<td>
    <p class="normal">
        2018
    </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
    <table>
        <tr>
            <td>
                <a href="datasets/ICMLA+2014+Accepted+Papers+Data+Set">
                    
                </a>
            </td>
            <td>
                <p class="normal">
                    <b>
                        <a href="datasets/ICMLA+2014+Accepted+Papers+Data+Set">
                            ICMLA 2014 Accepted Papers Data Set
                        </a>
                    </b>
                </p>
            </td>
        </tr>
    </table>
</td>
<!-- <td><p class="normal">This data set compromises the metadata for the 2014 ICMLA
conference's accepted papers, including ID, paper titles, author's keywords, abstracts and session
s in which they were exposed. </p></td> -->
<td>
    <p class="normal">
        Multivariate
    </p>
</td>
<td>
    <p class="normal">
        Classification, Clustering
    </p>
</td>

```

```

<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
105
</p>
</td>
<td>
<p class="normal">
5
</p>
</td>
<td>
<p class="normal">
2018
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation">
    BLE RSSI Dataset for Indoor localization and Navigation
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset contains RSSI readings gathered from an array of
Bluetooth Low Energy (BLE) iBeacons in a real-world and operational indoor environment for localiz
ation and navigation purposes.&nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate, Sequential, Time-Series
</p>
</td>
<td>
<p class="normal">
    Classification, Clustering
</p>
</td>
<td>
<p class="normal">
    Integer
</p>
</td>
<td>
<p class="normal">
    6611
</p>
</td>
<td>
<p class="normal">
    15
</p>
</td>
<td>
<p class="normal">
    2018
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>

```

```

<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Container+Crane+Controller+Data+Set">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Container+Crane+Controller+Data+Set">
    Container Crane Controller Data Set
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">A container crane has the function of transporting containers from one point to another point.&nbsp;</p></td> -->
<td>
<p class="normal">
    Univariate, Domain-Theory
</p>
</td>
<td>
<p class="normal">
    Classification, Regression
</p>
</td>
<td>
<p class="normal">
    Real
</p>
</td>
<td>
<p class="normal">
    15
</p>
</td>
<td>
<p class="normal">
    3
</p>
</td>
<td>
<p class="normal">
    2018
</p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Residential+Building+Data+Set">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Residential+Building+Data+Set">
    Residential Building Data Set
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Data set includes construction cost, sale prices, project

```

variables, and economic variables corresponding to real estate single-family residential apartments in Tehran, Iran. -->

<td>

<p class="normal">  
Multivariate  
</p>

</td>

<td>

<p class="normal">  
Regression  
</p>

</td>

<td>

<p class="normal">  
Real  
</p>

</td>

<td>

<p class="normal">  
372  
</p>

</td>

<td>

<p class="normal">  
105  
</p>

</td>

<td>

<p class="normal">  
2018  
</p>

</td>

<!-- <td><p class="normal">Computer</p></td> -->

</tr>

<tr>

<td>

<table>

<tr>

<td>

<a href="datasets/Health+News+in+Twitter">  


</a>

</td>

<td>

<p class="normal">  
<b>  
<a href="datasets/Health+News+in+Twitter">  
Health News in Twitter  
</a>  
</b>  
</p>

</td>

</tr>

</table>

</td>

<!-- <td><p class="normal">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></td> -->

<td>

<p class="normal">  
Text  
</p>

</td>

<td>

<p class="normal">  
Clustering  
</p>

</td>

<td>

<p class="normal">  
Real  
</p>

</td>

<td>

<p class="normal">  
58000  
</p>

```

</td>
<td>
<p class="normal">
  25000
</p>
</td>
<td>
<p class="normal">
  2018
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/chipseq">
  
</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/chipseq">
  chipseq
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">ChIP-seq experiments characterize protein modifications or binding at
specific genomic locations in specific samples. The machine learning
problem in these data is structured binary classification.&nbsp;</p></td> -->
<td>
<p class="normal">
  Sequential
</p>
</td>
<td>
<p class="normal">
  Classification
</p>
</td>
<td>
<p class="normal">
  Integer
</p>
</td>
<td>
<p class="normal">
  4960
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
  2018
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/SGEMM+GPU+kernel+performance">
  
</a>

```

```

        .. ~
    </td>
    <td>
        <p class="normal">
            <b>
                <a href="datasets/SGEMM+GPU+kernel+performance">
                    SGEMM GPU kernel performance
                </a>
            </b>
        </p>
    </td>
    </tr>
</table>
</td>
<!-- <td><p class="normal">Running times for multiplying two 2048 x 2048 matrices using a GPU OpenCL SGEMM kernel with varying parameters (using the library 'CLTune').&nbsp;</p></td> -->
<td>
    <p class="normal">
        Multivariate
    </p>
</td>
<td>
    <p class="normal">
        Regression
    </p>
</td>
<td>
    <p class="normal">
        Integer
    </p>
</td>
<td>
    <p class="normal">
        241600
    </p>
</td>
<td>
    <p class="normal">
        18
    </p>
</td>
<td>
    <p class="normal">
        2018
    </p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
    <td>
        <table>
            <tr>
                <td>
                    <a href="datasets/Repeat+Consumption+Matrices">
                        
                    </a>
                </td>
                <td>
                    <p class="normal">
                        <b>
                            <a href="datasets/Repeat+Consumption+Matrices">
                                Repeat Consumption Matrices
                            </a>
                        </b>
                    </p>
                </td>
            </tr>
        </table>
    </td>
    <!-- <td><p class="normal">The dataset contains 7 datasets of User - Item matrices, where each entry represents how many times a user consumed an item. Item is used as an umbrella term for various categories.&nbsp;</p></td> -->
    <td>
        <p class="normal">
            Multivariate
        </p>
    </td>
    <td>

```

```

<-->
<p class="normal">
    Clustering
</p>
</td>
<td>
<p class="normal">
    Real
</p>
</td>
<td>
<p class="normal">
    130000
</p>
</td>
<td>
<p class="normal">
    21000
</p>
</td>
<td>
<p class="normal">
    2018
</p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
    <a href="datasets/detection_of_IoT_botnet_attacks_N_BaIoT">
        
    </a>
</td>
<td>
    <p class="normal">
        <b>
            <a href="datasets/detection_of_IoT_botnet_attacks_N_BaIoT">
                detection_of_IoT_botnet_attacks_N_BaIoT
            </a>
        </b>
    </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset addresses the lack of public botnet datasets,
especially for the IoT. It suggests *real* traffic data, gathered from 9 commercial IoT devices au-
thentically infected by Mirai and BASHLITE.&nbsp;</p></td> -->
<td>
    <p class="normal">
        Multivariate, Sequential
    </p>
</td>
<td>
    <p class="normal">
        Classification, Clustering
    </p>
</td>
<td>
    <p class="normal">
        Real
    </p>
</td>
<td>
    <p class="normal">
        7062606
    </p>
</td>
<td>
    <p class="normal">
        115
    </p>
</td>
<td>
    <p class="normal">

```

```

--> class= "normal" >
2018
</p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Absenteeism+at+work">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Absenteeism+at+work">
Absenteeism at work
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The database was created with records of absenteeism at work
from July 2007 to July 2010 at a courier company in Brazil.&nbsp;</p></td> -->
<td>
<p class="normal">
Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
Classification, Clustering
</p>
</td>
<td>
<p class="normal">
Integer, Real
</p>
</td>
<td>
<p class="normal">
740
</p>
</td>
<td>
<p class="normal">
21
</p>
</td>
<td>
<p class="normal">
2018
</p>
</td>
<!-- <td><p class="normal">Business&nbsp;</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/SCADI">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/SCADI">
SCADI
</a>
</b>
</p>
</td>
</tr>

```

```

        ~/p/
    </td>
</tr>
</table>
</td>
<!-- &lt;td&gt;&lt;p class="normal"&gt;First self-care activities dataset based on ICF-CY.&amp;nbsp;&lt;/p&gt;
&lt;/td&gt; --&gt;
&lt;td&gt;
&lt;p class="normal"&gt;
    Multivariate
&lt;/p&gt;
&lt;/td&gt;
&lt;td&gt;
&lt;p class="normal"&gt;
    Classification, Clustering
&lt;/p&gt;
&lt;/td&gt;
&lt;td&gt;
&lt;p class="normal"&gt;
    &lt;/p&gt;
&lt;/td&gt;
&lt;td&gt;
&lt;p class="normal"&gt;
    70
&lt;/p&gt;
&lt;/td&gt;
&lt;td&gt;
&lt;p class="normal"&gt;
    206
&lt;/p&gt;
&lt;/td&gt;
&lt;td&gt;
&lt;p class="normal"&gt;
    2018
&lt;/p&gt;
&lt;/td&gt;
<!-- &lt;td&gt;&lt;p class="normal"&gt;Life&amp;nbsp;&lt;/p&gt;&lt;/td&gt; --&gt;
&lt;/tr&gt;
&lt;tr bgcolor="DDEEFF"&gt;
&lt;td&gt;
&lt;table&gt;
&lt;tr&gt;
&lt;td&gt;
&lt;a href="datasets/Condition+monitoring+of+hydraulic+systems"&gt;
    &lt;img border="1" src="assets/MLimages/SmallLargedefault.jpg"/&gt;
&lt;/a&gt;
&lt;/td&gt;
&lt;td&gt;
&lt;p class="normal"&gt;
    &lt;b&gt;
        &lt;a href="datasets/Condition+monitoring+of+hydraulic+systems"&gt;
            Condition monitoring of hydraulic systems
        &lt;/a&gt;
    &lt;/b&gt;
&lt;/p&gt;
&lt;/td&gt;
&lt;/tr&gt;
&lt;/table&gt;
&lt;/td&gt;
<!-- &lt;td&gt;&lt;p class="normal"&gt;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.&amp;nbsp;&lt;/p&gt;&lt;/td&gt; --&gt;
&lt;td&gt;
&lt;p class="normal"&gt;
    Multivariate, Time-Series
&lt;/p&gt;
&lt;/td&gt;
&lt;td&gt;
&lt;p class="normal"&gt;
    Classification, Regression
&lt;/p&gt;
&lt;/td&gt;
&lt;td&gt;
&lt;p class="normal"&gt;
    Real
&lt;/p&gt;
&lt;/td&gt;
&lt;td&gt;
</pre>

```

```
<!-->
<p class="normal">
  2205
</p>
</td>
<td>
<p class="normal">
  43680
</p>
</td>
<td>
<p class="normal">
  2018
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Carbon+Nanotubes">
  
</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Carbon+Nanotubes">
  Carbon Nanotubes
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset contains 10721 initial and calculated atomic
coordinates of carbon nanotubes. </p></td> -->
<td>
<p class="normal">
  Univariate
</p>
</td>
<td>
<p class="normal">
  Regression
</p>
</td>
<td>
<p class="normal">
  Real
</p>
</td>
<td>
<p class="normal">
  10721
</p>
</td>
<td>
<p class="normal">
  8
</p>
</td>
<td>
<p class="normal">
  2018
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
```

```

<a href="datasets/Optical+Interconnection+Network+">
  
</a>
</td>
<td>
  <p class="normal">
    <b>
      <a href="datasets/Optical+Interconnection+Network+">
        Optical Interconnection Network
      </a>
    </b>
  </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset contains 640 performance measurements from a simulation of 2-Dimensional Multiprocessor Optical Interconnection Network. &nbsp;</p></td> -->
<td>
  <p class="normal">
    Multivariate
  </p>
</td>
<td>
  <p class="normal">
    Classification, Regression
  </p>
</td>
<td>
  <p class="normal">
    Integer, Real
  </p>
</td>
<td>
  <p class="normal">
    640
  </p>
</td>
<td>
  <p class="normal">
    10
  </p>
</td>
<td>
  <p class="normal">
    2018
  </p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr>
<td>
  <table>
    <tr>
      <td>
        <a href="datasets/Sports+articles+for+objectivity+analysis">
          
        </a>
      </td>
      <td>
        <p class="normal">
          <b>
            <a href="datasets/Sports+articles+for+objectivity+analysis">
              Sports articles for objectivity analysis
            </a>
          </b>
        </p>
      </td>
    </tr>
  </table>
</td>
<!-- <td><p class="normal">1000 sports articles were labeled using Amazon Mechanical Turk as objective or subjective. The raw texts, extracted features, and the URLs from which the articles were retrieved are provided.&nbsp;</p></td> -->
<td>
  <p class="normal">
    Multivariate, Text
  <!--

```

```
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Integer
</p>
</td>
<td>
<p class="normal">
1000
</p>
</td>
<td>
<p class="normal">
59
</p>
</td>
<td>
<p class="normal">
2018
</p>
</td>
<!-- <td><p class="normal">Social &nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Breast+Cancer+Coimbra">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Breast+Cancer+Coimbra">
Breast Cancer Coimbra
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Clinical features were observed or measured for 64 patients with
breast cancer and 52 healthy controls. &nbsp;</p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Integer
</p>
</td>
<td>
<p class="normal">
116
</p>
</td>
<td>
<p class="normal">
10
</p>
</td>
<td>
```

```

<td>
<p class="normal">
2018
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data">
GNFUV Unmanned Surface Vehicles Sensor Data
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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). &nbsp;</p></td> -->
<td>
<p class="normal">
Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
Regression
</p>
</td>
<td>
<p class="normal">
Real
</p>
</td>
<td>
<p class="normal">
1672
</p>
</td>
<td>
<p class="normal">
5
</p>
</td>
<td>
<p class="normal">
2018
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Dishonest+Internet+users+Dataset">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Dishonest+Internet+users+Dataset">
Dishonest Internet users Dataset
</a>
</b>
</p>
</td>
</tr>
</table>
</td>

```

```

        </a>
    </b>
    </p>
    </td>
    </tr>
</table>
<td>
<!-- <td><p class="normal">The dataset was used to test an architecture based on a trust model capable to cope with the evaluation of the trustworthiness of users interacting in pervasive environments.&nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Classification, Clustering
</p>
</td>
<td>
<p class="normal">
    322
</p>
</td>
<td>
<p class="normal">
    5
</p>
</td>
<td>
<p class="normal">
    2018
</p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr>
<td>
<table>
    <tr>
        <td>
            <a href="datasets/Victorian+Era+Authorship+Attribution">
                
            </a>
        </td>
        <td>
            <p class="normal">
                <b>
                    <a href="datasets/Victorian+Era+Authorship+Attribution">
                        Victorian Era Authorship Attribution
                    </a>
                </b>
            </p>
        </td>
    </tr>
</table>
</td>
<!-- <td><p class="normal">To create the largest authorship attribution dataset, we extracted works of 50 well-known authors. To have a non-exhaustive learning, in training there are 45 authors whereas, in the testing, it's 50&nbsp;</p></td> -->
<td>
<p class="normal">
    Text
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">

```

```

        </p>
        </td>
        <td>
            <p class="normal">
                93600
            </p>
        </td>
        <td>
            <p class="normal">
                1000
            </p>
        </td>
        <td>
            <p class="normal">
                2018
            </p>
        </td>
        <!-- <td><p class="normal">Computer </p></td> -->
    </tr>
    <tr bgcolor="DDEEFF">
        <td>
            <table>
                <tr>
                    <td>
                        <a href="datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set">
                            
                        </a>
                    </td>
                    <td>
                        <p class="normal">
                            <b>
                                <a href="datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set">
                                    Simulated Falls and Daily Living Activities Data Set
                                </a>
                            </b>
                        </p>
                    </td>
                </tr>
            </table>
        </td>
        <!-- <td><p class="normal">20 falls and 16 daily living activities were performed by 17 volunteers with 5 repetitions while wearing 6 sensors (3.060 instances) that attached to their head, chest, waist, wrist, thigh and ankle. </p></td> -->
        <td>
            <p class="normal">
                Time-Series
            </p>
        </td>
        <td>
            <p class="normal">
                Classification
            </p>
        </td>
        <td>
            <p class="normal">
                Integer
            </p>
        </td>
        <td>
            <p class="normal">
                3060
            </p>
        </td>
        <td>
            <p class="normal">
                138
            </p>
        </td>
        <td>
            <p class="normal">
                2018
            </p>
        </td>
        <!-- <td><p class="normal">Life </p></td> -->
    </tr>
    <tr>
        <td>

```

```

<table>
<tr>
<td>
<a href="datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing">
Multimodal Damage Identification for Humanitarian Computing
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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.&nbsp;</p></td> -->
<td>
<p class="normal">
Multivariate, Text
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Integer
</p>
</td>
<td>
<p class="normal">
5879
</p>
</td>
<td>
<p class="normal">
2018
</p>
</td>
<!-- <td><p class="normal">Social&nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals">
EEG Steady-State Visual Evoked Potential Signals
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This database consists on 30 subjects performing Brain Computer
Interface for Steady State Visual Evoked Potentials (BCI-SSVEP). &nbsp;</p></td> -->
<td>

```

```
<p class="normal">
    Multivariate, Time-Series
</p>
</td>
<td>
    <p class="normal">
        Classification, Regression
    </p>
</td>
<td>
    <p class="normal">
        Integer
    </p>
</td>
<td>
    <p class="normal">
        9200
    </p>
</td>
<td>
    <p class="normal">
        16
    </p>
</td>
<td>
    <p class="normal">
        2018
    </p>
</td>
<!-- <td><p class="normal">Life &nbsp;</p></td> -->
</tr>
<tr>
<td>
    <table>
        <tr>
            <td>
                <a href="datasets/Roman+Urdu+Data+Set">
                    
                </a>
            </td>
            <td>
                <p class="normal">
                    <b>
                        <a href="datasets/Roman+Urdu+Data+Set">
                            Roman Urdu Data Set
                        </a>
                    </b>
                </p>
            </td>
        </tr>
    </table>
</td>
<!-- <td><p class="normal">Roman Urdu (the scripting style for Urdu language) is one of the
limited resource languages.A data corpus comprising of more than 20000 records was
collected.&nbsp;</p></td> -->
<td>
    <p class="normal">
        Text
    </p>
</td>
<td>
    <p class="normal">
        Classification
    </p>
</td>
<td>
    <p class="normal">
        20000
    </p>
</td>
<td>
    <p class="normal">
        2
    </p>

```

```

        </p>
    </td>
    <td>
        <p class="normal">
            2018
        </p>
    </td>
    <!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
    <td>
        <table>
            <tr>
                <td>
                    <a href="datasets/Avila">
                        
                    </a>
                </td>
                <td>
                    <p class="normal">
                        <b>
                            <a href="datasets/Avila">
                                Avila
                            </a>
                        </b>
                    </p>
                </td>
            </tr>
        </table>
    </td>
    <!-- <td><p class="normal">The Avila data set has been extracted from 800 images of the
'Avila Bible', an XII century giant Latin copy of the Bible. The prediction task consists in
associating each pattern to a copyist.&nbsp;</p></td> -->
    <td>
        <p class="normal">
            Multivariate
        </p>
    </td>
    <td>
        <p class="normal">
            Classification
        </p>
    </td>
    <td>
        <p class="normal">
            Real
        </p>
    </td>
    <td>
        <p class="normal">
            20867
        </p>
    </td>
    <td>
        <p class="normal">
            10
        </p>
    </td>
    <td>
        <p class="normal">
            2018
        </p>
    </td>
    <!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr>
    <td>
        <table>
            <tr>
                <td>
                    <a href="datasets/PANDOR">
                        
                    </a>
                </td>
                <td>
                    <p class="normal">
                        <b>

```

```

<a href="datasets/PANDOR">
    PANDOR
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">PANDOR is a novel and publicly available dataset for online recommendation provided by Purch (http://www.purch.com/). &nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Recommendation
</p>
</td>
<td>
<p class="normal">
    Categorical
</p>
</td>
<td>
<p class="normal">
    2018
</p>
</td>
<td>
<p class="normal">
    Life&nbsp;</p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>
    <a href="datasets/Drug+Review+Dataset+%28Druglib.com%29">
        
    </a>
</td>
<td>
    <p class="normal">
        <b>
            <a href="datasets/Drug+Review+Dataset+%28Druglib.com%29">
                Drug Review Dataset (Druglib.com)
            </a>
        </b>
    </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The dataset provides patient reviews on specific drugs along with related conditions. Reviews and ratings are grouped into reports on the three aspects benefits, side effects and overall comment.&nbsp;</p></td> -->
<td>
    <p class="normal">
        Multivariate, Text
    </p>
</td>
<td>
    <p class="normal">
        Classification, Regression, Clustering
    </p>
</td>
<td>
    <p class="normal">

```

```

    Integer
  </p>
</td>
<td>
<p class="normal">
  4143
</p>
</td>
<td>
<p class="normal">
  8
</p>
</td>
<td>
<p class="normal">
  2018
</p>
</td>
<!-- <td><p class="normal">&nbsp;</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Drug+Review+Dataset+%28Drugs.com%29">
  
</a>
</td>
<td>
<p class="normal">
<b>
  <a href="datasets/Drug+Review+Dataset+%28Drugs.com%29">
    Drug Review Dataset (Drugs.com)
  </a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The dataset provides patient reviews on specific drugs along with
related conditions and a 10 star patient rating reflecting overall patient satisfaction.&nbsp;</p>
-->
<td>
<p class="normal">
  Multivariate, Text
</p>
</td>
<td>
<p class="normal">
  Classification, Regression, Clustering
</p>
</td>
<td>
<p class="normal">
  Integer
</p>
</td>
<td>
<p class="normal">
  215063
</p>
</td>
<td>
<p class="normal">
  6
</p>
</td>
<td>
<p class="normal">
  2018
</p>
</td>
<!-- <td><p class="normal">Life&nbsp;</p></td> -->
</tr>
<tr bgcolor="#DDEEFF">

```

```

<td>
<table>
<tr>
<td>
<a href="datasets/Physical+Unclonale+Functions">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Physical+Unclonale+Functions">
    Physical Unclonale Functions
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The dataset is generated from Physical Unclonale Functions (PUFs ) simulation, specifically XOR Arbiter PUFs. PUFs are used for authentication purposes. For more i nfo, refer to our paper below.&nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">
    Integer
</p>
</td>
<td>
<p class="normal">
    6000000
</p>
</td>
<td>
<p class="normal">
    129
</p>
</td>
<td>
<p class="normal">
    2018
</p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Superconductivty+Data">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Superconductivty+Data">
    Superconductivty Data
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Two file s contain data on 21263 superconductors and their releva

```

```
nt features.&nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Regression
</p>
</td>
<td>
<p class="normal">
    Real
</p>
</td>
<td>
<p class="normal">
    21263
</p>
</td>
<td>
<p class="normal">
    81
</p>
</td>
<td>
<p class="normal">
    2018
</p>
</td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/WESAD%28Wearable+Stress+and+Affect+Detection%29">
    
</a>
</td>
<td>
<p class="normal">
<b>
    <a href="datasets/WESAD%28Wearable+Stress+and+Affect+Detection%29">
        WESAD (Wearable Stress and Affect Detection)
    </a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">WESAD (Wearable Stress and Affect Detection) contains data of 15
subjects during a stress-affect lab study, while wearing physiological and motion sensors.&nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate, Time-Series
</p>
</td>
<td>
<p class="normal">
    Classification, Regression
</p>
</td>
<td>
<p class="normal">
    Real
</p>
</td>
<td>
<p class="normal">
    63000000
</p>
</td>
```

```

., ..
<td>
<p class="normal">
12
</p>
</td>
<td>
<p class="normal">
2018
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2">
GNFUV Unmanned Surface Vehicles Sensor Data Set 2
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The data-set contains eight (2x4) data-sets of mobile sensor readings data (humidity, temperature) corresponding to a swarm of four Unmanned Surface Vehicles (USVs) in a test-bed, Athens, Greece.&ampnbsp</p></td> -->
<td>
<p class="normal">
Multivariate, Sequential, Time-Series
</p>
</td>
<td>
<p class="normal">
Regression
</p>
</td>
<td>
<p class="normal">
Real
</p>
</td>
<td>
<p class="normal">
10190
</p>
</td>
<td>
<p class="normal">
6
</p>
</td>
<td>
<p class="normal">
2018
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Student+Academics+Performance">

</a>
</td>

```

```

.. ..
<td>
<p class="normal">
<b>
<a href="datasets/Student+Academics+Performance">
    Student Academics Performance
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">The dataset tried to find the end semester percentage prediction
based on different social, economic and academic attributes. &nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">
    300
</p>
</td>
<td>
<p class="normal">
    22
</p>
</td>
<td>
<p class="normal">
    2018
</p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Online+Shoppers+Purchasing+Intention+Dataset">
    
</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Online+Shoppers+Purchasing+Intention+Dataset">
    Online Shoppers Purchasing Intention Dataset
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Of the 12,330 sessions in the dataset,
84.5% (10,422) were negative class samples that did not
end with shopping, and the rest (1908) were positive class
samples ending with shopping.&nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">

```

```

<p class="normal">
Classification, Clustering
</p>
</td>
<td>
<p class="normal">
Integer, Real
</p>
</td>
<td>
<p class="normal">
12330
</p>
</td>
<td>
<p class="normal">
18
</p>
</td>
<td>
<p class="normal">
2018
</p>
</td>
<!-- <td><p class="normal">Business&nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/PMU-UD">


```

```

    </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Parkinson%27s+Disease+Classification">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Parkinson%27s+Disease+Classification">
    Parkinson's Disease Classification
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">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). </p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">
    Integer, Real
</p>
</td>
<td>
<p class="normal">
    756
</p>
</td>
<td>
<p class="normal">
    754
</p>
</td>
<td>
<p class="normal">
    2018
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Electrical+Grid+Stability+Simulated+Data+">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Electrical+Grid+Stability+Simulated+Data+">
    Electrical Grid Stability Simulated Data
</a>
</b>
</p>
</td>
</tr>

```

```

        </tr>
</table>
</td>
<!-- <td><p class="normal">The local stability analysis of the 4-node star system
(electricity producer is in the center) implementing Decentral Smart Grid Control concept. &nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
<p class="normal">
    Classification, Regression
</p>
</td>
<td>
<p class="normal">
    Real
</p>
</td>
<td>
<p class="normal">
    10000
</p>
</td>
<td>
<p class="normal">
    14
</p>
</td>
<td>
<p class="normal">
    2018
</p>
</td>
<!-- <td><p class="normal">Physical&nbsp;</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
    <a href="datasets/Caesarian+Section+Classification+Dataset">
        
    </a>
</td>
<td>
    <p class="normal">
        <b>
            <a href="datasets/Caesarian+Section+Classification+Dataset">
                Caesarian Section Classification Dataset
            </a>
        </b>
    </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">This dataset contains information about caesarian section
results of 80 pregnant women with the most important characteristics of delivery problems in the
medical field.&nbsp;</p></td> -->
<td>
    <p class="normal">
        Univariate
    </p>
</td>
<td>
    <p class="normal">
        Classification
    </p>
</td>
<td>
    <p class="normal">
        Integer
    </p>
</td>
<td>
    <!-- -->

```

```

<td>
<p class="normal">
80
</p>
</td>
<td>
<p class="normal">
5
</p>
</td>
<td>
<p class="normal">
2018
</p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/BAUM-1">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/BAUM-1">
BAUM-1
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">BAUM-1 dataset contains 1184 multimodal facial video clips collected from 31 subjects. The 1184 video clips contain spontaneous facial expressions and speech of 13 emotional and mental states.&ampnbsp</p></td> -->
<td>
<p class="normal">
Time-Series
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
1184
</p>
</td>
<td>
<p class="normal">
</p>
</td>
<td>
<p class="normal">
2018
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/BAUM-2">

```

```


</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/BAUM-2">
    BAUM-2
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">A multilingual audio-visual affective face database consisting
of 1047 video clips of 286 subjects. &nbsp;</p></td> -->
<td>
<p class="normal">
    Time-Series
</p>
</td>
<td>
<p class="normal">
    Classification
</p>
</td>
<td>
<p class="normal">
    1047
</p>
</td>
<td>
<p class="normal">
    2018
</p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Audit+Data">
    
</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Audit+Data">
    Audit Data
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="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 s
uspicious firms.&nbsp;</p></td> -->
<td>
<p class="normal">
    Multivariate
</p>
</td>
<td>
    ..
    ..
    ..

```

```
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Real
</p>
</td>
<td>
<p class="normal">
777
</p>
</td>
<td>
<p class="normal">
18
</p>
</td>
<td>
<p class="normal">
2018
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/BuddyMove+Data+Set">

```

```

        </p>
    </td>
    <!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
    <td>
        <table>
            <tr>
                <td>
                    <a href="datasets/Real+estate+valuation+data+set">
                        
                    </a>
                </td>
                <td>
                    <p class="normal">
                        <b>
                            <a href="datasets/Real+estate+valuation+data+set">
                                Real estate valuation data set
                            </a>
                        </b>
                    </p>
                </td>
            </tr>
        </table>
    </td>
    <!-- <td><p class="normal">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. &nbsp;</p></td> -->
    <td>
        <p class="normal">
            Multivariate
        </p>
    </td>
    <td>
        <p class="normal">
            Regression
        </p>
    </td>
    <td>
        <p class="normal">
            Integer, Real
        </p>
    </td>
    <td>
        <p class="normal">
            414
        </p>
    </td>
    <td>
        <p class="normal">
            7
        </p>
    </td>
    <td>
        <p class="normal">
            2018
        </p>
    </td>
    <!-- <td><p class="normal">Business </p></td> -->
</tr>
<tr>
    <td>
        <table>
            <tr>
                <td>
                    <a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connected+speech+Dataset+*>
                        
                    </a>
                </td>
                <td>
                    <p class="normal">
                        <b>
                            <a href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connected+speech+Dataset+*>

```

Early biomarkers of Parkinson's disease based on natural connected speech Data Set

</a>

</b>

</p>

</td>

</tr>

</table>

</td>

<!-- <td><p class="normal">.nbsp;</p></td> -->

<td>

<p class="normal">

Multivariate

</p>

</td>

<td>

<p class="normal">

Classification

</p>

</td>

<td>

<p class="normal">

Real

</p>

</td>

<td>

<p class="normal">

</p>

</td>

<td>

<p class="normal">

</p>

</td>

<td>

<p class="normal">

2018

</p>

</td>

<!-- <td><p class="normal">Life&ampnbsp</p></td> -->

</tr>

<tr bgcolor="DDEEFF">

<td>

<table>

<tr>

<td>

<a href="datasets/Somerville+Happiness+Survey">



</a>

</td>

<td>

<p class="normal">

<b>

<a href="datasets/Somerville+Happiness+Survey">

Somerville Happiness Survey

</a>

</b>

</p>

</td>

</tr>

</table>

</td>

<!-- <td><p class="normal">A data extract of a non-federal dataset posted here

<https://catalog.data.gov/dataset/somerville-happiness-survey-responses-2011-2013-2015>&ampnbsp</p></td> -->

<td>

<p class="normal">

</p>

</td>

<td>

<p class="normal">

Classification

</p>

</td>

<td>

<p class="normal">

Integer

</p>

</td>

```

<td>
<p class="normal">
143
</p>
</td>
<td>
<p class="normal">
7
</p>
</td>
<td>
<p class="normal">
2018
</p>
</td>
<!-- <td><p class="normal">Life &nbsp;</p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/2.4+GHZ+Indoor+Channel+Measurements">

</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/2.4+GHZ+Indoor+Channel+Measurements">
2.4 GHZ Indoor Channel Measurements
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Measurement of the S21, consists of 10 sweeps, each sweep
contains 601 frequency points with spacing of 0.167MHz to cover a 100MHz band centered at
2.4GHz.&nbsp;</p></td> -->
<td>
<p class="normal">
Multivariate
</p>
</td>
<td>
<p class="normal">
Classification
</p>
</td>
<td>
<p class="normal">
Real
</p>
</td>
<td>
<p class="normal">
7840
</p>
</td>
<td>
<p class="normal">
5
</p>
</td>
<td>
<p class="normal">
2018
</p>
</td>
<!-- <td><p class="normal">Computer&nbsp;</p></td> -->
</tr>
<tr bgcolor="DDEEFF">
<td>
<table>
<tr>

```

```

<td>
  <a href="datasets/EMG+data+for+gestures">
    
  </a>
</td>
<td>
  <p class="normal">
    <b>
      <a href="datasets/EMG+data+for+gestures">
        EMG data for gestures
      </a>
    </b>
  </p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">These are files of raw EMG data recorded by MYO Thalmic
bracelet </p></td> -->
<td>
  <p class="normal">
    Time-Series
  </p>
</td>
<td>
  <p class="normal">
    Classification
  </p>
</td>
<td>
  <p class="normal">
    Real
  </p>
</td>
<td>
  <p class="normal">
    30000
  </p>
</td>
<td>
  <p class="normal">
    6
  </p>
</td>
<td>
  <p class="normal">
    2019
  </p>
</td>
<!-- <td><p class="normal">Life </p></td> -->
</tr>
<tr>
<td>
  <table>
    <tr>
      <td>
        <a href="datasets/Parking+Birmingham">
          
        </a>
      </td>
      <td>
        <p class="normal">
          <b>
            <a href="datasets/Parking+Birmingham">
              Parking Birmingham
            </a>
          </b>
        </p>
      </td>
    </tr>
  </table>
</td>
<!-- <td><p class="normal">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 </p></td> -->
<td>

```

```
<p class="normal">
    Multivariate, Univariate, Sequential, Time-Series
</p>
</td>
<td>
    <p class="normal">
        Classification, Regression, Clustering
    </p>
</td>
<td>
    <p class="normal">
        Real
    </p>
</td>
<td>
    <p class="normal">
        35717
    </p>
</td>
<td>
    <p class="normal">
        4
    </p>
</td>
<td>
    <p class="normal">
        2019
    </p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
    <table>
        <tr>
            <td>
                <a href="datasets/Behavior+of+the+urban+traffic+of+the+city+of+Sao+Paulo+in+Brazil">
                    
                </a>
            </td>
            <td>
                <p class="normal">
                    <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>
                </p>
            </td>
        </tr>
    </table>
</td>
<!-- <td><p class="normal">The database was created with records of behavior of the urban
traffic of the city of Sao Paulo in Brazil. </p></td> -->
<td>
    <p class="normal">
        Multivariate, Time-Series
    </p>
</td>
<td>
    <p class="normal">
        Classification, Regression
    </p>
</td>
<td>
    <p class="normal">
        Integer, Real
    </p>
</td>
<td>
    <p class="normal">
        135
    </p>
</td>
<td>
    <p class="normal">
        18
    </p>

```

```

</p>
</td>
<td>
<p class="normal">
  2018
</p>
</td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr>
<tr>
<td>
<table>
<tr>
<td>
<a href="datasets/Travel+Reviews">
  
</a>
</td>
<td>
<p class="normal">
<b>
<a href="datasets/Travel+Reviews">
  Travel Reviews
</a>
</b>
</p>
</td>
</tr>
</table>
</td>
<!-- <td><p class="normal">Reviews on destinations in 10 categories mentioned across East Asia. Each traveler rating is mapped as Excellent(4), Very Good(3), Average(2), Poor(1), and Terrible(0) and average rating is used. </p></td> -->
<td>
<p class="normal">
  Multivariate, Text
</p>
</td>
<td>
<p class="normal">
  Classification, Clustering
</p>
</td>
<td>
<p class="normal">
  Real
</p>
</td>
<td>
<p class="normal">
  980
</p>
</td>
<td>
<p class="normal">
  11
</p>
</td>
<td>
<p class="normal">
  2018
</p>
</td>
<!-- <td><p class="normal">Other </p></td> -->
</tr>
<tr bgcolor="#DDEEFF">
<td>
<table>
<tr>
<td>
<a href="datasets/Tarvel+Review+Ratings">
  
</a>
</td>
<td>
<p class="normal">
<b>

```

```

        <a href="datasets/Tarvel+Review+Ratings">
            Tarvel Review Ratings
        </a>
        </b>
        </p>
        </td>
        </tr>
        </table>
    </td>
    <!-- <td><p class="normal">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 calc
ulated.&nbsp;</p></td> -->
    <td>
        <p class="normal">
            Multivariate, Text
        </p>
    </td>
    <td>
        <p class="normal">
            Classification, Clustering
        </p>
    </td>
    <td>
        <p class="normal">
            Real
        </p>
    </td>
    <td>
        <p class="normal">
            5456
        </p>
    </td>
    <td>
        <p class="normal">
            25
        </p>
    </td>
    <td>
        <p class="normal">
            2018
        </p>
    </td>
    <!-- <td><p class="normal">Other&nbsp;</p></td> -->
</tr>
<tr>
    <td>
        <table>
            <tr>
                <td>
                    <a href="datasets/Rice+Leaf+Diseases">
                        
                    </a>
                </td>
                <td>
                    <p class="normal">
                        <b>
                            <a href="datasets/Rice+Leaf+Diseases">
                                Rice Leaf Diseases
                            </a>
                        </b>
                    </p>
                </td>
            </tr>
        </table>
    </td>
    <!-- <td><p class="normal">There are three classes/diseases: Bacterial leaf blight, Brown sp
ot, and Leaf smut, each having 40 images. The format of all images is jpg. &nbsp;</p></td> -->
    <td>
        <p class="normal">
            Multivariate
        </p>
    </td>
    <td>
        <p class="normal">
            Classification
        </p>
    </td>

```

```
<td>
<p class="normal">
    Integer
</p>
</td>
<td>
<p class="normal">
    120
</p>
</td>
<td>
<p class="normal">
    2019
</p>
</td>
<td>
<p class="normal">Computer&nbsp;</p></td> -->
</tr>
</table>
</td>
</tr>
</table>
<table align="center" cellpadding="5">
<tr valign="center">
<td>
<p class="normal">
    Supported By:
</p>
</td>
<td>

</td>
<td>
<p class="normal">
    In Collaboration With:
</p>
</td>
<td>

</td>
</tr>
</table>
<center>
<span class="normal">
<a href="about.html">
    About
</a>
|| 
<a href="citation_policy.html">
    Citation Policy
</a>
|| 
<a href="donation_policy.html">
    Donation Policy
</a>
|| 
<a href="contact.html">
    Contact
</a>
|| 
<a href="http://cml.ics.uci.edu">
    CML
</a>
</span>
</center>
</body>
</html>
```

## Título de la Base de Datos

In [13]:

```
print(database.title.string)
```

UCI Machine Learning Repository: Data Sets

In [14]:

```
links = database.find_all('a')
links
```

Out[14]:

```
<a alt="Home" href="index.html"></a>,
<a href="http://cml.ics.uci.edu"><font color="FFDD33">Center for Machine Learning and Intelligent
Systems</font></a>,
<a href="about.html">About</a>,
<a href="citation_policy.html">Citation Policy</a>,
<a href="donation_policy.html">Donate a Data Set</a>,
<a href="contact.html">Contact</a>,
<a href="http://www.google.com/search"><img align="middle" alt="Google" border="0" height="27" sr
c="http://www.google.com/logos/Logo_25blk.gif"/></a>,
<a href="datasets.php"><font color="#FFDD33" size="3"><b>View ALL Data Sets</b></font></a>,
<a href="datasets.php?
format=&task=cla&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=
e">Classification</a>,
<a href="datasets.php?
format=&task=reg&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=
e">Regression</a>,
<a href="datasets.php?
format=&task=clu&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=
e">Clustering</a>,
<a href="datasets.php?
format=&task=other&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=
e">Other</a>,
<a href="datasets.php?
format=&task=&att=cat&area=&numAtt=&numIns=&type=&sort=nameUp&view=
e">Categorical</a>,
<a href="datasets.php?
format=&task=&att=num&area=&numAtt=&numIns=&type=&sort=nameUp&view=
e">Numerical</a>,
<a href="datasets.php?
format=&task=&att=mix&area=&numAtt=&numIns=&type=&sort=nameUp&view=
e">Mixed</a>,
<a href="datasets.php?
format=&task=&att=multivariate&area=&numAtt=&numIns=&type=&sort=nameUp&view=
le">Multivariate</a>,
<a href="datasets.php?
format=&task=&att=uvar&area=&numAtt=&numIns=&type=&sort=nameUp&view=
le">Univariate</a>,
<a href="datasets.php?
format=&task=&att=seq&area=&numAtt=&numIns=&type=&sort=nameUp&view=
e">Sequential</a>,
<a href="datasets.php?
format=&task=&att=ts&area=&numAtt=&numIns=&type=&sort=nameUp&view=
t">Time-Series</a>,
<a href="datasets.php?
format=&task=&att=text&area=&numAtt=&numIns=&type=&sort=nameUp&view=
le">Text</a>,
<a href="datasets.php?
format=&task=&att=dt&area=&numAtt=&numIns=&type=&sort=nameUp&view=
t">Domain-Theory</a>,
<a href="datasets.php?
format=&task=&att=other&area=&numAtt=&numIns=&type=&sort=nameUp&view=
e">Other</a>,
<a href="datasets.php?
format=&task=&att=life&area=&numAtt=&numIns=&type=&sort=nameUp&view=
le">Life Sciences</a>,
<a href="datasets.php?
format=&task=&att=phys&area=&numAtt=&numIns=&type=&sort=nameUp&view=
le">Physical Sciences</a>,
<a href="datasets.php?
format=&task=&att=comp&area=&numAtt=&numIns=&type=&sort=nameUp&view=
le">CS / Engineering</a>,
<a href="datasets.php?
format=&task=&att=soc&area=&numAtt=&numIns=&type=&sort=nameUp&view=
le">
```

```
e">Social Sciences</a>,
<a href="datasets.php?format=&task=&att=&area=bus&numAtt=&numIns=&type=&sort=nameUp&view=e">Business</a>,
<a href="datasets.php?format=&task=&att=&area=game&numAtt=&numIns=&type=&sort=nameUp&viewle">Game</a>,
<a href="datasets.php?format=&task=&att=&area=other&numAtt=&numIns=&type=&sort=nameUp&viewble">Other</a>,
<a href="datasets.php?format=&task=&att=&area=less10&numAtt=&numIns=&type=&sort=nameUp&viewable">Less than 10</a>,
<a href="datasets.php?format=&task=&att=&area=10to100&numAtt=&numIns=&type=&sort=nameUp&viewtable">10 to 100</a>,
<a href="datasets.php?format=&task=&att=&area=greater100&numIns=&type=&sort=nameUp&view=table">Greater than 100</a>,
<a href="datasets.php?format=&task=&att=&area=less100&numIns=&type=&sort=nameUp&viewtable">Less than 100</a>,
<a href="datasets.php?format=&task=&att=&area=100to1000&numIns=&type=&sort=nameUp&vieww=table">100 to 1000</a>,
<a href="datasets.php?format=&task=&att=&area=greater1000&type=&sort=nameUp&view=table">Greater than 1000</a>,
<a href="datasets.php?format=&mat=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&viewe">Matrix</a>,
<a href="datasets.php?format=nonmat&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&viewable">Non-Matrix</a>,
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=li<st View</a>,
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameDown&view=t"><b>Name</b></a>,
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=taskUp&view=tak<b>Data Types</b></a>,
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=taskUp&view=tak<b>Default Task</b></a>,
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=attTypeUp&view=e"><b>Attribute Types</b></a>,
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=instUp&view=tak<b># Instances</b></a>,
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=attUp&view=tabl<b># Attributess</b></a>,
<a href="datasets.php?format=&task=&att=&area=&numAtt=&numIns=&type=&sort=dateUp&view=tak<b>Year</b></a>,
<a href="datasets/Abalone"></a>,
<a href="datasets/Abalone">Abalone</a>,
<a href="datasets/Adult"></a>,
<a href="datasets/Adult">Adult</a>,
<a href="datasets/Annealing"></a>,
<a href="datasets/Annealing">Annealing</a>,
<a href="datasets/Anonymous+Microsoft+Web+Data"></a>,
<a href="datasets/Anonymous+Microsoft+Web+Data">Anonymous Microsoft Web Data</a>,
<a href="datasets/Arrhythmia"></a>,
<a href="datasets/Arrhythmia">Arrhythmia</a>,
<a href="datasets/Artificial+Characters"></a>,
<a href="datasets/Artificial+Characters">Artificial Characters</a>,
<a href="datasets/Audiology+%28Original%29"></a>,
<a href="datasets/Audiology+%28Original%29">Audiology (Original)</a>,
<a href="datasets/Audiology+%28Standardized%29"></a>,
```

```
<a href="datasets/Audiology%28Standardized%29">Audiology (Standardized)</a>,
<a href="datasets/Auto+MPG"></a>,
<a href="datasets/Auto+MPG">Auto MPG</a>,
<a href="datasets/Automobile"></a>,
<a href="datasets/Automobile">Automobile</a>,
<a href="datasets/Badges"></a>,
<a href="datasets/Badges">Badges</a>,
<a href="datasets/Balance+Scale"></a>,
<a href="datasets/Balance+Scale">Balance Scale</a>,
<a href="datasets/Balloons"></a>,
<a href="datasets/Balloons">Balloons</a>,
<a href="datasets/Breast+Cancer"></a>,
<a href="datasets/Breast+Cancer">Breast Cancer</a>,
<a href="datasets/Breast+Cancer+Wisconsin%28Original%29"></a>,
<a href="datasets/Breast+Cancer+Wisconsin%28Original%29">Breast Cancer Wisconsin (Original)</a>,
<a href="datasets/Breast+Cancer+Wisconsin%28Prognostic%29"></a>,
<a href="datasets/Breast+Cancer+Wisconsin%28Prognostic%29">Breast Cancer Wisconsin (Prognostic)</a>,
<a href="datasets/Breast+Cancer+Wisconsin%28Diagnostic%29"></a>,
<a href="datasets/Breast+Cancer+Wisconsin%28Diagnostic%29">Breast Cancer Wisconsin (Diagnostic)</a>,
<a href="datasets/Pittsburgh+Bridges"></a>,
<a href="datasets/Pittsburgh+Bridges">Pittsburgh Bridges</a>,
<a href="datasets/Car+Evaluation"></a>,
<a href="datasets/Car+Evaluation">Car Evaluation</a>,
<a href="datasets/Census+Income"></a>,
<a href="datasets/Census+Income">Census Income</a>,
<a href="datasets/Chess%28King-Rook+vs.+King-Knight%29"></a>,
<a href="datasets/Chess%28King-Rook+vs.+King-Knight%29">Chess (King-Rook vs. King-Knight)</a>,
<a href="datasets/Chess%28King-Rook+vs.+King-Pawn%29"></a>,
<a href="datasets/Chess%28King-Rook+vs.+King-Pawn%29">Chess (King-Rook vs. King-Pawn)</a>,
<a href="datasets/Chess%28King-Rook+vs.+King%29"></a>,
<a href="datasets/Chess%28King-Rook+vs.+King%29">Chess (King-Rook vs. King)</a>,
<a href="datasets/Chess%28Domain+Theories%29"></a>,
<a href="datasets/Chess%28Domain+Theories%29">Chess (Domain Theories)</a>,
<a href="datasets/Bach+Chorales"></a>,
<a href="datasets/Bach+Chorales">Bach Chorales</a>,
<a href="datasets/Connect-4"></a>,
<a href="datasets/Connect-4">Connect-4</a>,
<a href="datasets/Credit+Approval"></a>,
<a href="datasets/Credit+Approval">Credit Approval</a>,
<a href="datasets/Japanese+Credit+Screening"></a>,
<a href="datasets/Japanese+Credit+Screening">Japanese Credit Screening</a>,
<a href="datasets/Computer+Hardware"></a>

<a href="datasets/Computer+Hardware">Computer Hardware</a>,
<a href="datasets/Contraceptive+Method+Choice"></a>,
<a href="datasets/Contraceptive+Method+Choice">Contraceptive Method Choice</a>,
<a href="datasets/Covertype"></a>,
<a href="datasets/Covertype">Covertype</a>,
<a href="datasets/Cylinder+Bands"></a>,
<a href="datasets/Cylinder+Bands">Cylinder Bands</a>,
<a href="datasets/Dermatology"></a>,
<a href="datasets/Dermatology">Dermatology</a>,
<a href="datasets/Diabetes"></a>,
<a href="datasets/Diabetes">Diabetes</a>,
<a href="datasets/DGP2++The+Second+Data+Generation+Program"></a>,
<a href="datasets/DGP2++The+Second+Data+Generation+Program">DGP2 - The Second Data Generation Program</a>,
<a href="datasets/Document+Understanding"></a>,
<a href="datasets/Document+Understanding">Document Understanding</a>,
<a href="datasets/EBL+Domain+Theories"></a>,
```

```
<a href="datasets/EBL+Domain+Theories">EBL Domain Theories</a>,
<a href="datasets/Echocardiogram"></a>,
<a href="datasets/Echocardiogram">Echocardiogram</a>,
<a href="datasets/Ecoli"></a>,
<a href="datasets/Ecoli">Ecoli</a>,
<a href="datasets/Flags"></a>,
<a href="datasets/Flags">Flags</a>,
<a href="datasets/Function+Finding"></a>,
<a href="datasets/Function+Finding">Function Finding</a>,
<a href="datasets/Glass+Identification"></a>,
<a href="datasets/Glass+Identification">Glass Identification</a>,
<a href="datasets/Haberman%27s+Survival"></a>,
<a href="datasets/Haberman%27s+Survival">Haberman's Survival</a>,
<a href="datasets/Hayes-Roth"></a>,
<a href="datasets/Hayes-Roth">Hayes-Roth</a>,
<a href="datasets/Heart+Disease"></a>,
<a href="datasets/Heart+Disease">Heart Disease</a>,
<a href="datasets/Hepatitis"></a>,
<a href="datasets/Hepatitis">Hepatitis</a>,
<a href="datasets/Horse+Colic"></a>,
<a href="datasets/Horse+Colic">Horse Colic</a>,
<a href="datasets/ICU"></a>,
<a href="datasets/ICU">ICU</a>,
<a href="datasets/Image+Segmentation"></a>,
<a href="datasets/Image+Segmentation">Image Segmentation</a>,
<a href="datasets/Internet+Advertisements"></a>,
<a href="datasets/Internet+Advertisements">Internet Advertisements</a>,
<a href="datasets/Ionosphere"></a>,
<a href="datasets/Ionosphere">Ionosphere</a>,
<a href="datasets/Iris"></a>,
<a href="datasets/Iris">Iris</a>,
<a href="datasets/ISOLET"></a>,
<a href="datasets/ISOLET">ISOLET</a>,
<a href="datasets/Kinship"></a>,
<a href="datasets/Kinship">Kinship</a>,
<a href="datasets/Labor+Relations"></a>,
<a href="datasets/Labor+Relations">Labor Relations</a>,
<a href="datasets/LED+Display+Domain"></a>,
<a href="datasets/LED+Display+Domain">LED Display Domain</a>,
<a href="datasets/Lenses"></a>,
<a href="datasets/Lenses">Lenses</a>,
<a href="datasets/Letter+Recognition"></a>,
<a href="datasets/Letter+Recognition">Letter Recognition</a>,
<a href="datasets/Liver+Disorders"></a>,
<a href="datasets/Liver+Disorders">Liver Disorders</a>,
<a href="datasets/Logic+Theorist"></a>,
<a href="datasets/Logic+Theorist">Logic Theorist</a>,
<a href="datasets/Lung+Cancer"></a>,
<a href="datasets/Lung+Cancer">Lung Cancer</a>,
<a href="datasets/Lymphography"></a>,
<a href="datasets/Lymphography">Lymphography</a>,
<a href="datasets/Mechanical+Analysis"></a>,
<a href="datasets/Mechanical+Analysis">Mechanical Analysis</a>,
<a href="datasets/Meta-data"></a>,
<a href="datasets/Meta-data">Meta-data</a>,
<a href="datasets/Mobile+Robots"></a>,
<a href="datasets/Mobile+Robots">Mobile Robots</a>,
<a href="datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29"></a>,
<a href="datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29">Molecular Biology (Promoter Ge
ne Sequences)</a>,
<a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%29"></a>,
<a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular Biology (Protein
Secondary Structure)</a>,
<a href="datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29"></a>,
<a href="datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29">Molecular Biology  
(Splice-junction Gene Sequences)</a>,
<a href="datasets/MONK%27s+Problems"></a>,
<a href="datasets/MONK%27s+Problems">MONK's Problems</a>,
<a href="datasets/Moral+Reasoner"></a>,
<a href="datasets/Moral+Reasoner">Moral Reasoner</a>,
<a href="datasets/Multiple+Features"></a>,
<a href="datasets/Multiple+Features">Multiple Features</a>,
<a href="datasets/Mushroom"></a>,
<a href="datasets/Mushroom">Mushroom</a>,
<a href="datasets/Musk+%28Version+1%29"></a>,
<a href="datasets/Musk+%28Version+1%29">Musk (Version 1)</a>,
<a href="datasets/Musk+%28Version+2%29"></a>,
<a href="datasets/Othello+Domain+Theory"></a>,
<a href="datasets/Othello+Domain+Theory">Othello Domain Theory</a>,
<a href="datasets/Page+Blocks+Classification"></a>,
<a href="datasets/Page+Blocks+Classification">Page Blocks Classification</a>,
<a href="datasets/Optical+Recognition+of+Handwritten+Digits"></a>,
<a href="datasets/Optical+Recognition+of+Handwritten+Digits">Optical Recognition of Handwritten Digits</a>,
<a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits"></a>,
<a href="datasets/Pen-Based+Recognition+of+Handwritten+Digits">Pen-Based Recognition of Handwritten Digits</a>,
<a href="datasets/Post-Operative+Patient"></a>,
<a href="datasets/Post-Operative+Patient">Post-Operative Patient</a>,
<a href="datasets/Primary+Tumor"></a>,
<a href="datasets/Primary+Tumor">Primary Tumor</a>,
<a href="datasets/Prodigy"></a>,
<a href="datasets/Prodigy">Prodigy</a>,
<a href="datasets/Qualitative+Structure+Activity+Relationships"></a>,
<a href="datasets/Qualitative+Structure+Activity+Relationships">Qualitative Structure Activity Relationships</a>,
<a href="datasets/Quadruped+Mammals"></a>,
<a href="datasets/Quadruped+Mammals">Quadruped Mammals</a>,
<a href="datasets/Servo"></a>,
<a href="datasets/Servo">Servo</a>,
<a href="datasets/Shuttle+Landing+Control"></a>,
<a href="datasets/Shuttle+Landing+Control">Shuttle Landing Control</a>,
<a href="datasets/Solar+Flare"></a>,
<a href="datasets/Solar+Flare">Solar Flare</a>,
<a href="datasets/Soybean+%28Large%29"></a>,
<a href="datasets/Soybean+%28Large%29">Soybean (Large)</a>,
<a href="datasets/Soybean+%28Small%29"></a>,
<a href="datasets/Soybean+%28Small%29">Soybean (Small)</a>,
<a href="datasets/Challenger+USA+Space+Shuttle+O-Ring"></a>,
<a href="datasets/Challenger+USA+Space+Shuttle+O-Ring">Challenger USA Space Shuttle O-Ring</a>,
<a href="datasets/Low+Resolution+Spectrometer"></a>,
<a href="datasets/Low+Resolution+Spectrometer">Low Resolution Spectrometer</a>,
<a href="datasets/Spambase"></a>,
<a href="datasets/Spambase">Spambase</a>,
<a href="datasets/SPECT+Heart"></a>,
<a href="datasets/SPECT+Heart">SPECT Heart</a>,
<a href="datasets/SPECTF+Heart"></a>,
<a href="datasets/SPECTF+Heart">SPECTF Heart</a>,
<a href="datasets/Sponge"></a>,
```

```
<a href="datasets/Sponge">Sponge</a>,
<a href="datasets/Statlog+Project"></a>,
<a href="datasets/Statlog+Project">Statlog Project</a>,
<a href="datasets/Student+Loan+Relational"></a>,
<a href="datasets/Student+Loan+Relational">Student Loan Relational</a>,
<a href="datasets/Teaching+Assistant+Evaluation"></a>,
<a href="datasets/Teaching+Assistant+Evaluation">Teaching Assistant Evaluation</a>,
<a href="datasets/Tic-Tac-Toe+Endgame"></a>,
<a href="datasets/Tic-Tac-Toe+Endgame">Tic-Tac-Toe Endgame</a>,
<a href="datasets/Thyroid+Disease"></a>,
<a href="datasets/Thyroid+Disease">Thyroid Disease</a>,
<a href="datasets/Trains"></a>,
<a href="datasets/Trains">Trains</a>,
<a href="datasets/University"></a>,
<a href="datasets/University">University</a>,
<a href="datasets/Congressional+Voting+Records"></a>,
<a href="datasets/Congressional+Voting+Records">Congressional Voting Records</a>,
<a href="datasets/Water+Treatment+Plant"></a>,
<a href="datasets/Water+Treatment+Plant">Water Treatment Plant</a>,
<a href="datasets/Waveform+Database+Generator+%28Version+1%29"></a>,
<a href="datasets/Waveform+Database+Generator+%28Version+1%29">Waveform Database Generator (Version 1)</a>,
<a href="datasets/Waveform+Database+Generator+%28Version+2%29"></a>,
<a href="datasets/Waveform+Database+Generator+%28Version+2%29">Waveform Database Generator (Version 2)</a>,
<a href="datasets/Wine"></a>,
<a href="datasets/Wine">Wine</a>,
<a href="datasets/Yeast"></a>,
<a href="datasets/Yeast">Yeast</a>,
<a href="datasets/Zoo"></a>,
<a href="datasets/Zoo">Zoo</a>,
<a href="datasets/Undocumented"></a>,
<a href="datasets/Undocumented">Undocumented</a>,
<a href="datasets/Twenty+Newsgroups"></a>,
<a href="datasets/Twenty+Newsgroups">Twenty Newsgroups</a>,
<a href="datasets/Australian+Sign+Language+signs"></a>,
<a href="datasets/Australian+Sign+Language+signs">Australian Sign Language signs</a>,
<a href="datasets/Australian+Sign+Language+signs+%28High+Quality%29"></a>,
<a href="datasets/Australian+Sign+Language+signs+%28High+Quality%29">Australian Sign Language signs (High Quality)</a>,
<a href="datasets/US+Census+Data+%281990%29"></a>,
<a href="datasets/US+Census+Data+%281990%29">US Census Data (1990)</a>,
<a href="datasets/Census-Income+%28KDD%29"></a>,
<a href="datasets/Census-Income+%28KDD%29">Census-Income (KDD)</a>,
<a href="datasets/Coil+1999+Competition+Data"></a>,
<a href="datasets/Coil+1999+Competition+Data">Coil 1999 Competition Data</a>,
<a href="datasets/Corel+Image+Features"></a>,
<a href="datasets/Corel+Image+Features">Corel Image Features</a>,
<a href="datasets/E.+Coli+Genes"></a>,
<a href="datasets/E.+Coli+Genes">E. Coli Genes</a>,
<a href="datasets/EEG+Database"></a>,
<a href="datasets/EEG+Database">EEG Database</a>,
<a href="datasets/El+Nino"></a>,
<a href="datasets/El+Nino">El Nino</a>,
<a href="datasets/Entree+Chicago+Recommendation+Data"></a>,
<a href="datasets/Entree+Chicago+Recommendation+Data">Entree Chicago Recommendation Data</a>,
<a href="datasets/CMU+Face+Images"></a>,
<a href="datasets/CMU+Face+Images">CMU Face Images</a>,
<a href="datasets/Insurance+Company+Benchmark+%28COIL+2000%29"></a>,
```

```
src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Insurance+Company+Benchmark+%28COIL+2000%29">Insurance Company Benchmark (COIL 2000)</a>,
<a href="datasets/Internet+Usage+Data"></a>,
<a href="datasets/Internet+Usage+Data">Internet Usage Data</a>,
<a href="datasets/IPUMS+Census+Database"></a>,
<a href="datasets/IPUMS+Census+Database">IPUMS Census Database</a>,
<a href="datasets/Japanese+Vowels"></a>,
<a href="datasets/Japanese+Vowels">Japanese Vowels</a>,
<a href="datasets/KDD+Cup+1998+Data"></a>,
<a href="datasets/KDD+Cup+1998+Data">KDD Cup 1998 Data</a>,
<a href="datasets/KDD+Cup+1999+Data"></a>,
<a href="datasets/KDD+Cup+1999+Data">KDD Cup 1999 Data</a>,
<a href="datasets/M.+Tuberculosis+Genes"></a>,
<a href="datasets/M.+Tuberculosis+Genes">M. Tuberculosis Genes</a>,
<a href="datasets/Movie"></a>,
<a href="datasets/Movie">Movie</a>,
<a href="datasets/MSNBC.com+Anonymous+Web+Data"></a>,
<a href="datasets/MSNBC.com+Anonymous+Web+Data">MSNBC.com Anonymous Web Data</a>,
<a href="datasets/NSF+Research+Award+Abstracts+1990-2003"></a>,
<a href="datasets/NSF+Research+Award+Abstracts+1990-2003">NSF Research Award Abstracts 1990-2003</a>,
<a href="datasets/Pioneer-1+Mobile+Robot+Data"></a>,
<a href="datasets/Pioneer-1+Mobile+Robot+Data">Pioneer-1 Mobile Robot Data</a>,
<a href="datasets/Pseudo+Periodic+Synthetic+Time+Series"></a>,
<a href="datasets/Pseudo+Periodic+Synthetic+Time+Series">Pseudo Periodic Synthetic Time Series</a>,
<a href="datasets/Reuters-21578+Text+Categorization+Collection"></a>,
<a href="datasets/Reuters-21578+Text+Categorization+Collection">Reuters-21578 Text Categorization Collection</a>,
<a href="datasets/Robot+Execution+Failures"></a>,
<a href="datasets/Robot+Execution+Failures">Robot Execution Failures</a>,
<a href="datasets/Synthetic+Control+Chart+Time+Series"></a>,
<a href="datasets/Synthetic+Control+Chart+Time+Series">Synthetic Control Chart Time Series</a>,
<a href="datasets/Syskill+and+Webert+Web+Page+Ratings"></a>,
<a href="datasets/Syskill+and+Webert+Web+Page+Ratings">Syskill and Webert Web Page Ratings</a>,
<a href="datasets/UNIX+User+Data"></a>,
<a href="datasets/UNIX+User+Data">UNIX User Data</a>,
<a href="datasets/Volcanoes+on+Venus+-+JARtool+experiment"></a>,
<a href="datasets/Volcanoes+on+Venus+-+JARtool+experiment">Volcanoes on Venus - JARtool experiment</a>,
<a href="datasets/Statlog%28Australian+Credit+Approval%29"></a>,
<a href="datasets/Statlog%28Australian+Credit+Approval%29">Statlog (Australian Credit Approval)</a>,
<a href="datasets/Statlog%28German+Credit+Data%29"></a>,
<a href="datasets/Statlog%28German+Credit+Data%29">Statlog (German Credit Data)</a>,
<a href="datasets/Statlog%28Heart%29"></a>,
<a href="datasets/Statlog%28Heart%29">Statlog (Heart)</a>,
<a href="datasets/Statlog%28Landsat+Satellite%29"></a>,
<a href="datasets/Statlog%28Landsat+Satellite%29">Statlog (Landsat Satellite)</a>,
<a href="datasets/Statlog%28Image+Segmentation%29"></a>,
<a href="datasets/Statlog%28Image+Segmentation%29">Statlog (Image Segmentation)</a>,
<a href="datasets/Statlog%28Shuttle%29"></a>,
<a href="datasets/Statlog%28Shuttle%29">Statlog (Shuttle)</a>,
<a href="datasets/Statlog%28Vehicle+Silhouettes%29"></a>,
```

<a href="datasets/Statlog%28Vehicle+Silhouettes%29">Statlog (Vehicle Silhouettes)</a>,  
<a href="datasets/Connectionist+Bench%28Nettalk+Corpus%29"></a>,  
<a href="datasets/Connectionist+Bench%28Nettalk+Corpus%29">Connectionist Bench (Nettalk Corpus)</a>,  
<a href="datasets/Connectionist+Bench%28Sonar%2C+Mines+vs.+Rocks%29"></a>,  
<a href="datasets/Connectionist+Bench%28Sonar%2C+Mines+vs.+Rocks%29">Connectionist Bench (Sonar, Mines vs. Rocks)</a>,  
<a href="datasets/Connectionist+Bench%28Vowel+Recognition+-+Deterding+Data%29"></a>,  
<a href="datasets/Connectionist+Bench%28Vowel+Recognition+-+Deterding+Data%29">Connectionist Bench (Vowel Recognition - Deterding Data)</a>,  
<a href="datasets/Economic+Sanctions"></a>,  
<a href="datasets/Economic+Sanctions">Economic Sanctions</a>,  
<a href="datasets/Protein+Data"></a>,  
<a href="datasets/Protein+Data">Protein Data</a>,  
<a href="datasets/Cloud"></a>,  
<a href="datasets/Cloud">Cloud</a>,  
<a href="datasets/CalIt2+Building+People+Counts"></a>,  
<a href="datasets/CalIt2+Building+People+Counts">CalIt2 Building People Counts</a>,  
<a href="datasets/Dodgers+Loop+Sensor"></a>,  
<a href="datasets/Dodgers+Loop+Sensor">Dodgers Loop Sensor</a>,  
<a href="datasets/Poker+Hand"></a>,  
<a href="datasets/Poker+Hand">Poker Hand</a>,  
<a href="datasets/MAGIC+Gamma+Telescope"></a>,  
<a href="datasets/MAGIC+Gamma+Telescope">MAGIC Gamma Telescope</a>,  
<a href="datasets/UJI+Pen+Characters"></a>,  
<a href="datasets/UJI+Pen+Characters">UJI Pen Characters</a>,  
<a href="datasets/Mammographic+Mass"></a>,  
<a href="datasets/Mammographic+Mass">Mammographic Mass</a>,  
<a href="datasets/Forest+Fires"></a>,  
<a href="datasets/Forest+Fires">Forest Fires</a>,  
<a href="datasets/Reuters+Transcribed+Subset"></a>,  
<a href="datasets/Reuters+Transcribed+Subset">Reuters Transcribed Subset</a>,  
<a href="datasets/Bag+of+Words"></a>,  
<a href="datasets/Bag+of+Words">Bag of Words</a>,  
<a href="datasets/Concrete+Compressive+Strength"></a>,  
<a href="datasets/Concrete+Compressive+Strength">Concrete Compressive Strength</a>,  
<a href="datasets/Hill-Valley"></a>,  
<a href="datasets/Hill-Valley">Hill-Valley</a>,  
<a href="datasets/Arcene"></a>,  
<a href="datasets/Arcene">Arcene</a>,  
<a href="datasets/Dexter"></a>,  
<a href="datasets/Dexter">Dexter</a>,  
<a href="datasets/Dorothea"></a>,  
<a href="datasets/Dorothea">Dorothea</a>,  
<a href="datasets/Gisette"></a>,  
<a href="datasets/Gisette">Gisette</a>,  
<a href="datasets/Madelon"></a>,  
<a href="datasets/Madelon">Madelon</a>,  
<a href="datasets/Ozone+Level+Detection"></a>,  
<a href="datasets/Ozone+Level+Detection">Ozone Level Detection</a>,  
<a href="datasets/Abscisic+Acid+Signaling+Network"></a>,  
<a href="datasets/Abscisic+Acid+Signaling+Network">Abscisic Acid Signaling Network</a>,  
<a href="datasets/Parkinsons"></a>,  
<a href="datasets/Parkinsons">Parkinsons</a>,  
<a href="datasets/Character+Trajectories"></a>,  
<a href="datasets/Character+Trajectories">Character Trajectories</a>,  
<a href="datasets/Blood+Transfusion+Service+Center"></a>,  
<a href="datasets/Blood+Transfusion+Service+Center">Blood Transfusion Service Center</a>,  
<a href="datasets/UJI+Pen+Characters+%28Version+2%29"></a>,  
<a href="datasets/UJI+Pen+Characters+%28Version+2%29">UJI Pen Characters (Version 2)</a>,

```
<a href="datasets/Semeion+Handwritten+Digit"></a>,
<a href="datasets/Semeion+Handwritten+Digit">Semeion Handwritten Digit</a>,
<a href="datasets/SECOM"></a>,
<a href="datasets/SECOM">SECOM</a>,
<a href="datasets/Plants"></a>,
<a href="datasets/Plants">Plants</a>,
<a href="datasets/Libras+Movement"></a>,
<a href="datasets/Libras+Movement">Libras Movement</a>,
<a href="datasets/Concrete+Slump+Test"></a>,
<a href="datasets/Concrete+Slump+Test">Concrete Slump Test</a>,
<a href="datasets/Communities+and+Crime"></a>,
<a href="datasets/Communities+and+Crime">Communities and Crime</a>,
<a href="datasets/Acute+Inflammations"></a>,
<a href="datasets/Acute+Inflammations">Acute Inflammations</a>,
<a href="datasets/Wine+Quality"></a>,
<a href="datasets/Wine+Quality">Wine Quality</a>,
<a href="datasets/URL+Reputation"></a>,
<a href="datasets/URL+Reputation">URL Reputation</a>,
<a href="datasets/p53+Mutants"></a>,
<a href="datasets/p53+Mutants">p53 Mutants</a>,
<a href="datasets/Parkinsons+Telemonitoring"></a>,
<a href="datasets/Parkinsons+Telemonitoring">Parkinsons Telemonitoring</a>,
<a href="datasets/Demospongiae"></a>,
<a href="datasets/Demospongiae">Demospongiae</a>,
<a href="datasets/Opinosis+Opinion+%26frasl%3B+Review"></a>,
<a href="datasets/Opinosis+Opinion+%26frasl%3B+Review">Opinosis Opinion / Review</a>,
<a href="datasets/Breast+Tissue"></a>,
<a href="datasets/Breast+Tissue">Breast Tissue</a>,
<a href="datasets/Cardiotocography"></a>,
<a href="datasets/Cardiotocography">Cardiotocography</a>,
<a href="datasets/Wall-Following+Robot+Navigation+Data"></a>,
<a href="datasets/Wall-Following+Robot+Navigation+Data">Wall-Following Robot Navigation Data</a>,
<a href="datasets/Spoken+Arabic+Digit"></a>,
<a href="datasets/Spoken+Arabic+Digit">Spoken Arabic Digit</a>,
<a href="datasets/Localization+Data+for+Person+Activity"></a>,
<a href="datasets/Localization+Data+for+Person+Activity">Localization Data for Person Activity</a>,
<a href="datasets/AutoUniv"></a>,
<a href="datasets/AutoUniv">AutoUniv</a>,
<a href="datasets/Steel+Plates+Faults"></a>,
<a href="datasets/Steel+Plates+Faults">Steel Plates Faults</a>,
<a href="datasets/MiniBooNE+particle+identification"></a>,
<a href="datasets/MiniBooNE+particle+identification">MiniBooNE particle identification</a>,
<a href="datasets/YearPredictionMSD"></a>,
<a href="datasets/YearPredictionMSD">YearPredictionMSD</a>,
<a href="datasets/PEMS-SF"></a>,
<a href="datasets/PEMS-SF">PEMS-SF</a>,
<a href="datasets/OpinRank+Review+Dataset"></a>,
<a href="datasets/OpinRank+Review+Dataset">OpinRank Review Dataset</a>,
<a href="datasets/Relative+location+of+CT+slices+on+axial+axis"></a>,
<a href="datasets/Relative+location+of+CT+slices+on+axial+axis">Relative location of CT slices on axial axis</a>,
<a href="datasets/Online+Handwritten+Assamese+Characters+Dataset"></a>,
<a href="datasets/Online+Handwritten+Assamese+Characters+Dataset">Online Handwritten Assamese Characters Dataset</a>,
<a href="datasets/PubChem+Bioassay+Data"></a>,
<a href="datasets/PubChem+Bioassay+Data">PubChem Bioassay Data</a>,
<a href="datasets/Record+Linkage+Comparison+Patterns"></a>,
```

<a href="datasets/Record+Linkage+Comparison+Patterns">Record Linkage Comparison Patterns</a>,  
<a href="datasets/Communities+and+Crime+Unnormalized"></a>,  
<a href="datasets/Communities+and+Crime+Unnormalized">Communities and Crime Unnormalized</a>,  
<a href="datasets/Vertebral+Column"></a>,  
<a href="datasets/Vertebral+Column">Vertebral Column</a>,  
<a href="datasets/EMG+Physical+Action+Data+Set"></a>,  
<a href="datasets/EMG+Physical+Action+Data+Set">EMG Physical Action Data Set</a>,  
<a href="datasets/Vicon+Physical+Action+Data+Set"></a>,  
<a href="datasets/Vicon+Physical+Action+Data+Set">Vicon Physical Action Data Set</a>,  
<a href="datasets/Amazon+Commerce+reviews+set"></a>,  
<a href="datasets/Amazon+Commerce+reviews+set">Amazon Commerce reviews set</a>,  
<a href="datasets/Amazon+Access+Samples"></a>,  
<a href="datasets/Amazon+Access+Samples">Amazon Access Samples</a>,  
<a href="datasets/Reuter\_50\_50"></a>,  
<a href="datasets/Reuter\_50\_50">Reuter\_50\_50</a>,  
<a href="datasets/Farm+Ads"></a>,  
<a href="datasets/Farm+Ads">Farm Ads</a>,  
<a href="datasets/DBWorld+e-mails"></a>,  
<a href="datasets/DBWorld+e-mails">DBWorld e-mails</a>,  
<a href="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29"></a>,  
<a href="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29">KEGG Metabolic Relation Network (Directed)</a>,  
<a href="datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29"></a>,  
<a href="datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29">KEGG Metabolic Reaction Network (Undirected)</a>,  
<a href="datasets/Bank+Marketing"></a>,  
<a href="datasets/Bank+Marketing">Bank Marketing</a>,  
<a href="datasets/YouTube+Comedy+Slam+Preference+Data"></a>,  
<a href="datasets/YouTube+Comedy+Slam+Preference+Data">YouTube Comedy Slam Preference Data</a>,  
<a href="datasets/Gas+Sensor+Array+Drift+Dataset"></a>,  
<a href="datasets/Gas+Sensor+Array+Drift+Dataset">Gas Sensor Array Drift Dataset</a>,  
<a href="datasets/ILPD+%28Indian+Liver+Patient+Dataset%29"></a>,  
<a href="datasets/ILPD+%28Indian+Liver+Patient+Dataset%29">ILPD (Indian Liver Patient Dataset)</a>,<a href="datasets/OPPORTUNITY+Activity+Recognition"></a>,  
<a href="datasets/OPPORTUNITY+Activity+Recognition">OPPORTUNITY Activity Recognition</a>,  
<a href="datasets/Nomao"></a>,  
<a href="datasets/Nomao">Nomao</a>,  
<a href="datasets/SMS+Spam+Collection"></a>,  
<a href="datasets/SMS+Spam+Collection">SMS Spam Collection</a>,  
<a href="datasets/Skin+Segmentation"></a>,  
<a href="datasets/Skin+Segmentation">Skin Segmentation</a>,  
<a href="datasets/Planning+Relax"></a>,  
<a href="datasets/Planning+Relax">Planning Relax</a>,  
<a href="datasets/PAMAP2+Physical+Activity+Monitoring"></a>,  
<a href="datasets/PAMAP2+Physical+Activity+Monitoring">PAMAP2 Physical Activity Monitoring</a>,  
<a href="datasets/Restaurant+%26+consumer+data"></a>,  
<a href="datasets/Restaurant+%26+consumer+data">Restaurant & consumer data</a>,  
<a href="datasets/CNAE-9"></a>,  
<a href="datasets/CNAE-9">CNAE-9</a>,  
<a href="datasets/Individual+household+electric+power+consumption"></a>,  
<a href="datasets/Individual+household+electric+power+consumption">Individual household electric power consumption</a>,  
<a href="datasets/seeds"></a>,  
<a href="datasets/seeds">seeds</a>,  
<a href="datasets/Northix"></a>.

<a href="datasets/Northix">Northix</a>,  
<a href="datasets/QtyT40I10D100K"></a>,  
<a href="datasets/QtyT40I10D100K">QtyT40I10D100K</a>,  
<a href="datasets/Legal+Case+Reports"></a>,  
<a href="datasets/Legal+Case+Reports">Legal Case Reports</a>,  
<a href="datasets/Human+Activity+Recognition+Using+Smartphones"></a>,  
<a href="datasets/Human+Activity+Recognition+Using+Smartphones">Human Activity Recognition Using Smartphones</a>,  
<a href="datasets/One-hundred+plant+species+leaves+data+set"></a>,  
<a href="datasets/One-hundred+plant+species+leaves+data+set">One-hundred plant species leaves data set</a>,  
<a href="datasets/Energy+efficiency"></a>,  
<a href="datasets/Energy+efficiency">Energy efficiency</a>,  
<a href="datasets/Yacht+Hydrodynamics"></a>,  
<a href="datasets/Yacht+Hydrodynamics">Yacht Hydrodynamics</a>,  
<a href="datasets/Fertility"></a>,  
<a href="datasets/Fertility">Fertility</a>,  
<a href="datasets/Daphnet+Freezing+of+Gait"></a>,  
<a href="datasets/Daphnet+Freezing+of+Gait">Daphnet Freezing of Gait</a>,  
<a href="datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29"></a>,  
<a href="datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29">3D Road Network (North Jutland, Denmark)</a>,  
<a href="datasets/ISTANBUL+STOCK+EXCHANGE"></a>,  
<a href="datasets/ISTANBUL+STOCK+EXCHANGE">ISTANBUL STOCK EXCHANGE</a>,  
<a href="datasets/Buzz+in+social+media+"></a>,  
<a href="datasets/Buzz+in+social+media+">Buzz in social media </a>,  
<a href="datasets/First-order+theorem+proving"></a>,  
<a href="datasets/First-order+theorem+proving">First-order theorem proving</a>,  
<a href="datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29"></a>,  
<a href="datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29">Wearable Computing: Classification of Body Postures and Movements (PUC-Rio)</a>,  
<a href="datasets/Gas+sensor+arrays+int+open+sampling+settings"></a>,  
<a href="datasets/Gas+sensor+arrays+int+open+sampling+settings">Gas sensor arrays in open sampling settings</a>,  
<a href="datasets/Climate+Model+Simulation+Crashes"></a>,  
<a href="datasets/Climate+Model+Simulation+Crashes">Climate Model Simulation Crashes</a>,  
<a href="datasets/MicroMass"></a>,  
<a href="datasets/MicroMass">MicroMass</a>,  
<a href="datasets/QSAR+biodegradation"></a>,  
<a href="datasets/QSAR+biodegradation">QSAR biodegradation</a>,  
<a href="datasets/BLOGGER"></a>,  
<a href="datasets/BLOGGER">BLOGGER</a>,  
<a href="datasets/Daily+and+Sports+Activities"></a>,  
<a href="datasets/Daily+and+Sports+Activities">Daily and Sports Activities</a>,  
<a href="datasets/User+Knowledge+Modeling"></a>,  
<a href="datasets/User+Knowledge+Modeling">User Knowledge Modeling</a>,  
<a href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection"></a>,  
<a href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection">Reuters RCV1 RCV2 Multilingual, Multiview Text Categorization Test collection</a>,  
<a href="datasets/NYSK"></a>,  
<a href="datasets/NYSK">NYSK</a>,  
<a href="datasets/Turkiye+Student+Evaluation"></a>,  
<a href="datasets/Turkiye+Student+Evaluation">Turkiye Student Evaluation</a>,  
<a href="datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machine"></a>.

</a>,  
<a href="datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machine">ser Knowledge Modeling Data (Students' Knowledge Levels on DC Electrical Machines)</a>,  
<a href="datasets/EEG+Eye+State"></a>,<a href="datasets/EEG+Eye+State">EEG Eye State</a>,  
<a href="datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure"></a>,  
<a href="datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure">Physicochemical Properties of Protein Tertiary Structure</a>,  
<a href="datasets/seismic-bumps"></a>,<a href="datasets/seismic-bumps">seismic-bumps</a>,  
<a href="datasets/banknote+authentication"></a>,  
<a href="datasets/banknote+authentication">banknote authentication</a>,  
<a href="datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Problem%3A+Pat"></a>,<a href="datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Problem%3A+Pat">USPTO Algorithm Challenge, run by NASA-Harvard Tournament Lab and TopCoder Problem: Pat</a>,<a href="datasets/YouTube+Multiview+Video+Games+Dataset"></a>,<a href="datasets/YouTube+Multiview+Video+Games+Dataset">YouTube Multiview Video Games Dataset</a>,<a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations"></a>,<a href="datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations">Gas Sensor Array Drift Dataset at Different Concentrations</a>,<a href="datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors"></a>,<a href="datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors">Activities of Daily Living (ADLs) Recognition Using Binary Sensors</a>,<a href="datasets/SkillCraft1+Master+Table+Dataset"></a>,<a href="datasets/SkillCraft1+Master+Table+Dataset">SkillCraft1 Master Table Dataset</a>,<a href="datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units"></a>,<a href="datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units">Weight Lifting Exercises monitored with Inertial Measurement Units</a>,<a href="datasets/SML2010"></a>,<a href="datasets/SML2010">SML2010</a>,<a href="datasets/Bike+Sharing+Dataset"></a>,<a href="datasets/Bike+Sharing+Dataset">Bike Sharing Dataset</a>,<a href="datasets/Predict+keywords+activities+in+a+online+social+media"></a>,<a href="datasets/Predict+keywords+activities+in+a+online+social+media">Predict keywords activities in a online social media</a>,<a href="datasets/Thoracic+Surgery+Data"></a>,<a href="datasets/Thoracic+Surgery+Data">Thoracic Surgery Data</a>,<a href="datasets/EMG+dataset+in+Lower+Limb"></a>,<a href="datasets/EMG+dataset+in+Lower+Limb">EMG dataset in Lower Limb</a>,<a href="datasets/SUSY"></a>,<a href="datasets/SUSY">SUSY</a>,<a href="datasets/HIGGS"></a>,<a href="datasets/HIGGS">HIGGS</a>,<a href="datasets/Qualitative\_Bankruptcy"></a>,<a href="datasets/Qualitative\_Bankruptcy">Qualitative\_Bankruptcy</a>,<a href="datasets/LSVT+Voice+Rehabilitation"></a>,<a href="datasets/LSVT+Voice+Rehabilitation">LSVT Voice Rehabilitation</a>,<a href="datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer"></a>,<a href="datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer">Dataset for ADL Recognition with Wrist-worn Accelerometer</a>,<a href="datasets/Wilt"></a>,<a href="datasets/Wilt">Wilt</a>,<a href="datasets/User+Identification+From+Walking+Activity"></a>,<a href="datasets/User+Identification+From+Walking+Activity">User Identification From Walking Activity</a>

+v+cy~/\~,  
<a href="datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer"><img border="1" s  
rc="assets/MLimages/SmallLargedefault.jpg"/></a>,  
<a href="datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer">Activity  
Recognition from Single Chest-Mounted Accelerometer</a>,  
<a href="datasets/Leaf"></a>,  
<a href="datasets/Leaf">Leaf</a>,  
<a href="datasets/Dresses\_Attribute\_Sales"></a>,  
<a href="datasets/Dresses\_Attribute\_Sales">Dresses\_Attribute\_Sales</a>,  
<a href="datasets/Tamilnadu+Electricity+Board+Hourly+Readings"></a>,  
<a href="datasets/Tamilnadu+Electricity+Board+Hourly+Readings">Tamilnadu Electricity Board Hourly  
Readings</a>,  
<a href="datasets/Airfoil+Self-Noise"></a>,  
<a href="datasets/Airfoil+Self-Noise">Airfoil Self-Noise</a>,  
<a href="datasets/Wholesale+customers"></a>,  
<a href="datasets/Wholesale+customers">Wholesale customers</a>,  
<a href="datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis"></a>,  
<a href="datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis">Twitter Data set for Arabic Sen  
timent Analysis</a>,  
<a href="datasets/Combined+Cycle+Power+Plant"></a>,  
<a href="datasets/Combined+Cycle+Power+Plant">Combined Cycle Power Plant</a>,  
<a href="datasets/Urban+Land+Cover">  
</a>,  
<a href="datasets/Urban+Land+Cover">Urban Land Cover</a>,  
<a href="datasets/Diabetes+130-US+hospitals+for+years+1999-2008"></a>,  
<a href="datasets/Diabetes+130-US+hospitals+for+years+1999-2008">Diabetes 130-US hospitals for ye  
ars 1999-2008</a>,  
<a href="datasets/Bach+Choral+Harmony"></a>,  
<a href="datasets/Bach+Choral+Harmony">Bach Choral Harmony</a>,  
<a href="datasets/StoneFlakes"></a>,  
<a href="datasets/StoneFlakes">StoneFlakes</a>,  
<a href="datasets/Tennis+Major+Tournament+Match+Statistics"></a>,  
<a href="datasets/Tennis+Major+Tournament+Match+Statistics">Tennis Major Tournament Match  
Statistics</a>,  
<a href="datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings"></a>,  
<a href="datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings">Parkinson Sp  
eech Dataset with Multiple Types of Sound Recordings</a>,  
<a href="datasets/Gesture+Phase+Segmentation"></a>,  
<a href="datasets/Gesture+Phase+Segmentation">Gesture Phase Segmentation</a>,  
<a href="datasets/Perfume+Data"></a>  
,  
<a href="datasets/Perfume+Data">Perfume Data</a>,  
<a href="datasets/BlogFeedback"></a>  
,  
<a href="datasets/BlogFeedback">BlogFeedback</a>,  
<a href="datasets/REALDISP+Activity+Recognition+Dataset"></a>,  
<a href="datasets/REALDISP+Activity+Recognition+Dataset">REALDISP Activity Recognition  
Dataset</a>,  
<a href="datasets/Newspaper+and+magazine+images+segmentation+dataset"></a>,  
<a href="datasets/Newspaper+and+magazine+images+segmentation+dataset">Newspaper and magazine imag  
es segmentation dataset</a>,  
<a href="datasets/AAAI+2014+Accepted+Papers"></a>,  
<a href="datasets/AAAI+2014+Accepted+Papers">AAAI 2014 Accepted Papers</a>,  
<a href="datasets/Gas+sensor+array+under+flow+modulation"></a>,  
<a href="datasets/Gas+sensor+array+under+flow+modulation">Gas sensor array under flow  
modulation</a>,  
<a href="datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures"></a>,  
<a href="datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures">Gas sensor array exposed to  
turbulent gas mixtures</a>,  
<a href="datasets/UJIIndoorLoc"></a>  
,

```

<a href="datasets/Classification+Classification"></a>,
<a href="datasets/Sentence+Classification"></a>,
<a href="datasets/Dow+Jones+Index"></a>,
<a href="datasets/Dow+Jones+Index">Dow Jones Index</a>,
<a href="datasets/sEMG+for+Basic+Hand+movements"></a>,
<a href="datasets/sEMG+for+Basic+Hand+movements">sEMG for Basic Hand movements</a>,
<a href="datasets/AAAI+2013+Accepted+Papers"></a>,
<a href="datasets/AAAI+2013+Accepted+Papers">AAAI 2013 Accepted Papers</a>,
<a href="datasets/Geographical+Original+of+Music"></a>,
<a href="datasets/Geographical+Original+of+Music">Geographical Original of Music</a>,
<a href="datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants"></a>,
<a href="datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants">Condition Based
Maintenance of Naval Propulsion Plants</a>,
<a href="datasets/Grammatical+Facial+Expressions"></a>,
<a href="datasets/Grammatical+Facial+Expressions">Grammatical Facial Expressions</a>,
<a href="datasets/NoisyOffice"></a>,
<a href="datasets/NoisyOffice">NoisyOffice</a>,
<a href="datasets/MHEALTH+Dataset"><
/a>,
<a href="datasets/MHEALTH+Dataset">MHEALTH Dataset</a>,
<a href="datasets/Student+Performance"></a>,
<a href="datasets/Student+Performance">Student Performance</a>,
<a href="datasets/ElectricityLoadDiagrams20112014"></a>,
<a href="datasets/ElectricityLoadDiagrams20112014">ElectricityLoadDiagrams20112014</a>,
<a href="datasets/Gas+sensor+array+under+dynamic+gas+mixtures"></a>,
<a href="datasets/Gas+sensor+array+under+dynamic+gas+mixtures">Gas sensor array under dynamic gas
mixtures</a>,
<a href="datasets/microblogPCU"></a>
,
<a href="datasets/microblogPCU">microblogPCU</a>,
<a href="datasets/Firm-Teacher_Clave-Direction_Classification"></a>,
<a href="datasets/Firm-Teacher_Clave-Direction_Classification">Firm-Teacher_Clave-
Direction_Classification</a>,
<a href="datasets/Dataset+for+Sensorless+Drive+Diagnosis"></a>,
<a href="datasets/Dataset+for+Sensorless+Drive+Diagnosis">Dataset for Sensorless Drive
Diagnosis</a>,
<a href="datasets/TV+News+Channel+Commercial+Detection+Dataset"></a>,
<a href="datasets/TV+News+Channel+Commercial+Detection+Dataset">TV News Channel Commercial
Detection Dataset</a>,
<a href="datasets/Phishing+Websites"></a>,
<a href="datasets/Phishing+Websites">Phishing Websites</a>,
<a href="datasets/Greenhouse+Gas+Observing+Network"></a>,
<a href="datasets/Greenhouse+Gas+Observing+Network">Greenhouse Gas Observing Network</a>,
<a href="datasets/Diabetic+Retinopathy+Debrecen+Data+Set"></a>,
<a href="datasets/Diabetic+Retinopathy+Debrecen+Data+Set">Diabetic Retinopathy Debrecen Data Set<
/a>,
<a href="datasets/HIV-1+protease+cleavage"></a>,
<a href="datasets/HIV-1+protease+cleavage">HIV-1 protease cleavage</a>,
<a href="datasets/Sentiment+Labelled+Sentences"></a>,
<a href="datasets/Sentiment+Labelled+Sentences">Sentiment Labelled Sentences</a>,
<a href="datasets/Online+News+Popularity"></a>,
<a href="datasets/Online+News+Popularity">Online News Popularity</a>,
<a href="datasets/Forest+type+mapping"></a>,
<a href="datasets/Forest+type+mapping">Forest type mapping</a>,
<a href="datasets/wiki4HE"></a>,
<a href="datasets/wiki4HE">wiki4HE</a>,
<a href="datasets/online+news+popularity+and+topic+classification+dataset"></a>,

```

<a href="datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset"></a>,  
<a href="datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset">Online Video Characteristics and Transcoding Time Dataset</a>,  
<a href="datasets/Chronic\_Kidney\_Disease"></a>,  
<a href="datasets/Chronic\_Kidney\_Disease">Chronic Kidney Disease</a>,  
<a href="datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014"></a>,  
<a href="datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014">Machine Learning based ZZAlpha Ltd. Stock Recommendations 2012-2014</a>,  
<a href="datasets/Folio"></a>,  
<a href="datasets/Folio">Folio</a>,  
<a href="datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015"></a>,  
<a href="datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015">Taxi Service Trajectory - Prediction Challenge, ECML PKDD 2015</a>,  
<a href="datasets/Cuff-Less+Blood+Pressure+Estimation"></a>,  
<a href="datasets/Cuff-Less+Blood+Pressure+Estimation">Cuff-Less Blood Pressure Estimation</a>,  
<a href="datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions"></a>,  
<a href="datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions">Smartphone-Based Recognition of Human Activities and Postural Transitions</a>,  
<a href="datasets/Mice+Protein+Expression"></a>,  
<a href="datasets/Mice+Protein+Expression">Mice Protein Expression</a>,  
<a href="datasets/UJIIndoorLoc-Mag"></a>,  
<a href="datasets/UJIIndoorLoc-Mag">UJI IndoorLoc-Mag</a>,  
<a href="datasets/Heterogeneity+Activity+Recognition"></a>,  
<a href="datasets/Heterogeneity+Activity+Recognition">Heterogeneity Activity Recognition</a>,  
<a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set"></a>,  
<a href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set">Educational Process Mining (EPM): A Learning Analytics Data Set</a>,  
<a href="datasets/HEPMASS"></a>,  
<a href="datasets/HEPMASS">HEPMASS</a>,  
<a href="datasets/Indoor+User+Movement+Prediction+from+RSS+data"></a>,  
<a href="datasets/Indoor+User+Movement+Prediction+from+RSS+data">Indoor User Movement Prediction from RSS data</a>,  
<a href="datasets/Open+University+Learning+Analytics+dataset"></a>,  
<a href="datasets/Open+University+Learning+Analytics+dataset">Open University Learning Analytics dataset</a>,  
<a href="datasets/default+of+credit+card+clients"></a>,  
<a href="datasets/default+of+credit+card+clients">default of credit card clients</a>,  
<a href="datasets/Mesothelioma%280%99s+disease+data+set+"></a>,  
<a href="datasets/Mesothelioma%280%99s+disease+data+set+">Mesothelioma's disease data set </a>,  
<a href="datasets/Online+Retail"></a>,<a href="datasets/Online+Retail">Online Retail</a>,  
<a href="datasets/SIFT10M"></a>,  
<a href="datasets/SIFT10M">SIFT10M</a>,  
<a href="datasets/GPS+Trajectories"></a>,  
<a href="datasets/GPS+Trajectories">GPS Trajectories</a>,  
<a href="datasets/Detect+Malicious+Executable%28AntiVirus%29"></a>,  
<a href="datasets/Detect+Malicious+Executable%28AntiVirus%29">Detect Malicious Executable (AntiVirus)</a>,  
<a href="datasets/Occupancy+Detection+"></a>,  
<a href="datasets/Occupancy+Detection+">Occupancy Detection </a>,  
<a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Parkinson%280%Disease"></a>,  
<a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Parkinson%280%Disease">Improved Spiral Test Using Digitized Graphics Tablet for Monitoring Parkinson's Disease</a>,

<a href="datasets/News+Aggregator"></a>,  
<a href="datasets/News+Aggregator">News Aggregator</a>,  
<a href="datasets/Air+Quality"></a>,  
<a href="datasets/Air+Quality">Air Quality</a>,  
<a href="datasets/Twin+gas+sensor+arrays"></a>,  
<a href="datasets/Twin+gas+sensor+arrays">Twin gas sensor arrays</a>,  
<a href="datasets/Gas+sensors+for+home+activity+monitoring"></a>,  
<a href="datasets/Gas+sensors+for+home+activity+monitoring">Gas sensors for home activity monitoring</a>,  
<a href="datasets/Facebook+Comment+Volume+Dataset"></a>,  
<a href="datasets/Facebook+Comment+Volume+Dataset">Facebook Comment Volume Dataset</a>,  
<a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Living+AAL%29"></a>,  
<a href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Living+AAL%29">Smartphone Dataset for Human Activity Recognition (HAR) in Ambient Assisted Living (AAL)</a>,  
<a href="datasets/Polish+companies+bankruptcy+data"></a>,  
<a href="datasets/Polish+companies+bankruptcy+data">Polish companies bankruptcy data</a>,  
<a href="datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29"></a>,  
<a href="datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29">Activity Recognition system based on Multisensor data fusion (AReM)</a>,  
<a href="datasets/Dota2+Games+Results"></a>,  
<a href="datasets/Dota2+Games+Results">Dota2 Games Results</a>,  
<a href="datasets/Facebook+metrics"></a>,  
<a href="datasets/Facebook+metrics">Facebook metrics</a>,  
<a href="datasets/UbiqLog+%28smartphone+lifelogging%29"></a>,  
<a href="datasets/UbiqLog+%28smartphone+lifelogging%29">UbiqLog (smartphone lifelogging)</a>,  
<a href="datasets/NIPS+Conference+Papers+1987-2015"></a>,  
<a href="datasets/NIPS+Conference+Papers+1987-2015">NIPS Conference Papers 1987-2015</a>,  
<a href="datasets/HTRU2"></a>,  
<a href="datasets/HTRU2">HTRU2</a>,  
<a href="datasets/Drug+consumption+%28quantified%29"></a>,  
<a href="datasets/Drug+consumption+%28quantified%29">Drug consumption (quantified)</a>,  
<a href="datasets/Appiances+energy+prediction"></a>,  
<a href="datasets/Appiances+energy+prediction">Appliances energy prediction</a>,  
<a href="datasets/Miskolc+IIS+Hybrid+IPS"></a>,  
<a href="datasets/Miskolc+IIS+Hybrid+IPS">Miskolc IIS Hybrid IPS</a>,  
<a href="datasets/KDC-4007+dataset+Collection"></a>,  
<a href="datasets/KDC-4007+dataset+Collection">KDC-4007 dataset Collection</a>,  
<a href="datasets/Geo-Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone"></a>,  
<a href="datasets/Geo-Magnetic field and WLAN dataset for indoor localisation from wristband and smartphone">Geo-Magnetic field and WLAN dataset for indoor localisation from wristband and smartphone</a>,  
<a href="datasets/DrivFace"></a>,  
<a href="datasets/DrivFace">DrivFace</a>,  
<a href="datasets/Website+Phishing"></a>,  
<a href="datasets/Website+Phishing">Website Phishing</a>,  
<a href="datasets/YouTube+Spam+Collection"></a>,  
<a href="datasets/YouTube+Spam+Collection">YouTube Spam Collection</a>,  
<a href="datasets/Beijing+PM2.5+Data"></a>,  
<a href="datasets/Beijing+PM2.5+Data">Beijing PM2.5 Data</a>,  
<a href="datasets/Cargo+2000+Freight+Tracking+and+Tracing"></a>,  
<a href="datasets/Cargo+2000+Freight+Tracking+and+Tracing">Cargo 2000 Freight Tracking and Tracing</a>,

<a href="datasets/Cervical+cancer+Risk+Factors%29"></a>,  
<a href="datasets/Cervical+cancer+28Risk+Factors%29">Cervical cancer (Risk Factors)</a>,  
<a href="datasets/Quality+Assessment+of+Digital+Colposcopies"></a>,  
<a href="datasets/Quality+Assessment+of+Digital+Colposcopies">Quality Assessment of Digital Colposcopies</a>,  
<a href="datasets/KASANDR"></a>,  
<a href="datasets/KASANDR">KASANDR</a>,  
<a href="datasets/FMA%3A+A+Dataset+For+Music+Analysis"></a>,  
<a href="datasets/FMA%3A+A+Dataset+For+Music+Analysis">FMA: A Dataset For Music Analysis</a>,  
<a href="datasets/Air+quality"></a>,  
<a href="datasets/Air+quality">Air quality</a>,  
<a href="datasets/Epileptic+Seizure+Recognition"></a>,  
<a href="datasets/Epileptic+Seizure+Recognition">Epileptic Seizure Recognition</a>,  
<a href="datasets/Devanagari+Handwritten+Character+Dataset"></a>,  
<a href="datasets/Devanagari+Handwritten+Character+Dataset">Devanagari Handwritten Character Data set</a>,  
<a href="datasets/Stock+portfolio+performance"></a>,  
<a href="datasets/Stock+portfolio+performance">Stock portfolio performance</a>,  
<a href="datasets/MoCap+Hand+Postures"></a>,  
<a href="datasets/MoCap+Hand+Postures">MoCap Hand Postures</a>,  
<a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech"></a>,  
<a href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech">Early biomarkers of Parkinson's disease based on natural connected speech</a>,  
<a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting"></a>,  
<a href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting">Data for Software Engineering Teamwork Assessment in Education Setting</a>,  
<a href="datasets/PM2.5+Data+of+Five+Chinese+Cities"></a>,  
<a href="datasets/PM2.5+Data+of+Five+Chinese+Cities">PM2.5 Data of Five Chinese Cities</a>,  
<a href="datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet"></a>,  
<a href="datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet">Parkinson Disease Spiral Drawings Using Digitized Graphics Tablet</a>,  
<a href="datasets/Sales\_Transactions\_Dataset\_Weekly"></a>,  
<a href="datasets/Sales\_Transactions\_Dataset\_Weekly">Sales\_Transactions\_Dataset\_Weekly</a>,  
<a href="datasets/Las+Vegas+Strip"></a>,  
<a href="datasets/Las+Vegas+Strip">Las Vegas Strip</a>,  
<a href="datasets/Eco-hotel"></a>,  
<a href="datasets/Eco-hotel">Eco-hotel</a>,  
<a href="datasets/MEU-Mobile+KSD"></a>,  
<a href="datasets/MEU-Mobile+KSD">MEU-Mobile KSD</a>,  
<a href="datasets/Crowdsourced+Mapping"></a>,  
<a href="datasets/Crowdsourced+Mapping">Crowdsourced Mapping</a>,  
<a href="datasets/gene+expression+cancer+RNA-Seq"></a>,  
<a href="datasets/gene+expression+cancer+RNA-Seq">gene expression cancer RNA-Seq</a>,  
<a href="datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer"></a>,  
<a href="datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer">Hybrid Indoor Positioning Dataset from WiFi RSSI, Bluetooth and magnetometer</a>,  
<a href="datasets/chestnut+%E2%80%93+LARVIC"></a>,  
<a href="datasets/chestnut+%E2%80%93+LARVIC">chestnut - LARVIC</a>,  
<a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%28OBS%29+Nrk"></a>,  
<a href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%28OBS%29+Nrk">Burst Header Packet (BHP) flooding attack on Optical Burst Switching (OBS) Network</a>,  
<a href="datasets/Motion+Capture+Hand+Postures"></a>,

```
<a href="datasets/Motion+Capture+Hand+Postures">Motion Capture Hand Postures</a>,
<a href="datasets/Anuran+Calls+%28MFCCs%29"></a>,
<a href="datasets/Anuran+Calls+%28MFCCs%29">Anuran Calls (MFCCs)</a>,
<a href="datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization"></a>,
<a href="datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization">TTC-3600: Benchmark dataset for Turkish text categorization</a>,
<a href="datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy"></a>,
<a href="datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy">Gastrointestinal Lesions in Regular Colonoscopy</a>,
<a href="datasets/Daily+Demand+Forecasting+Orders"></a>,
<a href="datasets/Daily+Demand+Forecasting+Orders">Daily Demand Forecasting Orders</a>,
<a href="datasets/Paper+Reviews"></a>,
<a href="datasets/Paper+Reviews">Paper Reviews</a>,
<a href="datasets/extention+of+Z-Alizadeh+sani+dataset"></a>,
<a href="datasets/extention+of+Z-Alizadeh+sani+dataset">extention of Z-Alizadeh sani dataset</a>,
<a href="datasets/Z-Alizadeh+Sani"></a>,
<a href="datasets/Z-Alizadeh+Sani">Z-Alizadeh Sani</a>,
<a href="datasets/Dynamic+Features+of+VirusShare+Executables"></a>,
<a href="datasets/Dynamic+Features+of+VirusShare+Executables">Dynamic Features of VirusShare Executables</a>,
<a href="datasets/IDA2016Challenge"></a>,
<a href="datasets/IDA2016Challenge">IDA2016Challenge</a>,
<a href="datasets/DSRC+Vehicle+Communications"></a>,
<a href="datasets/DSRC+Vehicle+Communications">DSRC Vehicle Communications</a>,
<a href="datasets/Mturk+User-Perceived+Clusters+over+Images"></a>,
<a href="datasets/Mturk+User-Perceived+Clusters+over+Images">Mturk User-Perceived Clusters over Images</a>,
<a href="datasets/Character+Font+Images"></a>,
<a href="datasets/Character+Font+Images">Character Font Images</a>,
<a href="datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Instance+Labels"></a>,
<a href="datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-Instance+Learning+with+Instance+Labels">DeliciousMIL: A Data Set for Multi-Label Multi-Instance Learning with Instance Labels</a>,
<a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++"></a>,
<a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++">Autistic Spectrum Disorder Screening Data for Children </a>,
<a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++"></a>,
<a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++">Autistic Spectrum Disorder Screening Data for Adolescent </a>,
<a href="datasets/APS+Failure+at+Scania+Trucks"></a>,
<a href="datasets/APS+Failure+at+Scania+Trucks">APS Failure at Scania Trucks</a>,
<a href="datasets/Wireless+Indoor+Localization"></a>,
<a href="datasets/Wireless+Indoor+Localization">Wireless Indoor Localization</a>,
<a href="datasets/HCC+Survival"></a>,
<a href="datasets/HCC+Survival">HCC Survival</a>,
<a href="datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015"></a>,
<a href="datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015">CSM (Conventional and Social Media Movies) Dataset 2014 and 2015</a>,
<a href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29"></a>,
<a href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29">University of Tehran Question Dataset 2016 (UTQD.2016)</a>,
<a href="datasets/Autism+Screening+Adult"></a>,
<a href="datasets/Autism+Screening+Adult">Autism Screening Adult</a>,
<a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor">
```

```
</a>,
<a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor">
    Activity recognition with healthy older people using a batteryless wearable sensor</a>,
    <a href="datasets/Immunotherapy+Dataset"></a>,
        <a href="datasets/Immunotherapy+Dataset">Immunotherapy Dataset</a>,
        <a href="datasets/Cryotherapy+Dataset"></a>,
            <a href="datasets/Cryotherapy+Dataset">Cryotherapy Dataset </a>,
            <a href="datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes"></a>,
                <a href="datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes">OCT data & Color F undus Images of Left & Right Eyes</a>,
                <a href="datasets/Discrete+Tone+Image+Dataset"></a>,
                    <a href="datasets/Discrete+Tone+Image+Dataset">Discrete Tone Image Dataset</a>,
                    <a href="datasets/News+Popularity+in+Multiple+Social+Media+Platforms"></a>,
                        <a href="datasets/News+Popularity+in+Multiple+Social+Media+Platforms">News Popularity in Multiple Social Media Platforms</a>,
                        <a href="datasets/Ultrasonic+flowmeter+diagnostics"></a>,
                            <a href="datasets/Ultrasonic+flowmeter+diagnostics">Ultrasonic flowmeter diagnostics</a>,
                            <a href="datasets/ICMLA+2014+Accepted+Papers+Data+Set"></a>,
                                <a href="datasets/ICMLA+2014+Accepted+Papers+Data+Set">ICMLA 2014 Accepted Papers Data Set</a>,
                                <a href="datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation"></a>,
                                    <a href="datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation">BLE RSSI Dataset for Indoor localization and Navigation</a>,
                                    <a href="datasets/Container+Crane+Controller+Data+Set"></a>,
                                        <a href="datasets/Container+Crane+Controller+Data+Set">Container Crane Controller Data Set</a>,
                                        <a href="datasets/Residential+Building+Data+Set"></a>,
                                            <a href="datasets/Residential+Building+Data+Set">Residential Building Data Set</a>,
                                            <a href="datasets/Health+News+in+Twitter"></a>,
                                                <a href="datasets/Health+News+in+Twitter">Health News in Twitter</a>,
                                                <a href="datasets/chipseq"></a>,
                                                    <a href="datasets/chipseq">chipseq</a>,
                                                    <a href="datasets/SGEMM+GPU+kernel+performance"></a>,
                                                        <a href="datasets/SGEMM+GPU+kernel+performance">SGEMM GPU kernel performance</a>,
                                                        <a href="datasets/Repeat+Consumption+Matrices"></a>,
                                                            <a href="datasets/Repeat+Consumption+Matrices">Repeat Consumption Matrices</a>,
                                                            <a href="datasets/detection_of_IoT_botnet_attacks_N_BaIoT"></a>,
                                                                <a href="datasets/detection_of_IoT_botnet_attacks_N_BaIoT">detection_of_IoT_botnet_attacks_N_BaIoT</a>

                                                                <a href="datasets/Absenteeism+at+work"></a>,
                                                                    <a href="datasets/Absenteeism+at+work">Absenteeism at work</a>,
                                                                    <a href="datasets/SCADI"></a>,
                                                                    <a href="datasets/SCADI">SCADI</a>,
                                                                    <a href="datasets/Condition+monitoring+of+hydraulic+systems"></a>,
                                                                        <a href="datasets/Condition+monitoring+of+hydraulic+systems">Condition monitoring of hydraulic systems</a>,
                                                                        <a href="datasets/Carbon+Nanotubes"></a>,
                                                                            <a href="datasets/Carbon+Nanotubes">Carbon Nanotubes</a>,
                                                                            <a href="datasets/Optical+Interconnection+Network+ "></a>,
                                                                                <a href="datasets/Optical+Interconnection+Network+ ">Optical Interconnection Network </a>,
                                                                                <a href="datasets/Sports+articles+for+objectivity+analysis"></a>,
                                                                                    <a href="datasets/Sports+articles+for+objectivity+analysis">Sports articles for objectivity analysis</a>,
                                                                                    <a href="datasets/Breast+Cancer+Coimbra"></a>,
                                                                                        <a href="datasets/Breast+Cancer+Coimbra">Breast Cancer Coimbra</a>,
                                                                                        <a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data"></a>,
```

<a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data">GNFUV Unmanned Surface Vehicles Sensor Data</a>,  
<a href="datasets/Dishonest+Internet+users+Dataset"></a>,  
<a href="datasets/Dishonest+Internet+users+Dataset">Dishonest Internet users Dataset</a>,  
<a href="datasets/Victorian+Era+Authorship+Attribution"></a>,  
<a href="datasets/Victorian+Era+Authorship+Attribution">Victorian Era Authorship Attribution</a>,  
<a href="datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set"></a>,  
<a href="datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set">Simulated Falls and Daily Living Activities Data Set</a>,  
<a href="datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing"></a>,  
<a href="datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing">Multimodal Damage Identification for Humanitarian Computing</a>,  
<a href="datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals"></a>,  
<a href="datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals">EEG Steady-State Visual Evoked Potential Signals</a>,  
<a href="datasets/Roman+Urdu+Data+Set"></a>,  
<a href="datasets/Roman+Urdu+Data+Set">Roman Urdu Data Set</a>,  
<a href="datasets/Avila"></a>,  
<a href="datasets/Avila">Avila</a>,  
<a href="datasets/PANDOR"></a>,  
<a href="datasets/PANDOR">PANDOR</a>,  
<a href="datasets/Drug+Review+Dataset+%28Druglib.com%29"></a>,  
<a href="datasets/Drug+Review+Dataset+%28Druglib.com%29">Drug Review Dataset (Druglib.com)</a>,  
<a href="datasets/Drug+Review+Dataset+%28Drugs.com%29"></a>,  
<a href="datasets/Drug+Review+Dataset+%28Drugs.com%29">Drug Review Dataset (Drugs.com)</a>,  
<a href="datasets/Physical+Unclonable+Functions"></a>,  
<a href="datasets/Physical+Unclonable+Functions">Physical Unclonable Functions</a>,  
<a href="datasets/Superc conductivity+Data"></a>,  
<a href="datasets/Superc conductivity+Data">Superconductivity Data</a>,  
<a href="datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29"></a>,  
<a href="datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29">WESAD (Wearable Stress and Affect Detection)</a>,  
<a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2"></a>,  
<a href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2">GNFUV Unmanned Surface Vehicles Sensor Data Set 2</a>,  
<a href="datasets/Student+Academics+Performance"></a>,  
<a href="datasets/Student+Academics+Performance">Student Academics Performance</a>,  
<a href="datasets/Online+Shoppers+Purchasing+Intention+Dataset"></a>,  
<a href="datasets/Online+Shoppers+Purchasing+Intention+Dataset">Online Shoppers Purchasing Intention Dataset</a>,  
<a href="datasets/PMU-UD"></a>,  
<a href="datasets/PMU-UD">PMU-UD</a>,  
<a href="datasets/Parkinson%27s+Disease+Classification"></a>,  
<a href="datasets/Parkinson%27s+Disease+Classification">Parkinson's Disease Classification</a>,  
<a href="datasets/Electrical+Grid+Stability+Simulated+Data+"></a>,  
<a href="datasets/Electrical+Grid+Stability+Simulated+Data+">Electrical Grid Stability Simulated Data </a>,  
<a href="datasets/Caesarian+Section+Classification+Dataset"></a>,  
<a href="datasets/Caesarian+Section+Classification+Dataset">Caesarian Section Classification Dataset</a>,  
<a href="datasets/BAUM-1"></a>,  
<a href="datasets/BAUM-1">BAUM-1</a>,  
<a href="datasets/BAUM-2"></a>,  
<a href="datasets/BAUM-2">BAUM-2</a>,  
<a href="datasets/Audit+Data"></a>,  
<a href="datasets/Audit+Data">Audit Data</a>,  
<a href="datasets/BuddyMove+Data+Set"></a>,  
<a href="datasets/BuddyMove+Data+Set">BuddyMove Data Set</a>,  
<a href="datasets/Real+estate+valuation+data+set"></a>

```

src="assets/MLimages/SmallLargedefault.jpg"/></a>,
<a href="datasets/Real+estate+valuation+data+set">Real estate valuation data set</a>,
<a href="datasets/Early+biomarkers+of+Parkinson%20%99s+disease+based+on+natural+connected+speech+Data+et+"></a>,
<a href="datasets/Early+biomarkers+of+Parkinson%20%99s+disease+based+on+natural+connected+speech+Data+et+">Early biomarkers of Parkinson's disease based on natural connected speech Data Set </a>,
<a href="datasets/Somerville+Happiness+Survey"></a>,
<a href="datasets/Somerville+Happiness+Survey">Somerville Happiness Survey</a>,
<a href="datasets/2.4+GHZ+Indoor+Channel+Measurements"></a>,
<a href="datasets/2.4+GHZ+Indoor+Channel+Measurements">2.4 GHZ Indoor Channel Measurements</a>,
<a href="datasets/EMG+data+for+gestures"></a>,
<a href="datasets/EMG+data+for+gestures">EMG data for gestures</a>,
<a href="datasets/Parking+Birmingham"></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"></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"></a>,
<a href="datasets/Travel+Reviews">Travel Reviews</a>,
<a href="datasets/Tarvel+Review+Ratings"></a>,
<a href="datasets/Tarvel+Review+Ratings">Tarvel Review Ratings</a>,
<a href="datasets/Rice+Leaf+Diseases"></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 [15]:

```
data = re.findall('"(datasets/.+?)"', str(links))
data
```

Out[15]:

```

['datasets/Abalone',
'datasets/Abalone',
'datasets/Adult',
'datasets/Adult',
'datasets/Annealing',
'datasets/Annealing',
'datasets/Anonymous+Microsoft+Web+Data',
'datasets/Anonymous+Microsoft+Web+Data',
'datasets/Arrhythmia',
'datasets/Arrhythmia',
'datasets/Artificial+Characters',
'datasets/Artificial+Characters',
'datasets/Audiology+%28Original%29',
'datasets/Audiology+%28Original%29',
'datasets/Audiology+%28Standardized%29',
'datasets/Audiology+%28Standardized%29',
'datasets/Auto+MPG',
'datasets/Auto+MPG',
'datasets/Automobile',
'datasets/Automobile',
'datasets/Badges',
'datasets/Badges',
'datasets/Balance+Scale',
'datasets/Balance+Scale',
'datasets/Balloons',
'datasets/Balloons',
'datasets/Breast+Cancer',
'datasets/Breast+Cancer',
'datasets/Breast+Cancer+Wisconsin+%28Original%29',
'datasets/Breast+Cancer+Wisconsin+%28Original%29',

```

```
'datasets/Breast+Cancer+Wisconsin+%28Prognostic%29',
'datasets/Breast+Cancer+Wisconsin+%28Prognostic%29',
'datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29',
'datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29',
'datasets/Pittsburgh+Bridges',
'datasets/Pittsburgh+Bridges',
'datasets/Car+Evaluation',
'datasets/Car+Evaluation',
'datasets/Census+Income',
'datasets/Census+Income',
'datasets/Chess+%28King-Rook+vs.+King-Knight%29',
'datasets/Chess+%28King-Rook+vs.+King-Knight%29',
'datasets/Chess+%28King-Rook+vs.+King-Pawn%29',
'datasets/Chess+%28King-Rook+vs.+King-Pawn%29',
'datasets/Chess+%28King-Rook+vs.+King%29',
'datasets/Chess+%28King-Rook+vs.+King%29',
'datasets/Chess+%28Domain+Theories%29',
'datasets/Chess+%28Domain+Theories%29',
'datasets/Bach+Chorales',
'datasets/Bach+Chorales',
'datasets/Connect-4',
'datasets/Connect-4',
'datasets/Credit+Approval',
'datasets/Credit+Approval',
'datasets/Japanese+Credit+Screening',
'datasets/Japanese+Credit+Screening',
'datasets/Computer+Hardware',
'datasets/Computer+Hardware',
'datasets/Contraceptive+Method+Choice',
'datasets/Contraceptive+Method+Choice',
'datasets/Covertype',
'datasets/Covertype',
'datasets/Cylinder+Bands',
'datasets/Cylinder+Bands',
'datasets/Dermatology',
'datasets/Dermatology',
'datasets/Diabetes',
'datasets/Diabetes',
'datasets/DGP2++The+Second+Data+Generation+Program',
'datasets/DGP2++The+Second+Data+Generation+Program',
'datasets/Document+Understanding',
'datasets/Document+Understanding',
'datasets/EBL+Domain+Theories',
'datasets/EBL+Domain+Theories',
'datasets/Echocardiogram',
'datasets/Echocardiogram',
'datasets/Ecoli',
'datasets/Ecoli',
'datasets/Flags',
'datasets/Flags',
'datasets/Function+Finding',
'datasets/Function+Finding',
'datasets/Glass+Identification',
'datasets/Glass+Identification',
'datasets/Haberman%27s+Survival',
'datasets/Haberman%27s+Survival',
'datasets/Hayes-Roth',
'datasets/Hayes-Roth',
'datasets/Heart+Disease',
'datasets/Heart+Disease',
'datasets/Hepatitis',
'datasets/Hepatitis',
'datasets/Horse+Colic',
'datasets/Horse+Colic',
'datasets/ICU',
'datasets/ICU',
'datasets/Image+Segmentation',
'datasets/Image+Segmentation',
'datasets/Internet+Advertisements',
'datasets/Internet+Advertisements',
'datasets/Ionosphere',
'datasets/Ionosphere',
'datasets/Iris',
'datasets/Iris',
'datasets/ISOLET',
'datasets/ISOLET',
'datasets/Kinship',
```

```
'datasets/Kinship',
'datasets/Labor+Relations',
'datasets/Labor+Relations',
'datasets/LED+Display+Domain',
'datasets/LED+Display+Domain',
'datasets/Lenses',
'datasets/Lenses',
'datasets/Letter+Recognition',
'datasets/Letter+Recognition',
'datasets/Liver+Disorders',
'datasets/Liver+Disorders',
'datasets/Logic+Theorist',
'datasets/Logic+Theorist',
'datasets/Lung+Cancer',
'datasets/Lung+Cancer',
'datasets/Lymphography',
'datasets/Lymphography',
'datasets/Mechanical+Analysis',
'datasets/Mechanical+Analysis',
'datasets/Meta-data',
'datasets/Meta-data',
'datasets/Mobile+Robots',
'datasets/Mobile+Robots',
'datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29',
'datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29',
'datasets/Molecular+Biology+%28Protein+Secondary+Structure%29',
'datasets/Molecular+Biology+%28Protein+Secondary+Structure%29',
'datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29',
'datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29',
'datasets/MONK%27s+Problems',
'datasets/MONK%27s+Problems',
'datasets/Moral+Reasoner',
'datasets/Moral+Reasoner',
'datasets/Multiple+Features',
'datasets/Multiple+Features',
'datasets/Mushroom',
'datasets/Mushroom',
'datasets/Musk+%28Version+1%29',
'datasets/Musk+%28Version+1%29',
'datasets/Musk+%28Version+2%29',
'datasets/Musk+%28Version+2%29',
'datasets/Nursery',
'datasets/Nursery',
'datasets/Othello+Domain+Theory',
'datasets/Othello+Domain+Theory',
'datasets/Page+Blocks+Classification',
'datasets/Page+Blocks+Classification',
'datasets/Optical+Recognition+of+Handwritten+Digits',
'datasets/Optical+Recognition+of+Handwritten+Digits',
'datasets/Pen-Based+Recognition+of+Handwritten+Digits',
'datasets/Pen-Based+Recognition+of+Handwritten+Digits',
'datasets/Post-Operative+Patient',
'datasets/Post-Operative+Patient',
'datasets/Primary+Tumor',
'datasets/Primary+Tumor',
'datasets/Prodigy',
'datasets/Prodigy',
'datasets/Qualitative+Structure+Activity+Relationships',
'datasets/Qualitative+Structure+Activity+Relationships',
'datasets/Quadruped+Mammals',
'datasets/Quadruped+Mammals',
'datasets/Servo',
'datasets/Servo',
'datasets/Shuttle+Landing+Control',
'datasets/Shuttle+Landing+Control',
'datasets/Solar+Flare',
'datasets/Solar+Flare',
'datasets/Soybean+%28Large%29',
'datasets/Soybean+%28Large%29',
'datasets/Soybean+%28Small%29',
'datasets/Soybean+%28Small%29',
'datasets/Challenger+USA+Space+Shuttle+O-Ring',
'datasets/Challenger+USA+Space+Shuttle+O-Ring',
'datasets/Low+Resolution+Spectrometer',
'datasets/Low+Resolution+Spectrometer',
'datasets/Spambase',
'datasets/Spambase'.
```

'datasets/Spambase',  
'datasets/SPECT+Heart',  
'datasets/SPECT+Heart',  
'datasets/SPECTF+Heart',  
'datasets/SPECTF+Heart',  
'datasets/Sponge',  
'datasets/Sponge',  
'datasets/Statlog+Project',  
'datasets/Statlog+Project',  
'datasets/Student+Loan+Relational',  
'datasets/Student+Loan+Relational',  
'datasets/Teaching+Assistant+Evaluation',  
'datasets/Teaching+Assistant+Evaluation',  
'datasets/Tic-Tac-Toe+Endgame',  
'datasets/Tic-Tac-Toe+Endgame',  
'datasets/Thyroid+Disease',  
'datasets/Thyroid+Disease',  
'datasets/Trains',  
'datasets/Trains',  
'datasets/University',  
'datasets/University',  
'datasets/Congressional+Voting+Records',  
'datasets/Congressional+Voting+Records',  
'datasets/Water+Treatment+Plant',  
'datasets/Water+Treatment+Plant',  
'datasets/Waveform+Database+Generator+%28Version+1%29',  
'datasets/Waveform+Database+Generator+%28Version+1%29',  
'datasets/Waveform+Database+Generator+%28Version+2%29',  
'datasets/Waveform+Database+Generator+%28Version+2%29',  
'datasets/Wine',  
'datasets/Wine',  
'datasets/Yeast',  
'datasets/Yeast',  
'datasets/Zoo',  
'datasets/Zoo',  
'datasets/Undocumented',  
'datasets/Undocumented',  
'datasets/Twenty+Newsgroups',  
'datasets/Twenty+Newsgroups',  
'datasets/Australian+Sign+Language+signs',  
'datasets/Australian+Sign+Language+signs',  
'datasets/Australian+Sign+Language+signs+%28High+Quality%29',  
'datasets/Australian+Sign+Language+signs+%28High+Quality%29',  
'datasets/US+Census+Data+%281990%29',  
'datasets/US+Census+Data+%281990%29',  
'datasets/Census-Income+%28KDD%29',  
'datasets/Census-Income+%28KDD%29',  
'datasets/Coil+1999+Competition+Data',  
'datasets/Coil+1999+Competition+Data',  
'datasets/Corel+Image+Features',  
'datasets/Corel+Image+Features',  
'datasets/E.+Coli+Genes',  
'datasets/E.+Coli+Genes',  
'datasets/EEG+Database',  
'datasets/EEG+Database',  
'datasets/El+Nino',  
'datasets/El+Nino',  
'datasets/Entree+Chicago+Recommendation+Data',  
'datasets/Entree+Chicago+Recommendation+Data',  
'datasets/CMU+Face+Images',  
'datasets/CMU+Face+Images',  
'datasets/Insurance+Company+Benchmark+%28COIL+2000%29',  
'datasets/Insurance+Company+Benchmark+%28COIL+2000%29',  
'datasets/Internet+Usage+Data',  
'datasets/Internet+Usage+Data',  
'datasets/IPUMS+Census+Database',  
'datasets/IPUMS+Census+Database',  
'datasets/Japanese+Vowels',  
'datasets/Japanese+Vowels',  
'datasets/KDD+Cup+1998+Data',  
'datasets/KDD+Cup+1998+Data',  
'datasets/KDD+Cup+1999+Data',  
'datasets/KDD+Cup+1999+Data',  
'datasets/M.+Tuberculosis+Genes',  
'datasets/M.+Tuberculosis+Genes',  
'datasets/Movie',  
'datasets/Movie',  
'datasets/MSNRG.com+Anonymous+Web+Data'.

'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+Weber+Web+Page+Ratings',  
'datasets/Syskill+and+Weber+Web+Page+Ratings',  
'datasets/UNIX+User+Data',  
'datasets/UNIX+User+Data',  
'datasets/Volcanoes+on+Venus+-+JARtool+experiment',  
'datasets/Volcanoes+on+Venus+-+JARtool+experiment',  
'datasets/Statlog%28Australian+Credit+Approval%29',  
'datasets/Statlog%28Australian+Credit+Approval%29',  
'datasets/Statlog%28German+Credit+Data%29',  
'datasets/Statlog%28German+Credit+Data%29',  
'datasets/Statlog%28Heart%29',  
'datasets/Statlog%28Heart%29',  
'datasets/Statlog%28Landsat+Satellite%29',  
'datasets/Statlog%28Landsat+Satellite%29',  
'datasets/Statlog%28Image+Segmentation%29',  
'datasets/Statlog%28Image+Segmentation%29',  
'datasets/Statlog%28Shuttle%29',  
'datasets/Statlog%28Shuttle%29',  
'datasets/Statlog%28Vehicle+Silhouettes%29',  
'datasets/Statlog%28Vehicle+Silhouettes%29',  
'datasets/Connectionist+Bench%28Nettalk+Corpus%29',  
'datasets/Connectionist+Bench%28Nettalk+Corpus%29',  
'datasets/Connectionist+Bench%28Sonar%2C+Mines+vs.+Rocks%29',  
'datasets/Connectionist+Bench%28Sonar%2C+Mines+vs.+Rocks%29',  
'datasets/Connectionist+Bench%28Vowel+Recognition+-+Deterding+Data%29',  
'datasets/Connectionist+Bench%28Vowel+Recognition+-+Deterding+Data%29',  
'datasets/Economic+Sanctions',  
'datasets/Economic+Sanctions',  
'datasets/Protein+Data',  
'datasets/Protein+Data',  
'datasets/Cloud',  
'datasets/Cloud',  
'datasets/CalIt2+Building+People+Counts',  
'datasets/CalIt2+Building+People+Counts',  
'datasets/Dodgers+Loop+Sensor',  
'datasets/Dodgers+Loop+Sensor',  
'datasets/Poker+Hand',  
'datasets/Poker+Hand',  
'datasets/MAGIC+Gamma+Telescope',  
'datasets/MAGIC+Gamma+Telescope',  
'datasets/UJI+Pen+Characters',  
'datasets/UJI+Pen+Characters',  
'datasets/Mammographic+Mass',  
'datasets/Mammographic+Mass',  
'datasets/Forest+Fires',  
'datasets/Forest+Fires',  
'datasets/Reuters+Transcribed+Subset',  
'datasets/Reuters+Transcribed+Subset',  
'datasets/Bag+of+Words',  
'datasets/Bag+of+Words',  
'datasets/Concrete+Compressive+Strength',  
'datasets/Concrete+Compressive+Strength',  
'datasets/Hill-Valley',  
'datasets/Hill-Valley',  
'datasets/Arcene',  
'datasets/Arcene',  
'datasets/Dexter',  
'datasets/Dexter',  
'datasets/Dorothea',  
'datasets/Dorothea',  
'datasets/Gisette',  
'datasets/Gisette',  
'datasets/Madelon',  
'datasets/Madelon'

```
'datasets/Madelon',
'datasets/Ozone+Level+Detection',
'datasets/Ozone+Level+Detection',
'datasets/Abscisic+Acid+Signaling+Network',
'datasets/Abscisic+Acid+Signaling+Network',
'datasets/Parkinsons',
'datasets/Parkinsons',
'datasets/Character+Trajectories',
'datasets/Character+Trajectories',
'datasets/Blood+Transfusion+Service+Center',
'datasets/Blood+Transfusion+Service+Center',
'datasets/UJI+Pen+Characters+%28Version+2%29',
'datasets/UJI+Pen+Characters+%28Version+2%29',
'datasets/Semeion+Handwritten+Digit',
'datasets/Semeion+Handwritten+Digit',
'datasets/SECOM',
'datasets/SECOM',
'datasets/Plants',
'datasets/Plants',
'datasets/Libras+Movement',
'datasets/Libras+Movement',
'datasets/Concrete+Slump+Test',
'datasets/Concrete+Slump+Test',
'datasets/Communities+and+Crime',
'datasets/Communities+and+Crime',
'datasets/Acute+Inflammations',
'datasets/Acute+Inflammations',
'datasets/Wine+Quality',
'datasets/Wine+Quality',
'datasets/URL+Reputation',
'datasets/URL+Reputation',
'datasets/p53+Mutants',
'datasets/p53+Mutants',
'datasets/Parkinsons+Telemonitoring',
'datasets/Parkinsons+Telemonitoring',
'datasets/Demospongiae',
'datasets/Demospongiae',
'datasets/Opinosis+Opinion+%26frasl%3B+Review',
'datasets/Opinosis+Opinion+%26frasl%3B+Review',
'datasets/Breast+Tissue',
'datasets/Breast+Tissue',
'datasets/Cardiotocography',
'datasets/Cardiotocography',
'datasets/Wall-Following+Robot+Navigation+Data',
'datasets/Wall-Following+Robot+Navigation+Data',
'datasets/Spoken+Arabic+Digit',
'datasets/Spoken+Arabic+Digit',
'datasets/Localization+Data+for+Person+Activity',
'datasets/Localization+Data+for+Person+Activity',
'datasets/AutoUniv',
'datasets/AutoUniv',
'datasets/Steel+Plates+Faults',
'datasets/Steel+Plates+Faults',
'datasets/MiniBooNE+particle+identification',
'datasets/MiniBooNE+particle+identification',
'datasets/YearPredictionMSD',
'datasets/YearPredictionMSD',
'datasets/PEMS-SF',
'datasets/PEMS-SF',
'datasets/OpinRank+Review+Dataset',
'datasets/OpinRank+Review+Dataset',
'datasets/Relative+location+of+CT+slices+on+axial+axis',
'datasets/Relative+location+of+CT+slices+on+axial+axis',
'datasets/Online+Handwritten+Assamese+Characters+Dataset',
'datasets/Online+Handwritten+Assamese+Characters+Dataset',
'datasets/PubChem+Bioassay+Data',
'datasets/PubChem+Bioassay+Data',
'datasets/Record+Linkage+Comparison+Patterns',
'datasets/Record+Linkage+Comparison+Patterns',
'datasets/Communities+and+Crime+Unnormalized',
'datasets/Communities+and+Crime+Unnormalized',
'datasets/Vertebral+Column',
'datasets/Vertebral+Column',
'datasets/EMG+Physical+Action+Data+Set',
'datasets/EMG+Physical+Action+Data+Set',
'datasets/Vicon+Physical+Action+Data+Set',
'datasets/Vicon+Physical+Action+Data+Set',
'datasets/Amazon+Commerce+review+set'
```

```
'datasets/Amazon+Commerce+reviews+set',
'datasets/Amazon+Commerce+reviews+set',
'datasets/Amazon+Access+Samples',
'datasets/Amazon+Access+Samples',
'datasets/Reuter_50_50',
'datasets/Reuter_50_50',
'datasets/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/QSAR+Bioactivity',
'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++++Problem%3A+Pat',
'datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Problem%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/Wholesale+customers',
'datasets/Wholesale+customers'
```

```
'datasets/Twitter+Data+set+rfor+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/UbiqLog+%28smartphone+lifelogging%29',
'datasets/UbiqLog+%28smartphone+lifelogging%29',
'datasets/NIPS+Conference+Papers+1987-2015',
'datasets/NIPS+Conference+Papers+1987-2015',
'datasets/HTRU2',
'datasets/HTRU2',
'datasets/Drug+consumption+%28quantified%29',
'datasets/Drug+consumption+%28quantified%29',
'datasets/Appliances+energy+prediction',
'datasets/Appliances+energy+prediction',
'datasets/Miskolc+IIS+Hybrid+IPS',
'datasets/Miskolc+IIS+Hybrid+IPS',
'datasets/KDC-4007+dataset+Collection',
'datasets/KDC-4007+dataset+Collection',
'datasets/Geo-
Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone',
'datasets/Geo-
Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone',
'datasets/DrivFace',
'datasets/DrivFace',
'datasets/Website+Phishing',
'datasets/Website+Phishing',
'datasets/YouTube+Spam+Collection',
'datasets/YouTube+Spam+Collection',
'datasets/Beijing+PM2.5+Data',
'datasets/Beijing+PM2.5+Data',
'datasets/Cargo+2000+Freight+Tracking+and+Tracing',
'datasets/Cargo+2000+Freight+Tracking+and+Tracing',
'datasets/Cervical+cancer+%28Risk+Factors%29',
'datasets/Cervical+cancer+%28Risk+Factors%29',
'datasets/Quality+Assessment+of+Digital+Colposcopies',
'datasets/Quality+Assessment+of+Digital+Colposcopies',
'datasets/KASANDR',
'datasets/KASANDR',
'datasets/FMA%3A+A+Dataset+For+Music+Analysis',
'datasets/FMA%3A+A+Dataset+For+Music+Analysis',
'datasets/Air+quality',
'datasets/Air+quality',
'datasets/Epileptic+Seizure+Recognition',
'datasets/Epileptic+Seizure+Recognition',
'datasets/Devanagari+Handwritten+Character+Dataset',
'datasets/Devanagari+Handwritten+Character+Dataset',
'datasets/Stock+portfolio+performance',
'datasets/Stock+portfolio+performance',
'datasets/MoCap+Hand+Postures',
'datasets/MoCap+Hand+Postures',
'datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech',
'datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech',
'datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting',
'datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting',
'datasets/PM2.5+Data+of+Five+Chinese+Cities',
'datasets/PM2.5+Data+of+Five+Chinese+Cities',
'datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet',
'datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet',
'datasets/Sales_Transactions_Dataset_Weekly',
'datasets/Sales_Transactions_Dataset_Weekly',
'datasets/Las+Vegas+Strip',
'datasets/Las+Vegas+Strip',
'datasets/Eco-hotel',
'datasets/Eco-hotel',
'datasets/MEU-Mobile+KSD',
'datasets/MEU-Mobile+KSD',
'datasets/Crowdsourced+Mapping',
'datasets/Crowdsourced+Mapping',
'datasets/genet+expression+cancer+RNA-Seq',
'datasets/genet+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+Networ
'datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%28OBS%29+Networ
'datasets/Motion+Capture+Hand+Postures',
```

'datasets/Motion+Capture+Hand+Postures',  
'datasets/Anuran+Calls+%28MFCCs%29',  
'datasets/Anuran+Calls+%28MFCCs%29',  
'datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization',  
'datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization',  
'datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy',  
'datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy',  
'datasets/Daily+Demand+Forecasting+Orders',  
'datasets/Daily+Demand+Forecasting+Orders',  
'datasets/Paper+Reviews',  
'datasets/Paper+Reviews',  
'datasets/extention+of+Z-Alizadeh+sani+dataset',  
'datasets/extention+of+Z-Alizadeh+sani+dataset',  
'datasets/Z-Alizadeh+Sani',  
'datasets/Z-Alizadeh+Sani',  
'datasets/Dynamic+Features+of+VirusShare+Executables',  
'datasets/Dynamic+Features+of+VirusShare+Executables',  
'datasets/IDA2016Challenge',  
'datasets/IDA2016Challenge',  
'datasets/DSRC+Vehicle+Communications',  
'datasets/DSRC+Vehicle+Communications',  
'datasets/Mturk+User-Perceived+Clusters+over+Images',  
'datasets/Mturk+User-Perceived+Clusters+over+Images',  
'datasets/Character+Font+Images',  
'datasets/Character+Font+Images',  
'datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-  
Instance+Learning+with+Instance+Labels',  
'datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-  
Instance+Learning+with+Instance+Labels',  
'datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++',  
'datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++',  
'datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++',  
'datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++',  
'datasets/APS+Failure+at+Scania+Trucks',  
'datasets/APS+Failure+at+Scania+Trucks',  
'datasets/Wireless+Indoor+Localization',  
'datasets/Wireless+Indoor+Localization',  
'datasets/HCC+Survival',  
'datasets/HCC+Survival',  
'datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015',  
'datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015',  
'datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29',  
'datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29',  
'datasets/Autism+Screening+Adult',  
'datasets/Autism+Screening+Adult',  
'datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor',  
'datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor',  
'datasets/Immunotherapy+Dataset',  
'datasets/Immunotherapy+Dataset',  
'datasets/Cryotherapy+Dataset+',  
'datasets/Cryotherapy+Dataset+',  
'datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes',  
'datasets/OCT+data+%26+Color+Fundus+Images+of+Left+%26+Right+Eyes',  
'datasets/Discrete+Tone+Image+Dataset',  
'datasets/Discrete+Tone+Image+Dataset',  
'datasets/News+Popularity+in+Multiple+Social+Media+Platforms',  
'datasets/News+Popularity+in+Multiple+Social+Media+Platforms',  
'datasets/Ultrasonic+flowmeter+diagnostics',  
'datasets/Ultrasonic+flowmeter+diagnostics',  
'datasets/ICMLA+2014+Accepted+Papers+Data+Set',  
'datasets/ICMLA+2014+Accepted+Papers+Data+Set',  
'datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation',  
'datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation',  
'datasets/Container+Crane+Controller+Data+Set',  
'datasets/Container+Crane+Controller+Data+Set',  
'datasets/Residential+Building+Data+Set',  
'datasets/Residential+Building+Data+Set',  
'datasets/Health+News+in+Twitter',  
'datasets/Health+News+in+Twitter',  
'datasets/chipseq',  
'datasets/chipseq',  
'datasets/SGEMM+GPU+kernel+performance',  
'datasets/SGEMM+GPU+kernel+performance',  
'datasets/Repeat+Consumption+Matrices',  
'datasets/Repeat+Consumption+Matrices',  
'datasets/detection\_of\_IoT\_botnet\_attacks\_N\_BaIoT',  
'datasets/detection\_of\_IoT\_botnet\_attacks\_N\_BaIoT',

'datasets/Absenteeism+at+work',  
'datasets/Absenteeism+at+work',  
'datasets/SCADI',  
'datasets/SCADI',  
'datasets/Condition+monitoring+of+hydraulic+systems',  
'datasets/Condition+monitoring+of+hydraulic+systems',  
'datasets/Carbon+Nanotubes',  
'datasets/Carbon+Nanotubes',  
'datasets/Optical+Interconnection+Network+',  
'datasets/Optical+Interconnection+Network+',  
'datasets/Sports+articles+for+objectivity+analysis',  
'datasets/Sports+articles+for+objectivity+analysis',  
'datasets/Breast+Cancer+Coimbra',  
'datasets/Breast+Cancer+Coimbra',  
'datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data',  
'datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data',  
'datasets/Dishonest+Internet+users+Dataset',  
'datasets/Dishonest+Internet+users+Dataset',  
'datasets/Victorian+Era+Authorship+Attribution',  
'datasets/Victorian+Era+Authorship+Attribution',  
'datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set',  
'datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set',  
'datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing',  
'datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing',  
'datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals',  
'datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals',  
'datasets/Roman+Urdu+Data+Set',  
'datasets/Roman+Urdu+Data+Set',  
'datasets/Avila',  
'datasets/Avila',  
'datasets/PANDOR',  
'datasets/PANDOR',  
'datasets/Drug+Review+Dataset+%28Druglib.com%29',  
'datasets/Drug+Review+Dataset+%28Druglib.com%29',  
'datasets/Drug+Review+Dataset+%28Drugs.com%29',  
'datasets/Drug+Review+Dataset+%28Drugs.com%29',  
'datasets/Physical+Unclonable+Functions',  
'datasets/Physical+Unclonable+Functions',  
'datasets/Superconductivity+Data',  
'datasets/Superconductivity+Data',  
'datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29',  
'datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29',  
'datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2',  
'datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2',  
'datasets/Student+Academics+Performance',  
'datasets/Student+Academics+Performance',  
'datasets/Online+Shoppers+Purchasing+Intention+Dataset',  
'datasets/Online+Shoppers+Purchasing+Intention+Dataset',  
'datasets/PMU-UD',  
'datasets/PMU-UD',  
'datasets/Parkinson%27s+Disease+Classification',  
'datasets/Parkinson%27s+Disease+Classification',  
'datasets/Electrical+Grid+Stability+Simulated+Data+',  
'datasets/Electrical+Grid+Stability+Simulated+Data+',  
'datasets/Caesarian+Section+Classification+Dataset',  
'datasets/Caesarian+Section+Classification+Dataset',  
'datasets/BAUM-1',  
'datasets/BAUM-1',  
'datasets/BAUM-2',  
'datasets/BAUM-2',  
'datasets/Audit+Data',  
'datasets/Audit+Data',  
'datasets/BuddyMove+Data+Set',  
'datasets/BuddyMove+Data+Set',  
'datasets/Real+estate+valuation+data+set',  
'datasets/Real+estate+valuation+data+set',  
  
'datasets/Early+biomarkers+of+Parkinson%28%99s+disease+based+on+natural+connected+speech+Data+Se  
  
'datasets/Early+biomarkers+of+Parkinson%28%99s+disease+based+on+natural+connected+speech+Data+Se  
'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 [16]:

```
len(set(data))
```

Out[16]:

469

## Nombre de la Base de Datos

In [17]:

```
def title(lista):
    ltitle = []
    for i in list(range(len(lista))):
        ltitle.append(re.search("\w+$", lista[i]).group())
        print(ltitle[i])
```

In [18]:

```
title(data)
```

Abalone  
Abalone  
Adult  
Adult  
Annealing  
Annealing  
Data  
Data  
Arrhythmia  
Arrhythmia  
Characters  
Characters  
29  
29  
29  
29  
MPG  
MPG  
Automobile  
Automobile  
Badges  
Badges  
Scale  
Scale  
Balloons  
Balloons  
Cancer  
Cancer  
29  
29  
29  
29  
29  
29  
29  
Bridges  
Bridges  
Evaluation  
Evaluation

## Evaluation

Income

Income

29

29

29

29

29

29

29

29

29

Chorales

Chorales

4

4

Approval

Approval

Screening

Screening

Hardware

Hardware

Choice

Choice

Covertype

Covertype

Bands

Bands

Dermatology

Dermatology

Diabetes

Diabetes

Program

Program

Understanding

Understanding

Theories

Theories

Echocardiogram

Echocardiogram

Ecoli

Ecoli

Flags

Flags

Finding

Finding

Identification

Identification

Survival

Survival

Roth

Roth

Disease

Disease

Hepatitis

Hepatitis

Colic

Colic

ICU

ICU

Segmentation

Segmentation

Advertisements

Advertisements

Ionosphere

Ionosphere

Iris

Iris

ISOLET

Kinship

Kinship

Relations

Relations

Domain

Domain

Lenses

Lenses

Recognition

recognition  
Recognition  
Disorders  
Disorders  
Theorist  
Theorist  
Cancer  
Cancer  
Lymphography  
Lymphography  
Analysis  
Analysis  
data  
data  
Robots  
Robots  
29  
29  
29  
29  
29  
29  
29  
29  
Problems  
Problems  
Reasoner  
Reasoner  
Features  
Features  
Mushroom  
Mushroom  
29  
29  
29  
29  
Nursery  
Nursery  
Theory  
Theory  
Classification  
Classification  
Digits  
Digits  
Digits  
Digits  
Patient  
Patient  
Tumor  
Tumor  
Prodigy  
Prodigy  
Relationships  
Relationships  
Mammals  
Mammals  
Servo  
Servo  
Control  
Control  
Flare  
Flare  
29  
29  
29  
29  
Ring  
Ring  
Spectrometer  
Spectrometer  
Spambase  
Spambase  
Heart  
Heart  
Heart  
Heart  
Heart  
Sponge  
Sponge  
Project  
~~~~~

Project
Relational
Relational
Evaluation
Evaluation
Endgame
Endgame
Disease
Disease
Trains
Trains
University
University
Records
Records
Plant
Plant
29
29
29
29
Wine
Wine
Yeast
Yeast
Zoo
Zoo
Undocumented
Undocumented
Newsgroups
Newsgroups
signs
signs
29
29
29
29
29
29
Data
Data
Features
Features
Genes
Genes
Database
Database
Nino
Nino
Data
Data
Images
Images
29
29
Data
Data
Database
Database
Vowels
Vowels
Data
Data
Data
Data
Genes
Genes
Movie
Movie
Data
Data
2003
2003
Data
Data
Series
Series

Trajectories
Center
Center
29
29
Digit
Digit
SECOM
SECOM
Plants
Plants
Movement
Movement
Test
Test
Crime
Crime
Inflammations
Inflammations
Quality
Quality
Reputation
Reputation
Mutants
Mutants
Telemonitoring
Telemonitoring
Demospongiae
Demospongiae
Review
Review
Tissue
Tissue
Cardiotocography
Cardiotocography
Data
Data
Digit
Digit
Activity
Activity
AutoUniv
AutoUniv
Faults
Faults
identification
identification
YearPredictionMSD
YearPredictionMSD
SF
SF
Dataset
Dataset
axis
axis
Dataset
Dataset
Data
Data
Patterns
Patterns
Unnormalized
Unnormalized
Column
Column
Set
Set
Set
Set
set
set
Samples
Samples
Reuter_50_50
Reuter_50_50
Ads
Ads
..

```
mails
mails
29
29
29
29
Marketing
Marketing
Data
Data
Dataset
Dataset
29
29
Recognition
Recognition
Nomao
Nomao
Collection
Collection
Segmentation
Segmentation
Relax
Relax
Monitoring
Monitoring
data
data
9
9
consumption
consumption
seeds
seeds
Northix
Northix
QtyT40I10D100K
QtyT40I10D100K
Reports
Reports
Smartphones
Smartphones
set
set
efficiency
efficiency
Hydrodynamics
Hydrodynamics
Fertility
Fertility
Gait
Gait
29
29
EXCHANGE
EXCHANGE
```

```
-----  
AttributeError                               Traceback (most recent call last)  
<ipython-input-18-0a1770b157e4> in <module>  
----> 1 title(data)  
  
<ipython-input-17-cb833418c6b7> in title(lista)
      2     ltitle = []
      3     for i in list(range(len(lista))):  
----> 4         ltitle.append(re.search("\w+$", lista[i]).group())
      5         print(ltitle[i])  
  
AttributeError: 'NoneType' object has no attribute 'group'
```

Link a la Base de Datos

In [19]:

```
enlace = "http://archive.ics.uci.edu/ml/datasets/"
```

In [20]:

```
def sumstr(a, b):
    l = a + b
    return l

def links(lista):
    for i in range(len(lista)):
        linksdata.append(sumstr(enlace, lista[i]))
        print(linksdata[i])
```

In [21]:

```
linksdata = []
links(data)
```

http://archive.ics.uci.edu/ml/datasets/datasets/Abalone
http://archive.ics.uci.edu/ml/datasets/datasets/Abalone
http://archive.ics.uci.edu/ml/datasets/datasets/Adult
http://archive.ics.uci.edu/ml/datasets/datasets/Adult
http://archive.ics.uci.edu/ml/datasets/datasets/Annealing
http://archive.ics.uci.edu/ml/datasets/datasets/Annealing
http://archive.ics.uci.edu/ml/datasets/datasets/Anonymous+Microsoft+Web+Data
http://archive.ics.uci.edu/ml/datasets/datasets/Anonymous+Microsoft+Web+Data
http://archive.ics.uci.edu/ml/datasets/datasets/Arrhythmia
http://archive.ics.uci.edu/ml/datasets/datasets/Arrhythmia
http://archive.ics.uci.edu/ml/datasets/datasets/Artificial+Characters
http://archive.ics.uci.edu/ml/datasets/datasets/Artificial+Characters
http://archive.ics.uci.edu/ml/datasets/datasets/Audiology+%28Original%29
http://archive.ics.uci.edu/ml/datasets/datasets/Audiology+%28Original%29
http://archive.ics.uci.edu/ml/datasets/datasets/Audiology+%28Standardized%29
http://archive.ics.uci.edu/ml/datasets/datasets/Audiology+%28Standardized%29
http://archive.ics.uci.edu/ml/datasets/datasets/Auto+MPG
http://archive.ics.uci.edu/ml/datasets/datasets/Auto+MPG
http://archive.ics.uci.edu/ml/datasets/datasets/Automobile
http://archive.ics.uci.edu/ml/datasets/datasets/Automobile
http://archive.ics.uci.edu/ml/datasets/datasets/Badges
http://archive.ics.uci.edu/ml/datasets/datasets/Badges
http://archive.ics.uci.edu/ml/datasets/datasets/Balance+Scale
http://archive.ics.uci.edu/ml/datasets/datasets/Balance+Scale
http://archive.ics.uci.edu/ml/datasets/datasets/Balloons
http://archive.ics.uci.edu/ml/datasets/datasets/Balloons
http://archive.ics.uci.edu/ml/datasets/datasets/Breast+Cancer
http://archive.ics.uci.edu/ml/datasets/datasets/Breast+Cancer
http://archive.ics.uci.edu/ml/datasets/datasets/Breast+Cancer+Wisconsin+%28Original%29
http://archive.ics.uci.edu/ml/datasets/datasets/Breast+Cancer+Wisconsin+%28Original%29
http://archive.ics.uci.edu/ml/datasets/datasets/Breast+Cancer+Wisconsin+%28Prognostic%29
http://archive.ics.uci.edu/ml/datasets/datasets/Breast+Cancer+Wisconsin+%28Prognostic%29
http://archive.ics.uci.edu/ml/datasets/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29
http://archive.ics.uci.edu/ml/datasets/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29
http://archive.ics.uci.edu/ml/datasets/datasets/Pittsburgh+Bridges
http://archive.ics.uci.edu/ml/datasets/datasets/Pittsburgh+Bridges
http://archive.ics.uci.edu/ml/datasets/datasets/Car+Evaluation
http://archive.ics.uci.edu/ml/datasets/datasets/Car+Evaluation
http://archive.ics.uci.edu/ml/datasets/datasets/Census+Income
http://archive.ics.uci.edu/ml/datasets/datasets/Census+Income
http://archive.ics.uci.edu/ml/datasets/datasets/Chess+%28King-Rook+vs.+King-Knight%29
http://archive.ics.uci.edu/ml/datasets/datasets/Chess+%28King-Rook+vs.+King-Knight%29
http://archive.ics.uci.edu/ml/datasets/datasets/Chess+%28King-Rook+vs.+King-Pawn%29
http://archive.ics.uci.edu/ml/datasets/datasets/Chess+%28King-Rook+vs.+King-Pawn%29
http://archive.ics.uci.edu/ml/datasets/datasets/Chess+%28King-Rook+vs.+King%29
http://archive.ics.uci.edu/ml/datasets/datasets/Chess+%28King-Rook+vs.+King%29
http://archive.ics.uci.edu/ml/datasets/datasets/Chess+%28Domain+Theories%29
http://archive.ics.uci.edu/ml/datasets/datasets/Chess+%28Domain+Theories%29
http://archive.ics.uci.edu/ml/datasets/datasets/Bach+Chorales
http://archive.ics.uci.edu/ml/datasets/datasets/Bach+Chorales
http://archive.ics.uci.edu/ml/datasets/datasets/Connect-4
http://archive.ics.uci.edu/ml/datasets/datasets/Connect-4
http://archive.ics.uci.edu/ml/datasets/datasets/Credit+Approval
http://archive.ics.uci.edu/ml/datasets/datasets/Credit+Approval
http://archive.ics.uci.edu/ml/datasets/datasets/Japanese+Credit+Screening
http://archive.ics.uci.edu/ml/datasets/datasets/Japanese+Credit+Screening

http://archive.ics.uci.edu/ml/datasets/datasets/Computer+Hardware
http://archive.ics.uci.edu/ml/datasets/datasets/Computer+Hardware
http://archive.ics.uci.edu/ml/datasets/datasets/Contraceptive+Method+Choice
http://archive.ics.uci.edu/ml/datasets/datasets/Contraceptive+Method+Choice
http://archive.ics.uci.edu/ml/datasets/datasets/Covertype
http://archive.ics.uci.edu/ml/datasets/datasets/Covertype
http://archive.ics.uci.edu/ml/datasets/datasets/Cylinder+Bands
http://archive.ics.uci.edu/ml/datasets/datasets/Cylinder+Bands
http://archive.ics.uci.edu/ml/datasets/datasets/Dermatology
http://archive.ics.uci.edu/ml/datasets/datasets/Dermatology
http://archive.ics.uci.edu/ml/datasets/datasets/Diabetes
http://archive.ics.uci.edu/ml/datasets/datasets/Diabetes
http://archive.ics.uci.edu/ml/datasets/datasets/DGP2+-+The+Second+Data+Generation+Program
http://archive.ics.uci.edu/ml/datasets/datasets/DGP2+-+The+Second+Data+Generation+Program
http://archive.ics.uci.edu/ml/datasets/datasets/Document+Understanding
http://archive.ics.uci.edu/ml/datasets/datasets/Document+Understanding
http://archive.ics.uci.edu/ml/datasets/datasets/EBL+Domain+Theories
http://archive.ics.uci.edu/ml/datasets/datasets/EBL+Domain+Theories
http://archive.ics.uci.edu/ml/datasets/datasets/Echocardiogram
http://archive.ics.uci.edu/ml/datasets/datasets/Echocardiogram
http://archive.ics.uci.edu/ml/datasets/datasets/Ecoli
http://archive.ics.uci.edu/ml/datasets/datasets/Ecoli
http://archive.ics.uci.edu/ml/datasets/datasets/Flags
http://archive.ics.uci.edu/ml/datasets/datasets/Flags
http://archive.ics.uci.edu/ml/datasets/datasets/Function+Finding
http://archive.ics.uci.edu/ml/datasets/datasets/Function+Finding
http://archive.ics.uci.edu/ml/datasets/datasets/Glass+Identification
http://archive.ics.uci.edu/ml/datasets/datasets/Glass+Identification
http://archive.ics.uci.edu/ml/datasets/datasets/Haberman%27s+Survival
http://archive.ics.uci.edu/ml/datasets/datasets/Haberman%27s+Survival
http://archive.ics.uci.edu/ml/datasets/datasets/Hayes-Roth
http://archive.ics.uci.edu/ml/datasets/datasets/Hayes-Roth
http://archive.ics.uci.edu/ml/datasets/datasets/Heart+Disease
http://archive.ics.uci.edu/ml/datasets/datasets/Heart+Disease
http://archive.ics.uci.edu/ml/datasets/datasets/Hepatitis
http://archive.ics.uci.edu/ml/datasets/datasets/Hepatitis
http://archive.ics.uci.edu/ml/datasets/datasets/Horse+Colic
http://archive.ics.uci.edu/ml/datasets/datasets/Horse+Colic
http://archive.ics.uci.edu/ml/datasets/datasets/ICU
http://archive.ics.uci.edu/ml/datasets/datasets/ICU
http://archive.ics.uci.edu/ml/datasets/datasets/Image+Segmentation
http://archive.ics.uci.edu/ml/datasets/datasets/Image+Segmentation
http://archive.ics.uci.edu/ml/datasets/datasets/Internet+Advertisements
http://archive.ics.uci.edu/ml/datasets/datasets/Internet+Advertisements
http://archive.ics.uci.edu/ml/datasets/datasets/Ionosphere
http://archive.ics.uci.edu/ml/datasets/datasets/Ionosphere
http://archive.ics.uci.edu/ml/datasets/datasets/Iris
http://archive.ics.uci.edu/ml/datasets/datasets/Iris
http://archive.ics.uci.edu/ml/datasets/datasets/ISOLET
http://archive.ics.uci.edu/ml/datasets/datasets/ISOLET
http://archive.ics.uci.edu/ml/datasets/datasets/Kinship
http://archive.ics.uci.edu/ml/datasets/datasets/Kinship
http://archive.ics.uci.edu/ml/datasets/datasets/Labor+Relations
http://archive.ics.uci.edu/ml/datasets/datasets/Labor+Relations
http://archive.ics.uci.edu/ml/datasets/datasets/LED+Display+Domain
http://archive.ics.uci.edu/ml/datasets/datasets/LED+Display+Domain
http://archive.ics.uci.edu/ml/datasets/datasets/Lenses
http://archive.ics.uci.edu/ml/datasets/datasets/Lenses
http://archive.ics.uci.edu/ml/datasets/datasets/Letter+Recognition
http://archive.ics.uci.edu/ml/datasets/datasets/Letter+Recognition
http://archive.ics.uci.edu/ml/datasets/datasets/Liver+Disorders
http://archive.ics.uci.edu/ml/datasets/datasets/Liver+Disorders
http://archive.ics.uci.edu/ml/datasets/datasets/Logic+Theorist
http://archive.ics.uci.edu/ml/datasets/datasets/Logic+Theorist
http://archive.ics.uci.edu/ml/datasets/datasets/Lung+Cancer
http://archive.ics.uci.edu/ml/datasets/datasets/Lung+Cancer
http://archive.ics.uci.edu/ml/datasets/datasets/Lymphography
http://archive.ics.uci.edu/ml/datasets/datasets/Lymphography
http://archive.ics.uci.edu/ml/datasets/datasets/Mechanical+Analysis
http://archive.ics.uci.edu/ml/datasets/datasets/Mechanical+Analysis
http://archive.ics.uci.edu/ml/datasets/datasets/Meta-data
http://archive.ics.uci.edu/ml/datasets/datasets/Meta-data
http://archive.ics.uci.edu/ml/datasets/datasets/Mobile+Robots
http://archive.ics.uci.edu/ml/datasets/datasets/Mobile+Robots
http://archive.ics.uci.edu/ml/datasets/datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29
http://archive.ics.uci.edu/ml/datasets/datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29
http://archive.ics.uci.edu/ml/datasets/datasets/Molecular+Biology+%28Protein+Secondary+Structure%29

http://archive.ics.uci.edu/ml/datasets/datasets/Molecular+Biology+%28Protein+Secondary+Structure%29
http://archive.ics.uci.edu/ml/datasets/datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29
http://archive.ics.uci.edu/ml/datasets/datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29
http://archive.ics.uci.edu/ml/datasets/datasets/MONK%27s+Problems
http://archive.ics.uci.edu/ml/datasets/datasets/MONK%27s+Problems
http://archive.ics.uci.edu/ml/datasets/datasets/Moral+Reasoner
http://archive.ics.uci.edu/ml/datasets/datasets/Moral+Reasoner
http://archive.ics.uci.edu/ml/datasets/datasets/Multiple+Features
http://archive.ics.uci.edu/ml/datasets/datasets/Multiple+Features
http://archive.ics.uci.edu/ml/datasets/datasets/Mushroom
http://archive.ics.uci.edu/ml/datasets/datasets/Mushroom
http://archive.ics.uci.edu/ml/datasets/datasets/Musk+%28Version+1%29
http://archive.ics.uci.edu/ml/datasets/datasets/Musk+%28Version+1%29
http://archive.ics.uci.edu/ml/datasets/datasets/Musk+%28Version+2%29
http://archive.ics.uci.edu/ml/datasets/datasets/Musk+%28Version+2%29
http://archive.ics.uci.edu/ml/datasets/datasets/Nursery
http://archive.ics.uci.edu/ml/datasets/datasets/Nursery
http://archive.ics.uci.edu/ml/datasets/datasets/Othello+Domain+Theory
http://archive.ics.uci.edu/ml/datasets/datasets/Othello+Domain+Theory
http://archive.ics.uci.edu/ml/datasets/datasets/Page+Blocks+Classification
http://archive.ics.uci.edu/ml/datasets/datasets/Page+Blocks+Classification
http://archive.ics.uci.edu/ml/datasets/datasets/Optical+Recognition+of+Handwritten+Digits
http://archive.ics.uci.edu/ml/datasets/datasets/Optical+Recognition+of+Handwritten+Digits
http://archive.ics.uci.edu/ml/datasets/datasets/Pen-Based+Recognition+of+Handwritten+Digits
http://archive.ics.uci.edu/ml/datasets/datasets/Pen-Based+Recognition+of+Handwritten+Digits
http://archive.ics.uci.edu/ml/datasets/datasets/Post-Operative+Patient
http://archive.ics.uci.edu/ml/datasets/datasets/Post-Operative+Patient
http://archive.ics.uci.edu/ml/datasets/datasets/Primary+Tumor
http://archive.ics.uci.edu/ml/datasets/datasets/Primary+Tumor
http://archive.ics.uci.edu/ml/datasets/datasets/Prodigy
http://archive.ics.uci.edu/ml/datasets/datasets/Prodigy
http://archive.ics.uci.edu/ml/datasets/datasets/Qualitative+Structure+Activity+Relationships
http://archive.ics.uci.edu/ml/datasets/datasets/Qualitative+Structure+Activity+Relationships
http://archive.ics.uci.edu/ml/datasets/datasets/Quadruped+Mammals
http://archive.ics.uci.edu/ml/datasets/datasets/Quadruped+Mammals
http://archive.ics.uci.edu/ml/datasets/datasets/Servo
http://archive.ics.uci.edu/ml/datasets/datasets/Servo
http://archive.ics.uci.edu/ml/datasets/datasets/Shuttle+Landing+Control
http://archive.ics.uci.edu/ml/datasets/datasets/Shuttle+Landing+Control
http://archive.ics.uci.edu/ml/datasets/datasets/Solar+Flare
http://archive.ics.uci.edu/ml/datasets/datasets/Solar+Flare
http://archive.ics.uci.edu/ml/datasets/datasets/Soybean+%28Large%29
http://archive.ics.uci.edu/ml/datasets/datasets/Soybean+%28Large%29
http://archive.ics.uci.edu/ml/datasets/datasets/Soybean+%28Small%29
http://archive.ics.uci.edu/ml/datasets/datasets/Soybean+%28Small%29
http://archive.ics.uci.edu/ml/datasets/datasets/Challenger+USA+Space+Shuttle+O-Ring
http://archive.ics.uci.edu/ml/datasets/datasets/Challenger+USA+Space+Shuttle+O-Ring
http://archive.ics.uci.edu/ml/datasets/datasets/Low+Resolution+Spectrometer
http://archive.ics.uci.edu/ml/datasets/datasets/Low+Resolution+Spectrometer
http://archive.ics.uci.edu/ml/datasets/datasets/Spambase
http://archive.ics.uci.edu/ml/datasets/datasets/Spambase
http://archive.ics.uci.edu/ml/datasets/datasets/SPECT+Heart
http://archive.ics.uci.edu/ml/datasets/datasets/SPECT+Heart
http://archive.ics.uci.edu/ml/datasets/datasets/SPECTF+Heart
http://archive.ics.uci.edu/ml/datasets/datasets/SPECTF+Heart
http://archive.ics.uci.edu/ml/datasets/datasets/Sponge
http://archive.ics.uci.edu/ml/datasets/datasets/Sponge
http://archive.ics.uci.edu/ml/datasets/datasets/Statlog+Project
http://archive.ics.uci.edu/ml/datasets/datasets/Statlog+Project
http://archive.ics.uci.edu/ml/datasets/datasets/Student+Loan+Relational
http://archive.ics.uci.edu/ml/datasets/datasets/Student+Loan+Relational
http://archive.ics.uci.edu/ml/datasets/datasets/Teaching+Assistant+Evaluation
http://archive.ics.uci.edu/ml/datasets/datasets/Teaching+Assistant+Evaluation
http://archive.ics.uci.edu/ml/datasets/datasets/Tic-Tac-Toe+Endgame
http://archive.ics.uci.edu/ml/datasets/datasets/Tic-Tac-Toe+Endgame
http://archive.ics.uci.edu/ml/datasets/datasets/Thyroid+Disease
http://archive.ics.uci.edu/ml/datasets/datasets/Thyroid+Disease
http://archive.ics.uci.edu/ml/datasets/datasets/Trains
http://archive.ics.uci.edu/ml/datasets/datasets/Trains
http://archive.ics.uci.edu/ml/datasets/datasets/University
http://archive.ics.uci.edu/ml/datasets/datasets/University
http://archive.ics.uci.edu/ml/datasets/datasets/Congressional+Voting+Records
http://archive.ics.uci.edu/ml/datasets/datasets/Congressional+Voting+Records

http://archive.ics.uci.edu/ml/datasets/datasets/Water+Treatment+Plant
http://archive.ics.uci.edu/ml/datasets/datasets/Water+Treatment+Plant
http://archive.ics.uci.edu/ml/datasets/datasets/Waveform+Database+Generator+%28Version+1%29
http://archive.ics.uci.edu/ml/datasets/datasets/Waveform+Database+Generator+%28Version+1%29
http://archive.ics.uci.edu/ml/datasets/datasets/Waveform+Database+Generator+%28Version+2%29
http://archive.ics.uci.edu/ml/datasets/datasets/Waveform+Database+Generator+%28Version+2%29
http://archive.ics.uci.edu/ml/datasets/datasets/Wine
http://archive.ics.uci.edu/ml/datasets/datasets/Wine
http://archive.ics.uci.edu/ml/datasets/datasets/Yeast
http://archive.ics.uci.edu/ml/datasets/datasets/Yeast
http://archive.ics.uci.edu/ml/datasets/datasets/Zoo
http://archive.ics.uci.edu/ml/datasets/datasets/Zoo
http://archive.ics.uci.edu/ml/datasets/datasets/Undocumented
http://archive.ics.uci.edu/ml/datasets/datasets/Undocumented
http://archive.ics.uci.edu/ml/datasets/datasets/Twenty+Newsgroups
http://archive.ics.uci.edu/ml/datasets/datasets/Twenty+Newsgroups
http://archive.ics.uci.edu/ml/datasets/datasets/Australian+Sign+Language+signs
http://archive.ics.uci.edu/ml/datasets/datasets/Australian+Sign+Language+signs
http://archive.ics.uci.edu/ml/datasets/datasets/Australian+Sign+Language+signs+High+Quality%29
http://archive.ics.uci.edu/ml/datasets/datasets/Australian+Sign+Language+signs+High+Quality%29
http://archive.ics.uci.edu/ml/datasets/datasets/US+Census+Data%281990%29
http://archive.ics.uci.edu/ml/datasets/datasets/US+Census+Data%281990%29
http://archive.ics.uci.edu/ml/datasets/datasets/Census-Income+%28KDD%29
http://archive.ics.uci.edu/ml/datasets/datasets/Census-Income+%28KDD%29
http://archive.ics.uci.edu/ml/datasets/datasets/Coil+1999+Competition+Data
http://archive.ics.uci.edu/ml/datasets/datasets/Coil+1999+Competition+Data
http://archive.ics.uci.edu/ml/datasets/datasets/Corel+Image+Features
http://archive.ics.uci.edu/ml/datasets/datasets/Corel+Image+Features
http://archive.ics.uci.edu/ml/datasets/datasets/E.+Coli+Genes
http://archive.ics.uci.edu/ml/datasets/datasets/E.+Coli+Genes
http://archive.ics.uci.edu/ml/datasets/datasets/EEG+Database
http://archive.ics.uci.edu/ml/datasets/datasets/EEG+Database
http://archive.ics.uci.edu/ml/datasets/datasets/El+Nino
http://archive.ics.uci.edu/ml/datasets/datasets/El+Nino
http://archive.ics.uci.edu/ml/datasets/datasets/Entree+Chicago+Recommendation+Data
http://archive.ics.uci.edu/ml/datasets/datasets/Entree+Chicago+Recommendation+Data
http://archive.ics.uci.edu/ml/datasets/datasets/CMU+Face+Images
http://archive.ics.uci.edu/ml/datasets/datasets/CMU+Face+Images
http://archive.ics.uci.edu/ml/datasets/datasets/Insurance+Company+Benchmark+%28COIL+2000%29
http://archive.ics.uci.edu/ml/datasets/datasets/Insurance+Company+Benchmark+%28COIL+2000%29
http://archive.ics.uci.edu/ml/datasets/datasets/Internet+Usage+Data
http://archive.ics.uci.edu/ml/datasets/datasets/Internet+Usage+Data
http://archive.ics.uci.edu/ml/datasets/datasets/IPUMS+Census+Database
http://archive.ics.uci.edu/ml/datasets/datasets/IPUMS+Census+Database
http://archive.ics.uci.edu/ml/datasets/datasets/Japanese+Vowels
http://archive.ics.uci.edu/ml/datasets/datasets/Japanese+Vowels
http://archive.ics.uci.edu/ml/datasets/datasets/KDD+Cup+1998+Data
http://archive.ics.uci.edu/ml/datasets/datasets/KDD+Cup+1998+Data
http://archive.ics.uci.edu/ml/datasets/datasets/KDD+Cup+1999+Data
http://archive.ics.uci.edu/ml/datasets/datasets/KDD+Cup+1999+Data
http://archive.ics.uci.edu/ml/datasets/datasets/M.+Tuberculosis+Genes
http://archive.ics.uci.edu/ml/datasets/datasets/M.+Tuberculosis+Genes
http://archive.ics.uci.edu/ml/datasets/datasets/Movie
http://archive.ics.uci.edu/ml/datasets/datasets/Movie
http://archive.ics.uci.edu/ml/datasets/datasets/MSNBC.com+Anonymous+Web+Data
http://archive.ics.uci.edu/ml/datasets/datasets/MSNBC.com+Anonymous+Web+Data
http://archive.ics.uci.edu/ml/datasets/datasets/NSF+Research+Award+Abstracts+1990-2003
http://archive.ics.uci.edu/ml/datasets/datasets/NSF+Research+Award+Abstracts+1990-2003
http://archive.ics.uci.edu/ml/datasets/datasets/Pioneer-1+Mobile+Robot+Data
http://archive.ics.uci.edu/ml/datasets/datasets/Pioneer-1+Mobile+Robot+Data
http://archive.ics.uci.edu/ml/datasets/datasets/Pseudo+Periodic+Synthetic+Time+Series
http://archive.ics.uci.edu/ml/datasets/datasets/Pseudo+Periodic+Synthetic+Time+Series
http://archive.ics.uci.edu/ml/datasets/datasets/Reuters-21578+Text+Categorization+Collection
http://archive.ics.uci.edu/ml/datasets/datasets/Reuters-21578+Text+Categorization+Collection
http://archive.ics.uci.edu/ml/datasets/datasets/Robot+Execution+Failures
http://archive.ics.uci.edu/ml/datasets/datasets/Robot+Execution+Failures
http://archive.ics.uci.edu/ml/datasets/datasets/Synthetic+Control+Chart+Time+Series
http://archive.ics.uci.edu/ml/datasets/datasets/Synthetic+Control+Chart+Time+Series
http://archive.ics.uci.edu/ml/datasets/datasets/Syskill+and+Weber+Web+Page+Ratings
http://archive.ics.uci.edu/ml/datasets/datasets/Syskill+and+Weber+Web+Page+Ratings
http://archive.ics.uci.edu/ml/datasets/datasets/UNIX+User+Data
http://archive.ics.uci.edu/ml/datasets/datasets/UNIX+User+Data
http://archive.ics.uci.edu/ml/datasets/datasets/Volcanoes+on+Venus+-+JARtool+experiment
http://archive.ics.uci.edu/ml/datasets/datasets/Volcanoes+on+Venus+-+JARtool+experiment
http://archive.ics.uci.edu/ml/datasets/datasets/Statlog%28Australian+Credit+Approval%29
http://archive.ics.uci.edu/ml/datasets/datasets/Statlog%28Australian+Credit+Approval%29
http://archive.ics.uci.edu/ml/datasets/datasets/Statlog%28German+Credit+Data%29

http://archive.ics.uci.edu/ml/datasets/datasets/Statlog%28German+Credit+Data%29
http://archive.ics.uci.edu/ml/datasets/datasets/Statlog%28Heart%29
http://archive.ics.uci.edu/ml/datasets/datasets/Statlog%28Heart%29
http://archive.ics.uci.edu/ml/datasets/datasets/Statlog%28Landsat+Satellite%29
http://archive.ics.uci.edu/ml/datasets/datasets/Statlog%28Landsat+Satellite%29
http://archive.ics.uci.edu/ml/datasets/datasets/Statlog%28Image+Segmentation%29
http://archive.ics.uci.edu/ml/datasets/datasets/Statlog%28Image+Segmentation%29
http://archive.ics.uci.edu/ml/datasets/datasets/Statlog%28Shuttle%29
http://archive.ics.uci.edu/ml/datasets/datasets/Statlog%28Shuttle%29
http://archive.ics.uci.edu/ml/datasets/datasets/Statlog%28Vehicle+Silhouettes%29
http://archive.ics.uci.edu/ml/datasets/datasets/Statlog%28Vehicle+Silhouettes%29
http://archive.ics.uci.edu/ml/datasets/datasets/Connectionist+Bench%28Nettalk+Corpus%29
http://archive.ics.uci.edu/ml/datasets/datasets/Connectionist+Bench%28Nettalk+Corpus%29
http://archive.ics.uci.edu/ml/datasets/datasets/Connectionist+Bench%28Sonar%2C+Mines+vs.+Rocks%29
http://archive.ics.uci.edu/ml/datasets/datasets/Connectionist+Bench%28Sonar%2C+Mines+vs.+Rocks%29
http://archive.ics.uci.edu/ml/datasets/datasets/Connectionist+Bench%28Vowel+Recognition+-+Deterding+Data%29
http://archive.ics.uci.edu/ml/datasets/datasets/Connectionist+Bench%28Vowel+Recognition+-+Deterding+Data%29
http://archive.ics.uci.edu/ml/datasets/datasets/Economic+Sanctions
http://archive.ics.uci.edu/ml/datasets/datasets/Economic+Sanctions
http://archive.ics.uci.edu/ml/datasets/datasets/Protein+Data
http://archive.ics.uci.edu/ml/datasets/datasets/Protein+Data
http://archive.ics.uci.edu/ml/datasets/datasets/Cloud
http://archive.ics.uci.edu/ml/datasets/datasets/Cloud
http://archive.ics.uci.edu/ml/datasets/datasets/CalIt2+Building+People+Counts
http://archive.ics.uci.edu/ml/datasets/datasets/CalIt2+Building+People+Counts
http://archive.ics.uci.edu/ml/datasets/datasets/Dodgers+Loop+Sensor
http://archive.ics.uci.edu/ml/datasets/datasets/Dodgers+Loop+Sensor
http://archive.ics.uci.edu/ml/datasets/datasets/Poker+Hand
http://archive.ics.uci.edu/ml/datasets/datasets/Poker+Hand
http://archive.ics.uci.edu/ml/datasets/datasets/MAGIC+Gamma+Telescope
http://archive.ics.uci.edu/ml/datasets/datasets/MAGIC+Gamma+Telescope
http://archive.ics.uci.edu/ml/datasets/datasets/UJI+Pen+Characters
http://archive.ics.uci.edu/ml/datasets/datasets/UJI+Pen+Characters
http://archive.ics.uci.edu/ml/datasets/datasets/Mammographic+Mass
http://archive.ics.uci.edu/ml/datasets/datasets/Mammographic+Mass
http://archive.ics.uci.edu/ml/datasets/datasets/Forest+Fires
http://archive.ics.uci.edu/ml/datasets/datasets/Forest+Fires
http://archive.ics.uci.edu/ml/datasets/datasets/Reuters+Transcribed+Subset
http://archive.ics.uci.edu/ml/datasets/datasets/Reuters+Transcribed+Subset
http://archive.ics.uci.edu/ml/datasets/datasets/Bag+of+Words
http://archive.ics.uci.edu/ml/datasets/datasets/Bag+of+Words
http://archive.ics.uci.edu/ml/datasets/datasets/Concrete+Compressive+Strength
http://archive.ics.uci.edu/ml/datasets/datasets/Concrete+Compressive+Strength
http://archive.ics.uci.edu/ml/datasets/datasets/Hill-Valley
http://archive.ics.uci.edu/ml/datasets/datasets/Hill-Valley
http://archive.ics.uci.edu/ml/datasets/datasets/Arcene
http://archive.ics.uci.edu/ml/datasets/datasets/Arcene
http://archive.ics.uci.edu/ml/datasets/datasets/Dexter
http://archive.ics.uci.edu/ml/datasets/datasets/Dexter
http://archive.ics.uci.edu/ml/datasets/datasets/Dorothea
http://archive.ics.uci.edu/ml/datasets/datasets/Dorothea
http://archive.ics.uci.edu/ml/datasets/datasets/Gisette
http://archive.ics.uci.edu/ml/datasets/datasets/Gisette
http://archive.ics.uci.edu/ml/datasets/datasets/Madelon
http://archive.ics.uci.edu/ml/datasets/datasets/Madelon
http://archive.ics.uci.edu/ml/datasets/datasets/Ozone+Level+Detection
http://archive.ics.uci.edu/ml/datasets/datasets/Ozone+Level+Detection
http://archive.ics.uci.edu/ml/datasets/datasets/Abscisic+Acid+Signaling+Network
http://archive.ics.uci.edu/ml/datasets/datasets/Abscisic+Acid+Signaling+Network
http://archive.ics.uci.edu/ml/datasets/datasets/Parkinsons
http://archive.ics.uci.edu/ml/datasets/datasets/Parkinsons
http://archive.ics.uci.edu/ml/datasets/datasets/Character+Trajectories
http://archive.ics.uci.edu/ml/datasets/datasets/Character+Trajectories
http://archive.ics.uci.edu/ml/datasets/datasets/Blood+Transfusion+Service+Center
http://archive.ics.uci.edu/ml/datasets/datasets/Blood+Transfusion+Service+Center
http://archive.ics.uci.edu/ml/datasets/datasets/UJI+Pen+Characters+%28Version+2%29
http://archive.ics.uci.edu/ml/datasets/datasets/UJI+Pen+Characters+%28Version+2%29
http://archive.ics.uci.edu/ml/datasets/datasets/Semeion+Handwritten+Digit
http://archive.ics.uci.edu/ml/datasets/datasets/Semeion+Handwritten+Digit
http://archive.ics.uci.edu/ml/datasets/datasets/SECOM
http://archive.ics.uci.edu/ml/datasets/datasets/SECOM
http://archive.ics.uci.edu/ml/datasets/datasets/Plants
http://archive.ics.uci.edu/ml/datasets/datasets/Plants
http://archive.ics.uci.edu/ml/datasets/datasets/Libras+Movement
http://archive.ics.uci.edu/ml/datasets/datasets/Libras+Movement

http://archive.ics.uci.edu/ml/datasets/datasets/Concrete+Slump+Test
http://archive.ics.uci.edu/ml/datasets/datasets/Concrete+Slump+Test
http://archive.ics.uci.edu/ml/datasets/datasets/Communities+and+Crime
http://archive.ics.uci.edu/ml/datasets/datasets/Communities+and+Crime
http://archive.ics.uci.edu/ml/datasets/datasets/Acute+Inflammations
http://archive.ics.uci.edu/ml/datasets/datasets/Acute+Inflammations
http://archive.ics.uci.edu/ml/datasets/datasets/Wine+Quality
http://archive.ics.uci.edu/ml/datasets/datasets/Wine+Quality
http://archive.ics.uci.edu/ml/datasets/datasets/URL+Reputation
http://archive.ics.uci.edu/ml/datasets/datasets/URL+Reputation
http://archive.ics.uci.edu/ml/datasets/datasets/p53+Mutants
http://archive.ics.uci.edu/ml/datasets/datasets/p53+Mutants
http://archive.ics.uci.edu/ml/datasets/datasets/Parkinsons+Telemonitoring
http://archive.ics.uci.edu/ml/datasets/datasets/Parkinsons+Telemonitoring
http://archive.ics.uci.edu/ml/datasets/datasets/Demospongiae
http://archive.ics.uci.edu/ml/datasets/datasets/Demospongiae
http://archive.ics.uci.edu/ml/datasets/datasets/Opinosis+Opinion+%26frasl%3B+Review
http://archive.ics.uci.edu/ml/datasets/datasets/Opinosis+Opinion+%26frasl%3B+Review
http://archive.ics.uci.edu/ml/datasets/datasets/Breast+Tissue
http://archive.ics.uci.edu/ml/datasets/datasets/Breast+Tissue
http://archive.ics.uci.edu/ml/datasets/datasets/Cardiotocography
http://archive.ics.uci.edu/ml/datasets/datasets/Cardiotocography
http://archive.ics.uci.edu/ml/datasets/datasets/Wall-Following+Robot+Navigation+Data
http://archive.ics.uci.edu/ml/datasets/datasets/Wall-Following+Robot+Navigation+Data
http://archive.ics.uci.edu/ml/datasets/datasets/Spoken+Arabic+Digit
http://archive.ics.uci.edu/ml/datasets/datasets/Spoken+Arabic+Digit
http://archive.ics.uci.edu/ml/datasets/datasets/Localization+Data+for+Person+Activity
http://archive.ics.uci.edu/ml/datasets/datasets/Localization+Data+for+Person+Activity
http://archive.ics.uci.edu/ml/datasets/datasets/AutoUniv
http://archive.ics.uci.edu/ml/datasets/datasets/AutoUniv
http://archive.ics.uci.edu/ml/datasets/datasets/Steel+Plates+Faults
http://archive.ics.uci.edu/ml/datasets/datasets/Steel+Plates+Faults
http://archive.ics.uci.edu/ml/datasets/datasets/MiniBooNE+particle+identification
http://archive.ics.uci.edu/ml/datasets/datasets/MiniBooNE+particle+identification
http://archive.ics.uci.edu/ml/datasets/datasets/YearPredictionMSD
http://archive.ics.uci.edu/ml/datasets/datasets/YearPredictionMSD
http://archive.ics.uci.edu/ml/datasets/datasets/PEMS-SF
http://archive.ics.uci.edu/ml/datasets/datasets/PEMS-SF
http://archive.ics.uci.edu/ml/datasets/datasets/OpinRank+Review+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/OpinRank+Review+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Relative+location+of+CT+slices+on+axial+axis
http://archive.ics.uci.edu/ml/datasets/datasets/Relative+location+of+CT+slices+on+axial+axis
http://archive.ics.uci.edu/ml/datasets/datasets/Online+Handwritten+Assamese+Characters+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Online+Handwritten+Assamese+Characters+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/PubChem+Bioassay+Data
http://archive.ics.uci.edu/ml/datasets/datasets/PubChem+Bioassay+Data
http://archive.ics.uci.edu/ml/datasets/datasets/Record+Linkage+Comparison+Patterns
http://archive.ics.uci.edu/ml/datasets/datasets/Record+Linkage+Comparison+Patterns
http://archive.ics.uci.edu/ml/datasets/datasets/Communities+and+Crime+Unnormalized
http://archive.ics.uci.edu/ml/datasets/datasets/Communities+and+Crime+Unnormalized
http://archive.ics.uci.edu/ml/datasets/datasets/Vertebral+Column
http://archive.ics.uci.edu/ml/datasets/datasets/Vertebral+Column
http://archive.ics.uci.edu/ml/datasets/datasets/EMG+Physical+Action+Data+Set
http://archive.ics.uci.edu/ml/datasets/datasets/EMG+Physical+Action+Data+Set
http://archive.ics.uci.edu/ml/datasets/datasets/Vicon+Physical+Action+Data+Set
http://archive.ics.uci.edu/ml/datasets/datasets/Vicon+Physical+Action+Data+Set
http://archive.ics.uci.edu/ml/datasets/datasets/Amazon+Commerce+reviews+set
http://archive.ics.uci.edu/ml/datasets/datasets/Amazon+Commerce+reviews+set
http://archive.ics.uci.edu/ml/datasets/datasets/Amazon+Access+Samples
http://archive.ics.uci.edu/ml/datasets/datasets/Amazon+Access+Samples
http://archive.ics.uci.edu/ml/datasets/datasets/Reuter_50_50
http://archive.ics.uci.edu/ml/datasets/datasets/Reuter_50_50
http://archive.ics.uci.edu/ml/datasets/datasets/Farm+Ads
http://archive.ics.uci.edu/ml/datasets/datasets/Farm+Ads
http://archive.ics.uci.edu/ml/datasets/datasets/DBWorld+e-mails
http://archive.ics.uci.edu/ml/datasets/datasets/DBWorld+e-mails
http://archive.ics.uci.edu/ml/datasets/datasets/KEGG+Metabolic+Relation+Network+%28Directed%29
http://archive.ics.uci.edu/ml/datasets/datasets/KEGG+Metabolic+Relation+Network+%28Directed%29
http://archive.ics.uci.edu/ml/datasets/datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29
http://archive.ics.uci.edu/ml/datasets/datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29
http://archive.ics.uci.edu/ml/datasets/datasets/Bank+Marketing
http://archive.ics.uci.edu/ml/datasets/datasets/Bank+Marketing
http://archive.ics.uci.edu/ml/datasets/datasets/YouTube+Comedy+Slam+Preference+Data
http://archive.ics.uci.edu/ml/datasets/datasets/YouTube+Comedy+Slam+Preference+Data
http://archive.ics.uci.edu/ml/datasets/datasets/Gas+Sensor+Array+Drift+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Gas+Sensor+Array+Drift+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/ILPD+%28Indian+Liver+Patient+Dataset%29

http://archive.ics.uci.edu/ml/datasets/datasets/ILPD+%28Indian+Liver+Patient+Dataset%29
http://archive.ics.uci.edu/ml/datasets/datasets/OPPORTUNITY+Activity+Recognition
http://archive.ics.uci.edu/ml/datasets/datasets/OPPORTUNITY+Activity+Recognition
http://archive.ics.uci.edu/ml/datasets/datasets/Nomao
http://archive.ics.uci.edu/ml/datasets/datasets/Nomao
http://archive.ics.uci.edu/ml/datasets/datasets/SMS+Spam+Collection
http://archive.ics.uci.edu/ml/datasets/datasets/SMS+Spam+Collection
http://archive.ics.uci.edu/ml/datasets/datasets/Skin+Segmentation
http://archive.ics.uci.edu/ml/datasets/datasets/Skin+Segmentation
http://archive.ics.uci.edu/ml/datasets/datasets/Planning+Relax
http://archive.ics.uci.edu/ml/datasets/datasets/Planning+Relax
http://archive.ics.uci.edu/ml/datasets/datasets/PAMAP2+Physical+Activity+Monitoring
http://archive.ics.uci.edu/ml/datasets/datasets/PAMAP2+Physical+Activity+Monitoring
http://archive.ics.uci.edu/ml/datasets/datasets/Restaurant+%26+consumer+data
http://archive.ics.uci.edu/ml/datasets/datasets/Restaurant+%26+consumer+data
http://archive.ics.uci.edu/ml/datasets/datasets/CNAE-9
http://archive.ics.uci.edu/ml/datasets/datasets/CNAE-9
http://archive.ics.uci.edu/ml/datasets/datasets/Individual+household+electric+power+consumption
http://archive.ics.uci.edu/ml/datasets/datasets/Individual+household+electric+power+consumption
http://archive.ics.uci.edu/ml/datasets/datasets/seeds
http://archive.ics.uci.edu/ml/datasets/datasets/seeds
http://archive.ics.uci.edu/ml/datasets/datasets/Northix
http://archive.ics.uci.edu/ml/datasets/datasets/Northix
http://archive.ics.uci.edu/ml/datasets/datasets/QtyT40I10D100K
http://archive.ics.uci.edu/ml/datasets/datasets/QtyT40I10D100K
http://archive.ics.uci.edu/ml/datasets/datasets/Legal+Case+Reports
http://archive.ics.uci.edu/ml/datasets/datasets/Legal+Case+Reports
http://archive.ics.uci.edu/ml/datasets/datasets/Human+Activity+Recognition+Using+Smartphones
http://archive.ics.uci.edu/ml/datasets/datasets/Human+Activity+Recognition+Using+Smartphones
http://archive.ics.uci.edu/ml/datasets/datasets/One-hundred+plant+species+leaves+data+set
http://archive.ics.uci.edu/ml/datasets/datasets/One-hundred+plant+species+leaves+data+set
http://archive.ics.uci.edu/ml/datasets/datasets/Energy+efficiency
http://archive.ics.uci.edu/ml/datasets/datasets/Energy+efficiency
http://archive.ics.uci.edu/ml/datasets/datasets/Yacht+Hydrodynamics
http://archive.ics.uci.edu/ml/datasets/datasets/Yacht+Hydrodynamics
http://archive.ics.uci.edu/ml/datasets/datasets/Fertility
http://archive.ics.uci.edu/ml/datasets/datasets/Fertility
http://archive.ics.uci.edu/ml/datasets/datasets/Daphnet+Freezing+of+Gait
http://archive.ics.uci.edu/ml/datasets/datasets/Daphnet+Freezing+of+Gait
http://archive.ics.uci.edu/ml/datasets/datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29
http://archive.ics.uci.edu/ml/datasets/datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29
http://archive.ics.uci.edu/ml/datasets/datasets/ISTANBUL+STOCK+EXCHANGE
http://archive.ics.uci.edu/ml/datasets/datasets/ISTANBUL+STOCK+EXCHANGE
http://archive.ics.uci.edu/ml/datasets/datasets/Buzz+in+social+media+
http://archive.ics.uci.edu/ml/datasets/datasets/Buzz+in+social+media+
http://archive.ics.uci.edu/ml/datasets/datasets/First-order+theorem+proving
http://archive.ics.uci.edu/ml/datasets/datasets/First-order+theorem+proving
http://archive.ics.uci.edu/ml/datasets/datasets/Wearable+Computing%3A+Classification+of+Body+Posture+Movements+%28PUC-Rio%29
http://archive.ics.uci.edu/ml/datasets/datasets/Wearable+Computing%3A+Classification+of+Body+Posture+Movements+%28PUC-Rio%29
http://archive.ics.uci.edu/ml/datasets/datasets/Gas+sensor+arrays+in+open+sampling+settings
http://archive.ics.uci.edu/ml/datasets/datasets/Gas+sensor+arrays+in+open+sampling+settings
http://archive.ics.uci.edu/ml/datasets/datasets/Climate+Model+Simulation+Crashes
http://archive.ics.uci.edu/ml/datasets/datasets/Climate+Model+Simulation+Crashes
http://archive.ics.uci.edu/ml/datasets/datasets/MicroMass
http://archive.ics.uci.edu/ml/datasets/datasets/MicroMass
http://archive.ics.uci.edu/ml/datasets/datasets/QSAR+biodegradation
http://archive.ics.uci.edu/ml/datasets/datasets/QSAR+biodegradation
http://archive.ics.uci.edu/ml/datasets/datasets/BLOGGER
http://archive.ics.uci.edu/ml/datasets/datasets/BLOGGER
http://archive.ics.uci.edu/ml/datasets/datasets/Daily+and+Sports+Activities
http://archive.ics.uci.edu/ml/datasets/datasets/Daily+and+Sports+Activities
http://archive.ics.uci.edu/ml/datasets/datasets/User+Knowledge+Modeling
http://archive.ics.uci.edu/ml/datasets/datasets/User+Knowledge+Modeling
http://archive.ics.uci.edu/ml/datasets/datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Classification+Test+collection
http://archive.ics.uci.edu/ml/datasets/datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Classification+Test+collection
http://archive.ics.uci.edu/ml/datasets/datasets/NYSK
http://archive.ics.uci.edu/ml/datasets/datasets/NYSK
http://archive.ics.uci.edu/ml/datasets/datasets/Turkiye+Student+Evaluation
http://archive.ics.uci.edu/ml/datasets/datasets/Turkiye+Student+Evaluation
http://archive.ics.uci.edu/ml/datasets/datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+on+DC+Electrical+Machines%29
http://archive.ics.uci.edu/ml/datasets/datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+on+DC+Electrical+Machines%29

http://archive.ics.uci.edu/ml/datasets/datasets/EEG+Eye+State
http://archive.ics.uci.edu/ml/datasets/datasets/EEG+Eye+State
http://archive.ics.uci.edu/ml/datasets/datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure
http://archive.ics.uci.edu/ml/datasets/datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure
http://archive.ics.uci.edu/ml/datasets/datasets/seismic-bumps
http://archive.ics.uci.edu/ml/datasets/datasets/seismic-bumps
http://archive.ics.uci.edu/ml/datasets/datasets/banknote+authentication
http://archive.ics.uci.edu/ml/datasets/datasets/banknote+authentication
http://archive.ics.uci.edu/ml/datasets/datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Problem%3A+Pat
http://archive.ics.uci.edu/ml/datasets/datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++++Problem%3A+Pat
http://archive.ics.uci.edu/ml/datasets/datasets/YouTube+Multiview+Video+Games+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/YouTube+Multiview+Video+Games+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations
http://archive.ics.uci.edu/ml/datasets/datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations
http://archive.ics.uci.edu/ml/datasets/datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors
http://archive.ics.uci.edu/ml/datasets/datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors
http://archive.ics.uci.edu/ml/datasets/datasets/SkillCraft1+Master+Table+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/SkillCraft1+Master+Table+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units
http://archive.ics.uci.edu/ml/datasets/datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units
http://archive.ics.uci.edu/ml/datasets/datasets/SML2010
http://archive.ics.uci.edu/ml/datasets/datasets/SML2010
http://archive.ics.uci.edu/ml/datasets/datasets/Bike+Sharing+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Bike+Sharing+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Predict+keywords+activities+in+a+online+social+media
http://archive.ics.uci.edu/ml/datasets/datasets/Predict+keywords+activities+in+a+online+social+media

http://archive.ics.uci.edu/ml/datasets/datasets/Thoracic+Surgery+Data
http://archive.ics.uci.edu/ml/datasets/datasets/Thoracic+Surgery+Data
http://archive.ics.uci.edu/ml/datasets/datasets/EMG+dataset+in+Lower+Limb
http://archive.ics.uci.edu/ml/datasets/datasets/EMG+dataset+in+Lower+Limb
http://archive.ics.uci.edu/ml/datasets/datasets/SUSY
http://archive.ics.uci.edu/ml/datasets/datasets/SUSY
http://archive.ics.uci.edu/ml/datasets/datasets/HIGGS
http://archive.ics.uci.edu/ml/datasets/datasets/HIGGS
http://archive.ics.uci.edu/ml/datasets/datasets/Qualitative_Bankruptcy
http://archive.ics.uci.edu/ml/datasets/datasets/Qualitative_Bankruptcy
http://archive.ics.uci.edu/ml/datasets/datasets/LSVT+Voice+Rehabilitation
http://archive.ics.uci.edu/ml/datasets/datasets/LSVT+Voice+Rehabilitation
http://archive.ics.uci.edu/ml/datasets/datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer
http://archive.ics.uci.edu/ml/datasets/datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer
http://archive.ics.uci.edu/ml/datasets/datasets/Wilt
http://archive.ics.uci.edu/ml/datasets/datasets/Wilt
http://archive.ics.uci.edu/ml/datasets/datasets/User+Identification+From+Walking+Activity
http://archive.ics.uci.edu/ml/datasets/datasets/User+Identification+From+Walking+Activity
http://archive.ics.uci.edu/ml/datasets/datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer
http://archive.ics.uci.edu/ml/datasets/datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer
http://archive.ics.uci.edu/ml/datasets/datasets/Leaf
http://archive.ics.uci.edu/ml/datasets/datasets/Leaf
http://archive.ics.uci.edu/ml/datasets/datasets/Dresses_Attribute_Sales
http://archive.ics.uci.edu/ml/datasets/datasets/Dresses_Attribute_Sales
http://archive.ics.uci.edu/ml/datasets/datasets/Tamilnadu+Electricity+Board+Hourly+Readings
http://archive.ics.uci.edu/ml/datasets/datasets/Tamilnadu+Electricity+Board+Hourly+Readings
http://archive.ics.uci.edu/ml/datasets/datasets/Airfoil+Self-Noise
http://archive.ics.uci.edu/ml/datasets/datasets/Airfoil+Self-Noise
http://archive.ics.uci.edu/ml/datasets/datasets/Wholesale+customers
http://archive.ics.uci.edu/ml/datasets/datasets/Wholesale+customers
http://archive.ics.uci.edu/ml/datasets/datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis
http://archive.ics.uci.edu/ml/datasets/datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis
http://archive.ics.uci.edu/ml/datasets/datasets/Combined+Cycle+Power+Plant
http://archive.ics.uci.edu/ml/datasets/datasets/Combined+Cycle+Power+Plant
http://archive.ics.uci.edu/ml/datasets/datasets/Urban+Land+Cover

http://archive.ics.uci.edu/ml/datasets/datasets/Urban+Land+Cover
http://archive.ics.uci.edu/ml/datasets/datasets/Diabetes+130-US+hospitals+for+years+1999-2008
http://archive.ics.uci.edu/ml/datasets/datasets/Diabetes+130-US+hospitals+for+years+1999-2008
http://archive.ics.uci.edu/ml/datasets/datasets/Bach+Choral+Harmony
http://archive.ics.uci.edu/ml/datasets/datasets/Bach+Choral+Harmony
http://archive.ics.uci.edu/ml/datasets/datasets/StoneFlakes
http://archive.ics.uci.edu/ml/datasets/datasets/StoneFlakes
http://archive.ics.uci.edu/ml/datasets/datasets/Tennis+Major+Tournament+Match+Statistics
http://archive.ics.uci.edu/ml/datasets/datasets/Tennis+Major+Tournament+Match+Statistics
http://archive.ics.uci.edu/ml/datasets/datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sc
Recordings
http://archive.ics.uci.edu/ml/datasets/datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sc
Recordings
http://archive.ics.uci.edu/ml/datasets/datasets/Gesture+Phase+Segmentation
http://archive.ics.uci.edu/ml/datasets/datasets/Gesture+Phase+Segmentation
http://archive.ics.uci.edu/ml/datasets/datasets/Perfume+Data
http://archive.ics.uci.edu/ml/datasets/datasets/Perfume+Data
http://archive.ics.uci.edu/ml/datasets/datasets/BlogFeedback
http://archive.ics.uci.edu/ml/datasets/datasets/BlogFeedback
http://archive.ics.uci.edu/ml/datasets/datasets/REALDISP+Activity+Recognition+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/REALDISP+Activity+Recognition+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Newspaper+and+magazine+images+segmentation+dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Newspaper+and+magazine+images+segmentation+dataset
http://archive.ics.uci.edu/ml/datasets/datasets/AAAI+2014+Accepted+Papers
http://archive.ics.uci.edu/ml/datasets/datasets/AAAI+2014+Accepted+Papers
http://archive.ics.uci.edu/ml/datasets/datasets/Gas+sensor+array+under+flow+modulation
http://archive.ics.uci.edu/ml/datasets/datasets/Gas+sensor+array+under+flow+modulation
http://archive.ics.uci.edu/ml/datasets/datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures
http://archive.ics.uci.edu/ml/datasets/datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures
http://archive.ics.uci.edu/ml/datasets/datasets/UJIIndoorLoc
http://archive.ics.uci.edu/ml/datasets/datasets/UJIIndoorLoc
http://archive.ics.uci.edu/ml/datasets/datasets/Sentence+Classification
http://archive.ics.uci.edu/ml/datasets/datasets/Sentence+Classification
http://archive.ics.uci.edu/ml/datasets/datasets/Dow+Jones+Index
http://archive.ics.uci.edu/ml/datasets/datasets/Dow+Jones+Index
http://archive.ics.uci.edu/ml/datasets/datasets/sEMG+for+Basic+Hand+movements
http://archive.ics.uci.edu/ml/datasets/datasets/sEMG+for+Basic+Hand+movements
http://archive.ics.uci.edu/ml/datasets/datasets/AAAI+2013+Accepted+Papers
http://archive.ics.uci.edu/ml/datasets/datasets/AAAI+2013+Accepted+Papers
http://archive.ics.uci.edu/ml/datasets/datasets/Geographical+Original+of+Music
http://archive.ics.uci.edu/ml/datasets/datasets/Geographical+Original+of+Music
http://archive.ics.uci.edu/ml/datasets/datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Pla

http://archive.ics.uci.edu/ml/datasets/datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Pla

http://archive.ics.uci.edu/ml/datasets/datasets/Grammatical+Facial+Expressions
http://archive.ics.uci.edu/ml/datasets/datasets/Grammatical+Facial+Expressions
http://archive.ics.uci.edu/ml/datasets/datasets/NoisyOffice
http://archive.ics.uci.edu/ml/datasets/datasets/NoisyOffice
http://archive.ics.uci.edu/ml/datasets/datasets/MHEALTH+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/MHEALTH+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Student+Performance
http://archive.ics.uci.edu/ml/datasets/datasets/Student+Performance
http://archive.ics.uci.edu/ml/datasets/datasets/ElectricityLoadDiagrams20112014
http://archive.ics.uci.edu/ml/datasets/datasets/ElectricityLoadDiagrams20112014
http://archive.ics.uci.edu/ml/datasets/datasets/Gas+sensor+array+under+dynamic+gas+mixtures
http://archive.ics.uci.edu/ml/datasets/datasets/Gas+sensor+array+under+dynamic+gas+mixtures
http://archive.ics.uci.edu/ml/datasets/datasets/microblogPCU
http://archive.ics.uci.edu/ml/datasets/datasets/microblogPCU
http://archive.ics.uci.edu/ml/datasets/datasets/Firm-Teacher_Clave-Direction_Classification
http://archive.ics.uci.edu/ml/datasets/datasets/Firm-Teacher_Clave-Direction_Classification
http://archive.ics.uci.edu/ml/datasets/datasets/Dataset+for+Sensorless+Drive+Diagnosis
http://archive.ics.uci.edu/ml/datasets/datasets/Dataset+for+Sensorless+Drive+Diagnosis
http://archive.ics.uci.edu/ml/datasets/datasets/TV+News+Channel+Commercial+Detection+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/TV+News+Channel+Commercial+Detection+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Phishing+Websites
http://archive.ics.uci.edu/ml/datasets/datasets/Phishing+Websites
http://archive.ics.uci.edu/ml/datasets/datasets/Greenhouse+Gas+Observing+Network
http://archive.ics.uci.edu/ml/datasets/datasets/Greenhouse+Gas+Observing+Network
http://archive.ics.uci.edu/ml/datasets/datasets/Diabetic+Retinopathy+Debrecen+Data+Set
http://archive.ics.uci.edu/ml/datasets/datasets/Diabetic+Retinopathy+Debrecen+Data+Set
http://archive.ics.uci.edu/ml/datasets/datasets/HIV-1+protease+cleavage
http://archive.ics.uci.edu/ml/datasets/datasets/HIV-1+protease+cleavage
http://archive.ics.uci.edu/ml/datasets/datasets/Sentiment+Labelled+Sentences
http://archive.ics.uci.edu/ml/datasets/datasets/Sentiment+Labelled+Sentences
http://archive.ics.uci.edu/ml/datasets/datasets/Online+News+Popularity
http://archive.ics.uci.edu/ml/datasets/datasets/Online+News+Popularity

http://archive.ics.uci.edu/ml/datasets/datasets/Forest+type+mapping
http://archive.ics.uci.edu/ml/datasets/datasets/Forest+type+mapping
http://archive.ics.uci.edu/ml/datasets/datasets/wiki4HE
http://archive.ics.uci.edu/ml/datasets/datasets/wiki4HE
http://archive.ics.uci.edu/ml/datasets/datasets/Online+Video+Characteristics+and+Transcoding+Time+Iet
http://archive.ics.uci.edu/ml/datasets/datasets/Online+Video+Characteristics+and+Transcoding+Time+Iet
http://archive.ics.uci.edu/ml/datasets/datasets/Chronic_Kidney_Disease
http://archive.ics.uci.edu/ml/datasets/datasets/Chronic_Kidney_Disease
http://archive.ics.uci.edu/ml/datasets/datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendns+2012-2014
http://archive.ics.uci.edu/ml/datasets/datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendns+2012-2014
http://archive.ics.uci.edu/ml/datasets/datasets/Folio
http://archive.ics.uci.edu/ml/datasets/datasets/Folio
http://archive.ics.uci.edu/ml/datasets/datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015
http://archive.ics.uci.edu/ml/datasets/datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015
http://archive.ics.uci.edu/ml/datasets/datasets/Cuff-Less+Blood+Pressure+Estimation
http://archive.ics.uci.edu/ml/datasets/datasets/Cuff-Less+Blood+Pressure+Estimation
http://archive.ics.uci.edu/ml/datasets/datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions
http://archive.ics.uci.edu/ml/datasets/datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions
http://archive.ics.uci.edu/ml/datasets/datasets/Mice+Protein+Expression
http://archive.ics.uci.edu/ml/datasets/datasets/Mice+Protein+Expression
http://archive.ics.uci.edu/ml/datasets/datasets/UJIIndoorLoc-Mag
http://archive.ics.uci.edu/ml/datasets/datasets/UJIIndoorLoc-Mag
http://archive.ics.uci.edu/ml/datasets/datasets/Heterogeneity+Activity+Recognition
http://archive.ics.uci.edu/ml/datasets/datasets/Heterogeneity+Activity+Recognition
http://archive.ics.uci.edu/ml/datasets/datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+ytics+Data+Set
http://archive.ics.uci.edu/ml/datasets/datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+ytics+Data+Set
http://archive.ics.uci.edu/ml/datasets/datasets/HEPMASS
http://archive.ics.uci.edu/ml/datasets/datasets/HEPMASS
http://archive.ics.uci.edu/ml/datasets/datasets/Indoor+User+Movement+Prediction+from+RSS+data
http://archive.ics.uci.edu/ml/datasets/datasets/Indoor+User+Movement+Prediction+from+RSS+data
http://archive.ics.uci.edu/ml/datasets/datasets/Open+University+Learning+Analytics+dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Open+University+Learning+Analytics+dataset
http://archive.ics.uci.edu/ml/datasets/datasets/default+of+credit+card+clients
http://archive.ics.uci.edu/ml/datasets/datasets/default+of+credit+card+clients
http://archive.ics.uci.edu/ml/datasets/datasets/Mesothelioma%2E2%80%99s+disease+data+set+
http://archive.ics.uci.edu/ml/datasets/datasets/Mesothelioma%2E2%80%99s+disease+data+set+
http://archive.ics.uci.edu/ml/datasets/datasets/Online+Retail
http://archive.ics.uci.edu/ml/datasets/datasets/Online+Retail
http://archive.ics.uci.edu/ml/datasets/datasets/SIFT10M
http://archive.ics.uci.edu/ml/datasets/datasets/SIFT10M
http://archive.ics.uci.edu/ml/datasets/datasets/GPS+Trajectories
http://archive.ics.uci.edu/ml/datasets/datasets/GPS+Trajectories
http://archive.ics.uci.edu/ml/datasets/datasets/Detect+Malicious+Executable%28AntiVirus%29
http://archive.ics.uci.edu/ml/datasets/datasets/Detect+Malicious+Executable%28AntiVirus%29
http://archive.ics.uci.edu/ml/datasets/datasets/Occupancy+Detection+
http://archive.ics.uci.edu/ml/datasets/datasets/Occupancy+Detection+
http://archive.ics.uci.edu/ml/datasets/datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Table r+Monitoring+Parkinson%2E2%80%99s+Disease
http://archive.ics.uci.edu/ml/datasets/datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Table r+Monitoring+Parkinson%2E2%80%99s+Disease
http://archive.ics.uci.edu/ml/datasets/datasets/News+Aggregator
http://archive.ics.uci.edu/ml/datasets/datasets/News+Aggregator
http://archive.ics.uci.edu/ml/datasets/datasets/Air+Quality
http://archive.ics.uci.edu/ml/datasets/datasets/Air+Quality
http://archive.ics.uci.edu/ml/datasets/datasets/Twin+gas+sensor+arrays
http://archive.ics.uci.edu/ml/datasets/datasets/Twin+gas+sensor+arrays
http://archive.ics.uci.edu/ml/datasets/datasets/Gas+sensors+for+home+activity+monitoring
http://archive.ics.uci.edu/ml/datasets/datasets/Gas+sensors+for+home+activity+monitoring
http://archive.ics.uci.edu/ml/datasets/datasets/Facebook+Comment+Volume+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Facebook+Comment+Volume+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Smartphone+Dataset+for+Human+Activity+Recognition+% R%29+in+Ambient+Assisted+Living+%28AAL%29
http://archive.ics.uci.edu/ml/datasets/datasets/Smartphone+Dataset+for+Human+Activity+Recognition+% R%29+in+Ambient+Assisted+Living+%28AAL%29
http://archive.ics.uci.edu/ml/datasets/datasets/Polish+companies+bankruptcy+data
http://archive.ics.uci.edu/ml/datasets/datasets/Polish+companies+bankruptcy+data
http://archive.ics.uci.edu/ml/datasets/datasets/Activity+Recoqnition+system+based+on+Multisensor+da

usion+%28AReM%29
http://archive.ics.uci.edu/ml/datasets/datasets/Activity+Recognition+system+based+on+Multisensor+da
usion+%28AReM%29
http://archive.ics.uci.edu/ml/datasets/datasets/Dota2+Games+Results
http://archive.ics.uci.edu/ml/datasets/datasets/Dota2+Games+Results
http://archive.ics.uci.edu/ml/datasets/datasets/Facebook+metrics
http://archive.ics.uci.edu/ml/datasets/datasets/Facebook+metrics
http://archive.ics.uci.edu/ml/datasets/datasets/UbiqLog+%28smartphone+lifelogging%29
http://archive.ics.uci.edu/ml/datasets/datasets/UbiqLog+%28smartphone+lifelogging%29
http://archive.ics.uci.edu/ml/datasets/datasets/NIPS+Conference+Papers+1987-2015
http://archive.ics.uci.edu/ml/datasets/datasets/NIPS+Conference+Papers+1987-2015
http://archive.ics.uci.edu/ml/datasets/datasets/HTRU2
http://archive.ics.uci.edu/ml/datasets/datasets/HTRU2
http://archive.ics.uci.edu/ml/datasets/datasets/Drug+consumption+%28quantified%29
http://archive.ics.uci.edu/ml/datasets/datasets/Drug+consumption+%28quantified%29
http://archive.ics.uci.edu/ml/datasets/datasets/Appliances+energy+prediction
http://archive.ics.uci.edu/ml/datasets/datasets/Appliances+energy+prediction
http://archive.ics.uci.edu/ml/datasets/datasets/Miskolc+IIS+Hybrid+IPS
http://archive.ics.uci.edu/ml/datasets/datasets/Miskolc+IIS+Hybrid+IPS
http://archive.ics.uci.edu/ml/datasets/datasets/KDC-4007+dataset+Collection
http://archive.ics.uci.edu/ml/datasets/datasets/KDC-4007+dataset+Collection
http://archive.ics.uci.edu/ml/datasets/datasets/Geo-
Magnetic+field+and+WLan+dataset+for+indoor+localisation+from+wristband+and+smartphone
http://archive.ics.uci.edu/ml/datasets/datasets/Geo-
Magnetic+field+and+WLan+dataset+for+indoor+localisation+from+wristband+and+smartphone
http://archive.ics.uci.edu/ml/datasets/datasets/DrivFace
http://archive.ics.uci.edu/ml/datasets/datasets/DrivFace
http://archive.ics.uci.edu/ml/datasets/datasets/Website+Phishing
http://archive.ics.uci.edu/ml/datasets/datasets/Website+Phishing
http://archive.ics.uci.edu/ml/datasets/datasets/YouTube+Spam+Collection
http://archive.ics.uci.edu/ml/datasets/datasets/YouTube+Spam+Collection
http://archive.ics.uci.edu/ml/datasets/datasets/Beijing+PM2.5+Data
http://archive.ics.uci.edu/ml/datasets/datasets/Beijing+PM2.5+Data
http://archive.ics.uci.edu/ml/datasets/datasets/Cargo+2000+Freight+Tracking+and+Tracing
http://archive.ics.uci.edu/ml/datasets/datasets/Cargo+2000+Freight+Tracking+and+Tracing
http://archive.ics.uci.edu/ml/datasets/datasets/Cervical+cancer+%28Risk+Factors%29
http://archive.ics.uci.edu/ml/datasets/datasets/Cervical+cancer+%28Risk+Factors%29
http://archive.ics.uci.edu/ml/datasets/datasets/Quality+Assessment+of+Digital+Colposcopies
http://archive.ics.uci.edu/ml/datasets/datasets/Quality+Assessment+of+Digital+Colposcopies
http://archive.ics.uci.edu/ml/datasets/datasets/KASANDR
http://archive.ics.uci.edu/ml/datasets/datasets/KASANDR
http://archive.ics.uci.edu/ml/datasets/datasets/FMA%3A+A+Dataset+For+Music+Analysis
http://archive.ics.uci.edu/ml/datasets/datasets/FMA%3A+A+Dataset+For+Music+Analysis
http://archive.ics.uci.edu/ml/datasets/datasets/Air+quality
http://archive.ics.uci.edu/ml/datasets/datasets/Air+quality
http://archive.ics.uci.edu/ml/datasets/datasets/Epileptic+Seizure+Recognition
http://archive.ics.uci.edu/ml/datasets/datasets/Epileptic+Seizure+Recognition
http://archive.ics.uci.edu/ml/datasets/datasets/Devanagari+Handwritten+Character+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Devanagari+Handwritten+Character+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Stock+portfolio+performance
http://archive.ics.uci.edu/ml/datasets/datasets/Stock+portfolio+performance
http://archive.ics.uci.edu/ml/datasets/datasets/MoCap+Hand+Postures
http://archive.ics.uci.edu/ml/datasets/datasets/MoCap+Hand+Postures
http://archive.ics.uci.edu/ml/datasets/datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+
ral+connected+speech
http://archive.ics.uci.edu/ml/datasets/datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+
ral+connected+speech
http://archive.ics.uci.edu/ml/datasets/datasets/Data+for+Software+Engineering+Teamwork+Assessment+i
ucation+Setting
http://archive.ics.uci.edu/ml/datasets/datasets/Data+for+Software+Engineering+Teamwork+Assessment+i
ucation+Setting
http://archive.ics.uci.edu/ml/datasets/datasets/PM2.5+Data+of+Five+Chinese+Cities
http://archive.ics.uci.edu/ml/datasets/datasets/PM2.5+Data+of+Five+Chinese+Cities
http://archive.ics.uci.edu/ml/datasets/datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+C
ics+Tablet
http://archive.ics.uci.edu/ml/datasets/datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+C
ics+Tablet
http://archive.ics.uci.edu/ml/datasets/datasets/Sales_Transactions_Dataset_Weekly
http://archive.ics.uci.edu/ml/datasets/datasets/Sales_Transactions_Dataset_Weekly
http://archive.ics.uci.edu/ml/datasets/datasets/Las+Vegas+Strip
http://archive.ics.uci.edu/ml/datasets/datasets/Las+Vegas+Strip
http://archive.ics.uci.edu/ml/datasets/datasets/Eco-hotel
http://archive.ics.uci.edu/ml/datasets/datasets/Eco-hotel
http://archive.ics.uci.edu/ml/datasets/datasets/MEU-Mobile+KSD
http://archive.ics.uci.edu/ml/datasets/datasets/MEU-Mobile+KSD
http://archive.ics.uci.edu/ml/datasets/datasets/Crowdsourced+Mapping
http://archive.ics.uci.edu/ml/datasets/datasets/Crowdsourced+Mapping

http://archive.ics.uci.edu/ml/datasets/datasets/gene+expression+cancer+RNA-Seq
http://archive.ics.uci.edu/ml/datasets/datasets/gene+expression+cancer+RNA-Seq
http://archive.ics.uci.edu/ml/datasets/datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C
etooth+and+magnetometer
http://archive.ics.uci.edu/ml/datasets/datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C
etooth+and+magnetometer
http://archive.ics.uci.edu/ml/datasets/datasets/chestnut+%E2%80%93+LARVIC
http://archive.ics.uci.edu/ml/datasets/datasets/chestnut+%E2%80%93+LARVIC
http://archive.ics.uci.edu/ml/datasets/datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Op
l+Burst+Switching+%28OBS%29+Network
http://archive.ics.uci.edu/ml/datasets/datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Op
l+Burst+Switching+%28OBS%29+Network
http://archive.ics.uci.edu/ml/datasets/datasets/Motion+Capture+Hand+Postures
http://archive.ics.uci.edu/ml/datasets/datasets/Motion+Capture+Hand+Postures
http://archive.ics.uci.edu/ml/datasets/datasets/Anuran+Calls+%28MFCCs%29
http://archive.ics.uci.edu/ml/datasets/datasets/Anuran+Calls+%28MFCCs%29
http://archive.ics.uci.edu/ml/datasets/datasets/TTC-
3600%3A+Benchmark+dataset+for+Turkish+text+categorization
http://archive.ics.uci.edu/ml/datasets/datasets/TTC-
3600%3A+Benchmark+dataset+for+Turkish+text+categorization
http://archive.ics.uci.edu/ml/datasets/datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy
http://archive.ics.uci.edu/ml/datasets/datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy
http://archive.ics.uci.edu/ml/datasets/datasets/Daily+Demand+Forecasting+Orders
http://archive.ics.uci.edu/ml/datasets/datasets/Daily+Demand+Forecasting+Orders
http://archive.ics.uci.edu/ml/datasets/datasets/Paper+Reviews
http://archive.ics.uci.edu/ml/datasets/datasets/Paper+Reviews
http://archive.ics.uci.edu/ml/datasets/datasets/extention+of+Z-Alizadeh+sani+dataset
http://archive.ics.uci.edu/ml/datasets/datasets/extention+of+Z-Alizadeh+sani+dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Z-Alizadeh+Sani
http://archive.ics.uci.edu/ml/datasets/datasets/Z-Alizadeh+Sani
http://archive.ics.uci.edu/ml/datasets/datasets/Dynamic+Features+of+VirusShare+Executables
http://archive.ics.uci.edu/ml/datasets/datasets/Dynamic+Features+of+VirusShare+Executables
http://archive.ics.uci.edu/ml/datasets/datasets/IDA2016Challenge
http://archive.ics.uci.edu/ml/datasets/datasets/IDA2016Challenge
http://archive.ics.uci.edu/ml/datasets/datasets/DSRC+Vehicle+Communications
http://archive.ics.uci.edu/ml/datasets/datasets/DSRC+Vehicle+Communications
http://archive.ics.uci.edu/ml/datasets/datasets/Mturk+User-Perceived+Clusters+over+Images
http://archive.ics.uci.edu/ml/datasets/datasets/Mturk+User-Perceived+Clusters+over+Images
http://archive.ics.uci.edu/ml/datasets/datasets/Character+Font+Images
http://archive.ics.uci.edu/ml/datasets/datasets/Character+Font+Images
http://archive.ics.uci.edu/ml/datasets/datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-I
nstance+Learning+with+Instance+Labels
http://archive.ics.uci.edu/ml/datasets/datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-I
nstance+Learning+with+Instance+Labels
http://archive.ics.uci.edu/ml/datasets/datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Chilc
+
http://archive.ics.uci.edu/ml/datasets/datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Chilc
+
http://archive.ics.uci.edu/ml/datasets/datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adole
t++
http://archive.ics.uci.edu/ml/datasets/datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adole
t++
http://archive.ics.uci.edu/ml/datasets/datasets/APS+Failure+at+Scania+Trucks
http://archive.ics.uci.edu/ml/datasets/datasets/APS+Failure+at+Scania+Trucks
http://archive.ics.uci.edu/ml/datasets/datasets/Wireless+Indoor+Localization
http://archive.ics.uci.edu/ml/datasets/datasets/Wireless+Indoor+Localization
http://archive.ics.uci.edu/ml/datasets/datasets/HCC+Survival
http://archive.ics.uci.edu/ml/datasets/datasets/HCC+Survival
http://archive.ics.uci.edu/ml/datasets/datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Data
2014+and+2015
http://archive.ics.uci.edu/ml/datasets/datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Data
2014+and+2015
http://archive.ics.uci.edu/ml/datasets/datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.
%29
http://archive.ics.uci.edu/ml/datasets/datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.
.%29
http://archive.ics.uci.edu/ml/datasets/datasets/Autism+Screening+Adult
http://archive.ics.uci.edu/ml/datasets/datasets/Autism+Screening+Adult
http://archive.ics.uci.edu/ml/datasets/datasets/Activity+recognition+with+healthy+older+people+usir
batteryless+wearable+sensor
http://archive.ics.uci.edu/ml/datasets/datasets/Activity+recognition+with+healthy+older+people+usir
batteryless+wearable+sensor
http://archive.ics.uci.edu/ml/datasets/datasets/Immunotherapy+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Immunotherapy+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Cryotherapy+Dataset+
http://archive.ics.uci.edu/ml/datasets/datasets/Cryotherapy+Dataset+
http://archive.ics.uci.edu/ml/datasets/datasets/OCT+data+%26+Color+Fundus+of+Left+%26+Right+

http://archive.ics.uci.edu/ml/datasets/datasets/Discrete+Tone+Image+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Discrete+Tone+Image+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/News+Popularity+in+Multiple+Social+Media+Platforms
http://archive.ics.uci.edu/ml/datasets/datasets/News+Popularity+in+Multiple+Social+Media+Platforms
http://archive.ics.uci.edu/ml/datasets/datasets/Ultrasonic+flowmeter+diagnostics
http://archive.ics.uci.edu/ml/datasets/datasets/Ultrasonic+flowmeter+diagnostics
http://archive.ics.uci.edu/ml/datasets/datasets/ICMLA+2014+Accepted+Papers+Data+Set
http://archive.ics.uci.edu/ml/datasets/datasets/ICMLA+2014+Accepted+Papers+Data+Set
http://archive.ics.uci.edu/ml/datasets/datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Naviga

http://archive.ics.uci.edu/ml/datasets/datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Naviga

http://archive.ics.uci.edu/ml/datasets/datasets/Container+Crane+Controller+Data+Set
http://archive.ics.uci.edu/ml/datasets/datasets/Container+Crane+Controller+Data+Set
http://archive.ics.uci.edu/ml/datasets/datasets/Residential+Building+Data+Set
http://archive.ics.uci.edu/ml/datasets/datasets/Residential+Building+Data+Set
http://archive.ics.uci.edu/ml/datasets/datasets/Health+News+in+Twitter
http://archive.ics.uci.edu/ml/datasets/datasets/Health+News+in+Twitter
http://archive.ics.uci.edu/ml/datasets/datasets/chipseq
http://archive.ics.uci.edu/ml/datasets/datasets/chipseq
http://archive.ics.uci.edu/ml/datasets/datasets/SGEMM+GPU+kernel+performance
http://archive.ics.uci.edu/ml/datasets/datasets/SGEMM+GPU+kernel+performance
http://archive.ics.uci.edu/ml/datasets/datasets/Repeat+Consumption+Matrices
http://archive.ics.uci.edu/ml/datasets/datasets/Repeat+Consumption+Matrices
http://archive.ics.uci.edu/ml/datasets/datasets/detection_of_IoT_botnet_attacks_N_BaIoT
http://archive.ics.uci.edu/ml/datasets/datasets/detection_of_IoT_botnet_attacks_N_BaIoT
http://archive.ics.uci.edu/ml/datasets/datasets/Absenteeism+at+work
http://archive.ics.uci.edu/ml/datasets/datasets/Absenteeism+at+work
http://archive.ics.uci.edu/ml/datasets/datasets/SCADI
http://archive.ics.uci.edu/ml/datasets/datasets/SCADI
http://archive.ics.uci.edu/ml/datasets/datasets/Condition+monitoring+of+hydraulic+systems
http://archive.ics.uci.edu/ml/datasets/datasets/Condition+monitoring+of+hydraulic+systems
http://archive.ics.uci.edu/ml/datasets/datasets/Carbon+Nanotubes
http://archive.ics.uci.edu/ml/datasets/datasets/Carbon+Nanotubes
http://archive.ics.uci.edu/ml/datasets/datasets/Optical+Interconnection+Network+
http://archive.ics.uci.edu/ml/datasets/datasets/Optical+Interconnection+Network+
http://archive.ics.uci.edu/ml/datasets/datasets/Sports+articles+for+objectivity+analysis
http://archive.ics.uci.edu/ml/datasets/datasets/Sports+articles+for+objectivity+analysis
http://archive.ics.uci.edu/ml/datasets/datasets/Breast+Cancer+Coimbra
http://archive.ics.uci.edu/ml/datasets/datasets/Breast+Cancer+Coimbra
http://archive.ics.uci.edu/ml/datasets/datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data
http://archive.ics.uci.edu/ml/datasets/datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data
http://archive.ics.uci.edu/ml/datasets/datasets/Dishonest+Internet+users+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Dishonest+Internet+users+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Victorian+Era+Authorship+Attribution
http://archive.ics.uci.edu/ml/datasets/datasets/Victorian+Era+Authorship+Attribution
http://archive.ics.uci.edu/ml/datasets/datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Se

http://archive.ics.uci.edu/ml/datasets/datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Se

http://archive.ics.uci.edu/ml/datasets/datasets/Multimodal+Damage+Identification+for+Humanitarian+Co
ting
http://archive.ics.uci.edu/ml/datasets/datasets/Multimodal+Damage+Identification+for+Humanitarian+Co
ting
http://archive.ics.uci.edu/ml/datasets/datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals
http://archive.ics.uci.edu/ml/datasets/datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals
http://archive.ics.uci.edu/ml/datasets/datasets/Roman+Urdu+Data+Set
http://archive.ics.uci.edu/ml/datasets/datasets/Roman+Urdu+Data+Set
http://archive.ics.uci.edu/ml/datasets/datasets/Avila
http://archive.ics.uci.edu/ml/datasets/datasets/Avila
http://archive.ics.uci.edu/ml/datasets/datasets/PANDOR
http://archive.ics.uci.edu/ml/datasets/datasets/PANDOR
http://archive.ics.uci.edu/ml/datasets/datasets/Drug+Review+Dataset+%28Druglib.com%29
http://archive.ics.uci.edu/ml/datasets/datasets/Drug+Review+Dataset+%28Druglib.com%29
http://archive.ics.uci.edu/ml/datasets/datasets/Drug+Review+Dataset+%28Drugs.com%29
http://archive.ics.uci.edu/ml/datasets/datasets/Drug+Review+Dataset+%28Drugs.com%29
http://archive.ics.uci.edu/ml/datasets/datasets/Physical+Unclonable+Functions
http://archive.ics.uci.edu/ml/datasets/datasets/Physical+Unclonable+Functions
http://archive.ics.uci.edu/ml/datasets/datasets/Superconductivty+Data
http://archive.ics.uci.edu/ml/datasets/datasets/Superconductivty+Data
http://archive.ics.uci.edu/ml/datasets/datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29
http://archive.ics.uci.edu/ml/datasets/datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29
http://archive.ics.uci.edu/ml/datasets/datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2
http://archive.ics.uci.edu/ml/datasets/datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data+Set+2

```
http://archive.ics.uci.edu/ml/datasets/datasets/Student+Academics+Performance
http://archive.ics.uci.edu/ml/datasets/datasets/Student+Academics+Performance
http://archive.ics.uci.edu/ml/datasets/datasets/Online+Shoppers+Purchasing+Intention+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Online+Shoppers+Purchasing+Intention+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/PMU-UD
http://archive.ics.uci.edu/ml/datasets/datasets/PMU-UD
http://archive.ics.uci.edu/ml/datasets/datasets/Parkinson%27s+Disease+Classification
http://archive.ics.uci.edu/ml/datasets/datasets/Parkinson%27s+Disease+Classification
http://archive.ics.uci.edu/ml/datasets/datasets/Electrical+Grid+Stability+Simulated+Data+
http://archive.ics.uci.edu/ml/datasets/datasets/Electrical+Grid+Stability+Simulated+Data+
http://archive.ics.uci.edu/ml/datasets/datasets/Caesarian+Section+Classification+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/Caesarian+Section+Classification+Dataset
http://archive.ics.uci.edu/ml/datasets/datasets/BAUM-1
http://archive.ics.uci.edu/ml/datasets/datasets/BAUM-1
http://archive.ics.uci.edu/ml/datasets/datasets/BAUM-2
http://archive.ics.uci.edu/ml/datasets/datasets/BAUM-2
http://archive.ics.uci.edu/ml/datasets/datasets/Audit+Data
http://archive.ics.uci.edu/ml/datasets/datasets/Audit+Data
http://archive.ics.uci.edu/ml/datasets/datasets/BuddyMove+Data+Set
http://archive.ics.uci.edu/ml/datasets/datasets/BuddyMove+Data+Set
http://archive.ics.uci.edu/ml/datasets/datasets/Real+estate+valuation+data+set
http://archive.ics.uci.edu/ml/datasets/datasets/Real+estate+valuation+data+set
http://archive.ics.uci.edu/ml/datasets/datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+bas
n+natural+connected+speech+Data+Set+
http://archive.ics.uci.edu/ml/datasets/datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+bas
n+natural+connected+speech+Data+Set+
http://archive.ics.uci.edu/ml/datasets/datasets/Somerville+Happiness+Survey
http://archive.ics.uci.edu/ml/datasets/datasets/Somerville+Happiness+Survey
http://archive.ics.uci.edu/ml/datasets/datasets/2.4+GHZ+Indoor+Channel+Measurements
http://archive.ics.uci.edu/ml/datasets/datasets/2.4+GHZ+Indoor+Channel+Measurements
http://archive.ics.uci.edu/ml/datasets/datasets/EMG+data+for+gestures
http://archive.ics.uci.edu/ml/datasets/datasets/EMG+data+for+gestures
http://archive.ics.uci.edu/ml/datasets/datasets/Parking+Birmingham
http://archive.ics.uci.edu/ml/datasets/datasets/Parking+Birmingham
http://archive.ics.uci.edu/ml/datasets/datasets/Behavior+of+the+urban+traffic+of+the+city+of+Sao+Pa
in+Brazil
http://archive.ics.uci.edu/ml/datasets/datasets/Behavior+of+the+urban+traffic+of+the+city+of+Sao+Pa
in+Brazil
http://archive.ics.uci.edu/ml/datasets/datasets/Travel+Reviews
http://archive.ics.uci.edu/ml/datasets/datasets/Travel+Reviews
http://archive.ics.uci.edu/ml/datasets/datasets/Tarvel+Review+Ratings
http://archive.ics.uci.edu/ml/datasets/datasets/Tarvel+Review+Ratings
http://archive.ics.uci.edu/ml/datasets/datasets/Rice+Leaf+Diseases
http://archive.ics.uci.edu/ml/datasets/datasets/Rice+Leaf+Diseases
```

In [22]:

```
len(set(linksdata))
```

Out[22]:

```
469
```

Tipo de datos

In [23]:

```
URL2 = "https://archive.ics.uci.edu/ml/datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list"
```

In [24]:

```
HTML2 = requests.get(URL2).text
```

In [25]:

```
datatype = BeautifulSoup(HTML2, "lxml")
```

In [26]:

```

print(datatype.prettify()) 

<!DOCTYPE HTML>
<html>
<body>
<p>
  "-//W3C//DTD HTML 4.01 Transitional//EN">;
</p>
<title>
  UCI Machine Learning Repository: Data Sets
</title>
<!-- Stylesheet link -->
<link href="assets/ml.css" rel="stylesheet" type="text/css"/>
<script language="JavaScript" type="text/javascript">
<!--
function checkform ( form )
{
  // see http://www.thesitewizard.com/archive/validation.shtml
  // for an explanation of this script and how to use it on your
  // own website

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

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

  // ** END **
  return true ;
}

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

//-->
</script>
<!-- SITE HEADER (INCLUDES LOGO AND SEARCH BOX) -->
<!-- SITE HEADER (INCLUDES LOGO AND SEARCH BOX) -->
<table bgcolor="#003366" width="100%">
<tr>
<td>
  <span class="normal">
    <a alt="Home" href="index.html">
      
    </a>
    <br/>
    <a href="http://cml.ics.uci.edu">
      
    </a>
  </span>
</td>
<td>
  <span class="normal">
    <a alt="ML Repository" href="http://ml.ics.uci.edu/ml">
      
    </a>
  </span>
</td>
</tr>
</table>

```

```

<font color="#FFDD33">
  Center for Machine Learning and Intelligent Systems
</font>
</a>
</span>
</td>
<td align="right" valign="top" width="100%">
  <span class="whitetext">
    <a href="about.html">
      About
    </a>
    <a href="citation_policy.html">
      Citation Policy
    </a>
    <a href="donation_policy.html">
      Donate a Data Set
    </a>
    <a href="contact.html">
      Contact
    </a>
  </span>
  <br/>
  <br/>
<!-- Search Google --&gt;
&lt;form action="http://www.google.com/custom" method="GET" onsubmit="return checkform(this);"&gt;
  &lt;input maxlength="255" name="q" size="30" type="text" value="" /&gt;
  &lt;input name="sa" type="submit" value="Search"/&gt;
  &lt;input name="cof" type="hidden" value="AH:center;LH:130;L:http://archive.ics.uci.edu/assets/logo.gif;LW:384;AWFID:869c0b2eaa8d518e;" /&gt;
  &lt;input name="domains" type="hidden" value="ics.uci.edu"/&gt;
  &lt;br/&gt;
  &lt;input checked="" name="sitesearch" type="radio" value="ics.uci.edu"/&gt;
  &lt;span class="whitetext"&gt;
    &lt;font size="1"&gt;
      Repository
    &lt;/font&gt;
  &lt;/span&gt;
  &lt;input name="sitesearch" type="radio" value="" /&gt;
  &lt;span class="whitetext"&gt;
    &lt;font size="1"&gt;
      Web
    &lt;/font&gt;
  &lt;/span&gt;
  &lt;a href="http://www.google.com/search"&gt;
    &lt;img align="middle" alt="Google" border="0" height="27" src="http://www.google.com/logos/Logo_25blk.gif"/&gt;
  &lt;/a&gt;
  &lt;br/&gt;
&lt;/form&gt;
<!-- Search Google --&gt;
&lt;span class="whitetext"&gt;
  &lt;a href="datasets.php"&gt;
    &lt;font color="#FFDD33" size="3"&gt;
      &lt;b&gt;
        View ALL Data Sets
      &lt;/b&gt;
    &lt;/font&gt;
  &lt;/a&gt;
&lt;/span&gt;
&lt;br/&gt;
&lt;/td&gt;
&lt;/tr&gt;
&lt;/table&gt;
&lt;br/&gt;
&lt;table cellpadding="3"&gt;
&lt;tr&gt;
  &lt;td valign="top"&gt;
    &lt;table&gt;
      &lt;tr&gt;
        &lt;td&gt;
          &lt;p align="center" class="big"&gt;
            Browse Through:
          &lt;/p&gt;
        &lt;/td&gt;
      &lt;/tr&gt;
    &lt;/table&gt;
  &lt;/td&gt;
</pre>

```

```

<table border="2" cellpadding="2" width="150">
<tr>
<td bgcolor="#003366">
<p class="whitetext">
<b>
Default Task
</b>
</p>
</td>
</tr>
<tr>
<td valign="top">
<p class="normal">
<a href="datasets.php?
format=&task=cla&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=
">
Classification
</a>
<font color="red">
(350)
</font>
<br/>
<a href="datasets.php?
format=&task=reg&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=
">
Regression
</a>
<font color="red">
(96)
</font>
<br/>
<a href="datasets.php?
format=&task=clu&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=
">
Clustering
</a>
<font color="red">
(84)
</font>
<br/>
<a href="datasets.php?
format=&task=other&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=
st">
Other
</a>
<font color="red">
(55)
</font>
</p>
</td>
</tr>
<tr>
<td bgcolor="#003366">
<p class="whitetext">
<b>
Attribute Type
</b>
</p>
</td>
</tr>
<tr>
<td valign="top">
<p class="normal">
<a href="datasets.php?
format=&task=&att=cat&area=&numAtt=&numIns=&type=&sort=nameUp&view=
">
Categorical
</a>
<font color="red">
(38)
</font>
<br/>
<a href="datasets.php?
format=&task=&att=num&area=&numAtt=&numIns=&type=&sort=nameUp&view=
">
Numerical
</a>
- - - - -

```

```

<font color="red">
(307)
</font>
<br/>
<a href="datasets.php?
format=&task=&att=mix&area=&numAtt=&numIns=&type=&sort=nameUp&view=
">
    Mixed
</a>
<font color="red">
(55)
</font>
</p>
</td>
</tr>
<tr>
<td bgcolor="#003366">
<p class="whitetext">
<b>
    Data Type
</b>
</p>
</td>
</tr>
<tr>
<td valign="top">
<p class="normal">
<a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=mvar&sort=nameUp&view=
t">
    Multivariate
</a>
<font color="red">
(357)
</font>
<br/>
<a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=uvar&sort=nameUp&view=
t">
    Univariate
</a>
<font color="red">
(23)
</font>
<br/>
<a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=seq&sort=nameUp&view=
">
    Sequential
</a>
<font color="red">
(47)
</font>
<br/>
<a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=ts&sort=nameUp&view=
t">
    Time-Series
</a>
<font color="red">
(91)
</font>
<br/>
<a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=text&sort=nameUp&view=
t">
    Text
</a>
<font color="red">
(53)
</font>
<br/>
<a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=dt&sort=nameUp&view=
t">
    Domain-Theory
</a>

```

```

<font color="red">
(23)
</font>
<br/>
<a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=other&sort=nameUp&view
t">
    Other
</a>
<font color="red">
(21)
</font>
<br/>
</p>
</td>
</tr>
<tr>
<td bgcolor="#003366">
<p class="whitetext">
<b>
    Area
</b>
</p>
</td>
</tr>
<tr>
<td valign="top">
<p class="normal">
    <a href="datasets.php?
format=&task=&att=&area=life&numAtt=&numIns=&type=&sort=nameUp&view
t">
        Life Sciences
    </a>
    <font color="red">
        (107)
    </font>
    <br/>
    <a href="datasets.php?
format=&task=&att=&area=phys&numAtt=&numIns=&type=&sort=nameUp&view
t">
        Physical Sciences
    </a>
    <font color="red">
        (49)
    </font>
    <br/>
    <a href="datasets.php?
format=&task=&att=&area=comp&numAtt=&numIns=&type=&sort=nameUp&view
t">
        CS / Engineering
    </a>
    <font color="red">
        (170)
    </font>
    <br/>
    <a href="datasets.php?
format=&task=&att=&area=soc&numAtt=&numIns=&type=&sort=nameUp&view
t">
        Social Sciences
    </a>
    <font color="red">
        (26)
    </font>
    <br/>
    <a href="datasets.php?
format=&task=&att=&area=bus&numAtt=&numIns=&type=&sort=nameUp&view
t">
        Business
    </a>
    <font color="red">
        (29)
    </font>
    <br/>
    <a href="datasets.php?
format=&task=&att=&area=game&numAtt=&numIns=&type=&sort=nameUp&view
t">
        Game
    </a>

```

```

        </a>
        <font color="red">
            (10)
        </font>
        <br/>
        <a href="datasets.php?
format=&task=&att=&area=other&numAtt=&numIns=&type=&sort=nameUp&viewlist">
            Other
        </a>
        <font color="red">
            (73)
        </font>
        </p>
    </td>
</tr>
<tr>
    <td bgcolor="#003366">
        <p class="whitetext">
            <b>
                # Attributes
            </b>
        </p>
    </td>
</tr>
<tr>
    <td valign="top">
        <p class="normal">
            <a href="datasets.php?
format=&task=&att=&area=&numAtt=less10&numIns=&type=&sort=nameUp&viewlist">
                Less than 10
            </a>
            <font color="red">
                (113)
            </font>
            <br/>
            <a href="datasets.php?
format=&task=&att=&area=&numAtt=10to100&numIns=&type=&sort=nameUp&viewlist">
                10 to 100
            </a>
            <font color="red">
                (210)
            </font>
            <br/>
            <a href="datasets.php?
format=&task=&att=&area=&numAtt=greater100&numIns=&type=&sort=nameUp&viewlist">
                Greater than 100
            </a>
            <font color="red">
                (84)
            </font>
        </p>
    </td>
</tr>
<tr>
    <td bgcolor="#003366">
        <p class="whitetext">
            <b>
                # Instances
            </b>
        </p>
    </td>
</tr>
<tr>
    <td valign="top">
        <p class="normal">
            <a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=less100&type=&sort=nameUp&viewlist">
                Less than 100
            </a>
            <font color="red">
                (27)
            </font>

```

```

<br/>
<a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=100to1000&type=&sort=nameUp&w=list">
    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=list">
    Greater than 1000
</a>
<font color="red">
    (246)
</font>
</p>
</td>
</tr>
<tr>
<td bgcolor="#003366">
<p class="whitetext">
<b>
    Format Type
</b>
</p>
</td>
</tr>
<tr>
<td valign="top">
<p class="normal">
    <a href="datasets.php?
format=mat&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=*>
        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>
    </p>
    </td>
    </tr>
    </table>
</td>
<td valign="top">
<table width="100%">
<tr>
<td>
<p class="big">
<b>
        469
    </b>
</p>
</td>
<td align="right">
<p class="normal">
    <a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=tak
Table View
</a>
<font color="gray">
    List View
</font>
</p>

```

```

</td>
</tr>
</table>
<table cellpadding="3">
<tr>
<td>
<hr/>
<p class="normal">
1.
<b>
<a href="datasets/2.4+GHZ+Indoor+Channel+Measurements">
2.4 GHZ Indoor Channel Measurements
</a>
</b>
: 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.
</p>
<p class="normal">
2.
<b>
<a href="datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29">
3D Road Network (North Jutland, Denmark)
</a>
</b>
: 3D road network with highly accurate elevation information (+-20cm) from Denmark used
in eco-routing and fuel/Co2-estimation routing algorithms.
</p>
<p 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), including paper titles, abstracts, and keywords of varying granularity.
</p>
<p class="normal">
4.
<b>
<a href="datasets/AAAI+2014+Accepted+Papers">
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>
<p class="normal">
5.
<b>
<a href="datasets/Abalone">
Abalone
</a>
</b>
: Predict the age of abalone from physical measurements
</p>
<p class="normal">
6.
<b>
<a href="datasets/Abscisic+Acid+Signaling+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 separate boolean pseudodynamic simulations using an asynchronous update scheme.
</p>
<p class="normal">
7.
<b>
<a href="datasets/Absenteeism+at+work">
Absenteeism at work
</a>
</b>
: The database was created with records of absenteeism at work from July 2007 to July 2010 at a courier company in Brazil.
</p>

```

```

<p class="normal">
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 the ADLs performed by two users on a daily basis in their own homes.
</p>
<p class="normal">
9.
<b>
<a href="datasets/Activity+Recognition+from+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 for Activity Recognition research purposes.
</p>
<p class="normal">
10.
<b>
<a href="datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29">
    Activity Recognition system based on Multisensor 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>
<p class="normal">
11.
<b>
<a href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor">
    Activity recognition with healthy older people using a batteryless wearable sensor
</a>
</b>
: 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>
<p class="normal">
12.
<b>
<a href="datasets/Acute+Inflammations">
    Acute Inflammations
</a>
</b>
: 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.
</p>
<p class="normal">
13.
<b>
<a href="datasets/Adult">
    Adult
</a>
</b>
: Predict whether income exceeds $50K/yr based on census data. Also known as "Census Income" dataset.
</p>
<p class="normal">
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.
</p>

```

```

<p class="normal">
15.
<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.
</p>
<p class="normal">
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.
</p>
<p class="normal">
17.
<b>
<a href="datasets/Amazon+Access+Samples">
  Amazon Access Samples
</a>
</b>
: Amazon's InfoSec is getting smarter about the way Access data is leveraged. This is an a
nonymized sample of access provisioned within the company.
</p>
<p class="normal">
18.
<b>
<a href="datasets/Amazon+Commerce+reviews+set">
  Amazon Commerce reviews set
</a>
</b>
: The dataset is used for authorship identification in online Writeprint which is a new re
search field of pattern recognition.
</p>
<p class="normal">
19.
<b>
<a href="datasets/Annealing">
  Annealing
</a>
</b>
: Steel annealing data
</p>
<p class="normal">
20.
<b>
<a href="datasets/Anonymous+Microsoft+Web+Data">
  Anonymous Microsoft Web Data
</a>
</b>
: Log of anonymous users of www.microsoft.com; predict areas of the web site a user visite
d based on data on other areas the user visited.
</p>
<p class="normal">
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 the species labels (multilabel).
</p>
<p class="normal">
22.
<b>
<a href="datasets/Appliances+energy+prediction">
  Appliances energy prediction
</a>
</b>
: Experimental data used to create regression models of appliances energy use in a low

```

energy building.

</p>

<p class="normal">

23.

 APS Failure at Scania Trucks

: 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.

</p>

<p class="normal">

24.

 Arcene

: ARCENE's task is to distinguish cancer versus normal patterns from mass-spectrometric data. This is a two-class classification problem with continuous input variables. This dataset is one of 5 datasets of the NIPS 2003 feature selection challenge.

</p>

<p class="normal">

25.

 Arrhythmia

: Distinguish between the presence and absence of cardiac arrhythmia and classify it in one of the 16 groups.

</p>

<p class="normal">

26.

 Artificial Characters

: Dataset artificially generated by using first order theory which describes structure of ten capital letters of English alphabet

</p>

<p class="normal">

27.

 Audiology (Original)

: Nominal audiology dataset from Baylor

</p>

<p class="normal">

28.

 Audiology (Standardized)

: Standardized version of the original audiology database

</p>

<p class="normal">

29.

 Audit Data

: 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.

</p>

<p class="normal">

30.

 Australian Sign Language signs

```

</a>
</b>
: 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.
</p>
<p class="normal">
31.
<b>
<a href="datasets/Australian+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 signs were captured from a native signer using high-quality position trackers
</p>
<p class="normal">
32.
<b>
<a href="datasets/Autism+Screening+Adult">
    Autism Screening Adult
</a>
</b>
: Autistic Spectrum Disorder Screening Data for Adult. This dataset is related to
classification and predictive tasks.
</p>
<p class="normal">
33.
<b>
<a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent++">
    Autistic Spectrum Disorder Screening Data for Adolescent
</a>
</b>
: Autistic Spectrum Disorder Screening Data for Adolescent. This dataset is related to
classification and predictive tasks.
</p>
<p class="normal">
34.
<b>
<a href="datasets/Autistic+Spectrum+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
</p>
<p class="normal">
35.
<b>
<a href="datasets/Auto+MPG">
    Auto MPG
</a>
</b>
: Revised from CMU StatLib library, data concerns city-cycle fuel consumption
</p>
<p class="normal">
36.
<b>
<a href="datasets/Automobile">
    Automobile
</a>
</b>
: From 1985 Ward's Automotive Yearbook
</p>
<p class="normal">
37.
<b>
<a href="datasets/AutoUniv">
    AutoUniv
</a>
</b>
: 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.
</p>
<p class="normal">
38.
<b>
<a href="datasets/Avila">
    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>
 <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.
 </p>
 <p class="normal">
 40.

 Bach Chorales

 : Time-series data based on chorales; challenge is to learn generative grammar; data in Li sp
 </p>
 <p class="normal">
 41.

 Badges

 : Badges labeled with a "+" or "-" as a function of a person's name
 </p>
 <p class="normal">
 42.

 Bag of Words

 : This data set contains five text collections in the form of bags-of-words.
 </p>
 <p class="normal">
 43.

 Balance Scale

 : Balance scale weight & distance database
 </p>
 <p class="normal">
 44.

 Balloons

 : Data previously used in cognitive psychology experiment; 4 data sets represent different conditions of an experiment
 </p>
 <p class="normal">
 45.

 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).
 </p>
 <p class="normal">
 46.

```

```


 banknote authentication

: Data were extracted from images that were taken for the evaluation of an authentication
procedure for banknotes.
</p>
<p class="normal">
47.

 BAUM-1

: BAUM-1 dataset contains 1184 multimodal facial video clips collected from 31 subjects. T
he 1184 video clips contain spontaneous facial expressions and speech of 13 emotional and mental s
tates.
</p>
<p class="normal">
48.

 BAUM-2

: A multilingual audio-visual affective face database consisting of 1047 video clips of 28
6 subjects.
</p>
<p class="normal">
49.

 Behavior of the urban traffic of the city of Sao Paulo in Brazil

: The database was created with records of behavior of the urban traffic of the city of
Sao Paulo in Brazil.
</p>
<p class="normal">
50.

 Beijing PM2.5 Data

: 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.
</p>
<p class="normal">
51.

 Bike Sharing Dataset

: 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>
<p class="normal">
52.

 BLE RSSI Dataset for Indoor localization and Navigation

: This dataset contains RSSI readings gathered from an array of Bluetooth Low Energy
(BLE) iBeacons in a real-world and operational indoor environment for localization and navigation
purposes.
</p>
<p class="normal">
53.

 BlogFeedback


```

```

~/*~
: Instances in this dataset contain features extracted from blog posts. The task
associated with the data is to predict how many comments the post will receive.
</p>
<p class="normal">
54.

BLOGGER

: In this paper, we look for to recognize the causes of users tend
to cyber space in Kohkiloye and Boyer Ahmad Province in
Iran
</p>
<p class="normal">
55.

Blood Transfusion Service Center

: Data taken from the Blood Transfusion Service Center in Hsin-Chu City in Taiwan -- this
is a classification problem.
</p>
<p class="normal">
56.

Breast Cancer

: Breast Cancer Data (Restricted Access)
</p>
<p class="normal">
57.

Breast Cancer Coimbra

: Clinical features were observed or measured for 64 patients with breast cancer and 52 h
ealthy controls.
</p>
<p class="normal">
58.

Breast Cancer Wisconsin (Diagnostic)

: Diagnostic Wisconsin Breast Cancer Database
</p>
<p class="normal">
59.

Breast Cancer Wisconsin (Original)

: Original Wisconsin Breast Cancer Database
</p>
<p class="normal">
60.

Breast Cancer Wisconsin (Prognostic)

: Prognostic Wisconsin Breast Cancer Database
</p>
<p class="normal">
61.

Breast Tissue


```

```
~\n>
 : Dataset with electrical impedance measurements of freshly excised tissue samples from the breast.
</p>
<p class="normal">
 62.

 BuddyMove Data Set

 : User interest information extracted from user reviews published in holidayiq.com about various types of point of interests in South India
</p>
<p class="normal">
 63.

 Burst Header Packet (BHP) flooding attack on Optical Burst Switching (OBS) Network

 : One of the primary challenges in identifying the risks of the Burst Header Packet (BHP) flood attacks in Optical Burst Switching networks (OBS) is the scarcity of reliable historical data.
</p>
<p class="normal">
 64.

 Buzz 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>
<p class="normal">
 65.

 Caesarian Section Classification Dataset

 : This dataset contains information about caesarian section results of 80 pregnant women with the most important characteristics of delivery problems in the medical field.
</p>
<p class="normal">
 66.

 CalIt2 Building People Counts

 : This data comes from the main door of the CalIt2 building at UCI.
</p>
<p class="normal">
 67.

 Car Evaluation

 : Derived from simple hierarchical decision model, this database may be useful for testing constructive induction and structure discovery methods.
</p>
<p class="normal">
 68.

 Carbon Nanotubes

 : This dataset contains 10721 initial and calculated atomic coordinates of carbon nanotubes.
</p>
<~ ~~~~"normal"~>
```

```

<p class="normal">
69.

Cardiotocography

: The dataset consists of measurements of fetal heart rate (FHR) and uterine contraction (UC) features on cardiotocograms classified by expert obstetricians.
</p>
<p class="normal">
70.

Cargo 2000 Freight Tracking and Tracing

: 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).
</p>
<p class="normal">
71.

Census Income

: Predict whether income exceeds $50K/yr based on census data. Also known as "Adult" data set.
</p>
<p class="normal">
72.

Census-Income (KDD)

: This data set contains weighted census data extracted from the 1994 and 1995 current population surveys conducted by the U.S. Census Bureau.
</p>
<p class="normal">
73.

Cervical cancer (Risk Factors)

: This dataset focuses on the prediction of indicators/diagnosis of cervical cancer. The features cover demographic information, habits, and historic medical records.
</p>
<p class="normal">
74.

Challenger USA Space Shuttle O-Ring

: Task: predict the number of O-rings that experience thermal distress on a flight at 31 degrees F given data on the previous 23 shuttle flights
</p>
<p class="normal">
75.

Character Font Images

: Character images from scanned and computer generated fonts.
</p>
<p class="normal">
76.

Character Trajectories


```

: Multiple, labelled samples of pen tip trajectories recorded whilst writing individual characters. All samples are from the same writer, for the purposes of primitive extraction. Only characters with a single pen-down segment were considered.

```
</p>
<p class="normal">
77.

 Chess (Domain Theories)

: 6 different domain theories for generating legal moves of chess
</p>
<p class="normal">
78.

 Chess (King-Rook vs. King)

: Chess Endgame Database for White King and Rook against Black King (KRK).
</p>
<p class="normal">
79.

 Chess (King-Rook vs. King-Knight)

: Knight Pin Chess End-Game Database Creator
</p>
<p class="normal">
80.

 Chess (King-Rook vs. King-Pawn)

: King+Rook versus King+Pawn on a7 (usually abbreviated KRKPQ7).
</p>
<p class="normal">
81.

 chestnut - LARVIC

: The research project presents this database, shows the images of chestnuts that will be processed to determine the presence or absence of defects
</p>
<p class="normal">
82.

 chipseq

: 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.
</p>
<p class="normal">
83.

 Chronic_Kidney_Disease

: This dataset can be used to predict the chronic kidney disease and it can be collected from the hospital nearly 2 months of period.
</p>
<p class="normal">
84.

 Climate Model Simulation Crashes

<...>
```

```

</p>
: Given Latin hypercube samples of 18 climate model input parameter values, predict climat
e model simulation crashes and determine the parameter value combinations that cause the failures.
</p>
<p class="normal">
85.

 Cloud

: Little Documentation
</p>
<p class="normal">
86.

 CMU Face Images

: 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
</p>
<p class="normal">
87.

 CNAE-9

: This is a data set containing 1080 documents of free text business descriptions of Brazi
lian companies categorized into a
subset of 9 categories
</p>
<p class="normal">
88.

 Coil 1999 Competition Data

: This data set is from the 1999 Computational Intelligence and Learning (COIL)
competition. The data contains measurements of river chemical concentrations and algae densities.
</p>
<p class="normal">
89.

 Combined Cycle Power Plant

: The dataset contains 9568 data points collected from a Combined Cycle Power Plant over
6 years (2006-2011), when the plant was set to work with full load.
</p>
<p class="normal">
90.

 Communities and Crime

: Communities within the United States. The data combines socio-economic data from the 19
90 US Census, law enforcement data from the 1990 US LEMAS survey, and crime data from the 1995 FBI
UCR.
</p>
<p class="normal">
91.

 Communities and Crime Unnormalized

: 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 f
rom the 1995 FBI UCR
</p>
 ...

```

```
<p class="normal">
92.

 Computer Hardware

: Relative CPU Performance Data, described in terms of its cycle time, memory size, etc.
</p>
<p class="normal">
93.

 Concrete Compressive Strength

: Concrete is the most important material in civil engineering. The concrete compressive strength is a highly nonlinear function of age and ingredients.
</p>
<p class="normal">
94.

 Concrete Slump Test

: Concrete is a highly complex material. The slump flow of concrete is not only determined by the water content, but that is also influenced by other concrete ingredients.
</p>
<p class="normal">
95.

 Condition Based Maintenance of Naval Propulsion Plants

: 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.
</p>
<p class="normal">
96.

 Condition monitoring of hydraulic systems

: The data set addresses the condition assessment of a hydraulic test rig based on multi sensor data. Four fault types are superimposed with several severity grades impeding selective quantification.
</p>
<p class="normal">
97.

 Congressional Voting Records

: 1984 United Stated Congressional Voting Records; Classify as Republican or Democrat
</p>
<p class="normal">
98.

 Connect-4

: Contains connect-4 positions
</p>
<p class="normal">
99.

 Connectionist Bench (Nettalk Corpus)

: 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
```

```
</p>
<p class="normal">
100.

 Connectionist Bench (Sonar, Mines vs. Rocks)

: 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.
</p>
<p class="normal">
101.

 Connectionist Bench (Vowel Recognition - Deterding Data)

: Speaker independent recognition of the eleven steady state vowels of British English
using a specified training set of lpc derived log area ratios.
</p>
<p class="normal">
102.

 Container Crane Controller Data Set

: A container crane has the function of transporting containers from one point to another
point.
</p>
<p class="normal">
103.

 Contraceptive Method Choice

: Dataset is a subset of the 1987 National Indonesia Contraceptive Prevalence Survey.
</p>
<p class="normal">
104.

 Corel Image 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-occurrence
</p>
<p class="normal">
105.

 Covertype

: Forest CoverType dataset
</p>
<p class="normal">
106.

 Credit Approval

: This data concerns credit card applications; good mix of attributes
</p>
<p class="normal">
107.

 Crowdsourced Mapping

: Crowdsourced data from OpenStreetMap is used to automate the classification of satellite
```

```

images into different land cover classes (impervious, farm, forest, grass, orchard, water).
</p>
<p class="normal">
108.

Cryotherapy Dataset

: This dataset contains information about wart treatment results of 90 patients using cryotherapy.
</p>
<p class="normal">
109.

CSM (Conventional and Social Media Movies) Dataset 2014 and 2015

: 12 features categorized as conventional and social media features. Both conventional features, collected from movies databases on Web as well as social media features (YouTube, Twitter).
</p>
<p class="normal">
110.

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.
</p>
<p class="normal">
111.

Cylinder Bands

: Used in decision tree induction for mitigating process delays known as "cylinder bands" in rotogravure printing
</p>
<p class="normal">
112.

Daily and Sports Activities

: The dataset comprises motion sensor data of 19 daily and sports activities each performed by 8 subjects in their own style for 5 minutes. Five Xsens MTx units are used on the torso, arms, and legs.
</p>
<p class="normal">
113.

Daily Demand Forecasting Orders

: The dataset was collected during 60 days, this is a real database of a brazilian logistics company.
</p>
<p class="normal">
114.

Daphnet Freezing of Gait

: This dataset contains the annotated readings of 3 acceleration sensors at the hip and leg of Parkinson's disease patients that experience freezing of gait (FoG) during walking tasks.
</p>
<p class="normal">
115.


```

```


 Data for Software Engineering Teamwork Assessment in Education Setting

 : Data include over 100 Team Activity Measures and outcomes (ML classes) obtained from ac-
tivities of 74 student teams during the creation of final class project in SW Eng. classes at
SFSU, Fulda, FAU
</p>
<p class="normal">
116.

 Dataset for ADL Recognition with Wrist-worn Accelerometer

 : Recordings of 16 volunteers performing 14 Activities of Daily Living (ADL) while carryir-
g a single wrist-worn tri-axial accelerometer.
</p>
<p class="normal">
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.
</p>
<p class="normal">
118.

 DBWorld e-mails

 : It contains 64 e-mails which I have manually collected from DBWorld mailing list. They a-
re classified in: 'announces of conferences' and 'everything else'.
</p>
<p class="normal">
119.

 default of credit card clients

 : 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.
</p>
<p class="normal">
120.

<a href="datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-Label+Multi-
Instance+Learning+with+Instance+Labels">
 DeliciousMIL: A Data Set for Multi-Label Multi-Instance Learning with Instance Labels

 : This dataset includes 1) 12234 documents (8251 training, 3983 test) extracted from Delic-
iousT140 dataset, 2) class labels for all documents, 3) labels for a subset of sentences of the te-
st documents.
</p>
<p class="normal">
121.

 Demospongiae

 : Marine sponges of the Demospongiae class classification domain.
</p>
<p class="normal">
122.

 Dermatology


```

: Aim for this dataset is to determine the type of Eryhemato-Squamous Disease.  
 </p>  
 <p class="normal">  
 123.  
 <b>  
   <a href="datasets/Detect+Malacious+Executable%28AntiVirus%29">  
     Detect Malacious Executable(AntiVirus)  
   </a>  
 </b>  
 : I extract features from malacious and non-malacious and create and training dataset to  
 teach svm classifier.Dataset made of unknown executable to detect if it is virus or normal safe  
 executable.  
 </p>  
 <p class="normal">  
 124.  
 <b>  
   <a href="datasets/detection\_of\_IoT\_botnet\_attacks\_N\_BaIoT">  
     detection\_of\_IoT\_botnet\_attacks\_N\_BaIoT  
   </a>  
 </b>  
 : 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 Mir  
 ai and BASHLITE.  
 </p>  
 <p class="normal">  
 125.  
 <b>  
   <a href="datasets/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 examples each. The dataset is split into training set(85%) and testing set(15%).  
 </p>  
 <p class="normal">  
 126.  
 <b>  
   <a href="datasets/Dexter">  
     Dexter  
   </a>  
 </b>  
 : DEXTER is a text classification problem in a bag-of-word representation. This is a two-class  
 classification problem with sparse continuous input variables. This dataset is one of five datasets of the NIPS 2003 feature selection challenge.  
 </p>  
 <p class="normal">  
 127.  
 <b>  
   <a href="datasets/DGP2++The+Second+Data+Generation+Program">  
     DGP2 - The Second Data Generation Program  
   </a>  
 </b>  
 : Generates application domains based on specific parameters, number of features, and  
 proportion of positive to negative examples  
 </p>  
 <p class="normal">  
 128.  
 <b>  
   <a href="datasets/Diabetes">  
     Diabetes  
   </a>  
 </b>  
 : This diabetes dataset is from AIM '94  
 </p>  
 <p class="normal">  
 129.  
 <b>  
   <a href="datasets/Diabetes+130-US+hospitals+for+years+1999-2008">  
     Diabetes 130-US hospitals for years 1999-2008  
   </a>  
 </b>  
 : This data has been prepared to analyze factors related to readmission as well as other  
 outcomes pertaining to patients with diabetes.  
 </p>  
 <p class="normal">

130.  
<b>  
  <a href="datasets/Diabetic+Retinopathy+Debrecen+Data+Set">  
    Diabetic Retinopathy Debrecen Data Set  
  </a>  
</b>  
  : This dataset contains features extracted from the Messidor image set to predict whether an image contains signs of diabetic retinopathy or not.  
</p>  
<p class="normal">  
131.  
<b>  
  <a href="datasets/Discrete+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.  
</p>  
<p class="normal">  
132.  
<b>  
  <a href="datasets/Dishonest+Internet+users+Dataset">  
    Dishonest Internet users Dataset  
  </a>  
</b>  
  : The dataset was used to test an architecture based on a trust model capable to cope with the evaluation of the trustworthiness of users interacting in pervasive environments.  
</p>  
<p class="normal">  
133.  
<b>  
  <a href="datasets/Document+Understanding">  
    Document Understanding  
  </a>  
</b>  
  : Five concepts, expressed as predicates, to be learned  
</p>  
<p class="normal">  
134.  
<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 Los Angeles  
</p>  
<p class="normal">  
135.  
<b>  
  <a href="datasets/Dorothea">  
    Dorothea  
  </a>  
</b>  
  : DOROTHEA is a drug discovery dataset. Chemical compounds represented by structural molecular features must be classified as active (binding to thrombin) or inactive. This is one of 5 datasets of the NIPS 2003 feature selection challenge.  
</p>  
<p class="normal">  
136.  
<b>  
  <a href="datasets/Dota2+Games+Results">  
    Dota2 Games Results  
  </a>  
</b>  
  : 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.  
</p>  
<p class="normal">  
137.  
<b>  
  <a href="datasets/Dow+Jones+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.

</p>

<p class="normal">

138.

<b>

<a href="datasets/Dresses\_Attribute\_Sales">

Dresses\_Attribute\_Sales

</a>

</b>

: This dataset contain Attributes of dresses and their recommendations according to their sales. Sales are monitor on the basis of alternate days.

</p>

<p class="normal">

139.

<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 drivers with several facial features.

</p>

<p class="normal">

140.

<b>

<a href="datasets/Drug+consumption%28quantified%29">

Drug consumption (quantified)

</a>

</b>

: Classify type of drug consumer by personality data

</p>

<p class="normal">

141.

<b>

<a href="datasets/Drug+Review+Dataset%28Druglib.com%29">

Drug Review Dataset (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>

<p class="normal">

142.

<b>

<a href="datasets/Drug+Review+Dataset%28Drugs.com%29">

Drug Review Dataset (Drugs.com)

</a>

</b>

: The dataset provides patient reviews on specific drugs along with related conditions and a 10 star patient rating reflecting overall patient satisfaction.

</p>

<p class="normal">

143.

<b>

<a href="datasets/DSRC+Vehicle+Communications">

DSRC Vehicle Communications

</a>

</b>

: This set Provides data regarding wireless communications between vehicles and road side units. two separate data sets are provided (normal scenario) and in the presence of attacker (jammer).

</p>

<p class="normal">

144.

<b>

<a href="datasets/Dynamic+Features+of+VirusShare+Executables">

Dynamic Features of VirusShare Executables

</a>

</b>

: This dataset contains the dynamic features of 107,888 executables, collected by VirusShare from Nov/2010 to Jul/2014.

</p>

<p class="normal">

145.

<b>

```


 E. Coli Genes

: 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 pro
vided.
</p>
<p class="normal">
146.

<a
href="datasets/Early+biomarkers+of+Parkinson%E2%80%99s+disease+based+on+natural+connected+speech+Da
set+">
 Early biomarkers of Parkinson's disease based on natural connected speech Data Set

:
</p>
<p class="normal">
147.

 Early biomarkers of Parkinson's disease based on natural connected speech

: 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.
</p>
<p class="normal">
148.

 EBL Domain Theories

: Assorted small-scale domain theories
</p>
<p class="normal">
149.

 Echocardiogram

: Data for classifying if patients will survive for at least one year after a heart
attack
</p>
<p class="normal">
150.

 Eco-hotel

: This dataset includes Online Textual Reviews from both online (e.g., TripAdvisor) and
offline (e.g., Guests' book) sources from the Areias do Seixo Eco-Resort.
</p>
<p class="normal">
151.

 Ecoli

: This data contains protein localization sites
</p>
<p class="normal">
152.

 Economic Sanctions

: Domain Theory on Economic Sanctions; Undocumented

```

```
</p>
<p class="normal">
153.

 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.
</p>
<p class="normal">
154.

 EEG Database

: 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
</p>
<p class="normal">
155.

 EEG Eye State

: The data set consists of 14 EEG values and a value indicating the eye state.
</p>
<p class="normal">
156.

 EEG Steady-State Visual Evoked Potential Signals

: This database consists on 30 subjects performing Brain Computer Interface for Steady State Visual Evoked Potentials (BCI-SSVEP).
</p>
<p class="normal">
157.

 El Nino

: The data set contains oceanographic and surface meteorological readings taken from a series of buoys positioned throughout the equatorial Pacific.
</p>
<p class="normal">
158.

 Electrical Grid Stability Simulated Data

: The local stability analysis of the 4-node star system (electricity producer is in the center) implementing Decentral Smart Grid Control concept.
</p>
<p class="normal">
159.

 ElectricityLoadDiagrams20112014

: This data set contains electricity consumption of 370 points/clients.
</p>
<p class="normal">
160.

 EMG data for gestures

: These are files of raw EMG data recorded by MYO Thalmic bracelet
```

```
</p>
<p class="normal">
161.

 EMG dataset in Lower Limb

: 3 different exercises: sitting, standing and walking in the muscles: biceps femoris, vastus medialis, rectus femoris and semitendinosus addition to goniometry in the exercises.
</p>
<p class="normal">
162.

 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.
</p>
<p class="normal">
163.

 Energy efficiency

: This study looked into assessing the heating load and cooling load requirements of buildings (that is, energy efficiency) as a function of building parameters.
</p>
<p class="normal">
164.

 Entree Chicago Recommendation Data

: This data contains a record of user interactions with the Entree Chicago restaurant recommendation system.
</p>
<p class="normal">
165.

 Epileptic Seizure Recognition

: This dataset is a pre-processed and re-structured/reshaped version of a very commonly used dataset featuring epileptic seizure detection.
</p>
<p class="normal">
166.

 extention of Z-Alizadeh sani dataset

: It was collected for CAD diagnosis.
</p>
<p class="normal">
167.

 Facebook Comment Volume Dataset

: 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.
</p>
<p class="normal">
168.

 Facebook metrics

```

..--  
</b>  
: Facebook performance metrics of a renowned cosmetic's brand Facebook page.  
</p>  
<p class="normal">  
169.  
<b>  
  <a href="datasets/Farm+Ads">  
    Farm Ads  
  </a>  
</b>  
: This data was collected from text ads found on twelve websites that deal with various farm animal related topics. The binary labels are based on whether or not the content owner approves of the ad.  
</p>  
<p class="normal">  
170.  
<b>  
  <a href="datasets/Fertility">  
    Fertility  
  </a>  
</b>  
: 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  
</p>  
<p class="normal">  
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.  
</p>  
<p class="normal">  
172.  
<b>  
  <a href="datasets/First-order+theorem+proving">  
    First-order theorem proving  
  </a>  
</b>  
: Given a theorem, predict which of five heuristics will give the fastest proof when used by a first-order prover. A sixth prediction declines to attempt a proof, should the theorem be too difficult.  
</p>  
<p class="normal">  
173.  
<b>  
  <a href="datasets/Flags">  
    Flags  
  </a>  
</b>  
: From Collins Gem Guide to Flags, 1986  
</p>  
<p class="normal">  
174.  
<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, genres; play counts, favorites, comments; description, biography, tags; together with audio (343 days, 917 GiB) and features.  
</p>  
<p class="normal">  
175.  
<b>  
  <a href="datasets/Folio">  
    Folio  
  </a>  
</b>  
: 20 photos of leaves for each of 32 different species.  
</p>  
<p class="normal">  
176.

```

+ + +

 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).
</p>
<p class="normal">
177.

 Forest type mapping

: Multi-temporal remote sensing data of a forested area in Japan. The goal is to map different forest types using spectral data.
</p>
<p class="normal">
178.

 Function Finding

: Cases collected mostly from investigations in physical science; intention is to evaluate function-finding algorithms
</p>
<p class="normal">
179.

 Gas Sensor Array Drift Dataset

: This archive contains 13910 measurements from 16 chemical sensors utilized in simulations for drift compensation in a discrimination task of 6 gases at various levels of concentrations.
</p>
<p class="normal">
180.

 Gas Sensor Array Drift Dataset at Different Concentrations

: This archive contains 13910 measurements from 16 chemical sensors exposed to 6 different gases at various concentration levels.
</p>
<p class="normal">
181.

 Gas sensor array exposed to turbulent gas mixtures

: 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.
</p>
<p class="normal">
182.

 Gas sensor array under dynamic gas mixtures

: The data set contains the recordings of 16 chemical sensors exposed to two dynamic gas mixtures at varying concentrations. For each mixture, signals were acquired continuously during 12 hours.
</p>
<p class="normal">
183.

 Gas sensor array under flow modulation


```

```

~/*~

: 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.
</p>
<p class="normal">
184.

 Gas sensor arrays in open sampling settings

: The dataset contains 18000 time-series recordings from a chemical detection platform at six different locations in a wind tunnel facility in response to ten high-priority chemical gaseous substances
</p>
<p class="normal">
185.

 Gas sensors for home activity monitoring

: 100 recordings of a sensor array under different conditions in a home setting: background, wine and banana presentations. The array includes 8 MOX gas sensors, and humidity and temperature sensors.
</p>
<p class="normal">
186.

 Gastrointestinal 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 lesions and 40 adenoma.
</p>
<p class="normal">
187.

 gene expression cancer RNA-Seq

: This collection of data is part of the RNA-Seq (HiSeq) PANCAN data set, it is a random extraction of gene expressions of patients having different types of tumor: BRCA, KIRC, COAD, LUAD and PRAD.
</p>
<p class="normal">
188.

 Geo-Magnetic field and WLAN dataset for indoor localisation from wristband and smartphone

: A multisource and multivariate dataset for indoor localisation methods based on WLAN and Geo-Magnetic field fingerprinting
</p>
<p class="normal">
189.

 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 geographical origin of music.
</p>
<p class="normal">
190.

 Gesture Phase Segmentation

~/*~
```

```

~\alpha

 : 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.
</p>
<p class="normal">
191.

 Gisette

 : GISETTE is a handwritten digit recognition problem. The problem is to separate the
 highly confusable digits '4' and '9'. This dataset is one of five datasets of the NIPS 2003 featur
 e selection challenge.
</p>
<p class="normal">
192.

 Glass Identification

 : From USA Forensic Science Service; 6 types of glass; defined in terms of their oxide cor
 tency (i.e. Na, Fe, K, etc)
</p>
<p class="normal">
193.

 GNFUV Unmanned 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 i
 n Athens (Greece).
</p>
<p class="normal">
194.

 GNFUV Unmanned Surface Vehicles Sensor Data Set 2

 : The data-set contains eight (2x4) data-sets of mobile sensor readings data (humidity, te
 mperature) corresponding to a swarm of four Unmanned Surface Vehicles (USVs) in a test-bed,
 Athens, Greece.
</p>
<p class="normal">
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>
<p class="normal">
196.

 Grammatical Facial Expressions

 : This dataset supports the development of models that make possible to interpret
 Grammatical Facial Expressions from Brazilian Sign Language (Libras).
</p>
<p class="normal">
197.

 Greenhouse Gas Observing Network

 : Design an observing network to monitor emissions of a greenhouse gas (GHG) in California
 using time series of quantitative observations and traces from weather model simulations

```

given time series of synthetic observations and tracers from weather model simulations.

</p>

<p class="normal">

198.

<b>

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

Haberman's Survival

</a>

</b>

: Dataset contains cases from study conducted on the survival of patients who had undergone surgery for breast cancer

</p>

<p class="normal">

199.

<b>

<a href="datasets/Hayes-Roth">

Hayes-Roth

</a>

</b>

: Topic: human subjects study

</p>

<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 diagnosed with HCC.

</p>

<p class="normal">

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 more than 15 major health news agencies such as BBC, CNN, and NYT.

</p>

<p class="normal">

202.

<b>

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

Heart Disease

</a>

</b>

: 4 databases: Cleveland, Hungary, Switzerland, and the VA Long Beach

</p>

<p class="normal">

203.

<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>

<p class="normal">

204.

<b>

<a href="datasets/HEPMASS">

HEPMASS

</a>

</b>

: 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>

<p class="normal">

205.

<b>

<a href="datasets/Heterogeneity+Activity+Recognition">

Heterogeneity Activity Recognition

</a>

</b>

```

</p>
: The Heterogeneity Human Activity Recognition (HHAR) dataset from Smartphones and
Smartwatches is a dataset devised to benchmark human activity recognition algorithms
(classification, automatic data segmentation, sensor fusion, feature extraction, etc.) in real-
world contexts; specifically, the dataset is gathered with a variety of different device models an
d use-scenarios, in order to reflect sensing heterogeneities to be expected in real deployments.
</p>
<p class="normal">
206.

 HIGGS

: This is a classification problem to distinguish between a signal process which produces
Higgs bosons and a background process which does not.
</p>
<p class="normal">
207.

 Hill-Valley

: Each record represents 100 points on a two-dimensional graph. When plotted in order (fr
m 1 through 100) as the Y co-ordinate, the points will create either a Hill (a bump in the terra
in) or a Valley (a dip in the terrain).
</p>
<p class="normal">
208.

 HIV-1 protease cleavage

: The data contains lists of octamers (8 amino acids) and a flag (-1 or 1) depending on wh
ether HIV-1 protease will cleave in the central position (between amino acids 4 and 5).
</p>
<p class="normal">
209.

 Horse Colic

: Well documented attributes; 368 instances with 28 attributes (continuous, discrete, and
nominal); 30% missing values
</p>
<p class="normal">
210.

 HTRU2

: Pulsar candidates collected during the HTRU survey. Pulsars are a type of star, of consi
derable scientific interest. Candidates must be classified in to pulsar and non-pulsar classes to
aid discovery.
</p>
<p class="normal">
211.

 Human Activity Recognition Using Smartphones

: Human Activity Recognition database built from the recordings of 30 subjects performing
activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial
sensors.
</p>
<p class="normal">
212.

 Hybrid Indoor Positioning Dataset from WiFi RSSI, Bluetooth and magnetometer

`..`
```

```

</p>
: 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.
</p>
<p class="normal">
213.

 ICMLA 2014 Accepted Papers Data Set

: This data set compromises the metadata for the 2014 ICMLA conference's accepted papers,
including ID, paper titles, author's keywords, abstracts and sessions in which they were exposed.
</p>
<p class="normal">
214.

 ICU

: Data set prepared for the use of participants for the 1994 AAAI Spring Symposium on Art
ificial Intelligence in Medicine.
</p>
<p class="normal">
215.

 IDA2016Challenge

: The dataset consists of data collected from heavy Scania trucks in everyday usage.
</p>
<p class="normal">
216.

 ILPD (Indian Liver Patient Dataset)

: This data set contains 10 variables that are age, gender, total Bilirubin, direct
Bilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT and Alkphos.
</p>
<p class="normal">
217.

 Image Segmentation

: Image data described by high-level numeric-valued attributes, 7 classes
</p>
<p class="normal">
218.

 Immunotherapy Dataset

: This dataset contains information about wart treatment results of 90 patients using immu
notherapy.
</p>
<p class="normal">
219.

<a href="datasets/Improved+Spiral+Test+Using+Digitized+Graphics+Tablet+for+Monitoring+Parkinson%E2%80%
Disease">
 Improved Spiral Test Using Digitized Graphics Tablet for Monitoring Parkinson's Disease

: Handwriting database consists of 25 PWP(People with Parkinson) and 15 healthy individual
s.Three types of recordings (Static Spiral Test, Dynamic Spiral Test and Stability Test) are
taken.
</p>
<p class="normal">
...

```

220.  
<b>  
  <a href="datasets/Individual+household+electric+power+consumption">  
    Individual household electric 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 values are available.  
</p>  
<p class="normal">  
221.  
<b>  
  <a href="datasets/Indoor+User+Movement+Prediction+from+RSS+data">  
    Indoor User Movement Prediction from RSS data  
  </a>  
</b>  
  : 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>  
<p class="normal">  
222.  
<b>  
  <a href="datasets/Insurance+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. The data consists of 86 variables and includes product usage data and socio-demographic data  
</p>  
<p class="normal">  
223.  
<b>  
  <a href="datasets/Internet+Advertisements">  
    Internet Advertisements  
  </a>  
</b>  
  : This dataset represents a set of possible advertisements on Internet pages.  
</p>  
<p class="normal">  
224.  
<b>  
  <a href="datasets/Internet+Usage+Data">  
    Internet Usage Data  
  </a>  
</b>  
  : This data contains general demographic information on internet users in 1997.  
</p>  
<p class="normal">  
225.  
<b>  
  <a href="datasets/Ionosphere">  
    Ionosphere  
  </a>  
</b>  
  : Classification of radar returns from the ionosphere  
</p>  
<p class="normal">  
226.  
<b>  
  <a href="datasets/IPUMS+Census+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.  
</p>  
<p class="normal">  
227.  
<b>  
  <a href="datasets/Iris">  
    Iris  
  </a>  
</b>  
  : Famous database; from Fisher, 1936  
</p>

```
</p>
<p class="normal">
228.

 ISOLET

: Goal: Predict which letter-name was spoken--a simple classification task.
</p>
<p class="normal">
229.

 ISTANBUL STOCK EXCHANGE

: Data sets includes returns of Istanbul Stock Exchange with seven other international inc
ex; SP, DAX, FTSE, NIKKEI, BOVESPA, MSCE_EU, MSCI_EM from Jun 5, 2009 to Feb 22, 2011.
</p>
<p class="normal">
230.

 Japanese Credit Screening

: Includes domain theory (generated by talking to Japanese domain experts); data in Lisp
</p>
<p class="normal">
231.

 Japanese Vowels

: This dataset records 640 time series of 12 LPC cepstrum coefficients taken from nine
male speakers.
</p>
<p class="normal">
232.

 KASANDR

: 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.
</p>
<p class="normal">
233.

 KDC-4007 dataset Collection

: KDC-4007 dataset Collection is the Kurdish Documents Classification text used in categor
ies regarding Kurdish Sorani news and articles.
</p>
<p class="normal">
234.

 KDD Cup 1998 Data

: This is the data set used for The Second International Knowledge Discovery and Data
Mining Tools Competition, which was held in conjunction with KDD-98
</p>
<p class="normal">
235.

 KDD Cup 1999 Data

: 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

</p>

<p class="normal">

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. Variety of graphical features presented.

</p>

<p class="normal">

237.

<b>

<a href="datasets/KEGG+Metabolic+Relation+Network+%28Directed%29">

KEGG Metabolic Relation Network (Directed)

</a>

</b>

: KEGG Metabolic pathways modeled as directed relation network. Variety of graphical features presented.

</p>

<p class="normal">

238.

<b>

<a href="datasets/Kinship">

Kinship

</a>

</b>

: Relational dataset

</p>

<p class="normal">

239.

<b>

<a href="datasets/Labor+Relations">

Labor Relations

</a>

</b>

: From Collective Bargaining Review

</p>

<p class="normal">

240.

<b>

<a href="datasets/Las+Vegas+Strip">

Las Vegas Strip

</a>

</b>

: 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>).

</p>

<p class="normal">

241.

<b>

<a href="datasets/Leaf">

Leaf

</a>

</b>

: This dataset consists in a collection of shape and texture features extracted from digital images of leaf specimens originating from a total of 40 different plant species.

</p>

<p class="normal">

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 databases

</p>

<p class="normal">

243.

<b>

<a href="datasets/Legal+Case+Reports">

Legal Case Reports

</a>

</b>

: 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.

```
</p>
<p class="normal">
244.

 Lenses

: Database for fitting contact lenses
</p>
<p class="normal">
245.

 Letter Recognition

: Database of character image features; try to identify the letter
</p>
<p class="normal">
246.

 Libras Movement

: The data set contains 15 classes of 24 instances each. Each class references to a hand movement type in LIBRAS (Portuguese name 'Língua BRAsileira de Sinais', oficial brazilian signal language).
</p>
<p class="normal">
247.

 Liver Disorders

: BUPA Medical Research Ltd. database donated by Richard S. Forsyth
</p>
<p class="normal">
248.

 Localization Data for Person Activity

: Data contains recordings of five people performing different activities. Each person wore four sensors (tags) while performing the same scenario five times.
</p>
<p class="normal">
249.

 Logic Theorist

: All code for Logic Theorist
</p>
<p class="normal">
250.

 Low Resolution Spectrometer

: From IRAS data -- NASA Ames Research Center
</p>
<p class="normal">
251.

 LSVT Voice Rehabilitation


```

: 126 samples from 14 participants, 309 features. Aim: assess whether voice rehabilitation treatment lead to phonations considered 'acceptable' or 'unacceptable' (binary class classification problem).

</p>
 <p class="normal">  
 252.  
 <b>  
   <a href="datasets/Lung+Cancer">  
     Lung Cancer  
   </a>  
 </b>  
   : Lung cancer data; no attribute definitions  
</p>
 <p class="normal">  
 253.  
 <b>  
   <a href="datasets/Lymphography">  
     Lymphography  
   </a>  
 </b>  
   : This lymphography domain was obtained from the University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia. (Restricted access)  
</p>
 <p class="normal">  
 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, homology (similarity to other genes) and structural information, and function (if known) are provided  
</p>
 <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 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.  
</p>
 <p class="normal">  
 256.  
 <b>  
   <a href="datasets/Madelon">  
     Madelon  
   </a>  
 </b>  
   : MADEON is an artificial dataset, which was part of the NIPS 2003 feature selection challenge. This is a two-class classification problem with continuous input variables. The difficulty is that the problem is multivariate and highly non-linear.  
</p>
 <p class="normal">  
 257.  
 <b>  
   <a href="datasets/MAGIC+Gamma+Telescope">  
     MAGIC Gamma Telescope  
   </a>  
 </b>  
   : Data are MC generated to simulate registration of high energy gamma particles in an atmospheric Cherenkov telescope  
</p>
 <p class="normal">  
 258.  
 <b>  
   <a href="datasets/Mammographic+Mass">  
     Mammographic Mass  
   </a>  
 </b>  
   : Discrimination of benign and malignant mammographic masses based on BI-RADS attributes and the patient's age.  
</p>
 <p class="normal">

259.  
<b>  
  <a href="datasets/Mechanical+Analysis">  
    Mechanical Analysis  
  </a>  
</b>  
  : Fault diagnosis problem of electromechanical devices; also PUMPS DATA SET is newer version with domain theory and results  
</p>  
<p class="normal">  
260.  
<b>  
  <a href="datasets/Mesothelioma%E2%80%99s+disease+data+set+">  
    Mesothelioma's disease data set  
  </a>  
</b>  
  : 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.  
</p>  
<p class="normal">  
261.  
<b>  
  <a href="datasets/Meta-data">  
    Meta-data  
  </a>  
</b>  
  : Meta-Data was used in order to give advice about which classification method is appropriate for a particular dataset (taken from results of Statlog project).  
</p>  
<p class="normal">  
262.  
<b>  
  <a href="datasets/MEU-Mobile+KSD">  
    MEU-Mobile KSD  
  </a>  
</b>  
  : 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.  
</p>  
<p class="normal">  
263.  
<b>  
  <a href="datasets/MHEALTH+Dataset">  
    MHEALTH Dataset  
  </a>  
</b>  
  : The MHEALTH (Mobile Health) dataset is devised to benchmark techniques dealing with human behavior analysis based on multimodal body sensing.  
</p>  
<p class="normal">  
264.  
<b>  
  <a href="datasets/Mice+Protein+Expression">  
    Mice Protein Expression  
  </a>  
</b>  
  : Expression levels of 77 proteins measured in the cerebral cortex of 8 classes of control and Down syndrome mice exposed to context fear conditioning, a task used to assess associative learning.  
</p>  
<p class="normal">  
265.  
<b>  
  <a href="datasets/microblogPCU">  
    microblogPCU  
  </a>  
</b>  
  : 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.  
</p>  
<p class="normal">  
266.  
<b>  
  <a href="datasets/MicroMass">  
    MicroMass  
  </a>

```


: A dataset to explore machine learning approaches for the identification of microorganisms from mass-spectrometry data.
</p>
<p class="normal">
267.

 MiniBooNE particle identification

: This dataset is taken from the MiniBooNE experiment and is used to distinguish electron neutrinos (signal) from muon neutrinos (background).
</p>
<p class="normal">
268.

 Miskolc IIS Hybrid IPS

: 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.
</p>
<p class="normal">
269.

 Mobile Robots

: Learning concepts from sensor data of a mobile robot; set of data sets
</p>
<p class="normal">
270.

 MoCap Hand Postures

: 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.
</p>
<p class="normal">
271.

 Molecular Biology (Promoter Gene Sequences)

: E. Coli promoter gene sequences (DNA) with partial domain theory
</p>
<p class="normal">
272.

 Molecular Biology (Protein Secondary Structure)

: From CMU connectionist bench repository; Classifies secondary structure of certain globular proteins
</p>
<p class="normal">
273.

 Molecular Biology (Splice-junction Gene Sequences)

: Primate splice-junction gene sequences (DNA) with associated imperfect domain theory
</p>
<p class="normal">
274.

```

```


 MONK's Problems

: A set of three artificial domains over the same attribute space; Used to test a wide
range of induction algorithms
</p>
<p class="normal">
275.

 Moral Reasoner

: Horn-clause model that qualitatively simulates moral reasoning; Theory includes negated
literals
</p>
<p class="normal">
276.

 Motion Capture Hand Postures

: 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 co
mmon.
</p>
<p class="normal">
277.

 Movie

: This data set contains a list of over 10000 films including many older, odd, and cult
films. There is information on actors, casts, directors, producers, studios, etc.
</p>
<p class="normal">
278.

 MSNBC.com Anonymous Web Data

: 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.
</p>
<p class="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.
</p>
<p class="normal">
280.

 Multimodal Damage 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 landscape,
Infrastructural, Human, Non-damage.
</p>
<p class="normal">
281.

 Multiple Features


```

: This dataset consists of features of handwritten numerals ('0'--'9') extracted from a collection of Dutch utility maps  
 </p>  
 <p class="normal">  
 282.  
 <b>  
   <a href="datasets/Mushroom">  
 Mushroom  
   </a>  
 </b>  
 : From Audobon Society Field Guide; mushrooms described in terms of physical characteristics; classification: poisonous or edible  
 </p>  
 <p class="normal">  
 283.  
 <b>  
   <a href="datasets/Musk+Version+1">  
 Musk (Version 1)  
   </a>  
 </b>  
 : The goal is to learn to predict whether new molecules will be musks or non-musks  
 </p>  
 <p class="normal">  
 284.  
 <b>  
   <a href="datasets/Musk+Version+2">  
 Musk (Version 2)  
   </a>  
 </b>  
 : The goal is to learn to predict whether new molecules will be musks or non-musks  
 </p>  
 <p class="normal">  
 285.  
 <b>  
   <a href="datasets/News+Aggregator">  
 News Aggregator  
   </a>  
 </b>  
 : References to news pages collected from an web aggregator in the period from 10-March-2014 to 10-August-2014. The resources are grouped into clusters that represent pages discussing the same story.  
 </p>  
 <p class="normal">  
 286.  
 <b>  
   <a href="datasets/News+Popularity+in+Multiple+Social+Media+Platforms">  
 News Popularity in Multiple Social Media Platforms  
   </a>  
 </b>  
 : Large data set of news items and their respective social feedback on multiple platforms: Facebook, Google+ and LinkedIn.  
 </p>  
 <p class="normal">  
 287.  
 <b>  
   <a href="datasets/Newspaper+and+magazine+images+segmentation+dataset">  
 Newspaper and magazine images segmentation dataset  
   </a>  
 </b>  
 : 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.  
 </p>  
 <p class="normal">  
 288.  
 <b>  
   <a href="datasets/NIPS+Conference+Papers+1987-2015">  
 NIPS Conference Papers 1987-2015  
   </a>  
 </b>  
 : This data set contains the distribution of words in the full text of the NIPS conference papers published from 1987 to 2015.  
 </p>  
 <p class="normal">  
 289.  
 <b>  
   <a href="datasets/NoisyOffice">  
 NoisyOffice  
   </a>  
 </b>

```


: Corpus intended to do cleaning (or binarization) and enhancement of noisy grayscale
printed text images using supervised learning methods. Noisy images and their corresponding ground
truth provided.
</p>
<p class="normal">
290.

Nomao

: Nomao collects data about places (name, phone, localization...) from many sources.
Deduplication consists in detecting what data refer to the same place.
Instances in the dataset compare 2 spots.
</p>
<p class="normal">
291.

Northix

: Northix is designed to be a schema matching benchmark problem for data integration of
two entity relationship databases.
</p>
<p class="normal">
292.

NSF Research Award Abstracts 1990-2003

: 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 in-
dexing the bag-of-word
</p>
<p class="normal">
293.

Nursery

: Nursery Database was derived from a hierarchical decision model originally developed to
rank applications for nursery schools.
</p>
<p class="normal">
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).
</p>
<p class="normal">
295.

Occupancy Detection

: Experimental data used for binary classification (room occupancy) from Temperature, Humid-
ity, Light and CO2. Ground-truth occupancy was obtained from time stamped pictures that were taken
every minute.
</p>
<p class="normal">
296.

OCT data & Color Fundus Images of Left & Right Eyes


```

..  
 : This dataset contains OCT data (in mat format) and color fundus data (in jpg format) of left & right eyes of 50 healthy persons.  
 </p>  
 <p class="normal">  
 297.  
 <b>  
   <a href="datasets/One-hundred+plant+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 are given.  
 </p>  
 <p class="normal">  
 298.  
 <b>  
   <a href="datasets/Online+Handwritten+Assamese+Characters+Dataset">  
     Online Handwritten Assamese Characters Dataset  
   </a>  
 </b>  
 : This is a dataset of 8235 online handwritten assamese characters. The "online" process involves capturing of data as text is written on a digitizing tablet with an electronic pen.  
 </p>  
 <p class="normal">  
 299.  
 <b>  
   <a href="datasets/Online+News+Popularity">  
     Online News Popularity  
   </a>  
 </b>  
 : 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).  
 </p>  
 <p class="normal">  
 300.  
 <b>  
   <a href="datasets/Online+Retail">  
     Online Retail  
   </a>  
 </b>  
 : This is a transnational data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail.  
 </p>  
 <p class="normal">  
 301.  
 <b>  
   <a href="datasets/Online+Shoppers+Purchasing+Intention+Dataset">  
     Online Shoppers Purchasing Intention Dataset  
   </a>  
 </b>  
 : 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.  
 </p>  
 <p class="normal">  
 302.  
 <b>  
   <a href="datasets/Online+Video+Characteristics+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 characteristics along with the YouTube video ID.  
 </p>  
 <p class="normal">  
 303.  
 <b>  
   <a href="datasets/Open+University+Learning+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 seven selected courses and more than 3000 0 students.  
 </p>

```
`````
<p class="normal">
304.
<b>
<a href="datasets/Opinosis+Opinion+%26frasl%3B+Review">
Opinosis Opinion / Review
</a>
</b>
: 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".
</p>
<p class="normal">
305.
<b>
<a href="datasets/OpinRank+Review+Dataset">
OpinRank Review Dataset
</a>
</b>
: This data set contains user reviews of cars and and hotels collected from Tripadvisor (~259,000 reviews) and Edmunds (~42,230 reviews).
</p>
<p class="normal">
306.
<b>
<a href="datasets/OPPORTUNITY+Activity+Recognition">
OPPORTUNITY Activity Recognition
</a>
</b>
: 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).
</p>
<p class="normal">
307.
<b>
<a href="datasets/Optical+Interconnection+Network+">
Optical Interconnection Network
</a>
</b>
: This dataset contains 640 performance measurements from a simulation of 2-Dimensional Multiprocessor Optical Interconnection Network.
</p>
<p class="normal">
308.
<b>
<a href="datasets/Optical+Recognition+of+Handwritten+Digits">
Optical Recognition of Handwritten Digits
</a>
</b>
: Two versions of this database available; see folder
</p>
<p class="normal">
309.
<b>
<a href="datasets/Othello+Domain+Theory">
Othello Domain Theory
</a>
</b>
: Used in research to generate features for an inductive learning system
</p>
<p class="normal">
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 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.
</p>
<p class="normal">
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 c
ata extracted from biophysical simulations.
</p>
<p class="normal">
312.
<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 tha
t has been detected by a segmentation process.
</p>
<p class="normal">
313.
<b>
<a href="datasets/PAMAP2+Physical+Activity+Monitoring">
    PAMAP2 Physical Activity Monitoring
</a>
</b>
: 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.
</p>
<p class="normal">
314.
<b>
<a href="datasets/PANDOR">
    PANDOR
</a>
</b>
: PANDOR is a novel and publicly available dataset for online recommendation provided by
Purch (http://www.purch.com/).
</p>
<p class="normal">
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 computing and informatics. The task is to predict the orientation or the evaluation
of a review.
</p>
<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
Birmingham City Council. UK Open Government Licence (OGL).
https://data.birmingham.gov.uk/dataset/birmingham-parking
</p>
<p class="normal">
317.
<b>
<a href="datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet">
    Parkinson Disease Spiral Drawings Using Digitized Graphics Tablet
</a>
</b>
: Handwriting database consists of 62 PWP(People with Parkinson) and 15 healthy individual
s. Three types of recordings (Static Spiral Test, Dynamic Spiral Test and Stability Test) are
taken.
</p>
<p class="normal">
318.
<b>
<a href="datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings">
    Parkinson Speech Dataset with Multiple Types of Sound Recordings
</a>
</b>
: The training data belongs to 20 Parkinson's Disease (PD) patients and 20 healthy
subjects. From all subjects, multiple types of sound recordings (26) are taken.
</p>

```

```
~\p~  
<p class="normal">  
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).  
</p>  
<p class="normal">  
320.  
<b>  
  <a href="datasets/Parkinsons">  
    Parkinsons  
  </a>  
</b>  
  : Oxford Parkinson's Disease Detection Dataset  
</p>  
<p class="normal">  
321.  
<b>  
  <a href="datasets/Parkinsons+Telemonitoring">  
    Parkinsons Telemonitoring  
  </a>  
</b>  
  : Oxford Parkinson's Disease Telemonitoring Dataset  
</p>  
<p class="normal">  
322.  
<b>  
  <a href="datasets/PEMS-SF">  
    PEMS-SF  
  </a>  
</b>  
  : 15 months worth of daily data (440 daily records) that describes the occupancy rate, bet  
ween 0 and 1, of different car lanes of the San Francisco bay area freeways across time.  
</p>  
<p class="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 writers  
</p>  
<p class="normal">  
324.  
<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 hand  
held odor meter (OMX-GR sensor) per second for 28 seconds period.  
</p>  
<p class="normal">  
325.  
<b>  
  <a href="datasets/Phishing+Websites">  
    Phishing Websites  
  </a>  
</b>  
  : This dataset collected mainly from: PhishTank archive, MillerSmiles archive, Google's  
searching operators.  
</p>  
<p class="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  
... ~~~~~ below...
```

our paper below.

</p>

<p class="normal">

327.

Physicochemical Properties of Protein Tertiary Structure

: This is a data set of Physicochemical Properties of Protein Tertiary Structure. The data set is taken from CASP 5-9. There are 45730 decoys and size varying from 0 to 21 armstrong.

</p>

<p class="normal">

328.

Pioneer-1 Mobile Robot Data

: 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>

<p class="normal">

329.

Pittsburgh Bridges

: Bridges database that has original and numeric-discretized datasets

</p>

<p class="normal">

330.

Planning Relax

: The dataset concerns with the classification of two mental stages from recorded EEG signals: Planning (during imagination of motor act) and Relax state.

</p>

<p class="normal">

331.

Plants

: 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>

<p class="normal">

332.

PM2.5 Data of Five Chinese Cities

: 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.

</p>

<p class="normal">

333.

PMU-UD

: 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).

</p>

<p class="normal">

334.

Poker Hand

```

    Poker Hand
  </a>
</b>
: Purpose is to predict poker hands
</p>
<p class="normal">
335.
<b>
<a href="datasets/Polish+companies+bankruptcy+data">
  Polish companies bankruptcy data
</a>
</b>
: The dataset is about bankruptcy prediction of Polish companies. The bankrupt companies were analyzed in the period 2000-2012, while the still operating companies were evaluated from 2007 to 2013.
</p>
<p class="normal">
336.
<b>
<a href="datasets/Post-Operative+Patient">
  Post-Operative Patient
</a>
</b>
: Dataset of patient features
</p>
<p class="normal">
337.
<b>
<a href="datasets/Predict+keywords+activities+in+a+online+social+media">
  Predict keywords activities in a online social media
</a>
</b>
: The data from Twitter was collected during 360 consecutive days. It was done by querying 1497 English keywords sampled from Wikipedia. This dataset is proposed in a Learning to rank setting.
</p>
<p class="normal">
338.
<b>
<a href="datasets/Primary+Tumor">
  Primary Tumor
</a>
</b>
: From Ljubljana Oncology Institute
</p>
<p class="normal">
339.
<b>
<a href="datasets/Prodigy">
  Prodigy
</a>
</b>
: Assorted domains like blocksworld, eightpuzzle, and schedworld.
</p>
<p class="normal">
340.
<b>
<a href="datasets/Protein+Data">
  Protein Data
</a>
</b>
: Undocumented
</p>
<p class="normal">
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.
</p>
<p class="normal">
342.
<b>
<a href="datasets/PubChem+Bioassay+Data">
  PubChem Bioassay Data
</a>
</b>

```

PubChem Bioassay Data

: These highly imbalanced bioassay datasets are from the differing types of screening that can be performed using HTS technology. 21 datasets were created from 12 bioassays.
</p>
<p class="normal">
343.

QSAR biodegradation

: Data set containing values for 41 attributes (molecular descriptors) used to classify 1055 chemicals into 2 classes (ready and not ready biodegradable).
</p>
<p class="normal">
344.

QtyT40I10D100K

: Since there is no numerical sequential data stream available in standard data sets, this data set is generated from the original T40I10D100K data set
</p>
<p class="normal">
345.

Quadruped Mammals

: The file animals.c is a data generator of structured instances representing quadruped animals
</p>
<p class="normal">
346.

Qualitative Structure Activity Relationships

: Two sets of datasets are given: pyrimidines and triazines
</p>
<p class="normal">
347.

Qualitative_Bankruptcy

: Predict the Bankruptcy from Qualitative parameters from experts.
</p>
<p class="normal">
348.

Quality Assessment of Digital Colposcopies

: This dataset explores the subjective quality assessment of digital colposcopies.
</p>
<p class="normal">
349.

Real estate valuation data set

: 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.
</p>
<p class="normal">
350.

REALDISP Activity Recognition Dataset

 : The REALDISP dataset is devised to evaluate techniques dealing with the effects of sensor displacement in wearable activity recognition as well as to benchmark general activity recognition algorithms
 </p>
 <p class="normal">
 351.

 Record Linkage Comparison Patterns

 : Element-wise comparison of records with personal data from a record linkage setting.
 The task is to decide from a comparison pattern whether the underlying records belong to one person.
 </p>
 <p class="normal">
 352.

 Relative location of CT slices on axial axis

 : 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.
 </p>
 <p class="normal">
 353.

 Repeat Consumption Matrices

 : 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.
 </p>
 <p class="normal">
 354.

 Residential Building Data Set

 : Data set includes construction cost, sale prices, project variables, and economic variables corresponding to real estate single-family residential apartments in Tehran, Iran.
 </p>
 <p class="normal">
 355.

 Restaurant & consumer data

 : 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.
 </p>
 <p class="normal">
 356.

 Reuters RCV1 RCV2 Multilingual, Multiview Text Categorization Test collection

 : This test collection contains feature characteristics of documents originally written in five different languages and their translations, over a common set of 6 categories.
 </p>
 <p class="normal">
 357.

 Reuters Transcribed Subset

 : 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.

```
</p>
<p class="normal">
358.
<b>
<a href="datasets/Reuters-21578+Text+Categorization+Collection">
    Reuters-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.
</p>
<p class="normal">
359.
<b>
<a href="datasets/Reuter_50_50">
    Reuter_50_50
</a>
</b>
: The dataset is used for authorship identification in online Writeprint which is a new research field of pattern recognition.
</p>
<p class="normal">
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>
<p class="normal">
361.
<b>
<a href="datasets/Robot+Execution+Failures">
    Robot Execution Failures
</a>
</b>
: This dataset contains force and torque measurements on a robot after failure detection. Each failure is characterized by 15 force/torque samples collected at regular time intervals
</p>
<p class="normal">
362.
<b>
<a href="datasets/Roman+Urdu+Data+Set">
    Roman Urdu Data Set
</a>
</b>
: Roman Urdu (the scripting style for Urdu language) is one of the limited resource languages. A data corpus comprising of more than 20000 records was collected.
</p>
<p class="normal">
363.
<b>
<a href="datasets/Sales_Transactions_Dataset_Weekly">
    Sales_Transactions_Dataset_Weekly
</a>
</b>
: Contains weekly purchased quantities of 800 over products over 52 weeks. Normalised values are provided too.
</p>
<p class="normal">
364.
<b>
<a href="datasets/SCADI">
    SCADI
</a>
</b>
: First self-care activities dataset based on ICF-CY.
</p>
<p class="normal">
365.
<b>
<a href="datasets/SECOM">
    SECOM
</b>
```

```

        </a>
    </b>
    : Data from a semi-conductor manufacturing process
</p>
<p class="normal">
366.
<b>
<a href="datasets/seeds">
    seeds
</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.
</p>
<p class="normal">
367.
<b>
<a href="datasets/seismic-bumps">
    seismic-bumps
</a>
</b>
: The data describe the problem of high energy (higher than 10^4 J) seismic bumps
forecasting in a coal
mine. Data come from two of longwalls located in a Polish coal mine.
</p>
<p class="normal">
368.
<b>
<a href="datasets/Semeion+Handwritten+Digit">
    Semeion Handwritten Digit
</a>
</b>
: 1593 handwritten digits from around 80 persons were scanned, stretched in a rectangular
box 16x16 in a gray scale of 256 values.
</p>
<p class="normal">
369.
<b>
<a href="datasets/sEMG+for+Basic+Hand+movements">
    sEMG for Basic Hand movements
</a>
</b>
: The "sEMG for Basic Hand movements" includes 2 databases of surface electromyographic si
gnals of 6 hand movements using Delsys' EMG System. Healthy subjects conducted six daily life gras
ps.
</p>
<p class="normal">
370.
<b>
<a href="datasets/Sentence+Classification">
    Sentence Classification
</a>
</b>
: Contains sentences from the abstract and introduction of 30 articles annotated with a m
odified Argumentative Zones annotation scheme. These articles come from biology, machine learning
and psychology.
</p>
<p class="normal">
371.
<b>
<a href="datasets/Sentiment+Labelled+Sentences">
    Sentiment Labelled Sentences
</a>
</b>
: The dataset contains sentences labelled with positive or negative sentiment.
</p>
<p class="normal">
372.
<b>
<a href="datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machine
"%
    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-courses and reported in my Ph.D. Thesis.

```
</p>
<p class="normal">
 373.
<b>
<a href="datasets/Servo">
  Servo
</a>
</b>
 : Data was from a simulation of a servo system
</p>
<p class="normal">
 374.
<b>
<a href="datasets/SGEMM+GPU+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').
</p>
<p class="normal">
 375.
<b>
<a href="datasets/Shuttle+Landing+Control">
  Shuttle Landing Control
</a>
</b>
 : Tiny database; all nominal values
</p>
<p class="normal">
 376.
<b>
<a href="datasets/SIFT10M">
  SIFT10M
</a>
</b>
 : 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>
<p class="normal">
 377.
<b>
<a href="datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set">
  Simulated Falls and Daily Living Activities Data Set
</a>
</b>
 : 20 falls and 16 daily living activities were performed by 17 volunteers with 5 repetitic
ns while wearing 6 sensors (3.060 instances) that attached to their head, chest, waist, wrist, thi
gh and ankle.
</p>
<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, Blair, Chen, & Henrey (2013).
</p>
<p class="normal">
 379.
<b>
<a href="datasets/Skin+Segmentation">
  Skin Segmentation
</a>
</b>
 : The Skin Segmentation dataset is constructed over B, G, R color space. Skin and Nonskin
dataset is generated using skin textures from face images of diversity of age, gender, and race pe
ople.
</p>
<p class="normal">
 380.
<b>
<a
```

```
href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Living+AAL%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 improve the accuracy of our human activity recognition algorithms applied in the domain of Ambient Assisted Living.
    </p>
    <p class="normal">
        381.
        <b>
            <a href="datasets/Smartphone-Based+Recognition+of+Human+Activities+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 postural transitions while carrying a waist-mounted smartphone with embedded inertial sensors.
    </p>
    <p class="normal">
        382.
        <b>
            <a href="datasets/SML2010">
                SML2010
            </a>
        </b>
        : This dataset is collected from a monitor system mounted in a domotic house. It corresponds to approximately 40 days of monitoring data.
    </p>
    <p class="normal">
        383.
        <b>
            <a href="datasets/SMS+Spam+Collection">
                SMS Spam Collection
            </a>
        </b>
        : The SMS Spam Collection is a public set of SMS labeled messages that have been collected for mobile phone spam research.
    </p>
    <p class="normal">
        384.
        <b>
            <a href="datasets/Solar+Flare">
                Solar Flare
            </a>
        </b>
        : Each class attribute counts the number of solar flares of a certain class that occur in a 24 hour period
    </p>
    <p class="normal">
        385.
        <b>
            <a href="datasets/Somerville+Happiness+Survey">
                Somerville Happiness Survey
            </a>
        </b>
        : A data extract of a non-federal dataset posted here
https://catalog.data.gov/dataset/somerville-happiness-survey-responses-2011-2013-2015
    </p>
    <p class="normal">
        386.
        <b>
            <a href="datasets/Soybean+%28Large%29">
                Soybean (Large)
            </a>
        </b>
        : Michalski's famous soybean disease database
    </p>
    <p class="normal">
        387.
        <b>
            <a href="datasets/Soybean+%28Small%29">
                Soybean (Small)
            </a>
        </b>
```

```

    : Michalski's famous soybean disease database
</p>
<p class="normal">
 388.
<b>
  <a href="datasets/Spambase">
    Spambase
  </a>
</b>
  : Classifying Email as Spam or Non-Spam
</p>
<p class="normal">
 389.
<b>
  <a href="datasets/SPECT+Heart">
    SPECT Heart
  </a>
</b>
  : Data on cardiac Single Proton Emission Computed Tomography (SPECT) images. Each patient
classified into two categories: normal and abnormal.
</p>
<p class="normal">
 390.
<b>
  <a href="datasets/SPECTF+Heart">
    SPECTF Heart
  </a>
</b>
  : Data on cardiac Single Proton Emission Computed Tomography (SPECT) images. Each patient
classified into two categories: normal and abnormal.
</p>
<p class="normal">
 391.
<b>
  <a href="datasets/Spoken+Arabic+Digit">
    Spoken Arabic Digit
  </a>
</b>
  : This dataset contains timeseries of mel-frequency cepstrum coefficients (MFCCs)
corresponding to spoken Arabic digits. Includes data from 44 male and 44 female native Arabic spe
akers.
</p>
<p class="normal">
 392.
<b>
  <a href="datasets/Sponge">
    Sponge
  </a>
</b>
  : Data on sponges; Attributes in spanish
</p>
<p class="normal">
 393.
<b>
  <a href="datasets/Sports+articles+for+objectivity+analysis">
    Sports articles for objectivity analysis
  </a>
</b>
  : 1000 sports articles were labeled using Amazon Mechanical Turk as objective or subjectiv
e. The raw texts, extracted features, and the URLs from which the articles were retrieved are prov
ided.
</p>
<p class="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 (Credit Screening Database) in a slightly different form
</p>
<p class="normal">
 395.
<b>
  <a href="datasets/Statlog%28German+Credit+Data%29">
    Statlog (German Credit Data)

```

```

        </a>
    </b>
    : 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
    </p>
    <p class="normal">
    396.
    <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 Disease databases) but in a slightly different form
    </p>
    <p class="normal">
    397.
    <b>
        <a href="datasets/Statlog+%28Image+Segmentation%29">
            Statlog (Image Segmentation)
        </a>
    </b>
    : 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.
    </p>
    <p class="normal">
    398.
    <b>
        <a href="datasets/Statlog+%28Landsat+Satellite%29">
            Statlog (Landsat Satellite)
        </a>
    </b>
    : Multi-spectral values of pixels in 3x3 neighbourhoods in a satellite image, and the classification associated with the central pixel in each neighbourhood
    </p>
    <p class="normal">
    399.
    <b>
        <a href="datasets/Statlog+%28Shuttle%29">
            Statlog (Shuttle)
        </a>
    </b>
    : The shuttle dataset contains 9 attributes all of which are numerical. Approximately 80% of the data belongs to class 1
    </p>
    <p class="normal">
    400.
    <b>
        <a href="datasets/Statlog+%28Vehicle+Silhouettes%29">
            Statlog (Vehicle Silhouettes)
        </a>
    </b>
    : 3D objects within a 2D image by application of an ensemble of shape feature extractors to the 2D silhouettes of the objects.
    </p>
    <p class="normal">
    401.
    <b>
        <a href="datasets/Statlog+Project">
            Statlog Project
        </a>
    </b>
    : Various Databases: Vehicle silhouettes, Landsat Sattelite, Shuttle, Australian Credit Approval, Heart Disease, Image Segmentation, German Credit
    </p>
    <p class="normal">
    402.
    <b>
        <a href="datasets/Steel+Plates+Faults">
            Steel Plates Faults
        </a>
    </b>
    : A dataset of steel plates' faults, classified into 7 different types.
The goal was to train machine learning for automatic pattern recognition.
    </p>
    <p class="normal">
    403.

```

```

<b>
<a href="datasets/Stock+portfolio+performance">
  Stock portfolio performance
</a>
</b>
: The data set of performances of weighted scoring stock portfolios are obtained with
mixture design from the US stock market historical database.
</p>
<p class="normal">
  404.
<b>
<a href="datasets/StoneFlakes">
  StoneFlakes
</a>
</b>
: 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.
</p>
<p class="normal">
  405.
<b>
<a href="datasets/Student+Academics+Performance">
  Student Academics Performance
</a>
</b>
: The dataset tried to find the end semester percentage prediction based on different soci
al, economic and academic attributes.
</p>
<p class="normal">
  406.
<b>
<a href="datasets/Student+Loan+Relational">
  Student Loan Relational
</a>
</b>
: Student Loan Relational Domain
</p>
<p class="normal">
  407.
<b>
<a href="datasets/Student+Performance">
  Student Performance
</a>
</b>
: Predict student performance in secondary education (high school).
</p>
<p class="normal">
  408.
<b>
<a href="datasets/Superconductivty+Data">
  Superconductivty Data
</a>
</b>
: Two file s contain data on 21263 superconductors and their relevant features.
</p>
<p class="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.
</p>
<p class="normal">
  410.
<b>
<a href="datasets/Synthetic+Control+Chart+Time+Series">
  Synthetic Control Chart Time Series
</a>
</b>
: This data consists of synthetically generated control charts.
</p>
<p class="normal">
  411.

```

```

<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 separate subjects (Bands- recording artists; Goats; Sheep;
and BioMedical)
</p>
<p class="normal">
412.
<b>
<a href="datasets/Tamilnadu+Electricity+Board+Hourly+Readings">
  Tamilnadu Electricity Board 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
</p>
<p class="normal">
413.
<b>
<a href="datasets/Tarvel+Review+Ratings">
  Tarvel Review Ratings
</a>
</b>
: 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.
</p>
<p class="normal">
414.
<b>
<a href="datasets/Taxi+Service+Trajectory+-+Prediction+Challenge%2C+ECML+PKDD+2015">
  Taxi Service Trajectory - Prediction Challenge, ECML PKDD 2015
</a>
</b>
: An accurate dataset describing trajectories performed by all the 442 taxis running in th
e city of Porto, in Portugal.
</p>
<p class="normal">
415.
<b>
<a href="datasets/Teaching+Assistant+Evaluation">
  Teaching Assistant Evaluation
</a>
</b>
: The data consist of evaluations of teaching performance; scores are "low", "medium", or
"high"
</p>
<p class="normal">
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 a
t the four major tennis tournaments of the year 2013. Each file has 42 columns and a minimum of 76
rows.
</p>
<p class="normal">
417.
<b>
<a href="datasets/Thoracic+Surgery+Data">
  Thoracic Surgery Data
</a>
</b>
: The data is dedicated to classification problem related to the post-operative life expec
tancy in the lung cancer patients: class 1 - death within one year after surgery, class 2 - surviv
al.
</p>
<p class="normal">
418.
<b>
<a href="datasets/Thyroid+Disease">
  Thyroid Disease
</a>
</b>

```

```

</b>
: 10 separate databases from Garavan Institute
</p>
<p class="normal">
419.
<b>
<a href="datasets/Tic-Tac-Toe+Endgame">
Tic-Tac-Toe Endgame
</a>
</b>
: Binary classification task on possible configurations of tic-tac-toe game
</p>
<p class="normal">
420.
<b>
<a href="datasets/Trains">
Trains
</a>
</b>
: 2 data formats (structured, one-instance-per-line)
</p>
<p class="normal">
421.
<b>
<a href="datasets/Travel+Reviews">
Travel Reviews
</a>
</b>
: Reviews on destinations in 10 categories mentioned across East Asia. Each traveler rating is mapped as Excellent(4), Very Good(3), Average(2), Poor(1), and Terrible(0) and average rating is used.
</p>
<p class="normal">
422.
<b>
<a href="datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+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 portals in Turkey. It has 4 different forms in ARFF Weka format.
</p>
<p class="normal">
423.
<b>
<a href="datasets/Turkiye+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 questions and additional 5 attributes.
</p>
<p class="normal">
424.
<b>
<a href="datasets/TV+News+Channel+Commercial+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 Indian and 2 international news channels ( 30 Hours each).
</p>
<p class="normal">
425.
<b>
<a href="datasets/Twenty+Newsgroups">
Twenty Newsgroups
</a>
</b>
: This data set consists of 20000 messages taken from 20 newsgroups.
</p>
<p class="normal">
426.
<b>

```

```
~~
<a href="datasets/Twin+gas+sensor+arrays">
  Twin gas sensor arrays
</a>
</b>
: 5 replicates of an 8-MOX gas sensor array were exposed to different gas conditions (4 vials at 10 concentration levels each).
</p>
<p class="normal">
  427.
<b>
<a href="datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis">
  Twitter Data set for Arabic Sentiment Analysis
</a>
</b>
: 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.
</p>
<p class="normal">
  428.
<b>
<a href="datasets/UbiqLog+%28smartphone+lifelogging%29">
  UbiqLog (smartphone lifelogging)
</a>
</b>
: UbiqLog is the smartphone lifelogging tool that runs on the smartphone of 35 users for about 2 months.
</p>
<p class="normal">
  429.
<b>
<a href="datasets/UJI+Pen+Characters">
  UJI Pen Characters
</a>
</b>
: Data consists of written characters in a UNIPEN-like format
</p>
<p class="normal">
  430.
<b>
<a href="datasets/UJI+Pen+Characters+%28Version+2%29">
  UJI Pen Characters (Version 2)
</a>
</b>
: A pen-based database with more than 11k isolated handwritten characters
</p>
<p class="normal">
  431.
<b>
<a href="datasets/UJIIndoorLoc">
  UJIIndoorLoc
</a>
</b>
: The UJIIndoorLoc is a Multi-Building Multi-Floor indoor localization database to test Indoor Positioning System that rely on WLAN/WiFi fingerprint.
</p>
<p class="normal">
  432.
<b>
<a href="datasets/UJIIndoorLoc-Mag">
  UJIIndoorLoc-Mag
</a>
</b>
: The UJIIndoorLoc-Mag is an indoor localization database to test Indoor Positioning System that rely on Earth's magnetic field variations.
</p>
<p class="normal">
  433.
<b>
<a href="datasets/Ultrasonic+flowmeter+diagnostics">
  Ultrasonic flowmeter diagnostics
</a>
</b>
: Fault diagnosis of four liquid ultrasonic flowmeters
</p>
<p class="normal">
  434.
<b>
```

```
~~~  
  <a href="datasets/Undocumented">  
    Undocumented  
  </a>  
  </b>  
  : Various datasets without documentation (feel free to explore!)  
  </p>  
  <p class="normal">  
    435.  
    <b>  
      <a href="datasets/University">  
        University  
      </a>  
    </b>  
    : Data in original (LISP-readable) form  
  </p>  
  <p class="normal">  
    436.  
    <b>  
      <a href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29">  
        University of Tehran Question Dataset 2016 (UTQD.2016)  
      </a>  
    </b>  
    : Persian questions gathered from a jeopardy game broadcasted on Iranian national  
television.  
  </p>  
  <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 years.  
  </p>  
  <p class="normal">  
    438.  
    <b>  
      <a href="datasets/Urban+Land+Cover">  
        Urban Land Cover  
      </a>  
    </b>  
    : Classification of urban land cover using high resolution aerial imagery. Intended to ass  
ist sustainable urban planning efforts.  
  </p>  
  <p class="normal">  
    439.  
    <b>  
      <a href="datasets/URL+Reputation">  
        URL Reputation  
      </a>  
    </b>  
    : Anonymized 120-day subset of the ICML-09 URL data containing 2.4 million examples and  
3.2 million features.  
  </p>  
  <p class="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.  
  </p>  
  <p class="normal">  
    441.  
    <b>  
      <a href="datasets/User+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.  
  </p>  
  <p class="normal">  
    442
```

```

442.
<b>
<a href="datasets/User+Knowledge+Modeling">
  User Knowledge Modeling
</a>
</b>
: 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.
</p>
<p class="normal">
443.
<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: Fat
</a>
</b>
: Data used for USPTO Algorithm Competition. Contains drawing pages from US patents with manually labeled figure and part labels.
</p>
<p class="normal">
444.
<b>
<a href="datasets/Vertebral+Column">
  Vertebral Column
</a>
</b>
: Data set containing values for six biomechanical features used to classify orthopaedic patients into 3 classes (normal, disk hernia or spondilolisthesis) or 2 classes (normal or abnormal).
</p>
<p class="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 human activity. The data have been collected by 10 subjects using the Vicon 3D tracker.
</p>
<p class="normal">
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-exhaustive learning, in training there are 45 authors whereas, in the testing, it's 50
</p>
<p class="normal">
447.
<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 volcanoes in the large set of Venus images returned by the Magellan spacecraft.
</p>
<p class="normal">
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 direction, for 4 rounds, using 24 ultrasound sensors arranged circularly around its 'waist'.
</p>
<p class="normal">
449.
<br>

```

```

<-->
<a href="datasets/Water+Treatment+Plant">
    Water Treatment Plant
</a>
</b>
: Multiple classes predict plant state
</p>
<p class="normal">
450.
<b>
<a href="datasets/Waveform+Database+Generator+%28Version+1%29">
    Waveform Database Generator (Version 1)
</a>
</b>
: CART book's waveform domains
</p>
<p class="normal">
451.
<b>
<a href="datasets/Waveform+Database+Generator+%28Version+2%29">
    Waveform Database Generator (Version 2)
</a>
</b>
: CART book's waveform domains
</p>
<p class="normal">
452.
<b>
<a href="datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29">
    Wearable Computing: Classification of Body Postures and Movements (PUC-Rio)
</a>
</b>
: A dataset with 5 classes (sitting-down, standing-up, standing, walking, and sitting) collected on 8 hours of activities of 4 healthy subjects. We also established a baseline performance index.
</p>
<p class="normal">
453.
<b>
<a href="datasets/Website+Phishing">
    Website Phishing
</a>
</b>
:
</p>
<p class="normal">
454.
<b>
<a href="datasets/Weight+Lifting+Exercises+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 exercise. One of the variations is the one predicted by the health professional.
</p>
<p class="normal">
455.
<b>
<a href="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 stress-affect lab study, while wearing physiological and motion sensors.
</p>
<p class="normal">
456.
<b>
<a href="datasets/Wholesale+customers">
    Wholesale customers
</a>
</b>
: The data set refers to clients of a wholesale distributor. It includes the annual spending in monetary units (m.u.) on diverse product categories
</p>
<p class="normal">

```

```

451.
<b>
<a href="datasets/wiki4HE">
  wiki4HE
</a>
</b>
: Survey of faculty members from two Spanish universities on teaching uses of Wikipedia
</p>
<p class="normal">
 458.
<b>
<a href="datasets/Wilt">
  Wilt
</a>
</b>
: High-resolution Remote Sensing data set (Quickbird). Small number of training samples of
diseased trees, large number for other land cover. Testing data set from stratified random sample
of image.
</p>
<p class="normal">
 459.
<b>
<a href="datasets/Wine">
  Wine
</a>
</b>
: Using chemical analysis determine the origin of wines
</p>
<p class="normal">
 460.
<b>
<a href="datasets/Wine+Quality">
  Wine Quality
</a>
</b>
: 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/).
</p>
<p class="normal">
 461.
<b>
<a href="datasets/Wireless+Indoor+Localization">
  Wireless Indoor Localization
</a>
</b>
: Collected in indoor space by observing signal strengths of seven WiFi signals visible or
a smartphone. The decision variable is one of the four rooms.
</p>
<p class="normal">
 462.
<b>
<a href="datasets/Yacht+Hydrodynamics">
  Yacht Hydrodynamics
</a>
</b>
: Delft data set, used to predict the hydrodynamic performance of sailing yachts from dimer
sions and velocity.
</p>
<p class="normal">
 463.
<b>
<a href="datasets/YearPredictionMSD">
  YearPredictionMSD
</a>
</b>
: 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.
</p>
<p class="normal">
 464.
<b>
<a href="datasets/Yeast">
  Yeast
</a>
</b>
: Predicting the Cellular Localization Sites of Proteins
<...>

```

```

</p>
<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.
</p>
<p class="normal">
466.
<b>
<a href="datasets/YouTube+Multiview+Video+Games+Dataset">
    YouTube Multiview Video 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>
<p class="normal">
467.
<b>
<a href="datasets/YouTube+Spam+Collection">
    YouTube Spam Collection
</a>
</b>
: It is a public set of comments collected for spam research. It has five datasets composed by 1,956 real messages extracted from five videos that were among the 10 most viewed on the collection period.
</p>
<p class="normal">
468.
<b>
<a href="datasets/Z-Alizadeh+Sani">
    Z-Alizadeh Sani
</a>
</b>
: It was collected for CAD diagnosis.
</p>
<p class="normal">
469.
<b>
<a href="datasets/Zoo">
    Zoo
</a>
</b>
: Artificial, 7 classes of animals
</p>
</td>
</tr>
</table>
<hr/>
</td>
</tr>
</table>
<table align="center" cellpadding="5">
<tr valign="center">
<td>
<p class="normal">
    Supported By:
</p>
</td>
<td>

</td>
<td>
<p class="normal">
    In Collaboration With:
</p>
</td>
<td>

</td>
...

```

```
</tr>
</table>
<center>
<span class="normal">
<a href="about.html">
    About
</a>
 ||
<a href="citation_policy.html">
    Citation Policy
</a>
 ||
<a href="donation_policy.html">
    Donation Policy
</a>
 ||
<a href="contact.html">
    Contact
</a>
 ||
<a href="http://cml.ics.uci.edu">
    CML
</a>
</span>
</center>
</body>
</html>
```

In [32]:

`str(datatype)`

Out [32] :

```

">>Classification</a> <font color="red">(350)</font><br/><a href="datasets.php?
format=&task=reg&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=
">Regression</a> <font color="red">(96)</font><br/><a href="datasets.php?
format=&task=clu&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=
">Clustering</a> <font color="red">(84)</font><br/><a href="datasets.php?
format=&task=other&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=
">Other</a> <font color="red">(55)</font> </p>\n</td>\n</tr> <td bgcolor="#003366"><p clas
s="whitetext"><b>Attribute Type</b> </p>\n</td>\n</tr>\n<tr>\n<td valign="top"><p class="normal"><a href="datasets.php?
format=&task=&att=cat&area=&numAtt=&numIns=&type=&sort=nameUp&view=
">Categorical</a> <font color="red">(38)</font><br/><a href="datasets.php?
format=&task=&att=num&area=&numAtt=&numIns=&type=&sort=nameUp&view=
">Numerical</a> <font color="red">(307)</font><br/><a href="datasets.php?
format=&task=&att=mix&area=&numAtt=&numIns=&type=&sort=nameUp&view=
">Mixed</a> <font color="red">(55)</font> </p>\n</td>\n</tr> <td bgcolor="#003366"><p
class="whitetext"><b>Data Type</b> </p>\n</td>\n</tr>\n<tr>\n<td valign="top"><p class="normal"><a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=mvar&sort=nameUp&view=
">Multivariate</a> <font color="red">(357)</font><br/><a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=uvar&sort=nameUp&view=
">Univariate</a> <font color="red">(23)</font><br/><a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=seq&sort=nameUp&view=
">Sequential</a> <font color="red">(47)</font><br/><a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=ts&sort=nameUp&view=
">Time-Series</a> <font color="red">(91)</font><br/><a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=text&sort=nameUp&view=
">Text</a> <font color="red">(53)</font><br/><a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=dt&sort=nameUp&view=
">Domain-Theory</a> <font color="red">(23)</font><br/><a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=other&sort=nameUp&view=
">Other</a> <font color="red">(21)</font><br/> </p>\n</td>\n</tr>\n<tr><td bgcolor="#003366"><p
class="whitetext"><b>Area</b> </p>\n</td>\n</tr>\n<tr>\n<td valign="top"><p class="normal"><a href="datasets.php?
format=&task=&att=&area=life&numAtt=&numIns=&type=&sort=nameUp&view=
">Life Sciences</a> <font color="red">(107)</font><br/><a href="datasets.php?
format=&task=&att=&area=phys&numAtt=&numIns=&type=&sort=nameUp&view=
">Physical Sciences</a> <font color="red">(49)</font><br/><a href="datasets.php?
format=&task=&att=&area=comp&numAtt=&numIns=&type=&sort=nameUp&view=
">CS / Engineering</a> <font color="red">(170)</font><br/><a href="datasets.php?
format=&task=&att=&area=soc&numAtt=&numIns=&type=&sort=nameUp&view=
">Social Sciences</a> <font color="red">(26)</font><br/><a href="datasets.php?
format=&task=&att=&area=bus&numAtt=&numIns=&type=&sort=nameUp&view=
">Business</a> <font color="red">(29)</font><br/><a href="datasets.php?
format=&task=&att=&area=game&numAtt=&numIns=&type=&sort=nameUp&view=
">Game</a> <font color="red">(10)</font><br/><a href="datasets.php?
format=&task=&att=&area=other&numAtt=&numIns=&type=&sort=nameUp&view=
">Other</a> <font color="red">(73)</font> </p>\n</td>\n</tr>\n<tr><td bgcolor="#003366"><p
class="whitetext"><b># Attributes</b> </p>\n</td>\n</tr>\n<tr>\n<td valign="top"><p
class="normal"><a href="datasets.php?
format=&task=&att=&area=&numAtt=less10&numIns=&type=&sort=nameUp&view=
">Less than 10</a> <font color="red">(113)</font><br/><a href="datasets.php?
format=&task=&att=&area=&numAtt=10to100&numIns=&type=&sort=nameUp&view=
">10 to 100</a> <font color="red">(210)</font><br/><a href="datasets.php?
format=&task=&att=&area=&numAtt=greater100&numIns=&type=&sort=nameUp&view=
">Greater than 100</a> <font color="red">(84)</font> </p>\n</td>\n</tr>\n<tr><td
bgcolor="#003366"><p class="whitetext"><b># Instances</b></p></td>\n</tr>\n<tr>\n<td valign="top"><p class="normal"><a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=less100&type=&sort=nameUp&view=
">Less than 100</a> <font color="red">(27)</font><br/><a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=100to1000&type=&sort=nameUp&view=
">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=
">Greater than 1000</a> <font color="red">(246)</font> </p>\n</td>\n</tr>\n<tr><td
bgcolor="#003366"><p class="whitetext"><b>Format Type</b> </p>\n</td>\n</tr>\n<tr>\n<td
valign="top"><p class="normal"><a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=
">Matrix</a> <font color="red">(324)</font><br/><a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=
">Non-Matrix</a> <font color="red">(145)</font> </p>\n</td>\n</tr></table>\n</td>\n<td
valign="top">\n<table width="100%">\n<tr>\n<td><p class="big"><b>469</b> Data Sets</p></td>\n<td align="right"><p class="normal"><a href="datasets.php?
format=&task=&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=
">Table View</a>\xa0\x00<font color="gray">List View</font></p></td>\n</tr>\n</table><table
cellpadding="3">\n<tr><td><hr/><p class="normal">1. <b><a href="datasets/2.4+GHZ+Indoor+Channel+Measurements">2.4 GHZ Indoor Channel Measurements</a></b>; M
easurement 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.</p><p class="normal">2. <b><a href="datasets/2.4+GHZ+Indoor+Channel+Measurements">2.4 GHZ Indoor Channel Measurements</a></b>; M
easurement 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.</p>

```

[3D Road Network \(North Jutland, Denmark\)](datasets/3D+Road+Network%28North+Jutland%2C+Denmark%29): 3D road network with highly accurate elevation information (+-20cm) from Denmark used in eco-routing and fuel/CO₂-estimation routing algorithms.

[AAAI 2013 Accepted Papers](datasets/AAAI+2013+Accepted+Papers): 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.

[AAAI 2014 Accepted Papers](datasets/AAAI+2014+Accepted+Papers): This data set compromises the metadata for the 2014 AAAI conference's accepted papers, including paper titles, authors, abstracts, and keywords of varying granularity.

[Abalone](datasets/Abalone): Predict the age of abalone from physical measurements

[Abscisic Acid Signaling Network](datasets/Abscisic+Acid+Signaling+Network): The objective is to determine the set of boolean rules that describe the interactions of the nodes within this plant signaling network. The dataset includes 300 separate boolean pseudodynamic simulations using an asynchronous update scheme.

[Absenteeism at work](datasets/Absenteeism+at+work): The database was created with records of absenteeism at work from July 2007 to July 2010 at a courier company in Brazil.

[Activities of Daily Living \(ADLs\) Recognition Using Binary Sensors](datasets/Activities+of+Daily+Living%28ADLs%29+Recognition+Using+Binary+Sensors): This dataset comprises information regarding the ADLs performed by two users on a daily basis in their own homes.

[Activity Recognition from Single Chest-Mounted Accelerometer](datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer): The dataset collects data from a wearable accelerometer mounted on the chest. The dataset is intended for Activity Recognition research purposes.

[Activity Recognition system based on Multisensor data fusion \(AReM\)](datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion%28AReM%29): This dataset contains temporal data from a Wireless Sensor Network worn by an actor performing the activities: bending, cycling, lying down, sitting, standing, walking.

[Activity recognition with healthy older people using a batteryless wearable sensor](datasets/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.

[Acute Inflammations](datasets/Acute+Inflammations): 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.

[Adult](datasets/Adult): Predict whether income exceeds \$50K/yr based on census data. Also known as "Census Income" dataset.

[Air Quality](datasets/Air+Quality): 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.

[Air quality](datasets/Air+quality): Contains the responses of a gas multisensor device deployed on the field in an Italian city.

[Airfoil Self-Noise](datasets/Airfoil+Self-Noise): 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.

[Amazon Access Samples](datasets/Amazon+Access+Samples): Amazon's InfoSec is getting smarter about the way Access data is leveraged. This is an anonymized sample of access provisioned within the company.

[Amazon Commerce reviews set](datasets/Amazon+Commerce+reviews+set): The dataset is used for authorship identification in online Writeprint which is a new research field of pattern recognition.

[Annealing](datasets/Annealing): Steel annealing data

[Anonymous Microsoft Web Data](datasets/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.

[Anuran Calls \(MFCCs\)](datasets/Anuran+Calls%28MFCCs%29): Acoustic features extracted from syllables of anuran (frogs) calls, including the family, the genus, and the species labels (multilabel).

[Appliances energy prediction](datasets/Appiances+energy+prediction): Experimental data used to create regression models of appliances energy use in a low energy building.

[APS Failure at Scania Trucks](datasets/APS+Failure+at+Scania+Trucks): 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.

[Arcene](datasets/Arcene): ARCENE's task is to distinguish cancer versus normal patterns from mass-spectrometric data. This is a two-class classification problem with continuous input variables. This dataset is one of 5 datasets of the NIPS 2003 feature selection challenge.

[Arrhythmia](datasets/Arrhythmia): Distinguish between the presence and absence of cardiac arrhythmia and classify it in one of the 16 groups.

[Artificial Characters](datasets/Artificial+Characters): Dataset artificially generated by using first order theory which describes structure of ten capital letters of English alphabet

[Audiology \(Original\)](datasets/Audiology%28Original%29): Nominal audiology dataset from Baylor

[Audiology \(Standardized\)](datasets/Audiology%28Standardized%29): Standardized version of the original audiology database

[Audit Data](datasets/Audit+Data): 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.

[Australian Sign Language signs](datasets/Australian+Sign+Language+signs): This data c

onsists of sample of Auslan (Australian Sign Language) signs. Examples of 95 signs were collected from five signers with a total of 6650 sign samples. This data consists of sample of Auslan (Australian Sign Language) signs. 27 examples of each of 95 Auslan signs were captured from a native signer using high-quality position trackers. Autistic Spectrum Disorder Screening Data for Adult. This dataset is related to classification and predictive tasks. Autistic Spectrum Disorder Screening Data for Adolescent. This dataset is related to classification and predictive tasks. Children screening data for autism suitable for classification and predictive tasks. Revised from CMU StatLib library, data concerns city-cycle fuel consumption. Auto MPG: Revised from CMU StatLib library, data concerns city-cycle fuel consumption. Auto Univ 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. 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. 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. Bach Chorales: Time-series data based on chorales; challenge is to learn generative grammar; data in Lisp. Badges: Badges labeled with a "+" or "-" as a function of a person's name. Bag of Words: This data set contains five text collections in the form of bags-of-words. Balance Scale: Balance scale weight & distance database. Balloons: Data previously used in cognitive psychology experiment; 4 data sets represent different conditions of an experiment. 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). Banknote authentication: Data were extracted from images that were taken for the evaluation of a authentication procedure for banknotes. BAUM-1: BAUM-1 dataset contains 1184 multimodal facial video clips collected from 31 subjects. The 1184 video clips contain spontaneous facial expressions and speech of 13 emotional and mental states. BAUM-2: A multilingual audio-visual affective face database consisting of 1047 video clips of 286 subjects. Behavior of the urban traffic of the city of Sao Paulo in Brazil: The database was created with records of behavior of the urban traffic of the city of Sao Paulo in Brazil. Beijing PM2.5 Data: This hourly data set contains the PM 2.5 data of US Embassy in Beijing. Meanwhile, meteorological data from Beijing Capital International Airport are also included. Bike Sharing Dataset: 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. BLE RSSI Dataset for Indoor localization and Navigation: This dataset contains RSSI readings gathered from an array of Bluetooth Low Energy (BLE) iBeacons in a real-world and operational indoor environment for localization and navigation purposes. BlogFeedback: Instances in this dataset contain features extracted from blog posts. The task associated with the data is to predict how many comments the post will receive. BLOGGER: In this paper, we look for to recognize the causes of users tend into cyber space in Kohkiloye and Boyer Ahmad Province in Iran. Blood Transfusion Service Center: Data taken from the Blood Transfusion Service Center in Hsin-Chu City in Taiwan -- this is a classification problem. Breast Cancer Data (Restricted Access): Breast Cancer Coimbra: Clinical features were observed or measured for 64 patients with breast cancer and 52 healthy controls. Breast Cancer Wisconsin (Diagnostic): Diagnostic Wisconsin Breast Cancer Database. Breast Cancer Wisconsin (Original): Original Wisconsin Breast Cancer Database. Breast Cancer Wisconsin (Prognostic): Prognostic Wisconsin Breast Cancer Database. Breast Tissue: Dataset with electrical impedance measurements of freshly excised tissue samples from the breast. BuddyMove Data Set: User interest information extracted from the BuddyMove Data Set.

om user reviews published in holidayiq.com about various types of point of interests in South India

</p><p class="normal">63. Burst Header Packet (BHP) flooding attack on Optical Burst Switching (OBS) Network: One of the primary challenges in identifying the risks of the Burst Header Packet (BHP) flood attacks in Optical Burst Switching networks (OBS) is the scarcity of reliable historical data. </p><p class="normal">64. Buzz in social media : This data-set contains examples of buzz events from two different social networks: Twitter, and To m\’s Hardware, a forum network focusing on new technology with more conservative dynamics.</p><p class="normal">65. Caesarian Section Classification Dataset: This dataset contains information about caesarian section results of 80 pregnant women with the most important characteristics of delivery problems in the medical field.</p><p class="normal">66. CalIt2 Building People Counts: This data comes from the main door of the CalIt2 building at UCI.</p><p class="normal">67. Car Evaluation: Derived from simple hierarchical decision model, this database may be useful for testing constructive induction and structure discovery methods.</p><p class="normal">68. Carbon Nanotubes: This dataset contains 10721 initial and calculated atomic coordinates of carbon nanotubes.</p><p class="normal">69. Cardiotocography: The dataset consists of measurements of fetal heart rate (FHR) and uterine contraction (UC) features on cardiotocograms classified by expert obstetricians.</p><p class="normal">70. Cargo 2000 Freight Tracking and Tracing: 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). </p><p class="normal">71. Census Income: Predict whether income exceeds \$50K/yr based on census data. Also known as "Adult" dataset.</p><p class="normal">72. Census-Income (KDD): This data set contains weighted census data extracted from the 1994 and 1995 current population surveys conducted by the U.S. Census Bureau.</p><p class="normal">73. Cervical cancer (Risk Factors): This dataset focuses on the prediction of indicators/diagnosis of cervical cancer. The features cover demographic information, habits, and historic medical records.</p><p class="normal">74. Challenger USA Space Shuttle O-Ring: Task: predict the number of O-rings that experience thermal distress on a flight at 31 degrees F given data on the previous 23 shuttle flights</p><p class="normal">75. Character Font Images: Character images from scanned and computer generated fonts.</p><p class="normal">76. Character Trajectories: Multiple, labelled samples of pen tip trajectories recorded whilst writing individual characters. All samples are from the same writer, for the purposes of primitive extraction. Only characters with a single pen-down segment were considered.</p><p class="normal">77. Chess (Domain Theories): 6 different domain theories for generating legal moves of chess</p><p class="normal">78. Chess (King+Rook vs. King): Chess Endgame Database for White King and Rook against Black King (KRK).</p><p class="normal">79. Chess (King+Rook vs. King-Knight): Knight Pin Chess End-Game Database Creator</p><p class="normal">80. Chess (King+Rook vs. King-Pawn): King+Rook versus King+Pawn on a7 (usually abbreviated KRKP A7).</p><p class="normal">81. chestnut - LARVIC: The research project presents this database, shows the images of chestnuts that will be processed to determine the presence or absence of defects</p><p class="normal">82. chipseq: ChIP-seq experiments characterize protein modifications or binding at non-specific genomic locations in specific samples. The machine learning problem in these data is structured binary classification.</p><p class="normal">83. Chronic_Kidney_Disease: This dataset can be used to predict the chronic kidney disease and it can be collected from the hospital nearly 2 months of period.</p><p class="normal">84. Climate Model Simulation Crashes: 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.</p><p class="normal">85. Cloud: Little Documentation</p><p class="normal">86. CMU Face Images: 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</p><p class="normal">87. CNAE-9: This is a dataset containing 1080 documents of free text business descriptions of Brazilian companies categorized into a subset of 9 categories</p><p class="normal">88. Coil 1999 Competition Data: This data set is from the 1999 Computational Intelligence and Learning (COIL) competition. The data contains measurements of river chemical concentrations and algae densities.</p><p class="normal">89. Combined Cycle Power Plant: The dataset contains 9568 data points collected from a Combined Cycle Power Plant over 6 years (2006-2011), when the plant was set to work with full load. </p><p class="normal">90. Communities and Crime: Communities within the United States. The data combines socio-economic data from the 1990 US Census, law enforcement data from the 1990 US LEMAS survey, and crime data from the 1995 FBI UCR.</p><p class="normal">91. Connexions Dataset: This dataset contains 10000 news articles from the Connexions website, which is a collection of open access educational resources. The dataset includes titles, URLs, and abstracts of the articles.

= "datasets/Communities+and+Crime+Unnormalized" > Communities and Crime Unnormalized :
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</p><p class="normal">92. Computer Hardware: Relative CPU Performance Data, described in terms of its cycle time, memory size, etc.</p><p class="normal">93. Concrete Compressive Strength: Concrete is the most important material in civil engineering. The concrete compressive strength is a highly nonlinear function of age and ingredients. </p><p class="normal">94. Concrete Slump Test: Concrete is a highly complex material. The slump flow of concrete is not only determined by the water content, but that is also influenced by other concrete ingredients.</p><p class="normal">95. Condition Based Maintenance of Naval Propulsion Plants: Data have been generated from a sophisticated simulator of a G as Turbines (GT), mounted on a Frigate characterized by a COmbed Diesel eLectric And Gas (CODLAG) propulsion plant type.</p><p class="normal">96. Condition monitoring of hydraulic systems: The data set addresses the condition assessment of a hydraulic test rig based on multi sensor data. Four fault types are superimposed with several severity grades impeding selective quantification.</p><p class="normal">97. Congressional Voting Records: 1984 United States Congressional Voting Records; Classify as Republican or Democrat</p><p class="normal">98. Connect-4: Contains connect-4 positions</p><p class="normal">99. Connectionist Bench (Nettalk Corpus): 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</p><p class="normal">100. Connectionist Bench (Sonar, Mines vs. Rocks): 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.</p><p class="normal">101. Connectionist Bench (Vowel Recognition - Deterding Data): Speaker independent recognition of the eleven steady state vowels of British English using a specified training set of lpc derived log area ratios.</p><p class="normal">102. Container Crane Controller Data Set: A container crane has the function of transporting containers from one point to another point.</p><p class="normal">103. Contraceptive Method Choice: Dataset is a subset of the 1987 National Indonesia Contraceptive Prevalence Survey.</p><p class="normal">104. Corel Image 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-occurrence</p><p class="normal">105. Covertype: Forest CoverType dataset</p><p class="normal">106. Credit Approval: This data concerns credit card applications; good mix of attributes</p><p class="normal">107. Crowdsourced Mapping: Crowdsourced data from OpenStreetMap is used to automate the classification of satellite images into different land cover classes (impervious, farm, forest, grass, orchard, water). </p><p class="normal">108. Cryotherapy Dataset: This dataset contains information about wart treatment results of 90 patients using cryotherapy.</p><p class="normal">109. CSM (Conventional and Social Media Movies) Dataset 2014 and 2015: 12 features categorized as conventional and social media features. Both conventional features, collected from movies databases on Web as well as social media features (YouTube, Twitter).</p><p class="normal">110. 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.</p><p class="normal">111. Cylinder Bands: Used in decision tree induction for mitigating process delays known as "cylinder bands" in rotogravure printing</p><p class="normal">112. Daily and Sports Activities: The dataset comprises motion sensor data of 19 daily and sports activities each performed by 8 subjects in their own style for 5 minutes. Five Xsens MTx units are used on the torso, arms, and legs.\r\n</p><p class="normal">113. Daily Demand Forecasting Orders: The dataset was collected during 60 days, this is a real database of a Brazilian logistics company.</p><p class="normal">114. Daphnet Freezing of Gait: This dataset contains the annotated readings of 3 acceleration sensors at the hip and leg of Parkinson's disease patients that experience freezing of gait (FoG) during walking tasks.\r\n</p><p class="normal">115. 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 teams during the creation of final class project in SW Eng. classes at SFSU, Fulda, FAU</p><p class="normal">116. Dataset for ADL Recognition with Wrist-worn Accelerometer: Recordings of 16 volunteers performing 14 Activities of Daily Living (ADL) while carrying a single wrist-worn tri-axial accelerometer.</p><p class="normal">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. </p><p class="normal">118. DBWorld e-mails: It contains 64 e-mails which I have manually collected from DBWorld mailing list. They are classified in: \'announces of conferences\' and \'everything else\'.</p><p class="normal">119. default of credit card clients: 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.</p><p class="normal">120. DeliciousMIL: A Data Set for Multi-Label Multi-Instance Learning with Instance Labels: This dataset includes 1) 12234 documents (8251 training, 3983 test) extracted from DeliciousT140 dataset, 2) class labels for all documents, 3) labels for a subset of sentences of the test documents.</p><p class="normal">121. Demospongiae: Marine sponges of the Demospongiae class classification domain.</p><p class="normal">122. Dermatology: Aim for this dataset is to determine the type of Eryhemato-Squamous Disease.</p><p class="normal">123. Detect Malicious Executable(AntiVirus): I extract features from malicious and non-malicious and create and training dataset to teach svm classifier. Dataset made of unknown executable to detect if it is virus or normal safe executable.</p><p class="normal">124. detection_of_IoT_botnet_attacks_N_BaIoT: This dataset addresses the lack of public botnet datasets, especially for the IoT. It suggests *real* traffic data, gathered from 9 commercial IoT devices authentically infected by Mirai and BASHLITE.</p><p class="normal">125. Devanagari Handwritten Character Dataset: This is an image database of Handwritten Devanagari characters. There are 46 classes of characters with 2000 examples each. The dataset is split into training set(85%) and testing set(15%). </p><p class="normal">126. Dexter: DEXTER is a text classification problem in a bag-of-word representation. This is a two-class classification problem with sparse continuous input variables. This dataset is one of five datasets of the NIPS 2003 feature selection challenge.<r/n><p class="normal">127. DGP2 - The Second Data Generation Program: Generates application domains based on specific parameters, number of features, and proportion of positive to negative examples.</p><p class="normal">128. Diabetes: This diabetes dataset is from AIM \'94.</p><p class="normal">129. Diabetes 130-US hospitals for years 1999-2008: This data has been prepared to analyze factors related to readmission as well as other r\nr\noutcomes pertaining to patients with diabetes.</p><p class="normal">130. Diabetic Retinopathy Debrecen Data Set: This dataset contains features extracted from the Messidor image set to predict whether an image contains signs of diabetic retinopathy or not. </p><p class="normal">131. Discrete Tone Image Dataset: 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. Dishonest Internet users Dataset: The dataset was used to test an architecture based on a trust model capable to cope with the evaluation of the trustworthiness of users interacting in pervasive environments.</p><p class="normal">133. Document Understanding: Five concepts, expressed as predicates, to be learned</p><p class="normal">134. Dodgers Loop Sensor: Loop sensor data was collected for the Glendale on ramp for the 101 North freeway in Los Angeles.</p><p class="normal">135. 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 challenge.</p><p class="normal">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.</p><p class="normal">137. Dow Jones Index: This dataset contains weekly data for the Dow Jones Industrial Index. It has been used in computational investing research.</p><p class="normal">138. Dresses_Attribute_Sales: This dataset contains Attributes of dresses and their recommendations according to their sales. Sales are monitored on the basis of alternate days. </p><p class="normal">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 drivers with several facial features.</p><p class="normal">140. Drug consumption (quantified): Classify type of drug consumer by personality data</p><p class="normal">141. Drug Review Dataset (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><p class="normal">142. Drug Review Dataset (Drugs.com): The dataset provides patient reviews on specific drugs along with related conditions and a 10 star patient rating reflecting overall patient satisfaction.</p><p class="normal">143. DSRC Vehicle Communications: This set provides data regarding wireless communications between vehicles and road side units. Two separate data sets are provided (normal scenario) and in the presence of attacker (jammer).</p><p class="normal">144. Dynamic Features of VirusShare: This set provides data regarding wireless communications between vehicles and road side units. Two separate data sets are provided (normal scenario) and in the presence of attacker (jammer).</p>

Executables: This dataset contains the dynamic features of 107,888 executables, collected by VirusShare from Nov/2010 to Jul/2014.</p><p class="normal">145. E. Coli Genes: Data giving characteristics of each ORF (potential gene) in the E. coli genome. Sequence, homology (similarity to other genes) and structural information, and function (if known) are provided.</p><p class="normal">146. Early biomarkers of Parkinson's disease based on natural connected speech Data Set : .</p><p class="normal">147. Early biomarkers of Parkinson's disease based on natural connected speech: 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.</p><p class="normal">148. EBL Domain Theories: Assorted small-scale domain theories</p><p class="normal">149. Echocardiogram: Data for classifying if patients will survive for at least one year after a heart attack</p><p class="normal">150. Eco-hotel: This dataset includes Online Textual Reviews from both online (e.g., Trip Advisor) and offline (e.g., Guests' book) sources from the Areias do Seixo Eco-Resort.</p><p class="normal">151. Ecoli: This data contains protein localization sites</p><p class="normal">152. Economic Sanctions: Domain Theory on Economic Sanctions; Undocumented</p><p class="normal">153. 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.</p><p class="normal">154. EEG Database: 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</p><p class="normal">155. EEG Eye State: The data set consists of 14 EEG values and a value indicating the eye state.</p><p class="normal">156. EEG Steady-State Visual Evoked Potential Signals: This database consists on 30 subjects performing Brain Computer Interface for Steady State Visual Evoked Potentials (BCI-SSVEP). </p><p class="normal">157. El Nino: The data set contains oceanographic and surface meteorological readings taken from a series of buoys positioned throughout the equatorial Pacific.</p><p class="normal">158. Electrical Grid Stability Simulated Data : The local stability analysis of the 4-node star system (electricity producer is in the center) implementing Decentral Smart Grid Control concept. </p><p class="normal">159. ElectricityLoadDiagrams20112014: This data set contains electricity consumption of 370 points/clients.\r\n</p><p class="normal">160. EMG data for gestures: These are files of raw EMG data recorded by MYO Thalmic bracelet</p><p class="normal">161. EMG dataset in Lower Limb: 3 different exercises: sitting, standing and walking in the muscles: biceps femoris, vastus medialis, rectus femoris and semitendinosus addition to goniometry in the exercises.</p><p class="normal">162. 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.</p><p class="normal">163. Energy efficiency: This study looked into assessing the heating load and cooling load requirements of buildings (that is, energy efficiency) as a function of building parameters.</p><p class="normal">164. Entree Chicago Recommendation Data: This data contains a record of user interactions with the Entree Chicago restaurant recommendation system.</p><p class="normal">165. Epileptic Seizure Recognition: This dataset is a pre-processed and re-structured/reshaped version of a very commonly used dataset featuring epileptic seizure detection. </p><p class="normal">166. extention of Z-Alizadeh sani dataset: It was collected for CAD diagnosis.</p><p class="normal">167. Facebook Comment Volume Dataset: 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.</p><p class="normal">168. Facebook metrics: Facebook performance metrics of a renowned cosmetic's brand Facebook page.</p><p class="normal">169. Farm Ads: This data was collected from text ads found on twelve websites that deal with various farm animal related topics. The binary labels are based on whether or not the content owner approves of the ad.</p><p class="normal">170. Fertility: 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</p><p class="normal">171. Firm-Teacher_Clave-Direction_Classification: The data are binary attack-point vectors and their clave-direction class(es) according to the partido-alto-based paradigm.</p><p class="normal">172. First-order theorem proving: Given a theorem, predict which of five heuristics will give the fastest proof when used by a first-order prover. A sixth prediction declines to attempt a proof, should the theorem be too difficult.</p><p class="normal">173. Flags: From Collins Gem Guide to Flags, 1986</p><p class="normal">174. FMA: A Data set For Music Analysis: FMA features 106,574 tracks and includes song title, album, artist

for music analysis, etc.). This features 100,000 tracks and includes song titles, album, artist, genres; play counts, favorites, comments; description, biography, tags; together with audio (343 days, 917 GiB) and features. 175. [Folio](#): 20 photos of leaves for each of 32 different species. 176. [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. [Forest type mapping](#): Multi-temporal remote sensing data of a forested area in Japan. The goal is to map different forest types using spectral data. 178. [Function+Finding](#): Function Finding: Cases collected mostly from investigations in physical science; intention is to evaluate function-finding algorithms. 179. [Gas Sensor Array Drift Dataset](#): This archive contains 13910 measurements from 16 chemical sensors utilized in simulations for drift compensation in a discrimination task of 6 gases at various levels of concentrations. 180. [Gas Sensor Array Drift Dataset at Different Concentrations](#): This archive contains 13910 measurements from 16 chemical sensors exposed to 6 different gases at various concentration levels. 181. [Gas sensor array exposed to turbulent gas mixtures](#): A chemical detection platform composed of 8 chemoresistive gas sensors was exposed to turbulent gas mixtures generated naturally in a wind tunnel. The acquired time series of the sensors are provided. 182. [Gas sensor array under dynamic gas mixtures](#): The data set contains the recordings of 16 chemical sensors exposed to two dynamic gas mixtures at varying concentrations. For each mixture, signals were acquired continuously during 12 hours. 183. [Gas sensor array under flow modulation](#): The data set contains 58 time series acquired from 16 chemical sensors under gas flow modulation conditions. The sensors were exposed to different gaseous binary mixtures of acetone and ethanol. 184. [Gas sensor arrays in open sampling settings](#): The dataset contains 18000 time-series recordings from a chemical detection platform at six different locations in a wind tunnel facility in response to ten high-priority chemical gaseous substances. 185. [Gas sensors for home activity monitoring](#): 100 recordings of a sensor array under different conditions in a home setting: background, wine and banana presentations. The array includes 8 MOX gas sensors, and humidity and temperature sensors. 186. [Gastrointestinal 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 lesions and 40 adenoma. 187. [Gene expression cancer RNA-Seq](#): This collection of data is part of the RNA-Seq (HiSeq) PANCAN data set, it is a random extraction of gene expressions of patients having different types of tumor: BRCA, KIRC, COAD, LUAD and PRAD. 188. [Geo-Magnetic field and WLAN dataset for indoor localisation from wristband and smartphone](#): A multivariate and multivariate dataset for indoor localisation methods based on WLAN and Geo-Magnetic field fingerprinting. 189. [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 geographical origin of music. 190. [Gesture Phase Segmentation](#): The dataset is composed by features extracted from 7 videos with people gesticulating, aiming at studying Gesture Phase Segmentation. It contains 50 attributes divided into two files for each video. 191. [Gisette](#): GISETTE is a handwritten digit recognition problem. The problem is to separate the highly confusable digits '\4\' and '\9\. This dataset is one of five datasets of the NIPS 2003 feature selection challenge. 192. [Glass Identification](#): From USA Forensic Science Service; 6 types of glass; defined in terms of their oxide content (i.e. Na, Fe, K, etc). 193. [GNFUV Unmanned 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. [GNFUV Unmanned Surface Vehicles Sensor Data Set 2](#): The data-set contains eight (2x4) data-sets of mobile sensor readings data (humidity, temperature) corresponding to a swarm of four Unmanned Surface Vehicles (USVs) in a test-bed, Athens, Greece. 195. [GPS Trajectories](#): The dataset has been feed by Android app called Go!Track. It is available at Google Play Store(<https://play.google.com/store/apps/details?id=com.go.router>). 196. [Grammatical Facial Expressions](#): This dataset supports the development of models that make possible to interpret Grammatical Facial Expressions from Brazilian Sign Language (Libras). 197. [Greenhouse Gas Observing Network](#): Design an observing network to monitor emissions of a greenhouse gas (CH₄) in California given time

design an observing network to monitor emissions of a greenhouse gas (CH₄) in California given time series of synthetic observations and tracers from weather model simulations.\r\n</p><p class="normal">198. Haberman's Survival: Data set contains cases from study conducted on the survival of patients who had undergone surgery for breast cancer</p><p class="normal">199. Hayes-Roth: Topic : human subjects study</p><p class="normal">200. HCC Survival: Hepatocellular Carcinoma dataset (HCC dataset) was collected at a University Hospital in Portugal. It contains real clinical data of 165 patients diagnosed with HCC.</p><p class="normal">201. 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><p class="normal">202. Heart Disease: 4 databases: Cleveland, Hungary, Switzerland, and the VA Long Beach</p><p class="normal">203. Hepatitis: From G.Gong: CMU; Mostly Boolean or numeric-valued attribute types; Includes cost data (donated by Peter Turney)</p><p class="normal">204. HEPMASS: 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><p class="normal">205. Heterogeneity Activity Recognition: The Heterogeneity Human Activity Recognition (HHAR) dataset from Smartphones and Smartwatches is a dataset devised to benchmark human activity recognition algorithms (classification, automatic data segmentation, sensor fusion, feature extraction, etc.) in real-world contexts; specifically, the dataset is gathered with a variety of different device models and use-scenarios, in order to reflect sensing heterogeneities to be expected in real deployments.</p><p class="normal">206. HIGGS: This is a classification problem to distinguish between a signal process which produces Higgs bosons and a background process which does not. </p><p class="normal">207. Hill-Valley: Each record represents 100 points on a two-dimensional graph. When plotted in order (from 1 through 100) as the Y co-ordinate, the points will create either a Hill (a bump in the terrain) or a Valley (a dip in the terrain).</p><p class="normal">208. HIV-1 protease cleavage: 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><p class="normal">209. Horse Colic: Well documented attributes; 368 instances with 28 attributes (continuous, discrete, and nominal); 30% missing values</p><p class="normal">210. HTRU2: Pulsar candidates collected during the HTRU survey. Pulsars are a type of star, of considerable scientific interest. Candidates must be classified into pulsar and non-pulsar classes to aid discovery.</p><p class="normal">211. Human Activity Recognition Using Smartphones: Human Activity Recognition database built from the recordings of 30 subjects performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors.</p><p class="normal">212. Hybrid Indoor 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 WLAN and Bluetooth interfaces, and Magnetometer. </p><p class="normal">213. ICMLA 2014 Accepted Papers Data Set: This data set compromises the metadata for the 2014 ICMLA conference's accepted papers, including ID, paper titles, author's keywords, abstracts and sessions in which they were exposed.</p><p class="normal">214. ICU: Data set prepared for the use of participants for the 1994 AAAI Spring Symposium on Artificial Intelligence in Medicine.</p><p class="normal">215. IDA2016Challenge: The dataset consists of data collected from heavy Scania trucks in everyday usage. </p><p class="normal">216. ILPD (Indian Liver Patient Dataset): This data set contains 10 variables that are age, gender, total Bilirubin, direct Bilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT and Alkphos.</p><p class="normal">217. Image Segmentation: Image data described by high-level numeric-valued attributes, 7 classes</p><p class="normal">218. Immunotherapy Dataset: This dataset contains information about wart treatment results of 90 patients using immunotherapy.</p><p class="normal">219. Improved Spiral Test Using Digitized Graphics Tablet for Monitoring Parkinson's Disease: 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.</p><p class="normal">220. Individual household electric power consumption: 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.</p><p class="normal">221. Indoor User Movement Prediction from RSS data: 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><p class="normal">222. Insurance Company Benchmark (COIL 2000): This data set used in the CoIL 2000 Challenge contains information on customers of an insurance company. The data consists of 86 variables and includes product usage data and socio-demographic data</p><p class="normal">223. Internet Advertisements: This dataset represents a set of possible advertisements on Internet pages.

advertisements/. This dataset represents a set of possible advertisements on internet pages.

</p><p class="normal">224. Internet Usage Data: This data contains general demographic information on internet users in 1997.</p><p class="normal">225. Ionosphere: Classification of radar returns from the ionosphere</p><p class="normal">226. IPUMS Census Database: This data set contains unweighted PUMS census data from the Los Angeles and Long Beach areas for the years 1970, 1980, and 1990.</p><p class="normal">227. Iris: Famous database; from Fisher, 1936</p><p class="normal">228. ISOLET: Goal: Predict which letter-name was spoken--a simple classification task.</p><p class="normal">229. ISTANBUL STOCK EXCHANGE: 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.</p><p class="normal">230. Japanese Credit Screening: Includes domain theory (generated by talking to Japanese domain experts); data in Lisp</p><p class="normal">231. Japanese Vowels: This dataset records 640 time series of 12 LPC cepstrum coefficients taken from nine male speakers.</p><p class="normal">232. KASANDR: 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.</p><p class="normal">233. KDC-4007 dataset Collection: KDC-4007 dataset Collection is the Kurdish Documents Classification text used in categories regarding Kurdish Sorani news and articles.</p><p class="normal">234. KDD Cup 1998 Data: This is the data set used for The Second International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-98</p><p class="normal">235. KDD Cup 1999 Data: 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</p><p class="normal">236. KEGG Metabolic Reaction Network (Undirected): KEGG Metabolic pathways modeled as un-directed reaction network. Variety of graphical features presented.</p><p class="normal">237. KEGG Metabolic Relation Network (Directed): KEGG Metabolic pathways modeled as directed relation network. Variety of graphical features presented.</p><p class="normal">238. Kinship: Relational dataset</p><p class="normal">239. Labor Relations: From Collective Bargaining Review</p><p class="normal">240. Las Vegas Strip: 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>).</p><p class="normal">241. Leaf: This dataset consists in a collection of shape and texture features extracted from digital images of leaf specimens originating from a total of 40 different plant species.</p><p class="normal">242. LED Display Domain: From Classification and Regression Trees book; We provide here 2 C programs for generating sample databases</p><p class="normal">243. Legal Case Reports: 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.</p><p class="normal">244. Lenses: Database for fitting contact lenses</p><p class="normal">245. Letter Recognition: Database of character image features; try to identify the letter</p><p class="normal">246. Libras Movement: The data set contains 15 classes of 24 instances each. Each class references to a hand movement type in LIBRAS (Portuguese\r\nname \'Língua Brasileira de Sinais\', oficial brazilian signal language).</p><p class="normal">247. Liver Disorders: BUPA Medical Research Ltd. database donated by Richard S. Forsyth</p><p class="normal">248. Localization Data for Person Activity: Data contains recordings of five people performing different activities. Each person wore four sensors (tags) while performing the same scenario five times.</p><p class="normal">249. Logic Theorist: All code for Logic Theorist</p><p class="normal">250. Low Resolution Spectrometer: From IRAS data -- NASA Ames Research Center</p><p class="normal">251. LSVT Voice Rehabilitation: 126 samples from 14 participants, 309 features. Aim: assess whether voice rehabilitation treatment lead to phonations considered \'acceptable\' or \'unacceptable\' (binary class classification problem).</p><p class="normal">252. Lung Cancer: Lung cancer data; no attribute definitions</p><p class="normal">253. Lymphography: This lymphography domain was obtained from the University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia. (Restricted access)</p><p class="normal">254. 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, and function (if known) are provided</p><p class="normal">255. Machine Learning based ZZAlpha Ltd. Stock Recommendations 2012-2014: 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. </p><p class="normal">256. Madelon: MADELON is an artificial dataset, which was part of the NIPS 2003 feature selection challenge. This is a two-class classification problem with continuous i

IPS 2005 feature selection challenge. This is a two-class classification problem with continuous input variables. The difficulty is that the problem is multivariate and highly non-linear.

MAGIC Gamma Telescope: Data are MC generated to simulate registration of high energy gamma particles in an atmospheric Cherenkov telescope.

Mammographic Mass: Discrimination of benign and malignant mammographic masses based on BI-RADS attributes and the patient's age.

Mechanical Analysis: Fault diagnosis problem of electromechanical devices; also PUMPS DATA SET is newer version with domain theory and results.

Mesothelioma's disease: Mesothelioma's disease data set were prepared at Dicle University Faculty of Medicine in Turkey.

MHEALTH Dataset: The MHEALTH (Mobile Health) dataset is devised to benchmark techniques dealing with human behavior analysis based on multimodal body sensing.

Mice Protein Expression: Expression levels of 77 proteins measured in the cerebral cortex of 8 classes of control and Down syndrome mice exposed to context fear conditioning, a task used to assess associative learning.

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.

MicroMass: A dataset to explore machine learning approaches for the identification of microorganisms from mass-spectrometry data.

MiniBooNE particle identification: This dataset is taken from the MiniBooNE experiment and is used to distinguish electron neutrinos (signal) from muon neutrinos (background).

Miskolc IIS Hybrid IPS: 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.

Mobile Robots: Learning concepts from sensor data of a mobile robot; set of data sets.

MoCap Hand Postures: 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.

Molecular Biology (Promoter Gene Sequences): E. Coli promoter gene sequences (DNA) with partial domain theory.

Molecular Biology (Protein Secondary Structure): From CMU connectionist bench repository; Classifies secondary structure of certain globular proteins.

Molecular Biology (Splice-junction Gene Sequences): Primate splice-junction gene sequences (DNA) with associated imperfect domain theory.

MONK's Problems: A set of three artificial domains over the same attribute space; Used to test a wide range of induction algorithms.

Moral Reasoner: Horn-clause model that qualitatively simulates moral reasoning; Theory includes negated literals.

Motion Capture Hand Postures: 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.

Movie: 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.

MSNBC.com Anonymous Web Data: 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.

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.

Multimodal Damage 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 landscape, Infrastructural, Human, Non-damage.

Multiple Features: This dataset consists of features of handwritten numerals ('0'--'9') extracted from a collection of Dutch utility maps.

Mushroom: From Audobon Society Field Guide; mushrooms described in terms of physical characteristics; classification: poisonous or edible.

Musk (Version 1): The goal is to learn to predict whether new molecules will be musks or non-musks.

Musk (Version 2): The goal

is to learn to predict whether new molecules will be musks or non-musks

News Aggregator: References to news pages collected from an web aggregator in the period from 10-March-2014 to 10-August-2014. The resources are grouped into clusters that represent pages discussing the same story.

</p><p class="normal">286. News Popularity in Multiple Social Media Platforms: Large data set of news items and their respective social feedback on multiple platforms: Facebook, Google+ and LinkedIn.

</p><p class="normal">287. Newspaper and magazine images segmentation dataset: 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.

</p><p class="normal">288. NIPS 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><p class="normal">289. NoisyOffice: Corpus intended to do cleaning (or binarization) and enhancement of noisy grayscale printed text images using supervised learning methods. Noisy images and their corresponding ground truth provided.

</p><p class="normal">290. Nomao: Nomao collects data about places (name, phone, localization...) from many sources.

\r\nDeduplication consists in detecting what data refer to the same place.

\r\nInstances in the dataset compare 2 spots.

</p><p class="normal">291. Northix: Northix is designed to be a schema matching benchmark problem for data integration of two entity relationship databases.

</p><p class="normal">292. NSF Research Award Abstracts 1990-2003: This data set consists of (a) 129,000 abstracts describing NSF awards for basic research, (b) bag-of-word data files extracted from the abstracts, (c) a list of words used for indexing the bag-of-word

</p><p class="normal">293. Nursery: Nursery Database was derived from a hierarchical decision model originally developed to rank applications for nursery schools.

</p><p class="normal">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).

</p><p class="normal">295. Occupancy Detection : Experimental data used for binary classification (room occupancy) from Temperature, Humidity, Light and CO₂. Ground-truth occupancy was obtained from time stamped pictures that were taken every minute.

</p><p class="normal">296. OCT data & Color Fundus Images of Left & Right Eyes: This dataset contains OCT data (in mat format) and color fundus data (in jpg format) of left & right eyes of 50 healthy persons.

</p><p class="normal">297. One-hundred plant species leaves data set: Sixteen samples of leaf each of one-hundred plant species. For each sample, a shape descriptor, fine scale margin and texture histogram are given.

</p><p class="normal">298. Online Handwritten Assamese Characters Dataset: This is a dataset of 8235 online handwritten assamese characters. The "online" process involves capturing of data as text is written on a digitizing tablet with an electronic pen.

</p><p class="normal">299. Online News Popularity: 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).

</p><p class="normal">300. Online Retail: This is a transnational data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail.

</p><p class="normal">301. Online Shoppers Purchasing Intention Dataset: Of the 12,330 sessions in the dataset, \r\n84.5% (10,422) were negative class samples that did not \r\nend with shopping, and the rest (1908) were positive class \r\nsamples ending with shopping.

</p><p class="normal">302. 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.

</p><p class="normal">303. Open University Learning Analytics dataset: Open University Learning Analytics Dataset contains data about courses, students and their interactions with Virtual Learning Environment for seven selected courses and more than 30000 students.

</p><p class="normal">304. 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 nano".

</p><p class="normal">305. OpinRank Review Dataset: This data set contains user reviews of cars and hotels collected from TripAdvisor (~259,000 \r\nreviews) and Edmunds (~42,230 reviews).

</p><p class="normal">306. OPPORTUNITY Activity Recognition: 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).

</p><p class="normal">307. Optical Interconnection Network : This dataset contains 640 performance measurements from a simulation of 2-Dimensional Multiprocessor Optical Interconnection Network.

</p><p class="normal">308. Optical Recognition of Handwritten Digits: Two versions of this database available; see folder

</p><p class="normal">309. Othello Domain Theory: Used in research to generate features for an inductive learning system

</p><p class="normal">310. Otter: A dataset of 1000000 chemical structures and their properties.

[Uzone Level Detection](#): Two ground ozone level data sets are included in this collection. One is the eight hour peak set (`eighthr.data`), the other is the one hour peak set (`onehr.data`). Those data were collected from 1998 to 2004 at the Houston, Galveston and Brazoria area.

[p53 Mutants](#): The goal is to model mutant p53 transcriptional activity (active vs inactive) based on data extracted from biophysical simulations.

[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.

[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.

[PANDOR](#): PANDOR is a novel and publicly available dataset for online recommendation provided by Purch (<http://www.purch.com/>).

[Paper Reviews](#): This sentiment analysis data set contains scientific paper reviews from an international conference on computing and informatics. The task is to predict the orientation or the evaluation of a review.

[Parking Birmingham](#): Data collected from car parks in Birmingham that are operated by NCP from Birmingham City Council. UK Open Government Licence (OGL).

[Parkinson Disease Spiral Drawings Using Digitized Graphics Tablet](#): 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.

[Parkinson Speech Dataset with Multiple Types of Sound Recordings](#): The training data belongs to 20 Parkinson's Disease (PD) patients and 20 healthy subjects. From all subjects, multiple types of sound recordings (26) are taken.

[Parkinson's Disease Classification](#): 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).

[Parkinsons](#): Oxford Parkinson's Disease Detection Dataset

[Parkinsons Telemonitoring](#): Oxford Parkinson's Disease Telemonitoring Dataset

[PEMS-SF](#): 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.

[Pen-Based Recognition of Handwritten Digits](#): Digit database of 250 samples from 44 writers

[Perfume Data](#): 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.

[Phishing Websites](#): This dataset collected mainly from: PhishTank archive, MillerSmiles archive, Google's searching operators.

[Physical Unclonable Functions](#): 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.

[Physicochemical Properties of Protein Tertiary Structure](#): This is a data set of Physicochemical Properties of Protein Tertiary Structure. The data set is taken from CASP 5-9. There are 45730 decoys and size varying from 0 to 21 armstrong.

[Pioneer-1 Mobile Robot Data](#): 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

[Pittsburgh Bridges](#): Bridges database that has original and numeric-discretized datasets

[Planning Relax](#): The dataset concerns with the classification of two mental stages from recorded EEG signals: Planning (during imagination of motor act) and Relax state.

[Plants](#): 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.

[PM2.5 Data of Five Chinese Cities](#): 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.

[PMU-UD](#): 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).

[Poker Hand](#): Purpose is to predict poker hands

[Polish companies bankruptcy data](#): The dataset is about bankruptcy prediction of Polish companies. The bankrupt companies were analyzed in the period 2000-2012, while the still operating companies were evaluated from 2007 to 2013.

[Post-Operative Patient](#): Dataset of patient features

[Predict keywords activities in online social media](#): The data from Twitter was collected during 360 consecutive days. It was done by querying 1497 English keywords sampled from Wikipedia. This dataset is proposed in

a Learning to rank setting.</p><p class="normal">338. Primary Tumor: From Ljubljana Oncology Institute</p><p class="normal">339. Prodigy: Assorted domains like blocksworld, eightpuzzle, and sched world.</p><p class="normal">340. Protein Data: Undocumented</p><p class="normal">341. Pseudo Periodic Synthetic Time Series: This data set is designed for testing indexing schemes in time series databases. The data appears highly periodic, but never exactly repeats itself.</p><p class="normal">342. PubChem Bioassay Data: These highly imbalanced bioassay datasets are from the differing types of screening that can be performed using HTS technology. 21 datasets were created from 12 bioassays.</p><p class="normal">343. QSAR biodegradation: Data set containing values for 41 attributes (molecular descriptors) used to classify 1055 chemicals into 2 classes (ready and not ready biodegradable).</p><p class="normal">344. Quadruped Mammals: The file animals.c is a data generator of structured instances representing quadruped animals</p><p class="normal">346. Qualitative Structure Activity Relationships: Two sets of datasets are given: pyrimidines and triazines</p><p class="normal">347. Qualitative_Bankruptcy: Predict the Bankruptcy from Qualitative parameters from experts.</p><p class="normal">348. Quality Assessment of Digital Colposcopies: This dataset explores the subjective quality assessment of digital colposcopies.</p><p class="normal">349. Real estate valuation data set: 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. </p><p class="normal">350. REALDISP Activity Recognition Dataset: The REALDISP dataset is devised to evaluate techniques dealing with the effects of sensor displacement in wearable activity recognition as well as to benchmark general activity recognition algorithms </p><p class="normal">351. Record Linkage Comparison Patterns: Element-wise comparison of records with personal data from a record linkage setting. The task is to decide from a comparison pattern whether the underlying records belong to one person.</p><p class="normal">352. Relative location of CT slices on axial axis: 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.</p><p class="normal">353. Repeat Consumption Matrices: 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.</p><p class="normal">354. Residential Building Data Set: Data set includes construction cost, sale prices, project variables, and economic variables corresponding to real estate single-family residential apartments in Tehran, Iran. </p><p class="normal">355. Restaurant & consumer data: 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. </p><p class="normal">356. Reuters RCV1 RCV2 Multilingual, Multiview Text Categorization Test collection: This test collection contains feature characteristics of documents originally written in five different languages and their translations, over a common set of 6 categories. </p><p class="normal">357. Reuters Transcribed Subset: This dataset is created by reading out 200 files from the 10 largest Reuters \r\n classes and using an Automatic Speech Recognition system to create \r\n ncorresponding transcriptions.</p><p class="normal">358. Reuters-21578 Text Categorization Collection: This is a collection of documents that appeared on Reuters newswire in 1987. The documents were assembled and indexed with categories.</p><p class="normal">359. Reuter_50_50: The dataset is used for authorship identification in online Writeprint which is a new research field of pattern recognition. </p><p class="normal">360. Rice Leaf Diseases: 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><p class="normal">361. Robot Execution Failures: This dataset contains force and torque measurements on a robot after failure detection. Each failure is characterized by 15 force/torque samples collected at regular time intervals</p><p class="normal">362. Roman Urdu Data Set: Roman Urdu (the scripting style for Urdu language) is one of the limited resource languages. A data corpus comprising of more than 2 0000 records was collected.</p><p class="normal">363. Sales_Transactions_Dataset_Weekly: Contains weekly purchased quantities of 800 over products over 52 weeks. Normalised values are provided too.</p><p class="normal">364. SCADI: First self-care activities dataset based on ICF-CY.</p><p class="normal">365. SECOM: Data from a semi-conductor manufacturing process</p><p class="normal">366. seeds: 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.</p><p class="normal">367. seismic-bumps: 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.</p><p class="normal">368. Semeion Handwritten Digit: 1593 handwritten digits from around 80 persons were scanned, stretched in a rectangular box 16x16 in a gray scale of 25 6 values.</p><p class="normal">369. sEMG for Basic Hand movements: The "sEMG for Basic Hand movements" includes 2 databases of surface electromyographic signals of 6 hand movements using Delsys\' EMG System. Healthy subjects conducted six daily life grasps.</p><p class="normal">370. Sentence Classification: 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.</p><p class="normal">371. Sentiment Labelled Sentences: The dataset contains sentences labelled with positive or negative sentiment.</p><p class="normal">372. ser Knowledge Modeling Data (Students\' Knowledge Levels on DC Electrical Machines): 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.</p><p class="normal">373. Servo: Data was from a simulation of a servo system</p><p class="normal">374. SGEMM GPU kernel performance: Running times for multiplying two 2048 x 2048 matrices using a GPU OpenCL SGEMM kernel with varying parameters (using the library \'CLTune\').</p><p class="normal">375. Shuttle Landing Control: Tiny database; all nominal values</p><p class="normal">376. 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><p class="normal">377. Simulated Falls and Daily Living Activities Data Set: 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.</p><p class="normal">378. SkillCraft1 Master Table Dataset: This data was used in Thompson et al. (2013). A list of possible game actions is discussed in Thompson, B lair, Chen, & Henrey (2013).</p><p class="normal">379. Skin Segmentation: The Skin Segmentation dataset is constructed over B, G, R color space. Skin and Nonskin dataset is generated using skin textures from face images of diversity of age, gender, and race people.</p><p class="normal">380. Smartphone Dataset for Human Activity Recognition (HAR) in Ambient Assisted Living (AAL): This data is an addition to an existing dataset on UCI. We collected more data to improve the accuracy of our human activity recognition algorithms applied in the domain of Ambient Assisted Living. </p><p class="normal">381. Smartphone-Based Recognition of Human Activities and Postural Transitions: Activity recognition data set built from the recordings of 30 subjects performing basic activities and postural transitions while carrying a waist-mounted smartphone with embedded inertial sensors.\r\n</p><p class="normal">382. SML2010: This dataset is collected from a monitor system mounted in a domotic house. It corresponds to approximately 40 days of monitoring data.</p><p class="normal">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.</p><p class="normal">384. Solar Flare: Each class attribute counts the number of solar flares of a certain class that occur in a 24 hour period</p><p class="normal">385. Somerville Happiness Survey: A data extract of a non-federal dataset posted here <https://catalog.data.gov/dataset/somerville-happiness-survey-responses-2011-2013-2015></p><p class="normal">386. Soybean (Large): Michalski\'s famous soybean disease database</p><p class="normal">387. Soybean (Small): Michalski\'s famous soybean disease database</p><p class="normal">388. Spambase: Classifying Email as Spam or Non-Spam</p><p class="normal">389. SPECT Heart: Data on cardiac Single Proton Emission Computed Tomography (SPECT) images. Each patient classified into two categories: normal and abnormal.</p><p class="normal">390. SPECTF Heart: Data on cardiac Single Proton Emission Computed Tomography (SPECT) images. Each patient classified into two categories: normal and abnormal.</p><p class="normal">391. Spoken Arabic Digit: This dataset contains timeseries of mel-frequency cepstrum coefficients (MFCCs) corresponding to spoken Arabic digits. Includes data from 44 male and 44 female native Arabic speakers.</p><p class="normal">392. Sponge: Data on sponges; Attributes in spanish</p><p class="normal">393. Sports articles for objectivity analysis: 1000 sports articles were labeled using Amazon Mechanical Turk as objective or subjective. The raw texts, extracted features, and the URLs from which the articles were retrieved are provided.</p><p class="normal">394. Statlog (Australian Credit Approval)

/b>: This file concerns credit card applications. This database exists elsewhere in the repository (Credit Screening Database) in a slightly different form</p><p class="normal">395. Statlog (German Credit Data): This dataset classifies people described by a set of attributes as good or bad credit risks. Comes in two forms (one all numeric). Also comes with a cost matrix</p><p class="normal">396. Statlog (Heart): 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</p><p class="normal">397. Statlog (Image Segmentation): 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.</p><p class="normal">398. Statlog (Landsat Satellite): Multi-spectral values of pixels in 3x3 neighbourhoods in a satellite image, and the classification associated with the central pixel in each neighbourhood</p><p class="normal">399. Statlog (Shuttle): The shuttle dataset contains 9 attributes all of which are numerical. Approximately 80% of the data belongs to class 1</p><p class="normal">400. Statlog (Vehicle Silhouettes): 3D objects within a 2D image by application of an ensemble of shape feature extractors to the 2D silhouettes of the objects.</p><p class="normal">401. Statlog Project: Various Databases: Vehicle silhouettes, Landsat Sattelite, Shuttle, Australian Credit Approval, Heart Disease, Image Segmentation, German Credit</p><p class="normal">402. Steel Plates Faults: A dataset of steel plates' faults, classified into 7 different types. \r\nThe goal was to train machine learning for automatic pattern recognition.\r\n</p><p class="normal">403. Stock portfolio performance: The data set of performances of weighted scoring stock portfolios are obtained with mixture design from the US stock market historical database.</p><p class="normal">404. StoneFlakes: Stone flakes are waste products of the stone tool production in\r\nthe prehistoric era. The variables are means of geometric and\r\nstylistic features of the flakes contained in different inventories.</p><p class="normal">405. Student Academics Performance: The dataset tried to find the end semester percentage prediction based on different social, economic and academic attributes. </p><p class="normal">406. Student Loan Relational: Student Loan Relational Domain</p><p class="normal">407. Student Performance: Predict student performance in secondary education (high school). </p><p class="normal">408. Superconductivity Data: Two files contain data on 21 263 superconductors and their relevant features.</p><p class="normal">409. SUSY: This is a classification problem to distinguish between a signal process which produces supersymmetric particles and a background process which does not.</p><p class="normal">410. Synthetic Control Chart Time Series: This data consists of synthetically generated control charts.</p><p class="normal">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 separate subjects (Bands- recording artists; Goats; Sheep; and BioMedical)</p><p class="normal">412. Tamilnadu Electricity Board Hourly Readings: This data can be effectively produced the result to fewer parameter of the Load profile can be reduced in the Database </p><p class="normal">413. Tarvel Review Ratings: 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.</p><p class="normal">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.\r\n</p><p class="normal">415. Teaching Assistant Evaluation: The data consist of evaluations of teaching performance; scores are "low", "medium", or "high"</p><p class="normal">416. Tennis Major Tournament 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.</p><p class="normal">417. Thoracic Surgery Data: 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><p class="normal">418. Thyroid Disease: 10 separate databases from Garavan Institute</p><p class="normal">419. Tic-Tac-Toe Endgame: Binary classification task on possible configurations of tic-tac-toe game</p><p class="normal">420. Trains: 2 data formats (structured, one-instance-per-line)</p><p class="normal">421. Travel Reviews: Reviews on destinations in 10 categories mentioned across East Asia. Each traveler rating is mapped as Excellent(4), Very Good(3), Average(2), Poor(1), and Terrible(0) and average rating is used.</p><p class="normal">422. 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 documents from 6 well-known portals in Turkey. It has 4 different forms in ARFF Weka format.</p><p class="normal">423. TTCC-3600: Benchmark dataset for Turkish text categorization: The TTCC-3600 data set is a collection of Turkish news and articles including categorized 3,600 documents from 6 well-known portals in Turkey. It has 4 different forms in ARFF Weka format.

[Turkiye Student Evaluation](datasets/Turkiye+Student+Evaluation): 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.

[TV News Channel Commercial Detection Dataset](datasets/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).

[Twenty Newsgroups](datasets/Twenty+Newsgroups): This data set consists of 20000 messages taken from 20 newsgroups.

[Twin gas sensor arrays](datasets/Twin+gas+sensor+arrays): 5 replicates of an 8-MOX gas sensor array were exposed to different gas conditions (4 volatiles at 10 concentration levels each).

[Twitter Data set for Arabic Sentiment Analysis](datasets/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.

[UbiqLog \(smartphone lifelogging\)](datasets/UbiqLog+%28smartphone+lifelogging%29): UbiqLog is the smartphone lifelogging tool that runs on the smartphone of 35 users for about 2 months.

[UJI Pen Characters](datasets/UJI+Pen+Characters): Data consists of written characters in a UNIPEN-like format.

[UJI Pen Characters \(Version 2\)](datasets/UJI+Pen+Characters+%28Version+2%29): A pen-based database with more than 11k isolated handwritten characters.

[UJIIIndoorLoc](datasets/UJIIIndoorLoc): The UJIIIndoorLoc is a Multi-Building Multi-Floor indoor localization database to test Indoor Positioning System that rely on WLAN/WiFi fingerprint.

[UJIIIndoorLoc-Mag](datasets/UJIIIndoorLoc-Mag): The UJIIIndoorLoc-Mag is an indoor localization database to test Indoor Positioning System that rely on Earth's magnetic field variations.

[Ultrasonic flowmeter diagnostics](datasets/Ultrasonic+flowmeter+diagnostics): Fault diagnosis of four liquid ultrasonic flowmeters.

[Undocumented](datasets/Undocumented): Various datasets without documentation (feel free to explore!).

[University](datasets/University): Data in original (LISP-readable) form.

[University of Tehran Question Dataset 2016 \(UTQD.2016\)](datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29): Persian questions gathered from a jeopardy game broadcasted on Iranian national television.

[UNIX User Data](datasets/UNIX+User+Data): 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.

[Urban Land Cover](datasets/Urban+Land+Cover): Classification of urban land cover using high resolution aerial imagery. Intended to assist sustainable urban planning efforts.

[URL Reputation](datasets/URL+Reputation): Anonymized 120-day subset of the ICML-09 URL data containing 2.4 million examples and 3.2 million features.

[US Census Data \(1990\)](datasets/US+Census+Data+%281990%29): 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.

[User Identification From Walking Activity](datasets/User+Identification+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.

[User Knowledge Modeling](datasets/User+Knowledge+Modeling): 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.

[USPTO Algorithm Challenge, run by NASA-Harvard Tournament Lab and TopCoder Problem: Pat](datasets/USPTO+Algorithm+Challenge%2C+run+by+NASA-Harvard+Tournament+Lab+and+TopCoder++Problem%3A+Pat): Data used for USPTO Algorithm Competition. Contains drawing pages from US patents with manually labeled figure and part labels.

[Vertebral Column](datasets/Vertebral+Column): Data set containing values for six biomechanical features used to classify orthopaedic patients into 3 classes (normal, disk hernia or spondylolisthesis) or 2 classes (normal or abnormal).

[Vicon Physical Action Data Set](datasets/Vicon+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 10 subjects using the Vicon 3D tracker.

[Victorian Era Authorship Attribution](datasets/Victorian+Era+Authorship+Attribution): To create the largest authorship attribution dataset, we extracted works of 50 well-known authors. To have a non-exhaustive learning, in training there are 45 authors whereas, in the testing, it's 50.

[Volcanoes on Venus - JARtool experiment](datasets/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 of Venus images returned by the Magellan spacecraft.

[Wall-Following Robot Navigation Data](datasets/Wall-Following+Robot+Navigation+Data): The data were collected as the SCITOS G5 robot navigates through the room following the wall in a clockwise direction, for 4 rounds, using 24 ultrasound sensors arranged circularly around its 'waist'.

[Water Treatment Plant](datasets/Water+Treatment+Plant): Multiple classes predict plant state.

[Waveform Database Generator \(Version 1\)](datasets/Waveform+Database+Generator+%28Version+1%29): CART book's waveform domains.

[Waveform Database Generator \(Version 2\)](datasets/Waveform+Database+Generator+%28Version+2%29)

```

)</a></b>: CART book's waveform domains</p><p class="normal">452. <b><a href="datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29">Wearable Computing: Classification of Body Postures and Movements (PUC-Rio)</a></b>: A dataset with 5 classes (sitting-down, standing-up, standing, walking, and sitting) collected on 8 hours of activities of 4 healthy subjects. We also established a baseline performance index.</p><p class="normal">453. <b><a href="datasets/Website+Phishing">Website Phishing</a></b>: \r\n\r\n</p><p class="normal">454. <b><a href="datasets/Weight+Lifting+Exercises+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 exercise. One of the variations is the one predicted by the health professional.</p><p class="normal">455. <b><a href="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 stress-affect lab study, while wearing physiological and motion sensors.</p><p class="normal">456. <b><a href="datasets/Wholesale+customers">Wholesale customers</a></b>: The dataset refers to clients of a wholesale distributor. It includes the annual spending in monetary units (m.u.) on diverse product categories</p><p class="normal">457. <b><a href="datasets/wiki4HE">wiki4HE</a></b>: Survey of faculty members from two Spanish universities on teaching uses of Wikipedia</p><p class="normal">458. <b><a href="datasets/Wilt">Wilt</a></b>: High-resolution Remote Sensing data set (Quickbird). Small number of training samples of diseased trees, large number for other land cover. Testing data set from stratified random sample of image.</p><p class="normal">459. <b><a href="datasets/Wine">Wine</a></b>: Using chemical analysis determine the origin of wines</p><p class="normal">460. <b><a href="datasets/Wine+Quality">Wine Quality</a></b>: 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/).</p><p class="normal">461. <b><a href="datasets/Wireless+Indoor+Localization">Wireless Indoor Localization</a></b>: 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. </p><p class="normal">462. <b><a href="datasets/Yacht+Hydrodynamics">Yacht Hydrodynamics</a></b>: Delft data set, used to predict the hydrodynamic performance of sailing yachts from dimensions and velocity.</p><p class="normal">463. <b><a href="datasets/YearPredictionMSD">YearPredictionMSD</a></b>: 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.</p><p class="normal">464. <b><a href="datasets/Yeast">Yeast</a></b>: Predicting the Cellular Localization Sites of Proteins</p><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.</p><p class="normal">466. <b><a href="datasets/YouTube+Multiview+Video+Games+Dataset">YouTube Multiview Video 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><p class="normal">467. <b><a href="datasets/YouTube+Spam+Collection">YouTube Spam Collection</a></b>: It is a public set of comments collected for spam research. It has five datasets composed by 1,956 real messages extracted from five videos that were among the 10 most viewed on the collection period.</p><p class="normal">468. <b><a href="datasets/Z-Alizadeh+Sani">Z-Alizadeh Sani</a></b>: It was collected for CAD diagnosis.</p><p class="normal">469. <b><a href="datasets/Zoo">Zoo</a></b>: Artificial, 7 classes of animals</p></td></tr></table>
<hr/>\n</td>\n</tr>\n</table>\n<table align="center" cellpadding="5"><tr valign="center">\n<td><p class="normal">Supported By:</p></td>\n<td></td>\n<td><p class="normal">\xa0In Collaboration With:</p></td>\n<td></td>\n</tr></table>\n<center>\n<span class="normal">\n<a href="about.html">About</a>\xa0\x0||\xa0\n<a href="citation_policy.html">Citation Policy</a>\xa0\x0||\xa0\n<a href="donation_policy.html">Donation Policy</a>\xa0\x0||\xa0\n<a href="contact.html">Contact</a>\xa0\x0||\xa0\n<a href="http://cml.ics.uci.edu">CML</a>\n</span>\n</center>\n</body></html>'
```

In [105]:

```
tittle = re.findall('<a href="datasets/.+>(.+)</a></b>', str(datatype))
```

In [106]:

```
len(tittle) #intente de todas las formas posibles pero no puede encontrar el error para extraer los datos.
```

Out[106]:

33

In [107]:

```
tittle
```

Out[107]:

```
['Activities of Daily Living (ADLs) Recognition Using Binary Sensors',
 'Acute Inflammations',
 'Bach Choral Harmony',
 'BLOGGER',
 'chipseq',
 'CNAE-9',
 'Daily and Sports Activities',
 'Daphnet Freezing of Gait',
 'Dexter',
 'Diabetes 130-US hospitals for years 1999-2008',
 'Discrete Tone Image Dataset',
 'ElectricityLoadDiagrams20112014',
 'Gas sensors for home activity monitoring',
 'Geographical Original of Music',
 'Gisette',
 'Greenhouse Gas Observing Network',
 'Libras Movement',
 'Mesothelioma's disease data set ',
 'Nomao',
 'Online Shoppers Purchasing Intention Dataset',
 'OpinRank Review Dataset',
 'p53 Mutants',
 'Parking Birmingham',
 'Reuters Transcribed Subset',
 'seismic-bumps',
 'Smartphone-Based Recognition of Human Activities and Postural Transitions',
 'Steel Plates Faults',
 'StoneFlakes',
 'Taxi Service Trajectory - Prediction Challenge, ECML PKDD 2015',
 'UbiqLog (smartphone lifelogging)',
 'User Identification From Walking Activity',
 'Website Phishing',
 'Zoo']
```

In [108]:

```
text = re.findall('<td><p class="normal"(>.+\|\S)\xa0</p></td>', str(datatype))
text
#Sin poder extraer esta información es imposible continuar con el ejercicio.
```

Out[108]:

```
[]
```