Taller 8

Métodos Computacionales para Políticas Públicas - URosario

Entrega: viernes 12-oct-2018 11:59 PM

[Su nombre acá]

[Su e-mail acá]

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.

Vamos a hacer "scraping" a esta página: http://archive.ics.uci.edu/ml/datasets.html, que contiene un listado de 360 bases de datos que hacen parte del repositorio de la Universidad de California, Irvine.

Su tarea consiste en crear un "Pandas dataframe" que contenga 360 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
- Descricpción de la base (Pista: Utilice la opción list view)

Diviértase.

In [1]:

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

In [3]:

```
from requests import get

url = 'http://archive.ics.uci.edu/ml/datasets.html'

response = get(url)
```

```
print(response.text)
```

```
<!DOCTYPE HTML PUBLIC \"-//W3C//DTD HTML 4.01 Transitional//EN\">
<html>
<head>
<title>UCI Machine Learning Repository: Data Sets</title>

<!-- Stylesheet link -->
<link rel="stylesheet" type="text/css" href="assets/ml.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>

</head>

<body>


<!-- SITE HEADER (INCLUDES LOGO AND SEARCH BOX) -->

<!-- SITE HEADER (INCLUDES LOGO AND SEARCH BOX) -->

<table width=100% bgcolor="#003366">
<tr>
 <td>
  <span class="normal"><a href="index.html" alt="Home"><img src="assets/logo.gif" border=0></img><
```

```
/a><br>         <a href="http://cml.ics.uci.edu"><fon
t color="FFDD33">Center for Machine Learning and Intelligent Systems</font></a></span>
 </td>
 <td width=100% valign=top align="right">
  <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 method=GET action=http://www.google.com/custom onsubmit="return checkform(this);">
  <INPUT TYPE=text name=q size=30 maxlength=255 value="">
  <INPUT type=submit name=sa VALUE="Search">
  <INPUT type=hidden name=cof
VALUE="AH:center;LH:130;L:http://archive.ics.uci.edu/assets/logo.gif;LW:384;AWFID:869c0b2eaa8d518e;

  <input type=hidden name=domains value="ics.uci.edu">
  <br>
  <input type=radio name=sitesearch value="ics.uci.edu" checked> <span class="whitetext"><font
size="1">Repository</font></span>
  <input type=radio name=sitesearch value=""> <span class="whitetext"><font size="1">Web</font>
</span>
            
  <A HREF=http://www.google.com/search><IMG SRC=http://www.google.com/logos/Logo_25blk.gif
border=0 ALT=Google align=middle height=27></A>
  <br>
  </FORM>
  <!-- Search Google -->


  <span class="whitetext"><a href="datasets.html"><font size="3" color="#FFDD33"><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 class="big" align=center>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.html?
format=&task=cla&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table'>Classification</a> <font
color=red>(330)</font><br><a href='datasets.html?
format=&task=reg&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table'>Regression</a> <font
color=red>(89)</font><br><a href='datasets.html?
format=&task=clu&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table'>Clustering</a> <font
color=red>(79)</font><br><a href='datasets.html?
format=&task=other&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table'>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.html?
format=&task=&att=cat&area=&numAtt=&numIns=&type=&sort=nameUp&view=table'>Categorical</a> <font co
lor=red>(38)</font><br><a href='datasets.html?
format=&task=&att=num&area=&numAtt=&numIns=&type=&sort=nameUp&view=table'>Numerical</a> <font
color=red>(288)</font><br><a href='datasets.html?
```

```
color=red>(200)</font><br><a href='datasets.html?
format=&task=&att=mix&area=&numAtt=&numIns=&type=&sort=nameUp&view=table'>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.html?
format=&task=&att=&area=&numAtt=&numIns=&type=mvar&sort=nameUp&view=table'>Multivariate</a> <font
color=red>(340)</font><br><a href='datasets.html?
format=&task=&att=&area=&numAtt=&numIns=&type=uvar&sort=nameUp&view=table'>Univariate</a> <font co
lor=red>(20)</font><br><a href='datasets.html?
format=&task=&att=&area=&numAtt=&numIns=&type=seq&sort=nameUp&view=table'>Sequential</a> <font
color=red>(45)</font><br><a href='datasets.html?
format=&task=&att=&area=&numAtt=&numIns=&type=ts&sort=nameUp&view=table'>Time-Series</a> <font
color=red>(84)</font><br><a href='datasets.html?
format=&task=&att=&area=&numAtt=&numIns=&type=text&sort=nameUp&view=table'>Text</a> <font
color=red>(50)</font><br><a href='datasets.html?
format=&task=&att=&area=&numAtt=&numIns=&type=dt&sort=nameUp&view=table'>Domain-Theory</a> <font c
olor=red>(23)</font><br><a href='datasets.html?
format=&task=&att=&area=&numAtt=&numIns=&type=other&sort=nameUp&view=table'>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.html?
format=&task=&att=&area=life&numAtt=&numIns=&type=&sort=nameUp&view=table'>Life Sciences</a> <font
color=red>(103)</font><br><a href='datasets.html?
format=&task=&att=&area=phys&numAtt=&numIns=&type=&sort=nameUp&view=table'>Physical Sciences</a> <
font color=red>(47)</font><br><a href='datasets.html?
format=&task=&att=&area=comp&numAtt=&numIns=&type=&sort=nameUp&view=table'>CS / Engineering</a> <f
ont color=red>(159)</font><br><a href='datasets.html?
format=&task=&att=&area=soc&numAtt=&numIns=&type=&sort=nameUp&view=table'>Social Sciences</a>
<font color=red>(26)</font><br><a href='datasets.html?
format=&task=&att=&area=bus&numAtt=&numIns=&type=&sort=nameUp&view=table'>Business</a> <font
color=red>(27)</font><br><a href='datasets.html?
format=&task=&att=&area=game&numAtt=&numIns=&type=&sort=nameUp&view=table'>Game</a> <font
color=red>(10)</font><br><a href='datasets.html?
format=&task=&att=&area=other&numAtt=&numIns=&type=&sort=nameUp&view=table'>Other</a> <font
color=red>(69)</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.html?
format=&task=&att=&area=&numAtt=less10&numIns=&type=&sort=nameUp&view=table'>Less than 10</a> <fon
t color=red>(104)</font><br><a href='datasets.html?
format=&task=&att=&area=&numAtt=10to100&numIns=&type=&sort=nameUp&view=table'>10 to 100</a> <font
color=red>(201)</font><br><a href='datasets.html?
format=&task=&att=&area=&numAtt=greater100&numIns=&type=&sort=nameUp&view=table'>Greater than
100</a> <font color=red>(83)</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.html?
format=&task=&att=&area=&numAtt=&numIns=less100&type=&sort=nameUp&view=table'>Less than 100</a> <f
ont color=red>(26)</font><br><a href='datasets.html?
format=&task=&att=&area=&numAtt=&numIns=100to1000&type=&sort=nameUp&view=table'>100 to 1000</a> <f
ont color=red>(153)</font><br><a href='datasets.html?
format=&task=&att=&area=&numAtt=&numIns=greater1000&type=&sort=nameUp&view=table'>Greater than 100
0</a> <font color=red>(234)</font> </p>
  </td>
 </tr>
 <tr><td bgcolor="#003366"><p class="whitetext"><b>Format Type</b> </p>
  </td>
 </tr>
 <tr>
```

<td valign=top> Matrix (309)
Non-Matrix (137)
</td>
</tr></table>
</td>

<td valign=top>

<table width=100%>
<tr>
<td> 446 Data Sets </td>
<td align="right"> Table View List View </td>
</tr>
</table><table border = 1 cellpadding = 5><tr bgcolor="#003366">
<td class="normal, whitetext"> Name </td>
<!-- <td> Abstract </td> -->
<td> Data Types </td>
<td> Default Task </td>
<td> Attribute Types </td>
<td> # Instances </td>
<td> # Attributes </td>
<td> Year </td>

<!-- <td> Area </td> -->

</tr><tr>
<td><table><tr><td> </td><td> Abalone </td></tr></table></td>
<!-- <td> Predict the age of abalone from physical measurements </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer, Real </td>
<td> 4177 </td>
<td> 8 </td>
<td> 1995 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> Adult </td></tr></table></td>
<!-- <td> Predict whether income exceeds $50K/yr based on census data. Also known as "Census Income" dataset. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 48842 </td>
<td> 14 </td>
<td> 1996 </td>
<!-- <td> Social </td> -->
</tr><tr>
<td><table><tr><td> </td><td> Annealing </td></tr></table></td>
<!-- <td> Steel annealing data </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer, Real </td>

<td> Categorical, Integer, Real </td>
<td> 798 </td>
<td> 38 </td>
<td>   </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Anonymous+Microsoft+Web+Data">Anonymous Microsoft Web Data </td></tr></t
able></td>
<!-- <td> Log of anonymous users of www.microsoft.com; predict areas of the
web site a user visited based on data on other areas the user visited. </td> -->
<td>   </td>
<td> Recommender-Systems </td>
<td> Categorical </td>
<td> 37711 </td>
<td> 294 </td>
<td> 1998 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img src="assets/MLimages/SmallLarge5.jpg" bor
der=1 /> </td><td> Arrhythmia</
p></td></tr></table></td>
<!-- <td> Distinguish between the presence and absence of cardiac arrhythmia a
nd classify it in one of the 16 groups. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer, Real </td>
<td> 452 </td>
<td> 279 </td>
<td> 1998 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLarge6.jpg" border=1 /> </td><td> <a
href="datasets/Artificial+Characters">Artificial Characters </td></tr></table></td>
<!-- <td> Dataset artificially generated by using first order theory which
describes structure of ten capital letters of English alphabet </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer, Real </td>
<td> 6000 </td>
<td> 7 </td>
<td> 1992 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLarge7.jpg" border=1 /> </td><td> <a
href="datasets/Audiology+%28Original%29">Audiology (Original) </td></tr></table></td>
<!-- <td> Nominal audiology dataset from Baylor </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 226 </td>
<td>   </td>
<td> 1987 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLarge7.jpg" border=1 /> </td><td> <a
href="datasets/Audiology+%28Standardized%29">Audiology (Standardized) </td></tr></table
></td>
<!-- <td> Standardized version of the original audiology database </t
d> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 226 </td>
<td> 69 </td>
<td> 1992 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><img src="assets/MLimages/SmallLarge9.jpg" borde
r=1 /> </td><td> Auto MPG </td
></tr></table></td>
<!-- <td> Revised from CMU StatLib library, data concerns city-cycle fuel cons
umption </td> -->
<td> Multivariate </td>

<td> Multivariate </td>
<td> Regression </td>
<td> Categorical, Real </td>
<td> 398 </td>
<td> 8 </td>
<td> 1993 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img src="assets/MLimages/SmallLarge10.jpg" bo
rder=1 /> </td><td> Automobile<
/p></td></tr></table></td>
<!-- <td> From 1985 Ward's Automotive Yearbook </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Categorical, Integer, Real </td>
<td> 205 </td>
<td> 26 </td>
<td> 1987 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><img src="assets/MLimages/SmallLargedefault.jpg" b
order=1 /> </td><td> Badges </td
></tr></table></td>
<!-- <td> Badges labeled with a "+" or "-" as a function of a person's
name </td> -->
<td> Univariate, Text </td>
<td> Classification </td>
<td>   </td>
<td> 294 </td>
<td> 1 </td>
<td> 1994 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img src="assets/MLimages/SmallLarge12.jpg"
border=1 /> </td><td> Balance
Scale </td></tr></table></td>
<!-- <td> Balance scale weight & distance database </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 625 </td>
<td> 4 </td>
<td> 1994 </td>
<!-- <td> Social </td> -->
</tr><tr>
<td><table><tr><td><img src="assets/MLimages/SmallLarge13.jpg" bord
er=1 /> </td><td> Balloons </t
d></tr></table></td>
<!-- <td> Data previously used in cognitive psychology experiment; 4 data sets
represent different conditions of an experiment </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 16 </td>
<td> 4 </td>
<td>   </td>
<!-- <td> Social </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img src="assets/MLimages/SmallLarge14.jpg"
border=1 /> </td><td> Breast
Cancer </td></tr></table></td>
<!-- <td> Breast Cancer Data (Restricted Access) </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 286 </td>
<td> 9 </td>
<td> 1988 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLarge14.jpg" border=1 /> </td><td> <a href=
"datasets/Breast+Cancer+Wisconsin+%28Original%29">Breast Cancer Wisconsin (Original) </
td></tr></table></td>
<!-- <td> Original Wisconsin Breast Cancer Database </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>

<td> Integer </td>
<td> 699 </td>
<td> 10 </td>
<td> 1992 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLarge14.jpg" border=1 /> </td><td> <a href=
"datasets/Breast+Cancer+Wisconsin+%28Prognostic%29">Breast Cancer Wisconsin (Prognostic)</
p></td></tr></table></td>
<!-- <td> Prognostic Wisconsin Breast Cancer Database </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Real </td>
<td> 198 </td>
<td> 34 </td>
<td> 1995 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLarge14.jpg" border=1 /> </td><td> <a href=
"datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29">Breast Cancer Wisconsin (Diagnostic)</
p></td></tr></table></td>
<!-- <td> Diagnostic Wisconsin Breast Cancer Database </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 569 </td>
<td> 32 </td>
<td> 1995 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLarge18.jpg" border=1 /> </td><td> <a href=
"datasets/Pittsburgh+Bridges">Pittsburgh Bridges </td></tr></table></td>
<!-- <td> Bridges database that has original and numeric-discretized
datasets </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 108 </td>
<td> 13 </td>
<td> 1990 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLarge19.jpg" border=1 /> </td><td> <a href=
"datasets/Car+Evaluation">Car Evaluation </td></tr></table></td>
<!-- <td> Derived from simple hierarchical decision model, this database may b
e useful for testing constructive induction and structure discovery methods. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 1728 </td>
<td> 6 </td>
<td> 1997 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img src="assets/MLimages/SmallLarge2.jpg"
border=1 /> </td><td> Census
Income </td></tr></table></td>
<!-- <td> Predict whether income exceeds $50K/yr based on census data. Also
known as "Adult" dataset. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 48842 </td>
<td> 14 </td>
<td> 1996 </td>
<!-- <td> Social </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLarge24.jpg" border=1 /> </td><td> <a href=
"datasets/Chess+%28King-Rook+vs.+King-Knight%29">Chess (King-Rook vs. King-Knight) </td
></tr></table></td>
<!-- <td> Knight Pin Chess End-Game Database Creator </td> -->
<td> Multivariate, Data-Generator </td>

```
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Categorical, Integer </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal">22 </p></td>
   <td><p class="normal">1988 </p></td>
   <!-- <td><p class="normal">Game </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Chess+%28King-Rook+vs.+King-Pawn%29"><img
src="assets/MLimages/SmallLarge24.jpg" border=1 /></a> </td><td><p class="normal"><b><a href=
"datasets/Chess+%28King-Rook+vs.+King-Pawn%29">Chess (King-Rook vs. King-Pawn)</a></b></p></td></t
r></table></td>
   <!-- <td><p class="normal">King+Rook versus King+Pawn on a7 (usually abbreviated KRKPA7). 
</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><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><a href="datasets/Chess+%28King-Rook+vs.+King%29"><img
src="assets/MLimages/SmallLarge24.jpg" border=1 /></a> </td><td><p class="normal"><b><a href=
"datasets/Chess+%28King-Rook+vs.+King%29">Chess (King-Rook vs. King)</a></b></p></td></tr></table>
</td>
   <!-- <td><p class="normal">Chess Endgame Database for White King and Rook against Black King (K
RK). </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><a href="datasets/Chess+%28Domain+Theories%29"><img
src="assets/MLimages/SmallLarge24.jpg" border=1 /></a> </td><td><p class="normal"><b><a href=
"datasets/Chess+%28Domain+Theories%29">Chess (Domain Theories)</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">6 different domain theories for generating legal moves of
chess </p></td> -->
   <td><p class="normal">Domain-Theory </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <!-- <td><p class="normal">Game </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Bach+Chorales"><img src="assets/MLimages/SmallLarge25.jpg"
border=1 /></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"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Connect-4">Connect-4</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">Contains connect-4 positions </p></td> -->
   <td><p class="normal">Multivariate, Spatial </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Categorical </p></td>
   <td><p class="normal">67557 </p></td>
   <td><p class="normal">42 </p></td>
   <td><p class="normal">1995 </p></td>
   <!-- <td><p class="normal">Game </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Credit+Approval"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Credit+Approval">Credit Approval</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">This data concerns credit card applications; good mix of
```

```
attributes </p></td> -->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Categorical, Integer, Real </p></td>
   <td><p class="normal">690 </p></td>
   <td><p class="normal">15 </p></td>
   <td><p class="normal"> </p></td>
   <!-- <td><p class="normal">Financial </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Japanese+Credit+Screening"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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><tr>
   <td><table><tr><td><a href="datasets/Computer+Hardware"><img
src="assets/MLimages/SmallLarge29.jpg" border=1 /></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"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Contraceptive+Method+Choice">Contraceptive Method Choice</a></b></p></td></tr></tab
le></td>
   <!-- <td><p class="normal">Dataset is a subset of the 1987 National Indonesia Contraceptive Pre
valence 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><a href="datasets/Covertype"><img src="assets/MLimages/SmallLarge31.jpg" bor
der=1 /></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 </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><a href="datasets/Cylinder+Bands"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 </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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Dermatology"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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. </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><a href="datasets/Diabetes"><img src="assets/MLimages/SmallLargedefault.jpg"
border=1 /></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 </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><a href="datasets/DGP2+-+The+Second+Data+Generation+Program"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 o
f 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">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><a href="datasets/Document+Understanding"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 </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">1994 </p></td>
   <!-- <td><p class="normal">Other </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/EBL+Domain+Theories"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 </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">Computer </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Echocardiogram"><img
src="assets/MLimages/SmallLarge38.jpg" border=1 /></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 </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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Ecoli"><img src="assets/MLimages/SmallLarge120.jpg" border
=1 /></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 </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><a href="datasets/Flags"><img src="assets/MLimages/SmallLarge40.jpg"
border=1 /></a> </td><td><p class="normal"><b><a href="datasets/Flags">Flags</a></b></p></td>
</tr></table></td>
   <!-- <td><p class="normal">From Collins Gem Guide to Flags, 1986 </p></td> -->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Categorical, Integer </p></td>
   <td><p class="normal">194 </p></td>
   <td><p class="normal">30 </p></td>
   <td><p class="normal">1990 </p></td>
   <!-- <td><p class="normal">Other </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Function+Finding"><img
src="assets/MLimages/SmallLarge41.jpg" border=1 /></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; inte
ntion is to evaluate function-finding algorithms </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 </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Glass+Identification"><img
src="assets/MLimages/SmallLarge42.jpg" border=1 /></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 term
s 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><a href="datasets/Haberman%27s+Survival"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Haberman%27s+Survival">Haberman's Survival</a></b></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><a href="datasets/Hayes-Roth"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Hayes-Roth">Hayes-Roth</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">Topic: human subjects study </p></td> -->
   <td><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><a href="datasets/Heart+Disease"><img src="assets/MLimages/SmallLarge45.jpg"
border=1 /></a> </td><td><p class="normal"><b><a href="datasets/Heart+Disease">Heart
Disease</a></b></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><a href="datasets/Hepatitis"><img src="assets/MLimages/SmallLarge46.jpg" bor
der=1 /></a> </td><td><p class="normal"><b><a href="datasets/Hepatitis">Hepatitis</a></b></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><a href="datasets/Horse+Colic"><img src="assets/MLimages/SmallLarge47.jpg" b
order=1 /></a> </td><td><p class="normal"><b><a href="datasets/Horse+Colic">Horse Colic</a></
b></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><a href="datasets/ICU"><img src="assets/MLimages/SmallLarge49.jpg" border=1
/></a> </td><td><p class="normal"><b><a href="datasets/ICU">ICU</a></b></p></td></tr></table>
</td>
   <!-- <td><p class="normal">Data set prepared for the use of participants for the 1994 AAAI Spri
ng 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><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <!-- <td><p class="normal">Life </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Image+Segmentation"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Image+Segmentation">Image Segmentation</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">Image data described by high-level numeric-valued attributes, 7
classes </p></td> -->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Real </p></td>
   <td><p class="normal">2310 </p></td>
   <td><p class="normal">19 </p></td>
   <td><p class="normal">1990 </p></td>
   <!-- <td><p class="normal">Other </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Internet+Advertisements"><img
src="assets/MLimages/SmallLarge51.jpg" border=1 /></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. </p></td> -->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Categorical, Integer, Real </p></td>
   <td><p class="normal">3279 </p></td>
   <td><p class="normal">1558 </p></td>
   <td><p class="normal">1998 </p></td>
   <!-- <td><p class="normal">Computer </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Ionosphere"><img src="assets/MLimages/SmallLarge52.jpg" bo
rder=1 /></a> </td><td><p class="normal"><b><a href="datasets/Ionosphere">Ionosphere</a></b><
/p></td></tr></table></td>
```

```
   <!-- <td><p class="normal">Classification of radar returns from the ionosphere </p></td> -
->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Integer, Real </p></td>
   <td><p class="normal">351 </p></td>
   <td><p class="normal">34 </p></td>
   <td><p class="normal">1989 </p></td>
   <!-- <td><p class="normal">Physical </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Iris"><img src="assets/MLimages/SmallLarge53.jpg" border=1
/></a> </td><td><p class="normal"><b><a href="datasets/Iris">Iris</a></b></p></td></tr></tabl
e></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>
   <td><p class="normal">1988 </p></td>
   <!-- <td><p class="normal">Life </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/ISOLET"><img src="assets/MLimages/SmallLargedefault.jpg" b
order=1 /></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. </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><a href="datasets/Kinship"><img src="assets/MLimages/SmallLarge55.jpg" borde
r=1 /></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 </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><a href="datasets/Labor+Relations"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Labor+Relations">Labor Relations</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">From Collective Bargaining Review </p></td> -->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal">Categorical, Integer, Real </p></td>
   <td><p class="normal">57 </p></td>
   <td><p class="normal">16 </p></td>
   <td><p class="normal">1988 </p></td>
   <!-- <td><p class="normal">Social </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/LED+Display+Domain"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 p
rograms for generating sample databases </p></td> -->
   <td><p class="normal">Multivariate, Data-Generator </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Categorical </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal">7 </p></td>
   <td><p class="normal">1988 </p></td>
   <!-- <td><p class="normal">Computer </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Lenses"><img src="assets/MLimages/SmallLargedefault.jpg" b
order=1 /></a> </td><td><p class="normal"><b><a href="datasets/Lenses">Lenses</a></b></p></td
></tr></table></td>
   <!-- <td><p class="normal">Database for fitting contact lenses </p></td> -->
   <td><p class="normal">Multivariate </p></td>
```

```
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Categorical </p></td>
   <td><p class="normal">24 </p></td>
   <td><p class="normal">4 </p></td>
   <td><p class="normal">1990 </p></td>
   <!-- <td><p class="normal">Other </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Letter+Recognition"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 </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><a href="datasets/Liver+Disorders"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Liver+Disorders">Liver Disorders</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">BUPA Medical Research Ltd. database donated by Richard S.
Forsyth </p></td> -->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal">Categorical, Integer, Real </p></td>
   <td><p class="normal">345 </p></td>
   <td><p class="normal">7 </p></td>
   <td><p class="normal">1990 </p></td>
   <!-- <td><p class="normal">Life </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Logic+Theorist"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Logic+Theorist">Logic Theorist</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">All code for Logic Theorist </p></td> -->
   <td><p class="normal">Domain-Theory </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <!-- <td><p class="normal">Computer </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Lung+Cancer"><img src="assets/MLimages/SmallLarge62.jpg" b
order=1 /></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 </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><a href="datasets/Lymphography"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 Ce
ntre, 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><a href="datasets/Mechanical+Analysis"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 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><a href="datasets/Meta-data"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Mobile+Robots"><img src="assets/MLimages/SmallLarge66.jpg"
border=1 /></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 se
ts </p></td> -->
   <td><p class="normal">Domain-Theory </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal">Categorical, Integer, Real </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal">1995 </p></td>
   <!-- <td><p class="normal">Computer </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Molecular+Biology+%28Promoter+Gene+Sequences%29"><img src=
"assets/MLimages/SmallLarge67.jpg" border=1 /></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 </p></td> -->
   <td><p class="normal">Sequential, Domain-Theory </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Categorical </p></td>
   <td><p class="normal">106 </p></td>
   <td><p class="normal">58 </p></td>
   <td><p class="normal">1990 </p></td>
   <!-- <td><p class="normal">Life </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Molecular+Biology+%28Protein+Secondary+Structure%29"><img
src="assets/MLimages/SmallLarge67.jpg" border=1 /></a> </td><td><p class="normal"><b><a href=
"datasets/Molecular+Biology+%28Protein+Secondary+Structure%29">Molecular Biology (Protein
Secondary Structure)</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">From CMU connectionist bench repository; Classifies secondary
structure of 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><a href="datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29"><i
mg src="assets/MLimages/SmallLarge67.jpg" border=1 /></a> </td><td><p class="normal"><b><a hr
ef="datasets/Molecular+Biology+%28Splice-junction+Gene+Sequences%29">Molecular Biology (Splice-jun
ction Gene Sequences)</a></b></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><a href="datasets/MONK%27s+Problems"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
```

```
href="datasets/MONK%27s+Problems">MONK's Problems</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">A set of three artificial domains over the same attribute space; Use
d 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><a href="datasets/Moral+Reasoner"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Moral+Reasoner">Moral Reasoner</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">Horn-clause model that qualitatively simulates moral reasoning; Theor
y 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><a href="datasets/Multiple+Features"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Multiple+Features">Multiple Features</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">This dataset consists of features of handwritten numerals (`0'--`9')
extracted 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><a href="datasets/Mushroom"><img src="assets/MLimages/SmallLarge73.jpg" bord
er=1 /></a> </td><td><p class="normal"><b><a href="datasets/Mushroom">Mushroom</a></b></p></t
d></tr></table></td>
   <!-- <td><p class="normal">From Audobon Society Field Guide; mushrooms described in terms of
physical characteristics; classification: poisonous or edible </p></td> -->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Categorical </p></td>
   <td><p class="normal">8124 </p></td>
   <td><p class="normal">22 </p></td>
   <td><p class="normal">1987 </p></td>
   <!-- <td><p class="normal">Life </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Musk+%28Version+1%29"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Musk+%28Version+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 </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><a href="datasets/Musk+%28Version+2%29"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Musk+%28Version+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 </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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Nursery"><img src="assets/MLimages/SmallLargedefault.jpg"
```

```
border=1 /></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 ori
ginally developed to rank applications for nursery schools. </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 </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Othello+Domain+Theory"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 syst
em </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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Page+Blocks+Classification"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 layou
t 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><a href="datasets/Optical+Recognition+of+Handwritten+Digits"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Optical+Recognition+of+Handwritten+Digits">Optical Recognition of Handwritten
Digits</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">Two versions of this database available; see folder </p></td> -
->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Integer </p></td>
   <td><p class="normal">5620 </p></td>
   <td><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"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Pen-Based+Recognition+of+Handwritten+Digits">Pen-Based Recognition of Handwritten D
igits</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"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Post-Operative+Patient">Post-Operative Patient</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">Dataset of patient features </p></td> -->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Categorical, Integer </p></td>
   <td><p class="normal">90 </p></td>
   <td><p class="normal">8 </p></td>
   <td><p class="normal">1993 </p></td>
   <!-- <td><p class="normal">Life </p></td> -->
```

```
   </tr><tr>
   <td><table><tr><td><a href="datasets/Primary+Tumor"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Primary+Tumor">Primary Tumor</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">From Ljubljana Oncology Institute </p></td> -->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Categorical </p></td>
   <td><p class="normal">339 </p></td>
   <td><p class="normal">17 </p></td>
   <td><p class="normal">1988 </p></td>
   <!-- <td><p class="normal">Life </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Prodigy"><img src="assets/MLimages/SmallLargedefault.jpg"
border=1 /></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"> </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <!-- <td><p class="normal">Other </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Qualitative+Structure+Activity+Relationships"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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"> </p></td>
   <td><p class="normal"> </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"><img
src="assets/MLimages/SmallLarge86.jpg" border=1 /></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"><img src="assets/MLimages/SmallLarge87.jpg"
border=1 /></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"><img
src="assets/MLimages/SmallLarge92.jpg" border=1 /></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><img src="assets/MLimages/SmallLarge89.jpg" b
order=1 /> </td><td> Solar Flare</
b> </td></tr></table></td>
<!-- <td> Each class attribute counts the number of solar flares of a certain
class that occur in a 24 hour period </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Categorical </td>
<td> 1389 </td>
<td> 10 </td>
<td> 1989 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLarge90.jpg" border=1 /> </td><td> <a href=
"datasets/Soybean+%28Large%29">Soybean (Large) </td></tr></table></td>
<!-- <td> Michalski's famous soybean disease database </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 307 </td>
<td> 35 </td>
<td> 1988 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLarge90.jpg" border=1 /> </td><td> <a href=
"datasets/Soybean+%28Small%29">Soybean (Small) </td></tr></table></td>
<!-- <td> Michalski's famous soybean disease database </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 47 </td>
<td> 35 </td>
<td> 1987 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLarge92.jpg" border=1 /> </td><td> <a href=
"datasets/Challenger+USA+Space+Shuttle+O-Ring">Challenger USA Space Shuttle O-Ring </td
></tr></table></td>
<!-- <td> Task: predict the number of O-rings that experience thermal distress
on a flight at 31 degrees F given data on the previous 23 shuttle flights </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Integer </td>
<td> 23 </td>
<td> 4 </td>
<td> 1993 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Low+Resolution+Spectrometer">Low Resolution Spectrometer </td></tr></tab
le></td>
<!-- <td> From IRAS data -- NASA Ames Research Center </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 531 </td>
<td> 102 </td>
<td> 1988 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img src="assets/MLimages/SmallLarge94.jpg" bord
er=1 /> </td><td> Spambase </t
d></tr></table></td>
<!-- <td> Classifying Email as Spam or Non-Spam </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 4601 </td>
<td> 57 </td>
<td> 1999 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/SPECT+Heart"><img src="assets/MLimages/SmallLarge45.jpg" b

```
<td><table><tr><td><a href="datasets/SPECT+Heart"><img src="assets/MLimages/SmallLarge45.jpg" b
order=1 /></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) i
mages. 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">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><a href="datasets/SPECTF+Heart"><img src="assets/MLimages/SmallLarge45.jpg"
border=1 /></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) i
mages. 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><a href="datasets/Sponge"><img src="assets/MLimages/SmallLarge97.jpg"
border=1 /></a> </td><td><p class="normal"><b><a href="datasets/Sponge">Sponge</a></b></p></t
d></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">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 </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Statlog+Project"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 silhouttes, Landsat Sattelite, Shuttle, A
ustralian Credit Approval, Heart Disease, Image Segmentation, German Credit </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">1992 </p></td>
   <!-- <td><p class="normal">Other </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Student+Loan+Relational"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 </p></td> -->
   <td><p class="normal">Domain-Theory </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal">1000 </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal">1993 </p></td>
   <!-- <td><p class="normal">Social </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Teaching+Assistant+Evaluation"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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" </p></td> -->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Categorical, Integer </p></td>
   <td><p class="normal">151 </p></td>
   <td><p class="normal">5 </p></td>
   <td><p class="normal">1997 </p></td>
   <!-- <td><p class="normal">Other </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Tic-Tac-Toe+Endgame"><img
```

<td><table><tr><td><img
src="assets/MLimages/SmallLarge101.jpg" border=1 /> </td><td> <a href
="datasets/Tic-Tac-Toe+Endgame">Tic-Tac-Toe Endgame </td></tr></table></td>
<!-- <td> Binary classification task on possible configurations of tic-tac-toe
game </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 958 </td>
<td> 9 </td>
<td> 1991 </td>
<!-- <td> Game </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Thyroid+Disease">Thyroid Disease </td></tr></table></td>
<!-- <td> 10 separate databases from Garavan Institute </td> -->
<td> Multivariate, Domain-Theory </td>
<td> Classification </td>
<td> Categorical, Real </td>
<td> 7200 </td>
<td> 21 </td>
<td> 1987 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><img src="assets/MLimages/SmallLarge103.jpg" borde
r=1 /> </td><td> Trains </td></t
r></table></td>
<!-- <td> 2 data formats (structured, one-instance-per-line) </td> --
>
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 10 </td>
<td> 32 </td>
<td> 1994 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img src="assets/MLimages/SmallLarge104.jpg" b
order=1 /> </td><td> University
</td></tr></table></td>
<!-- <td> Data in original (LISP-readable) form </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 285 </td>
<td> 17 </td>
<td> 1988 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLarge105.jpg" border=1 /> </td><td> <a href
="datasets/Congressional+Voting+Records">Congressional Voting Records </td></tr></table
></td>
<!-- <td> 1984 United Stated Congressional Voting Records; Classify as
Republican or Democrat </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 435 </td>
<td> 16 </td>
<td> 1987 </td>
<!-- <td> Social </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Water+Treatment+Plant">Water Treatment Plant </td></tr></table></td>
<!-- <td> Multiple classes predict plant state </td> -->
<td> Multivariate </td>
<td> Clustering </td>
<td> Integer, Real </td>
<td> 527 </td>
<td> 38 </td>
<td> 1993 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a

```
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Waveform+Database+Generator+%28Version+1%29">Waveform Database Generator (Version 1
)</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">CART book's waveform domains </p></td> -->
   <td><p class="normal">Multivariate, Data-Generator </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Real </p></td>
   <td><p class="normal">5000 </p></td>
   <td><p class="normal">21 </p></td>
   <td><p class="normal">1988 </p></td>
   <!-- <td><p class="normal">Physical </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Waveform+Database+Generator+%28Version+2%29"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Waveform+Database+Generator+%28Version+2%29">Waveform Database Generator (Version 2
)</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">CART book's waveform domains </p></td> -->
   <td><p class="normal">Multivariate, Data-Generator </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Real </p></td>
   <td><p class="normal">5000 </p></td>
   <td><p class="normal">40 </p></td>
   <td><p class="normal">1988 </p></td>
   <!-- <td><p class="normal">Physical </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Wine"><img src="assets/MLimages/SmallLarge109.jpg" border=
1 /></a> </td><td><p class="normal"><b><a href="datasets/Wine">Wine</a></b></p></td></tr></ta
ble></td>
   <!-- <td><p class="normal">Using chemical analysis determine the origin of wines </p></td>
-->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal">Classification </p></td>
   <td><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 </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Yeast"><img src="assets/MLimages/SmallLarge110.jpg" border
=1 /></a> </td><td><p class="normal"><b><a href="datasets/Yeast">Yeast</a></b></p></td></tr><
/table></td>
   <!-- <td><p class="normal">Predicting the Cellular Localization Sites of Proteins </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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Zoo"><img src="assets/MLimages/SmallLarge111.jpg" border=1
/></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 </p></td> -->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Categorical, Integer </p></td>
   <td><p class="normal">101 </p></td>
   <td><p class="normal">17 </p></td>
   <td><p class="normal">1990 </p></td>
   <!-- <td><p class="normal">Life </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Undocumented"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Undocumented">Undocumented</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">Various datasets without documentation (feel free to explore!) 
</p></td> -->
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <!-- <td><p class="normal">Other </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Twenty+Newsgroups"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
```

src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Twenty+Newsgroups">Twenty Newsgroups </td></tr></table></td>
<!-- <td> This data set consists of 20000 messages taken from 20
newsgroups. </td> -->
<td> Text </td>
<td>   </td>
<td>   </td>
<td> 20000 </td>
<td>   </td>
<td> 1999 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLarge114.jpg" border=1 /> </td><td> <a href
="datasets/Australian+Sign+Language+signs">Australian Sign Language signs </td></tr></t
able></td>
<!-- <td> This data consists of sample of Auslan (Australian Sign Language) si
gns. Examples of 95 signs were collected from five signers with a total of 6650 sign
samples. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Categorical, Real </td>
<td> 6650 </td>
<td> 15 </td>
<td> 1999 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><img sr
c="assets/MLimages/SmallLarge114.jpg" border=1 /> </td><td> <a
href="datasets/Australian+Sign+Language+signs+%28High+Quality%29">Australian Sign Language signs (
High Quality) </td></tr></table></td>
<!-- <td> This data consists of sample of Auslan (Australian Sign Language) si
gns. 27 examples of each of 95 Auslan signs were captured from a native signer using high-quality
position trackers </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 2565 </td>
<td> 22 </td>
<td> 2002 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLarge2.jpg" border=1 /> </td><td> <a
href="datasets/US+Census+Data+%281990%29">US Census Data (1990) </td></tr></table></td>
<!-- <td> The USCensus1990raw data set contains a one percent sample of the Pu
blic Use Microdata Samples (PUMS) person records drawn from the full 1990 census sample.
</td> -->
<td> Multivariate </td>
<td> Clustering </td>
<td> Categorical </td>
<td> 2458285 </td>
<td> 68 </td>
<td>   </td>
<!-- <td> Social </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLarge2.jpg" border=1 /> </td><td> <a
href="datasets/Census-Income+%28KDD%29">Census-Income (KDD) </td></tr></table></td>
<!-- <td> This data set contains weighted census data extracted from the 1994
and 1995 current population surveys conducted by the U.S. Census Bureau. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 299285 </td>
<td> 40 </td>
<td> 2000 </td>
<!-- <td> Social </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLarge118.jpg" border=1 /> </td><td> <a href
="datasets/Coil+1999+Competition+Data">Coil 1999 Competition Data </td></tr></table></t
d>
<!-- <td> 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. </td> -->
<td> Multivariate </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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Corel+Image+Features"><img
src="assets/MLimages/SmallLarge119.jpg" border=1 /></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 co
llection. Four sets of features are available based on the color histogram, color histogram
layout, color moments, and co-occurence </p></td> -->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal">Real </p></td>
   <td><p class="normal">68040 </p></td>
   <td><p class="normal">89 </p></td>
   <td><p class="normal">1999 </p></td>
   <!-- <td><p class="normal">Other </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/E.+Coli+Genes"><img
src="assets/MLimages/SmallLarge120.jpg" border=1 /></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. c
oli genome. Sequence, homology (similarity to other genes) and structural information, and
function (if known) are provided. </p></td> -->
   <td><p class="normal">Relational </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal">2001 </p></td>
   <!-- <td><p class="normal">Life </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/EEG+Database"><img src="assets/MLimages/SmallLarge121.jpg"
border=1 /></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 </p></td> -->
   <td><p class="normal">Multivariate, Time-Series </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal">Categorical, Integer, Real </p></td>
   <td><p class="normal">122 </p></td>
   <td><p class="normal">4 </p></td>
   <td><p class="normal">1999 </p></td>
   <!-- <td><p class="normal">Life </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/El+Nino"><img src="assets/MLimages/SmallLargedefault.jpg"
border=1 /></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 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><a href="datasets/Entree+Chicago+Recommendation+Data"><img
src="assets/MLimages/SmallLarge123.jpg" border=1 /></a> </td><td><p class="normal"><b><a href
="datasets/Entree+Chicago+Recommendation+Data">Entree Chicago Recommendation Data</a></b></p></td>
</tr></table></td>
   <!-- <td><p class="normal">This data contains a record of user interactions with the Entree Chi
cago 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"> </p></td>
   <td><p class="normal">2000 </p></td>
   <!-- <td><p class="normal">Other </p></td> -->
   </tr><tr bgcolor="DDEEFF">
```

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

<td> 191779 </td>
<td> 481 </td>
<td> 1998 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/KDD+Cup+1999+Data">KDD Cup 1999 Data </td></tr></table></td>
<!-- <td> This is the data set used for The Third International Knowledge Disc
overy and Data Mining Tools Competition, which was held in conjunction with KDD-99 </td>
-->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 4000000 </td>
<td> 42 </td>
<td> 1999 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLarge131.jpg" border=1 /> </td><td> <a href
="datasets/M.+Tuberculosis+Genes">M. Tuberculosis Genes </td></tr></table></td>
<!-- <td> Data giving characteristics of each ORF (potential gene) in the M.
tuberculosis bacterium. Sequence, homology (similarity to other genes) and structural information,
and function (if known) are provided </td> -->
<td> Relational </td>
<td>   </td>
<td>   </td>
<td>   </td>
<td>   </td>
<td> 2001 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img src="assets/MLimages/SmallLarge132.jpg" border
=1 /> </td><td> Movie </td></tr><
/table></td>
<!-- <td> This data set contains a list of over 10000 films including many old
er, odd, and cult films. There is information on actors, casts, directors, producers, studios, etc
. </td> -->
<td> Multivariate, Relational </td>
<td>   </td>
<td>   </td>
<td> 10000 </td>
<td>   </td>
<td> 1999 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/MSNBC.com+Anonymous+Web+Data">MSNBC.com Anonymous Web Data </td></tr></t
able></td>
<!-- <td> This data describes the page visits of users who visited msnbc.com o
n September 28, 1999. Visits are recorded at the level of URL category (see description) and are r
ecorded in time order. </td> -->
<td> Sequential </td>
<td>   </td>
<td> Categorical </td>
<td> 989818 </td>
<td>   </td>
<td>   </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLarge134.jpg" border=1 /> </td><td> <a href
="datasets/NSF+Research+Award+Abstracts+1990-2003">NSF Research Award Abstracts 1990-2003<
/p></td></tr></table></td>
<!-- <td> This data set consists of (a) 129,000 abstracts describing NSF award
s for basic research, (b) bag-of-word data files extracted from the abstracts, (c) a list of words
used for indexing the bag-of-word </td> -->
<td> Text </td>
<td>   </td>
<td>   </td>
<td> 129000 </td>
<td>   </td>
<td> 2003 </td>
<!-- <td> Other </td> -->
</tr><tr>

<td><table><tr><td> </td><td> Pioneer-1 Mobile Robot Data </td></tr></table></td>
<!-- <td> This dataset contains time series sensor readings of the Pioneer-1 mobile robot. The data is broken into "experiences" in which the robot takes action for some period of time and experiences a control </td> -->
<td> Multivariate, Time-Series </td>
<td>   </td>
<td> Categorical, Real </td>
<td>   </td>
<td>   </td>
<td> 1999 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> Pseudo Periodic Synthetic Time Series </td></tr></table></td>
<!-- <td> This data set is designed for testing indexing schemes in time series databases. The data appears highly periodic, but never exactly repeats itself. </td> -->
<td> Univariate, Time-Series </td>
<td>   </td>
<td>   </td>
<td> 100000 </td>
<td>   </td>
<td> 1999 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td> </td><td> Reuters-21578 Text Categorization Collection </td></tr></table></td>
<!-- <td> This is a collection of documents that appeared on Reuters newswire in 1987. The documents were assembled and indexed with categories. </td> -->
<td> Text </td>
<td> Classification </td>
<td> Categorical </td>
<td> 21578 </td>
<td> 5 </td>
<td> 1997 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> Robot Execution Failures </td></tr></table></td>
<!-- <td> 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 </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Integer </td>
<td> 463 </td>
<td> 90 </td>
<td> 1999 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td> </td><td> Synthetic Control Chart Time Series </td></tr></table></td>
<!-- <td> This data consists of synthetically generated control charts. </td> -->
<td> Time-Series </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 600 </td>
<td>   </td>
<td> 1999 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> Syskill and Webert Web Page Ratings </td></tr></table></td>
<!-- <td> This database contains HTML source of web pages plus the ratings of

```
a single user on these web pages. Web pages are on four seperate subjects (Bands- recording artist
s; 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><a href="datasets/UNIX+User+Data"><img
src="assets/MLimages/SmallLarge141.jpg" border=1 /></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 comm
and 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>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <!-- <td><p class="normal">Computer </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Volcanoes+on+Venus+-+JARtool+experiment"><img
src="assets/MLimages/SmallLarge142.jpg" border=1 /></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 sp
acecraft. </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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Statlog+%28Australian+Credit+Approval%29"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 el
sewhere 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><a href="datasets/Statlog+%28German+Credit+Data%29"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Statlog+%28German+Credit+Data%29">Statlog (German Credit Data)</a></b></p></td></tr
></table></td>
   <!-- <td><p class="normal">This dataset classifies people described by a set of attributes as g
ood 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><a href="datasets/Statlog+%28Heart%29"><img
src="assets/MLimages/SmallLarge45.jpg" border=1 /></a> </td><td><p class="normal"><b><a href=
"datasets/Statlog+%28Heart%29">Statlog (Heart)</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">This dataset is a heart disease database similar to a database
already present in the repository (Heart Disease databases) but in a slightly different form 
</p></td> -->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal">Classification </p></td>
```

<td> Categorical, Real </td>
<td> 270 </td>
<td> 13 </td>
<td>   </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLarge146.jpg" border=1 /> </td><td> <a href
="datasets/Statlog+%28Landsat+Satellite%29">Statlog (Landsat Satellite) </td></tr></tab
le></td>
<!-- <td> Multi-spectral values of pixels in 3x3 neighbourhoods in a satellite
image, and the classification associated with the central pixel in each neighbourhood </t
d> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 6435 </td>
<td> 36 </td>
<td> 1993 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Statlog+%28Image+Segmentation%29">Statlog (Image Segmentation) </td></tr
></table></td>
<!-- <td> This dataset is an image segmentation database similar to a database
already present in the repository (Image segmentation database) but in a slightly different form.&
nbsp; </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 2310 </td>
<td> 19 </td>
<td> 1990 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLarge92.jpg" border=1 /> </td><td> <a href=
"datasets/Statlog+%28Shuttle%29">Statlog (Shuttle) </td></tr></table></td>
<!-- <td> The shuttle dataset contains 9 attributes all of which are
numerical. Approximately 80% of the data belongs to class 1 </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 58000 </td>
<td> 9 </td>
<td>   </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLarge149.jpg" border=1 /> </td><td> <a href
="datasets/Statlog+%28Vehicle+Silhouettes%29">Statlog (Vehicle Silhouettes) </td></tr><
/table></td>
<!-- <td> 3D objects within a 2D image by application of an ensemble of shape
feature extractors to the 2D silhouettes of the objects. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 946 </td>
<td> 18 </td>
<td>   </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Connectionist+Bench+%28Nettalk+Corpus%29">Connectionist Bench (Nettalk Corpus)<
/b> </td></tr></table></td>
<!-- <td> The file "nettalk.data" contains a list of 20,008 English words, alo
ng with a phonetic transcription for each word. The task is to train a network to produce the prop
er phonemes </td> -->
<td> Multivariate </td>
<td>   </td>
<td> Categorical </td>
<td> 20008 </td>
<td> 4 </td>
<td>   </td>
<!-- <td> Other </td> -->

```
   </tr><tr>
   <td><table><tr><td><a href="datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29"><img s
rc="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a h
ref="datasets/Connectionist+Bench+%28Sonar%2C+Mines+vs.+Rocks%29">Connectionist Bench (Sonar,
Mines vs. Rocks)</a></b></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. </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 </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Connectionist+Bench+%28Vowel+Recognition+-
+Deterding+Data%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td
><p class="normal"><b><a href="datasets/Connectionist+Bench+%28Vowel+Recognition+-
+Deterding+Data%29">Connectionist Bench (Vowel Recognition - Deterding Data)</a></b></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. </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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Economic+Sanctions"><img
src="assets/MLimages/SmallLarge153.jpg" border=1 /></a> </td><td><p class="normal"><b><a href
="datasets/Economic+Sanctions">Economic Sanctions</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">Domain Theory on Economic Sanctions; Undocumented </p></td> -->
   <td><p class="normal">Domain-Theory </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal"> </p></td>
   <!-- <td><p class="normal">Financial </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Protein+Data"><img src="assets/MLimages/SmallLarge154.jpg"
border=1 /></a> </td><td><p class="normal"><b><a href="datasets/Protein+Data">Protein
Data</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">Undocumented </p></td> -->
   <td><p class="normal"> </p></td>
   <td><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">Life </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Cloud"><img src="assets/MLimages/SmallLarge155.jpg" border
=1 /></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 </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><a href="datasets/CalIt2+Building+People+Counts"><img
src="assets/MLimages/SmallLarge156.jpg" border=1 /></a> </td><td><p class="normal"><b><a href
="datasets/CalIt2+Building+People+Counts">CalIt2 Building People Counts</a></b></p></td></tr></tab
le></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><a href="datasets/Dodgers+Loop+Sensor"><img
src="assets/MLimages/SmallLarge157.jpg" border=1 /></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 </p></td> -->
   <td><p class="normal">Multivariate, Time-Series </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal">Categorical, Integer </p></td>
   <td><p class="normal">50400 </p></td>
   <td><p class="normal">3 </p></td>
   <td><p class="normal">2006 </p></td>
   <!-- <td><p class="normal">Other </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Poker+Hand"><img src="assets/MLimages/SmallLarge158.jpg" b
order=1 /></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 </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><a href="datasets/MAGIC+Gamma+Telescope"><img
src="assets/MLimages/SmallLarge159.jpg" border=1 /></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 </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><a href="datasets/UJI+Pen+Characters"><img
src="assets/MLimages/SmallLarge160.jpg" border=1 /></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 </p
></td> -->
   <td><p class="normal">Multivariate, Sequential </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Integer </p></td>
   <td><p class="normal">1364 </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal">2007 </p></td>
   <!-- <td><p class="normal">Computer </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Mammographic+Mass"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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. </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 </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Forest+Fires"><img src="assets/MLimages/SmallLarge162.jpg"
border=1 /></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 othe
r 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><a href="datasets/Reuters+Transcribed+Subset"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Reuters+Transcribed+Subset">Reuters Transcribed Subset</a></b></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. </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 </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Bag+of+Words"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Bag+of+Words">Bag of Words</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">This data set contains five text collections in the form of bags-of-
words. </p></td> -->
   <td><p class="normal">Text </p></td>
   <td><p class="normal">Clustering </p></td>
   <td><p class="normal">Integer </p></td>
   <td><p class="normal">8000000 </p></td>
   <td><p class="normal">100000 </p></td>
   <td><p class="normal">2008 </p></td>
   <!-- <td><p class="normal">Other </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Concrete+Compressive+Strength"><img
src="assets/MLimages/SmallLarge165.jpg" border=1 /></a> </td><td><p class="normal"><b><a href
="datasets/Concrete+Compressive+Strength">Concrete Compressive Strength</a></b></p></td></tr></tab
le></td>
   <!-- <td><p class="normal">Concrete is the most important material in civil engineering. The co
ncrete compressive strength is a highly nonlinear function of age and ingredients.  </p></td>
-->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal">Regression </p></td>
   <td><p class="normal">Real </p></td>
   <td><p class="normal">1030 </p></td>
   <td><p class="normal">9 </p></td>
   <td><p class="normal">2007 </p></td>
   <!-- <td><p class="normal">Physical </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Hill-Valley"><img src="assets/MLimages/SmallLarge166.jpg"
border=1 /></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 p
lotted 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></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><a href="datasets/Arcene"><img src="assets/MLimages/SmallLarge167.jpg" borde
r=1 /></a> </td><td><p class="normal"><b><a href="datasets/Arcene">Arcene</a></b></p></td></t
r></table></td>
   <!-- <td><p class="normal">ARCENE's task is to distinguish cancer versus normal patterns from m
ass-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><img src="assets/MLimages/SmallLarge168.jpg" borde
r=1 /> </td><td> Dexter </td></t
r></table></td>
<!-- <td> 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.
  </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 2600 </td>
<td> 20000 </td>
<td> 2008 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><img src="assets/MLimages/SmallLarge169.jpg" bor
der=1 /> </td><td> Dorothea </
td></tr></table></td>
<!-- <td> DOROTHEA is a drug discovery dataset. Chemical compounds represented
by structural molecular features must be classified as active (binding to thrombin) or inactive. T
his is one of 5 datasets of the NIPS 2003 feature selection challenge. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 1950 </td>
<td> 100000 </td>
<td> 2008 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img src="assets/MLimages/SmallLarge170.jpg" bord
er=1 /> </td><td> Gisette </td>
</tr></table></td>
<!-- <td> GISETTE is a handwritten digit recognition problem. The problem is t
o separate the highly confusible digits '4' and '9'. This dataset is one of five datasets of the N
IPS 2003 feature selection challenge.
  </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 13500 </td>
<td> 5000 </td>
<td> 2008 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img src="assets/MLimages/SmallLarge171.jpg" bord
er=1 /> </td><td> Madelon </td>
</tr></table></td>
<!-- <td> MADELON is an artificial dataset, which was part of the NIPS 2003 fe
ature 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.   </t
d> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 4400 </td>
<td> 500 </td>
<td> 2008 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLarge172.jpg" border=1 /> </td><td> <a href
="datasets/Ozone+Level+Detection">Ozone Level Detection </td></tr></table></td>
<!-- <td> Two ground ozone level data sets are included in this collection. On
e is the eight hour peak set (eighthr.data), the other is the one hour peak set (onehr.data). Thos
e data were collected from 1998 to 2004 at the Houston, Galveston and Brazoria area. </td
> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 2536 </td>
<td> 73 </td>
<td> 2008 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Abscisic+Acid+Signaling+Network">Abscisic Acid Signaling Network </td></

tr></table></td>
<!-- <td> The objective is to determine the set of boolean rules that describe the interactions of the nodes within this plant signaling network. The dataset includes 300 separate boolean pseudodynamic simulations using an asynchronous update scheme.   </td> -->
<td> Multivariate </td>
<td> Causal-Discovery </td>
<td> Integer </td>
<td> 300 </td>
<td> 43 </td>
<td> 2008 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> Parkinsons </td></tr></table></td>
<!-- <td> Oxford Parkinson's Disease Detection Dataset </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 197 </td>
<td> 23 </td>
<td> 2008 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td> </td><td> Character Trajectories </td></tr></table></td>
<!-- <td> Multiple, labelled samples of pen tip trajectories recorded whilst writing individual characters. All samples are from the same writer, for the purposes of primitive extraction. Only characters with a single pen-down segment were considered. </td> -->
<td> Time-Series </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 2858 </td>
<td> 3 </td>
<td> 2008 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> Blood Transfusion Service Center </td></tr></table></td>
<!-- <td> Data taken from the Blood Transfusion Service Center in Hsin-Chu City in Taiwan -- this is a classification problem.   </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 748 </td>
<td> 5 </td>
<td> 2008 </td>
<!-- <td> Business </td> -->
</tr><tr>
<td><table><tr><td> </td><td> UJI Pen Characters (Version 2) </td></tr></table></td>
<!-- <td> A pen-based database with more than 11k isolated handwritten characters </td> -->
<td> Multivariate, Sequential </td>
<td> Classification </td>
<td> Integer </td>
<td> 11640 </td>
<td>   </td>
<td> 2009 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> Semeion Handwritten Digit </td></tr></table></td>
<!-- <td> 1593 handwritten digits from around 80 persons were scanned, stretched in a rectangular box 16x16 in a gray scale of 256 values. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 1593 </td>
<td> 256 </td>
<td> 2008 </td>

<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img src="assets/MLimages/SmallLarge179.jpg" border
=1 /> </td><td> SECOM </td></tr><
/table></td>
<!-- <td> Data from a semi-conductor manufacturing process </td> -->
<td> Multivariate </td>
<td> Classification, Causal-Discovery </td>
<td> Real </td>
<td> 1567 </td>
<td> 591 </td>
<td> 2008 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img src="assets/MLimages/SmallLarge180.jpg" borde
r=1 /> </td><td> Plants </td></t
r></table></td>
<!-- <td> Data has been extracted from the USDA plants database. It contains a
ll plants (species and genera) in the database and the states of USA and Canada where they occur.&
nbsp; </td> -->
<td> Multivariate </td>
<td> Clustering </td>
<td> Categorical </td>
<td> 22632 </td>
<td> 70 </td>
<td> 2008 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLarge181.jpg" border=1 /> </td><td> <a href
="datasets/Libras+Movement">Libras Movement </td></tr></table></td>
<!-- <td> The data set contains 15 classes of 24 instances each. Each class
references to a hand movement type in LIBRAS (Portuguese
name 'L[illegible]ngua BRAsileira de Sinais', oficial brazilian signal language). </td> -->
<td> Multivariate, Sequential </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 360 </td>
<td> 91 </td>
<td> 2009 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLarge165.jpg" border=1 /> </td><td> <a href
="datasets/Concrete+Slump+Test">Concrete Slump Test </td></tr></table></td>
<!-- <td> Concrete is a highly complex material. The slump flow of concrete is
not only determined by the water content, but that is also influenced by other concrete
ingredients. </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 103 </td>
<td> 10 </td>
<td> 2009 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLarge183.jpg" border=1 /> </td><td> <a href
="datasets/Communities+and+Crime">Communities and Crime </td></tr></table></td>
<!-- <td> Communities within the United States. The data combines socio-econom
ic data from the 1990 US Census, law enforcement data from the 1990 US LEMAS survey, and crime dat
a from the 1995 FBI UCR. </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 1994 </td>
<td> 128 </td>
<td> 2009 </td>
<!-- <td> Social </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLarge184.jpg" border=1 /> </td><td> <a href
="datasets/Acute+Inflammations">Acute Inflammations </td></tr></table></td>
<!-- <td> The data was created by a medical expert as a data set to test the e
xpert system,
which will perform the presumptive diagnosis of two diseases of the urinary system.
  </td> -->

<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer </td>
<td> 120 </td>
<td> 6 </td>
<td> 2009 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td> </td><td> Wine Quality </td></tr></table></td>
<!-- <td> Two datasets are included, related to red and white vinho verde wine samples, from the north of Portugal. The goal is to model wine quality based on physicochemical tests (see [Cortez et al., 2009], http://www3.dsi.uminho.pt/pcortez/wine/). </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Real </td>
<td> 4898 </td>
<td> 12 </td>
<td> 2009 </td>
<!-- <td> Business </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> URL Reputation </td></tr></table></td>
<!-- <td> Anonymized 120-day subset of the ICML-09 URL data containing 2.4 million examples and 3.2 million features. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 2396130 </td>
<td> 3231961 </td>
<td> 2009 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td> </td><td> p53 Mutants </td></tr></table></td>
<!-- <td> The goal is to model mutant p53 transcriptional activity (active vs inactive) based on data extracted from biophysical simulations.
  </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 16772 </td>
<td> 5409 </td>
<td> 2010 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> Parkinsons Telemonitoring </td></tr></table></td>
<!-- <td> Oxford Parkinson's Disease Telemonitoring Dataset </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Integer, Real </td>
<td> 5875 </td>
<td> 26 </td>
<td> 2009 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td> </td><td> Demospongiae </td></tr></table></td>
<!-- <td> Marine sponges of the Demospongiae class classification domain. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 503 </td>
<td>   </td>
<td> 2010 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> Opinosis Opinion ⁄ Review </td></tr>

```
<td><table><tr><td><a href="datasets/Opinosis+Opinion+%20Review">Opinosis Opinion &#x2013; Review</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">This dataset contains sentences extracted from user reviews on a giv
en topic. Example topics are “performance of Toyota Camry” and “sound quality of ipod nano”.  
;</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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Breast+Tissue"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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. </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><a href="datasets/Cardiotocography"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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. </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><a href="datasets/Wall-Following+Robot+Navigation+Data"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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'. </p></td> -->
   <td><p class="normal">Multivariate, Sequential </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Real </p></td>
   <td><p class="normal">5456 </p></td>
   <td><p class="normal">24 </p></td>
   <td><p class="normal">2010 </p></td>
   <!-- <td><p class="normal">Computer </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Spoken+Arabic+Digit"><img
src="assets/MLimages/SmallLarge195.jpg" border=1 /></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 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>
   <td><p class="normal">2010 </p></td>
   <!-- <td><p class="normal">Other </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Localization+Data+for+Person+Activity"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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. &n
bsp;</p></td> -->
   <td><p class="normal">Univariate, Sequential, Time-Series </p></td>
   <td><p class="normal">Classification </p></td>
```

<td> Classification </td>
<td> Real </td>
<td> 164860 </td>
<td> 8 </td>
<td> 2010 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> AutoUniv </td></tr></table></td>
<!-- <td> AutoUniv is an advanced data generator for classifications tasks. The aim is to reflect the nuances and heterogeneity of real data. Data can be generated in .csv, ARFF or C4.5 formats. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical, Integer, Real </td>
<td>   </td>
<td>   </td>
<td> 2010 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td> </td><td> Steel Plates Faults </td></tr></table></td>
<!-- <td> A dataset of steel plates’ faults, classified into 7 different types.
The goal was to train machine learning for automatic pattern recognition.
  </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 1941 </td>
<td> 27 </td>
<td> 2010 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> MiniBooNE particle identification </td></tr></table></td>
<!-- <td> This dataset is taken from the MiniBooNE experiment and is used to distinguish electron neutrinos (signal) from muon neutrinos (background). </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 130065 </td>
<td> 50 </td>
<td> 2010 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td> </td><td> YearPredictionMSD </td></tr></table></td>
<!-- <td> Prediction of the release year of a song from audio features. Songs are mostly western, commercial tracks ranging from 1922 to 2011, with a peak in the year 2000s. </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 515345 </td>
<td> 90 </td>
<td> 2011 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> PEMS-SF </td></tr></table></td>
<!-- <td> 15 months worth of daily data (440 daily records) that describes the occupancy rate, between 0 and 1, of different car lanes of the San Francisco bay area freeways across time. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 440 </td>
<td> 138672 </td>
<td> 2011 </td>
<!-- <td> Computer </td> -->
</tr><tr>

</td></tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/OpinRank+Review+Dataset">OpinRank Review Dataset </td></tr></table></td>
<!-- <td> This data set contains user reviews of cars and and hotels collected
from Tripadvisor (~259,000
reviews) and Edmunds (~42,230 reviews).   </td> -->
<td> Text </td>
<td>   </td>
<td>   </td>
<td>   </td>
<td> 2011 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Relative+location+of+CT+slices+on+axial+axis">Relative location of CT slices on axi
al axis </td></tr></table></td>
<!-- <td> The dataset consists of 384 features extracted from CT images. The c
lass variable is numeric and denotes the relative location of the CT slice on the axial axis of th
e human body. </td> -->
<td> Domain-Theory </td>
<td> Regression </td>
<td> Real </td>
<td> 53500 </td>
<td> 386 </td>
<td> 2011 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img src="
assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a href=
"datasets/Online+Handwritten+Assamese+Characters+Dataset">Online Handwritten Assamese Characters D
ataset </td></tr></table></td>
<!-- <td> This is a dataset of 8235 online handwritten assamese characters. Th
e “online” process involves capturing of data as text is written on a digitizing tablet with an el
ectronic pen. </td> -->
<td> Multivariate, Sequential </td>
<td> Classification </td>
<td> Integer </td>
<td> 8235 </td>
<td>   </td>
<td> 2011 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/PubChem+Bioassay+Data">PubChem Bioassay Data </td></tr></table></td>
<!-- <td> These highly imbalanced bioassay datasets are from the differing typ
es of screening that can be performed using HTS technology. 21 datasets were created from 12 bioas
says. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td>   </td>
<td>   </td>
<td> 2011 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Record+Linkage+Comparison+Patterns">Record Linkage Comparison Patterns <
/td></tr></table></td>
<!-- <td> Element-wise comparison of records with personal data from a record
linkage setting. The task is to decide from a comparison pattern whether the underlying records be
long to one person. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 5749132 </td>
<td> 12 </td>
<td> 2011 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Communities+and+Crime+Unnormalized">Communities and Crime Unnormalized <
/td></tr></table></td>

</td></tr></table></td>
<!-- <td> Communities in the US. Data combines socio-economic data from the '90 Census, law enforcement data from the 1990 Law Enforcement Management and Admin Stats survey, and crime data from the 1995 FBI UCR </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 2215 </td>
<td> 147 </td>
<td> 2011 </td>
<!-- <td> Social </td> -->
</tr><tr>
<td><table><tr><td> </td><td> Vertebral Column </td></tr></table></td>
<!-- <td> Data set containing values for six biomechanical features used to classify orthopaedic patients into 3 classes (normal, disk hernia or spondilolysthesis) or 2 classes (normal or abnormal). </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 310 </td>
<td> 6 </td>
<td> 2011 </td>
<!-- <td>   </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> EMG Physical Action Data Set </td></tr></table></td>
<!-- <td> The Physical Action Data Set includes 10 normal and 10 aggressive physical actions that measure the human activity. The data have been collected by 4 subjects using the Delsys EMG wireless apparatus. </td> -->
<td> Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 10000 </td>
<td> 8 </td>
<td> 2011 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td> </td><td> Vicon Physical Action Data Set </td></tr></table></td>
<!-- <td> The Physical Action Data Set includes 10 normal and 10 aggressive physical actions that measure the human activity. The data have been collected by 10 subjects using the Vicon 3D tracker. </td> -->
<td> Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 3000 </td>
<td> 27 </td>
<td> 2011 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> Amazon Commerce reviews set </td></tr></table></td>
<!-- <td> The dataset is used for authorship identification in online Writeprint which is a new research field of pattern recognition.   </td> -->
<td> Multivariate, Text, Domain-Theory </td>
<td> Classification </td>
<td> Real </td>
<td> 1500 </td>
<td> 10000 </td>
<td> 2011 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td> </td><td> Amazon Access Samples </td></tr></table></td>
<!-- <td> Amazon's InfoSec is getting smarter about the way Access data is leveraged. This is an anonymized sample of access provisioned within the company. </td> -->
<td> Time-Series, Domain-Theory </td>
<td> Regression, Clustering, Causal-Discovery </td>

<td> Regression, Clustering, Causal-Discovery </td>
<td>   </td>
<td> 30000 </td>
<td> 20000 </td>
<td> 2011 </td>
<!-- <td> Business </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Reuter_50_50">Reuter_50_50 </td></tr></table></td>
<!-- <td> The dataset is used for authorship identification in online
Writeprint which is a new research field of pattern recognition.   </td> -->
<td> Multivariate, Text, Domain-Theory </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 2500 </td>
<td> 10000 </td>
<td> 2011 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img src="assets/MLimages/SmallLargedefault.jpg"
border=1 /> </td><td> Farm Ads</p
></td></tr></table></td>
<!-- <td> This data was collected from text ads found on twelve websites that
deal with various farm animal related topics. The binary labels are based on whether or not the c
ontent owner approves of the ad. </td> -->
<td> Text </td>
<td> Classification </td>
<td>   </td>
<td> 4143 </td>
<td> 54877 </td>
<td> 2011 </td>
<!-- <td> Business </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/DBWorld+e-mails">DBWorld e-mails </td></tr></table></td>
<!-- <td> It contains 64 e-mails which I have manually collected from DBWorld
mailing list. They are classified in: 'announces of conferences' and 'everything else'. <
/td> -->
<td> Text </td>
<td> Classification </td>
<td>   </td>
<td> 64 </td>
<td> 4702 </td>
<td> 2011 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img src="
assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a href=
"datasets/KEGG+Metabolic+Relation+Network+%28Directed%29">KEGG Metabolic Relation Network
(Directed) </td></tr></table></td>
<!-- <td> KEGG Metabolic pathways modeled as directed relation network. Variet
y of graphical features presented. </td> -->
<td> Multivariate, Univariate, Text </td>
<td> Classification, Regression, Clustering </td>
<td> Integer, Real </td>
<td> 53414 </td>
<td> 24 </td>
<td> 2011 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img src
="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a hre
f="datasets/KEGG+Metabolic+Reaction+Network+%28Undirected%29">KEGG Metabolic Reaction Network (Und
irected) </td></tr></table></td>
<!-- <td> KEGG Metabolic pathways modeled as un-directed reaction network. Var
iety of graphical features presented. </td> -->
<td> Multivariate, Univariate, Text </td>
<td> Classification, Regression, Clustering </td>
<td> Integer, Real </td>
<td> 65554 </td>
<td> 29 </td>
<td> 2011 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Bank+Marketing"><img

```
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Bank+Marketing">Bank Marketing</a></b></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><a href="datasets/YouTube+Comedy+Slam+Preference+Data"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/YouTube+Comedy+Slam+Preference+Data">YouTube Comedy Slam Preference Data</a></b></p
></td></tr></table></td>
   <!-- <td><p class="normal">This dataset provides user vote data on which video from a pair of v
ideos 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>
   <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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Gas+Sensor+Array+Drift+Dataset"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Gas+Sensor+Array+Drift+Dataset">Gas Sensor Array Drift Dataset</a></b></p></td></tr
></table></td>
   <!-- <td><p class="normal">This archive contains 13910 measurements from 16 chemical sensors ut
ilized in simulations for drift compensation in a discrimination task of 6 gases at various 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></td>
   <!-- <td><p class="normal">Computer </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/ILPD+%28Indian+Liver+Patient+Dataset%29"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/ILPD+%28Indian+Liver+Patient+Dataset%29">ILPD (Indian Liver Patient Dataset)</a></b
></p></td></tr></table></td>
   <!-- <td><p class="normal">This data set contains 10 variables that are age, gender, total Bili
rubin, direct Bilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT and Alkphos. </p></td
> -->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Integer, Real </p></td>
   <td><p class="normal">583 </p></td>
   <td><p class="normal">10 </p></td>
   <td><p class="normal">2012 </p></td>
   <!-- <td><p class="normal">Life </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/OPPORTUNITY+Activity+Recognition"><img
src="assets/MLimages/SmallLarge226.jpg" border=1 /></a> </td><td><p class="normal"><b><a href
="datasets/OPPORTUNITY+Activity+Recognition">OPPORTUNITY Activity Recognition</a></b></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><a href="datasets/Nomao"><img src="assets/MLimages/SmallLargedefault.jpg" bo
rder=1 /></a> </td><td><p class="normal"><b><a href="datasets/Nomao">Nomao</a></b></p></td></
tr></table></td>
```

<!-- <td> Nomao collects data about places (name, phone, localization...) from many sources.
Deduplication consists in detecting what data refer to the same place.
Instances in the dataset compare 2 spots. </td> -->
<td> Univariate </td>
<td> Classification </td>
<td> Real </td>
<td> 34465 </td>
<td> 120 </td>
<td> 2012 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td> </td><td> SMS Spam Collection </td></tr></table></td>
<!-- <td> The SMS Spam Collection is a public set of SMS labeled messages that have been collected for mobile phone spam research. </td> -->
<td> Multivariate, Text, Domain-Theory </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 5574 </td>
<td>   </td>
<td> 2012 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> Skin Segmentation </td></tr></table></td>
<!-- <td> The Skin Segmentation dataset is constructed over B, G, R color space. Skin and Nonskin dataset is generated using skin textures from face images of diversity of age, gender, and race people. </td> -->
<td> Univariate </td>
<td> Classification </td>
<td> Real </td>
<td> 245057 </td>
<td> 4 </td>
<td> 2012 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td> </td><td> Planning Relax </td></tr></table></td>
<!-- <td> The dataset concerns with the classification of two mental stages fr om recorded EEG signals: Planning (during imagination of motor act) and Relax state.   </td> -->
<td> Univariate </td>
<td> Classification </td>
<td> Real </td>
<td> 182 </td>
<td> 13 </td>
<td> 2012 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> PAMAP2 Physical Activity Monitoring </td></tr></table></td>
<!-- <td> The PAMAP2 Physical Activity Monitoring dataset contains data of 18 different physical activities, performed by 9 subjects wearing 3 inertial measurement units and a heart rate monitor. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 3850505 </td>
<td> 52 </td>
<td> 2012 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td> </td><td> Restaurant & consumer data </td></tr></table></td>
<!-- <td> The dataset was obtained from a recommender system prototype. The task was to generate a top-n list of restaurants according to the consumer preferences.   </td> -->
<td> Multivariate </td>
<td>   </td>

```
   <td><p class="normal"> </p></td>
   <td><p class="normal">138 </p></td>
   <td><p class="normal">47 </p></td>
   <td><p class="normal">2012 </p></td>
   <!-- <td><p class="normal">Computer </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/CNAE-9"><img src="assets/MLimages/SmallLargedefault.jpg" b
order=1 /></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 d
escriptions of Brazilian companies categorized into a
subset of 9 categories </p></td> -->
   <td><p class="normal">Multivariate, Text </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Integer </p></td>
   <td><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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Individual+household+electric+power+consumption"><img src=
"assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a href
="datasets/Individual+household+electric+power+consumption">Individual household electric power co
nsumption</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">Measurements of electric power consumption in one household with a o
ne-minute sampling rate over a period of almost 4 years. Different electrical quantities and some
sub-metering values are available. </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><a href="datasets/seeds"><img src="assets/MLimages/SmallLargedefault.jpg" bo
rder=1 /></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. </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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Northix"><img src="assets/MLimages/SmallLargedefault.jpg"
border=1 /></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 da
ta 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><a href="datasets/QtyT40I10D100K"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 </p></t
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"> </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Legal+Case+Reports"><img
```

```
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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. </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 </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Human+Activity+Recognition+Using+Smartphones"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Human+Activity+Recognition+Using+Smartphones">Human Activity Recognition Using Smar
tphones</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. </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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/One-hundred+plant+species+leaves+data+set"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/One-hundred+plant+species+leaves+data+set">One-hundred plant species leaves data se
t</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. </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><a href="datasets/Energy+efficiency"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 r
equirements 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><a href="datasets/Yacht+Hydrodynamics"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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, h
ealth 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"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Daphnet+Freezing+of+Gait">Daphnet Freezing of Gait</a></b></p></td></tr></table></t
d>
   <!-- <td><p class="normal">This dataset contains the annotated readings of 3 acceleration senso
rs at the hip and leg of Parkinson's disease patients that experience freezing of gait (FoG) durin
g 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><a href="datasets/3D+Road+Network+%28North+Jutland%2C+Denmark%29"><img src="
assets/MLimages/SmallLargedefault.jpg" border=1 /></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"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 oth
er international index; SP, DAX, FTSE, NIKKEI, BOVESPA, MSCE_EU, MSCI_EM from Jun 5, 2009 to Feb 2
2, 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><a href="datasets/Buzz+in+social+media+"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 so
cial networks: Twitter, and Tom's Hardware, a forum network focusing on new technology with more c
onservative 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><a href="datasets/First-order+theorem+proving"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/First-order+theorem+proving">First-order theorem proving</a></b></p></td></tr></tab
le></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, s
hould the theorem be too difficult. </p></td> -->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Real </p></td>
   <td><p class="normal">6118 </p></td>
   <td><p class="normal">51 </p></td>
   <td><p class="normal">2013 </p></td>
```

```
   <!-- <td><p class="normal">Computer </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a
href="datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio%29">
<img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><
b><a href="datasets/Wearable+Computing%3A+Classification+of+Body+Postures+and+Movements+%28PUC-Rio
%29">Wearable Computing: Classification of Body Postures and Movements (PUC-Rio)</a></b></p></td><
/tr></table></td>
   <!-- <td><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><a href="datasets/Gas+sensor+arrays+in+open+sampling+settings"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Gas+sensor+arrays+in+open+sampling+settings">Gas sensor arrays in open sampling set
tings</a></b></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><a href="datasets/Climate+Model+Simulation+Crashes"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Climate+Model+Simulation+Crashes">Climate Model Simulation Crashes</a></b></p></td>
</tr></table></td>
   <!-- <td><p class="normal">Given Latin hypercube samples of 18 climate model input parameter
values, predict climate model simulation crashes and determine the parameter value combinations th
at 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><a href="datasets/MicroMass"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/MicroMass">MicroMass</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">A dataset to explore machine learning approaches for the
identification of 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><a href="datasets/QSAR+biodegradation"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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). </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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/BLOGGER"><img src="assets/MLimages/SmallLargedefault.jpg"
```

```
border=1 /></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 </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 </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Daily+and+Sports+Activities"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Daily+and+Sports+Activities">Daily and Sports Activities</a></b></p></td></tr></tab
le></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><a href="datasets/User+Knowledge+Modeling"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 th
e subject of Electrical DC Machines. The dataset had been obtained from Ph.D. Thesis. </p></t
d> -->
   <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><a
href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection"><i
mg src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b>
<a
href="datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection">Reu
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 o
riginally written in five different languages and their translations, over a common set of 6 categ
ories.  </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><a href="datasets/NYSK"><img src="assets/MLimages/SmallLargedefault.jpg" bor
der=1 /></a> </td><td><p class="normal"><b><a href="datasets/NYSK">NYSK</a></b></p></td></tr>
</table></td>
   <!-- <td><p class="normal">NYSK (New York v. Strauss-Kahn) is a collection of English news
articles about the case relating to allegations of sexual assault against the former IMF director
Dominique Strauss-Kahn (May 2011). </p></td> -->
   <td><p class="normal">Multivariate, Sequential, Text </p></td>
   <td><p class="normal">Clustering </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal">10421 </p></td>
   <td><p class="normal">7 </p></td>
   <td><p class="normal">2013 </p></td>
   <!-- <td><p class="normal">Social </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Turkiye+Student+Evaluation"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 st
udents from Gazi University in Ankara (Turkey). There is a total of 28 course specific questions a
nd additional 5 attributes. </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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a
href="datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machine
"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 le
vels 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><a href="datasets/EEG+Eye+State"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 ey
e 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><a href="datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure">
<img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><
b><a href="datasets/Physicochemical+Properties+of+Protein+Tertiary+Structure">Physicochemical Prop
erties 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. </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><a href="datasets/seismic-bumps"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/seismic-bumps">seismic-bumps</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">The data describe the problem of high energy (higher than 10^4 J) se
ismic 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><a href="datasets/banknote+authentication"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/banknote+authentication">banknote authentication</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">Data were extracted from images that were taken for the evaluation o
```

f an authentication procedure for banknotes. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 1372 </td>
<td> 5 </td>
<td> 2013 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> USPTO Algorithm Challenge, run by NASA-Harvard Tournament Lab and TopCoder Problem: Pat </td></tr></table></td>
<!-- <td> Data used for USPTO Algorithm Competition. Contains drawing pages fr om US patents with manually labeled figure and part labels. </td> -->
<td> Domain-Theory </td>
<td> Classification </td>
<td> Integer </td>
<td> 306 </td>
<td> 5 </td>
<td> 2013 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td> </td><td> YouTube Multiview Video Games Dataset </td></tr></table></td>
<!-- <td> This dataset contains about 120k instances, each described by 13 feature types, with class information, specially useful for exploring multiview topics (cotraining, ensembles, clustering,..). </td> -->
<td> Multivariate, Text </td>
<td> Classification, Clustering </td>
<td> Integer, Real </td>
<td> 120000 </td>
<td> 1000000 </td>
<td> 2013 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> Gas Sensor Array Drift Dataset at Different Concentrations </td></tr></table></td>
<!-- <td> This archive contains 13910 measurements from 16 chemical sensors ex posed to 6 different gases at various concentration levels. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Regression, Clustering, Causa </td>
<td> Real </td>
<td> 13910 </td>
<td> 129 </td>
<td> 2013 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td> </td><td> Activities of Daily Living (ADLs) Recognition Using Binary Sensors </td></tr></table></td>
<!-- <td> This dataset comprises information regarding the ADLs performed by t wo users on a daily basis in their own homes.   </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Classification, Clustering </td>
<td>   </td>
<td> 2747 </td>
<td>   </td>
<td> 2013 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> SkillCraft1 Master Table Dataset </td></tr></table></td>
<!-- <td> This data was used in Thompson et al. (2013). A list of possible gam e actions is discussed in Thompson, Blair, Chen, & Henrey (2013). </td> -->

<td> Multivariate </td>
<td> Regression </td>
<td> Integer, Real </td>
<td> 3395 </td>
<td> 20 </td>
<td> 2013 </td>
<!-- <td> Game </td> -->
</tr><tr>
<td><table><tr><td><a
href="datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units">Weight Lifting
Exercises monitored with Inertial Measurement Units </td></tr></table></td>
<!-- <td> Six young health subjects were asked to perform 5 variations of the
biceps curl weight lifting exercise. One of the variations is the one predicted by the health prof
essional. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 39242 </td>
<td> 152 </td>
<td> 2013 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img src="assets/MLimages/SmallLargedefault.jpg"
border=1 /> </td><td> SML2010 <
/td></tr></table></td>
<!-- <td> This dataset is collected from a monitor system mounted in a domotic
house. It corresponds to approximately 40 days of monitoring data. </td> -->
<td> Multivariate, Sequential, Time-Series, Text </td>
<td> Regression </td>
<td> Real </td>
<td> 4137 </td>
<td> 24 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Bike+Sharing+Dataset">Bike Sharing Dataset </td></tr></table></td>
<!-- <td> This dataset contains the hourly and daily count of rental bikes bet
ween years 2011 and 2012 in Capital bikeshare system with the corresponding weather and seasonal i
nformation. </td> -->
<td> Univariate </td>
<td> Regression </td>
<td> Integer, Real </td>
<td> 17389 </td>
<td> 16 </td>
<td> 2013 </td>
<!-- <td> Social </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Predict+keywords+activities+in+a+online+social+media">Predict keywords activities i
n a online social media </td></tr></table></td>
<!-- <td> The data from Twitter was collected during 360 consecutive days. It
was done by querying 1497 English keywords sampled from Wikipedia. This dataset is proposed in a L
earning to rank setting. </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td>   </td>
<td> Integer, Real </td>
<td> 51 </td>
<td> 35 </td>
<td> 2013 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Thoracic+Surgery+Data">Thoracic Surgery Data </td></tr></table></td>
<!-- <td> The data is dedicated to classification problem related to the post-
operative life expectancy in the lung cancer patients: class 1 - death within one year after surge
ry, class 2 - survival. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 470 </td>
<td> 17 </td>

<td> 2013 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/EMG+dataset+in+Lower+Limb">EMG dataset in Lower Limb </td></tr></table><
/td>
<!-- <td> 3 different exercises: sitting, standing and walking in the muscles:
biceps femoris, vastus medialis, rectus femoris and semitendinosus addition to goniometry in the e
xercises. </td> -->
<td> Multivariate, Time-Series </td>
<td>   </td>
<td> Real </td>
<td> 132 </td>
<td> 5 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img src="assets/MLimages/SmallLargedefault.jpg" bor
der=1 /> </td><td> SUSY </td></tr>
</table></td>
<!-- <td> This is a classification problem to distinguish between a signal pro
cess which produces supersymmetric particles and a background process which does not. </t
d> -->
<td>   </td>
<td> Classification </td>
<td> Real </td>
<td> 5000000 </td>
<td> 18 </td>
<td> 2014 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img src="assets/MLimages/SmallLargedefault.jpg" bo
rder=1 /> </td><td> HIGGS </td></
tr></table></td>
<!-- <td> This is a classification problem to distinguish between a signal pro
cess which produces Higgs bosons and a background process which does not.   </td> -->
<td>   </td>
<td> Classification </td>
<td> Real </td>
<td> 11000000 </td>
<td> 28 </td>
<td> 2014 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Qualitative_Bankruptcy">Qualitative_Bankruptcy </td></tr></table></td>
<!-- <td> Predict the Bankruptcy from Qualitative parameters from
experts. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td>   </td>
<td> 250 </td>
<td> 7 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/LSVT+Voice+Rehabilitation">LSVT Voice Rehabilitation </td></tr></table><
/td>
<!-- <td> 126 samples from 14 participants, 309 features. Aim: assess whether
voice rehabilitation treatment lead to phonations considered 'acceptable' or 'unacceptable'
(binary class classification problem). </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 126 </td>
<td> 309 </td>
<td> 2014 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Dataset+for+ADL+Recognition+with+Wrist-
worn+Accelerometer"> </td><td
> <a href="datasets/Dataset+for+ADL+Recognition+with+Wrist-
worn+Accelerometer">Dataset for ADL Recognition with Wrist-worn Accelerometer </td></tr

```
Wrist-Accelerometer Database 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"><img src="assets/MLimages/SmallLargedefault.jpg" bor
der=1 /></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 strat
ified 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><a href="datasets/User+Identification+From+Walking+Activity"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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.
 </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"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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></t
d></tr></table></td>
   <!-- <td><p class="normal">The dataset collects data from a wearable accelerometer mounted on t
he 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"> </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"><img src="assets/MLimages/SmallLargedefault.jpg" bor
der=1 /></a> </td><td><p class="normal"><b><a href="datasets/Leaf">Leaf</a></b></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 spe
cies. </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><a href="datasets/Dresses_Attribute_Sales"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Dresses_Attribute_Sales">Dresses_Attribute_Sales</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">This dataset contain Attributes of dresses and their recommendations
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> 501 </td>
<td> 13 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Tamilnadu+Electricity+Board+Hourly+Readings">Tamilnadu Electricity Board Hourly Rea
dings </td></tr></table></td>
<!-- <td> This data can be effectively produced the result to fewer parameter
of the Load profile can be reduced in the Database   </td> -->
<td> Multivariate </td>
<td> Classification, Regression, Clustering </td>
<td> Real </td>
<td> 45781 </td>
<td> 5 </td>
<td> 2013 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Airfoil+Self-Noise">Airfoil Self-Noise </td></tr></table></td>
<!-- <td> NASA data set, obtained from a series of aerodynamic and acoustic te
sts of two and three-dimensional airfoil blade sections conducted in an anechoic wind
tunnel. </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 1503 </td>
<td> 6 </td>
<td> 2014 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Wholesale+customers">Wholesale customers </td></tr></table></td>
<!-- <td> The data set refers to clients of a wholesale distributor. It
includes the annual spending in monetary units (m.u.) on diverse product categories </td>
-->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> Integer </td>
<td> 440 </td>
<td> 8 </td>
<td> 2014 </td>
<!-- <td> Business </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img src="
assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a href=
"datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis">Twitter Data set for Arabic Sentiment An
alysis </td></tr></table></td>
<!-- <td> This problem of Sentiment Analysis (SA) has been studied well on the
English language but not Arabic one. Two main approaches have been devised: corpus-based and lexic
on-based.   </td> -->
<td> Text </td>
<td> Classification </td>
<td>   </td>
<td> 2000 </td>
<td> 2 </td>
<td> 2014 </td>
<!-- <td> Social </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Combined+Cycle+Power+Plant">Combined Cycle Power Plant </td></tr>
</table></td>
<!-- <td> The dataset contains 9568 data points collected from a Combined Cycl
e Power Plant over 6 years (2006-2011), when the plant was set to work with full load.   <
/td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 9568 </td>
<td> 4 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">

<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Urban+Land+Cover">Urban Land Cover </td></tr></table></td>
<!-- <td> Classification of urban land cover using high resolution aerial
imagery. Intended to assist sustainable urban planning efforts. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td>   </td>
<td> 168 </td>
<td> 148 </td>
<td> 2014 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Diabetes+130-US+hospitals+for+years+1999-2008">Diabetes 130-US hospitals for years
1999-2008 </td></tr></table></td>
<!-- <td> This data has been prepared to analyze factors related to
readmission as well as other

outcomes pertaining to patients with diabetes. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> Integer </td>
<td> 100000 </td>
<td> 55 </td>
<td> 2014 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Bach+Choral+Harmony">Bach Choral Harmony </td></tr></table></td>
<!-- <td> The data set is composed of 60 chorales (5665 events) by J.S. Bach (
1675-1750).
Each event of each chorale is labelled using 1 among 101 chord labels and described
through 14 features. </td> -->
<td> Sequential </td>
<td> Classification </td>
<td>   </td>
<td> 5665 </td>
<td> 17 </td>
<td> 2014 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/StoneFlakes">StoneFlakes </td></tr></table></td>
<!-- <td> Stone flakes are waste products of the stone tool production in
the prehistoric era. The variables are means of geometric and
stylistic features of the flakes contained in different inventories. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering, Causal-Discovery </td>
<td> Real </td>
<td> 79 </td>
<td> 8 </td>
<td> 2014 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Tennis+Major+Tournament+Match+Statistics">Tennis Major Tournament Match
Statistics </td></tr></table></td>
<!-- <td> This is a collection of 8 files containing the match statistics for
both women and men at the four major tennis tournaments of the year 2013. Each file has 42 columns
and a minimum of 76 rows. </td> -->
<td> Multivariate </td>
<td> Classification, Regression, Clustering </td>
<td> Integer, Real </td>
<td> 127 </td>
<td> 42 </td>
<td> 2014 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a
href="datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings">Parkinson Speech

href= datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings >Parkinson Speech Dataset with Multiple Types of Sound Recordings </td></tr></table></td>
<!-- <td> The training data belongs to 20 Parkinson's Disease (PD) patients and 20 healthy subjects. From all subjects, multiple types of sound recordings (26) are taken. </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Integer, Real </td>
<td> 1040 </td>
<td> 26 </td>
<td> 2014 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> Gesture Phase Segmentation </td></tr></table></td>
<!-- <td> The dataset is composed by features extracted from 7 videos with people gesticulating, aiming at studying Gesture Phase Segmentation. It contains 50 attributes divided into two files for each video. </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 9900 </td>
<td> 50 </td>
<td> 2014 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td> </td><td> Perfume Data </td></tr></table></td>
<!-- <td> This data consists of odors of 20 different perfumes. Data was obtained by using a handheld odor meter (OMX-GR sensor) per second for 28 seconds period. </td> -->
<td> Univariate, Domain-Theory </td>
<td> Classification, Clustering </td>
<td> Integer </td>
<td> 560 </td>
<td> 2 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> BlogFeedback </td></tr></table></td>
<!-- <td> Instances in this dataset contain features extracted from blog posts. The task associated with the data is to predict how many comments the post will receive. </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Integer, Real </td>
<td> 60021 </td>
<td> 281 </td>
<td> 2014 </td>
<!-- <td> Social </td> -->
</tr><tr>
<td><table><tr><td> </td><td> REALDISP Activity Recognition Dataset </td></tr></table></td>
<!-- <td> The REALDISP dataset is devised to evaluate techniques dealing with the effects of sensor displacement in wearable activity recognition as well as to benchmark general activity recognition algorithms   </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 1419 </td>
<td> 120 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> Newspaper and magazine images segmentation dataset </td></tr></table></td>
<!-- <td> Dataset is well suited for segmentation tasks. It contains 101 scanned pages from different newspapers and magazines in Russian with ground truth pixel-based masks. </td> -->

masks. </td> -->
<td>   </td>
<td> Classification </td>
<td>   </td>
<td> 101 </td>
<td>   </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/AAAI+2014+Accepted+Papers">AAAI 2014 Accepted Papers </td></tr></table><
/td>
<!-- <td> This data set compromises the metadata for the 2014 AAAI
conference's accepted papers, including paper titles, authors, abstracts, and keywords of varying
granularity. </td> -->
<td> Multivariate </td>
<td> Clustering </td>
<td>   </td>
<td> 399 </td>
<td> 6 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Gas+sensor+array+under+flow+modulation">Gas sensor array under flow modulation<
/b> </td></tr></table></td>
<!-- <td> The data set contains 58 time series acquired from 16 chemical senso
rs under gas flow modulation conditions. The sensors were exposed to different gaseous binary mixt
ures of acetone and ethanol. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Regression </td>
<td> Real </td>
<td> 58 </td>
<td> 120432 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img s
rc="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a h
ref="datasets/Gas+sensor+array+exposed+to+turbulent+gas+mixtures">Gas sensor array exposed to
turbulent gas mixtures </td></tr></table></td>
<!-- <td> A chemical detection platform composed of 8 chemoresistive gas senso
rs was exposed to turbulent gas mixtures generated naturally in a wind tunnel. The acquired time s
eries of the sensors are provided. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Regression </td>
<td> Real </td>
<td> 180 </td>
<td> 150000 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/UJIIndoorLoc">UJIIndoorLoc </td></tr></table></td>
<!-- <td> The UJIIndoorLoc is a Multi-Building Multi-Floor indoor localization
database to test Indoor Positioning System that rely on WLAN/WiFi fingerprint. </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Integer, Real </td>
<td> 21048 </td>
<td> 529 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Sentence+Classification">Sentence Classification </td></tr></table></td>
<!-- <td> Contains sentences from the abstract and introduction of 30 articles
annotated with a modified Argumentative Zones annotation scheme. These articles come from biology,
machine learning and psychology. </td> -->
<td> Text </td>
<td> Classification </td>
<td> Integer </td>
<td>   </td>

```
   <td><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/Dow+Jones+Index"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Dow+Jones+Index">Dow Jones Index</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">This dataset contains weekly data for the Dow Jones Industrial Index
.  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><a href="datasets/sEMG+for+Basic+Hand+movements"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/sEMG+for+Basic+Hand+movements">sEMG for Basic Hand movements</a></b></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><a href="datasets/AAAI+2013+Accepted+Papers"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/AAAI+2013+Accepted+Papers">AAAI 2013 Accepted Papers</a></b></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"> </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><a href="datasets/Geographical+Original+of+Music"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Geographical+Original+of+Music">Geographical Original of Music</a></b></p></td></tr
></table></td>
   <!-- <td><p class="normal">Instances in this dataset contain audio features extracted from 1059
wave files. The task associated with the data is to predict the geographical origin of music.
 </p></td> -->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal">Classification, Regression </p></td>
   <td><p class="normal">Real </p></td>
   <td><p class="normal">1059 </p></td>
   <td><p class="normal">68 </p></td>
   <td><p class="normal">2014 </p></td>
   <!-- <td><p class="normal">Other </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants"><i
mg src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 Tur
bines (GT), mounted on a Frigate characterized by a COmbined Diesel eLectric And Gas (CODLAG) prop
ulsion 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><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Grammatical+Facial+Expressions">Grammatical Facial Expressions </td></tr
></table></td>
<!-- <td> This dataset supports the development of models that make possible t
o interpret Grammatical Facial Expressions from Brazilian Sign Language (Libras). </td> -
->
<td> Multivariate, Sequential </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 27965 </td>
<td> 100 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/NoisyOffice">NoisyOffice </td></tr></table></td>
<!-- <td> Corpus intended to do cleaning (or binarization) and enhancement of
noisy grayscale printed text images using supervised learning methods. Noisy images and their corr
esponding ground truth provided. </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Real </td>
<td> 216 </td>
<td> 216 </td>
<td> 2015 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/MHEALTH+Dataset">MHEALTH Dataset </td></tr></table></td>
<!-- <td> The MHEALTH (Mobile Health) dataset is devised to benchmark
techniques dealing with human behavior analysis based on multimodal body sensing. </td> -
->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 120 </td>
<td> 23 </td>
<td> 2014 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Student+Performance">Student Performance </td></tr></table></td>
<!-- <td> Predict student performance in secondary education (high school). &n
bsp; </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Integer </td>
<td> 649 </td>
<td> 33 </td>
<td> 2014 </td>
<!-- <td> Social </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/ElectricityLoadDiagrams20112014">ElectricityLoadDiagrams20112014 </td></
tr></table></td>
<!-- <td> This data set contains electricity consumption of 370
points/clients.
  </td> -->
<td> Time-Series </td>
<td> Regression, Clustering </td>
<td> Real </td>
<td> 370 </td>
<td> 140256 </td>
<td> 2015 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Gas+sensor+array+under+dynamic+gas+mixtures">Gas sensor array under dynamic gas mix
tures </td></tr></table></td>
<!-- <td> The data set contains the recordings of 16 chemical sensors exposed
to two dynamic gas mixtures at varying concentrations. For each mixture, signals were acquired con

```
tinuously 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>
   <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><a href="datasets/microblogPCU"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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.  </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><table><tr><td><a href="datasets/Firm-Teacher_Clave-Direction_Classification"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 c
lass(es) according to the partido-alto-based paradigm. </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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Dataset+for+Sensorless+Drive+Diagnosis"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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.  </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 </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/TV+News+Channel+Commercial+Detection+Dataset"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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).   </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><a href="datasets/Phishing+Websites"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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. </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> 2015 </td>
<!-- <td> Computer Security </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Greenhouse+Gas+Observing+Network">Greenhouse Gas Observing Network </td>
</tr></table></td>
<!-- <td> Design an observing network to monitor emissions of a greenhouse gas
(GHG) in California given time series of synthetic observations and tracers from weather model
simulations.
  </td> -->
<td> Multivariate, Time-Series </td>
<td> Regression </td>
<td> Real </td>
<td> 2921 </td>
<td> 5232 </td>
<td> 2015 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Diabetic+Retinopathy+Debrecen+Data+Set">Diabetic Retinopathy Debrecen Data Set<
/b> </td></tr></table></td>
<!-- <td> This dataset contains features extracted from the Messidor image set
to predict whether an image contains signs of diabetic retinopathy or not.   </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 1151 </td>
<td> 20 </td>
<td> 2014 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/HIV-1+protease+cleavage">HIV-1 protease cleavage </td></tr></table></td>
<!-- <td> The data contains lists of octamers (8 amino acids) and a flag (-1 o
r 1) depending on whether HIV-1 protease will cleave in the central position (between amino acids
4 and 5). </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Categorical </td>
<td> 6590 </td>
<td> 1 </td>
<td> 2015 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Sentiment+Labelled+Sentences">Sentiment Labelled Sentences </td></tr></t
able></td>
<!-- <td> The dataset contains sentences labelled with positive or negative se
ntiment. </td> -->
<td> Text </td>
<td> Classification </td>
<td>   </td>
<td> 3000 </td>
<td>   </td>
<td> 2015 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Online+News+Popularity">Online News Popularity </td></tr></table></td>
<!-- <td> This dataset summarizes a heterogeneous set of features about
articles published by Mashable in a period of two years. The goal is to predict the number of shar
es in social networks (popularity). </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Integer, Real </td>
<td> 39797 </td>
<td> 61 </td>
<td> 2015 </td>
<!-- <td> Business </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a

```
href="datasets/Forest+type+mapping">Forest type mapping</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">Multi-temporal remote sensing data of a forested area in Japan. The
goal is to map different forest types using spectral data. </p></td> -->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal">326 </p></td>
   <td><p class="normal">27 </p></td>
   <td><p class="normal">2015 </p></td>
   <!-- <td><p class="normal">Life </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/wiki4HE"><img src="assets/MLimages/SmallLargedefault.jpg"
border=1 /></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 </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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a
href="datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><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.  </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><a href="datasets/Chronic_Kidney_Disease"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 i
t 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><a
href="datasets/Machine+Learning+based+ZZAlpha+Ltd.+Stock+Recommendations+2012-2014"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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, 201
2 - 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><a href="datasets/Folio"><img src="assets/MLimages/SmallLargedefault.jpg" bo
rder=1 /></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. </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> 20 </td>
<td> 2015 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Taxi+Service+Trajectory+-
+Prediction+Challenge%2C+ECML+PKDD+2015"><img src="assets/MLimages/SmallLargedefault.jpg" border=1
/> </td><td> <a href="datasets/Taxi+Service+Trajectory+-
+Prediction+Challenge%2C+ECML+PKDD+2015">Taxi Service Trajectory - Prediction Challenge, ECML PKDD
2015 </td></tr></table></td>
<!-- <td> An accurate dataset describing trajectories performed by all the 442
taxis running in the city of Porto, in Portugal.
  </td> -->
<td> Multivariate, Sequential, Time-Series, Domain-Theory </td>
<td> Clustering, Causal-Discovery </td>
<td> Real </td>
<td> 1710671 </td>
<td> 9 </td>
<td> 2015 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Cuff-Less+Blood+Pressure+Estimation">Cuff-Less Blood Pressure Estimation</p
></td></tr></table></td>
<!-- <td> This Data set provides preprocessed and cleaned vital signals which
can be used in designing algorithms for cuff-less estimation of the blood pressure. </td>
-->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Real </td>
<td> 12000 </td>
<td> 3 </td>
<td> 2015 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Smartphone-
Based+Recognition+of+Human+Activities+and+Postural+Transitions"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Smartphone-
Based+Recognition+of+Human+Activities+and+Postural+Transitions">Smartphone-Based Recognition of Hu
man Activities and Postural Transitions </td></tr></table></td>
<!-- <td> Activity recognition data set built from the recordings of 30
subjects performing basic activities and postural transitions while carrying a waist-mounted smart
phone with embedded inertial sensors.
  </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 10929 </td>
<td> 561 </td>
<td> 2015 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Mice+Protein+Expression">Mice Protein Expression </td></tr></table></td>
<!-- <td> Expression levels of 77 proteins measured in the cerebral cortex of
8 classes of control and Down syndrome mice exposed to context fear conditioning, a task used to a
ssess associative learning. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 1080 </td>
<td> 82 </td>
<td> 2015 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/UJIIndoorLoc-Mag">UJIIndoorLoc-Mag </td></tr></table></td>
<!-- <td> The UJIIndoorLoc-Mag is an indoor localization database to test
Indoor Positioning System that rely on Earth's magnetic field variations. </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Classification, Regression, Clustering </td>
<td> Integer, Real </td>
<td> 40000 </td>
<td> 13 </td>

<td> 2015 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Heterogeneity+Activity+Recognition">Heterogeneity Activity Recognition <
/td></tr></table></td>
<!-- <td> The Heterogeneity Human Activity Recognition (HHAR) dataset from
Smartphones and Smartwatches is a dataset devised to benchmark human activity recognition
algorithms (classification, automatic data segmentation, sensor fusion, feature extraction, etc.)
in real-world contexts; specifically, the dataset is gathered with a variety of different device m
odels and use-scenarios, in order to reflect sensing heterogeneities to be expected in real
deployments. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 43930257 </td>
<td> 16 </td>
<td> 2015 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a
href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Educational+Process+Mining+%28EPM%29%3A+A+Learning+Analytics+Data+Set">Educational
Process Mining (EPM): A Learning Analytics Data Set </td></tr></table></td>
<!-- <td> Educational Process Mining data set is built from the recordings of
115 subjects' activities through a logging application while learning with an educational
simulator. </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Classification, Regression, Clustering </td>
<td> Integer </td>
<td> 230318 </td>
<td> 13 </td>
<td> 2015 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img src="assets/MLimages/SmallLargedefault.jpg"
border=1 /> </td><td> HEPMASS <
/td></tr></table></td>
<!-- <td> The search for exotic particles requires sorting through a large num
ber of collisions to find the events of interest. This data set challenges one to detect a new par
ticle of unknown mass. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 10500000 </td>
<td> 28 </td>
<td> 2016 </td>
<!-- <td> Physical </td> -->
</tr><tr>
<td><table><tr><td><img src="a
ssets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Indoor+User+Movement+Prediction+from+RSS+data">Indoor User Movement Prediction from
RSS data </td></tr></table></td>
<!-- <td> This dataset contains temporal data from a Wireless Sensor Network d
eployed in real-world office environments. The task is intended as real-life benchmark in the area
of Ambient Assisted Living. </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 13197 </td>
<td> 4 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Open+University+Learning+Analytics+dataset">Open University Learning Analytics data
set </td></tr></table></td>
<!-- <td> Open University Learning Analytics Dataset contains data about
courses, students and their interactions with Virtual Learning Environment for seven selected cour
ses and more than 30000 students. </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Classification, Regression, Clustering </td>
<td> Integer </td>
<td>   </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
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/default+of+credit+card+clients">default of credit card clients</a></b></p></td></tr
></table></td>
   <!-- <td><p class="normal">This research aimed at the case of customers’ default payments in Ta
iwan and compares the predictive accuracy of probability of default among six data mining methods.
 </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><a href="datasets/Mesothelioma%E2%80%99s+disease+data+set+"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 Fa
culty 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><a href="datasets/Online+Retail"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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
occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online 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">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><a href="datasets/SIFT10M"><img src="assets/MLimages/SmallLargedefault.jpg"
border=1 /></a> </td><td><p class="normal"><b><a href="datasets/SIFT10M">SIFT10M</a></b></p><
/td></tr></table></td>
   <!-- <td><p class="normal">In SIFT10M, each data point is a SIFT feature which is extracted
from Caltech-256 by the open source VLFeat library. The corresponding patches of the SIFT features
are provided. </p></td> -->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal">Causal-Discovery </p></td>
   <td><p class="normal">Integer </p></td>
   <td><p class="normal">11164866 </p></td>
   <td><p class="normal">128 </p></td>
   <td><p class="normal">2016 </p></td>
   <!-- <td><p class="normal">Computer </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/GPS+Trajectories"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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).  
;</p></td> -->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal">Classification, Regression </p></td>
   <td><p class="normal">Real </p></td>
   <td><p class="normal">163 </p></td>
   <td><p class="normal">15 </p></td>
   <td><p class="normal">2016 </p></td>
   <!-- <td><p class="normal">Computer </p></td> -->
   </tr><tr bgcolor="DDEEFF">
```

<td><table><tr><td> </td><td> Detect Malacious Executable(AntiVirus) </td></tr></table></td>
<!-- <td> I extract features from malacious and non-malacious and create and training dataset to teach svm classifier.Dataset made of unknown executable to detect if it is virus or normal safe executable. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 373 </td>
<td> 513 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td> </td><td> Occupancy Detection </td></tr></table></td>
<!-- <td> Experimental data used for binary classification (room occupancy) from Temperature,Humidity,Light and CO2. Ground-truth occupancy was obtained from time stamped pictures that were taken every minute. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 20560 </td>
<td> 7 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> Improved Spiral Test Using Digitized Graphics Tablet for Monitoring Parkinson’s Disease </td></tr></table></td>
<!-- <td> Handwriting database consists of 25 PWP(People with Parkinson) and 15 healthy individuals.Three types of recordings (Static Spiral Test, Dynamic Spiral Test and Stability Test) are taken. </td> -->
<td> Multivariate </td>
<td> Classification, Regression, Clustering </td>
<td> Real </td>
<td> 40 </td>
<td> 7 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td> </td><td> News Aggregator </td></tr></table></td>
<!-- <td> References to news pages collected from an web aggregator in the period from 10-March-2014 to 10-August-2014. The resources are grouped into clusters that represent pages discussing the same story. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td>   </td>
<td> 422937 </td>
<td> 5 </td>
<td> 2016 </td>
<!-- <td> Other </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> Air Quality </td></tr></table></td>
<!-- <td> Contains the responses of a gas multisensor device deployed on the field in an Italian city. Hourly responses averages are recorded along with gas concentrations references from a certified analyzer.   </td> -->
<td> Multivariate, Time-Series </td>
<td> Regression </td>
<td> Real </td>
<td> 9358 </td>
<td> 15 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td> </td><td> <a

```
href="datasets/Twin+gas+sensor+arrays">Twin gas sensor arrays</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">5 replicates of an 8-MOX gas sensor array were exposed to different
gas 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><a href="datasets/Gas+sensors+for+home+activity+monitoring"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Gas+sensors+for+home+activity+monitoring">Gas sensors for home activity
monitoring</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">100 recordings of a sensor array under different conditions in a
home setting: 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><a href="datasets/Facebook+Comment+Volume+Dataset"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Facebook+Comment+Volume+Dataset">Facebook Comment Volume Dataset</a></b></p></td></
tr></table></td>
   <!-- <td><p class="normal">Instances in this dataset contain features extracted from facebook p
osts. 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><a
href="datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+in+Ambient+Assisted+Livi
28AAL%29"><img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class=
"normal"><b><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 (AAL)
</a></b></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 doma
in 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><a href="datasets/Polish+companies+bankruptcy+data"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Polish+companies+bankruptcy+data">Polish companies bankruptcy data</a></b></p></td>
</tr></table></td>
   <!-- <td><p class="normal">The dataset is about bankruptcy prediction of Polish companies.The ba
nkrupt companies were analyzed in the period 2000-2012, while the still operating companies were e
valuated 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><a
href="datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29"><img
```

src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+%28AReM%29">Activity R
ecognition system based on Multisensor data fusion (AReM) </td></tr></table></td>
<!-- <td> This dataset contains temporal data from a Wireless Sensor Network w
orn by an actor performing the activities: bending, cycling, lying down, sitting, standing,
walking. </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Classification </td>
<td> Real </td>
<td> 42240 </td>
<td> 6 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Dota2+Games+Results">Dota2 Games Results </td></tr></table></td>
<!-- <td> Dota 2 is a popular computer game with two teams of 5 players. At th
e start of the game each player chooses a unique hero with different strengths and
weaknesses. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td>   </td>
<td> 102944 </td>
<td> 116 </td>
<td> 2016 </td>
<!-- <td> Game </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Facebook+metrics">Facebook metrics </td></tr></table></td>
<!-- <td> Facebook performance metrics of a renowned cosmetic's brand Facebook
page. </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Integer </td>
<td> 500 </td>
<td> 19 </td>
<td> 2016 </td>
<!-- <td> Business </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/UbiqLog+%28smartphone+lifelogging%29">UbiqLog (smartphone lifelogging) <
/td></tr></table></td>
<!-- <td> UbiqLog is the smartphone lifelogging tool that runs on the
smartphone of 35 users for about 2 months.
  </td> -->
<td> Multivariate </td>
<td> Causal-Discovery </td>
<td>   </td>
<td> 9782222 </td>
<td>   </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/NIPS+Conference+Papers+1987-2015">NIPS Conference Papers 1987-2015 </td>
</tr></table></td>
<!-- <td> This data set contains the distribution of words in the full text of
the NIPS conference papers published from 1987 to 2015. </td> -->
<td> Text </td>
<td> Clustering </td>
<td> Integer </td>
<td> 11463 </td>
<td> 5812 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img src="assets/MLimages/SmallLargedefault.jpg" bo
rder=1 /> </td><td> HTRU2 </td></
tr></table></td>
<!-- <td> Pulsar candidates collected during the HTRU survey. Pulsars are a ty
pe of star, of considerable scientific interest. Candidates must be classified in to pulsar and
non-pulsar classes to aid discovery. </td> -->
<td> Multivariate </td>

<td> Classification, Clustering </td>
<td> Real </td>
<td> 17898 </td>
<td> 9 </td>
<td> 2017 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Drug+consumption+%28quantified%29">Drug consumption (quantified) </td></
tr></table></td>
<!-- <td> Classify type of drug consumer by personality data </td> --
>
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 1885 </td>
<td> 32 </td>
<td> 2016 </td>
<!-- <td> Social </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Appliances+energy+prediction">Appliances energy prediction </td></tr></t
able></td>
<!-- <td> Experimental data used to create regression models of appliances
energy use in a low energy building. </td> -->
<td> Multivariate, Time-Series </td>
<td> Regression </td>
<td> Real </td>
<td> 19735 </td>
<td> 29 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Miskolc+IIS+Hybrid+IPS">Miskolc IIS Hybrid IPS </td></tr></table></td>
<!-- <td> The dataset was created for the comparison and evaluation of hybrid
indoor positioning methods. The dataset presented contains data from W-LAN and Bluetooth
interfaces, and Magnetometer.   </td> -->
<td> Text </td>
<td> Classification, Clustering, Causal-Discovery </td>
<td> Integer </td>
<td> 1540 </td>
<td> 67 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/KDC-4007+dataset+Collection">KDC-4007 dataset Collection </td></tr></tab
le></td>
<!-- <td> KDC-4007 dataset Collection is the Kurdish Documents Classification
text used in categories regarding Kurdish Sorani news and articles. </td> -->
<td> Multivariate, Text </td>
<td> Classification, Regression </td>
<td> Integer </td>
<td> 4007 </td>
<td>   </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/Geo-
Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Geo-
Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone">Geo-
Magnetic field and WLAN dataset for indoor localisation from wristband and smartphone <
/td></tr></table></td>
<!-- <td> A multisource and multivariate dataset for indoor localisation metho
ds based on WLAN and Geo-Magnetic field fingerprinting </td> -->
<td> Multivariate, Sequential, Time-Series </td>
<td> Classification, Regression, Clustering </td>
<td> Integer, Real </td>
<td> 153540 </td>
<td> 25 </td>

<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td> </td><td> DrivFace </td></tr></table></td>
<!-- <td> The DrivFace contains images sequences of subjects while driving in real scenarios. It is composed of 606 samples of 640×480, acquired over different days from 4 drivers with several facial features. </td> -->
<td> Multivariate </td>
<td> Classification, Regression, Clustering </td>
<td> Real </td>
<td> 606 </td>
<td> 6400 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> Website Phishing </td></tr></table></td>
<!-- <td>

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

href="datasets/Cervical+cancer%2C+Risk+Factors">Cervical cancer (Risk Factors) </td>
</tr></table></td>
<!-- <td> This dataset focuses on the prediction of indicators/diagnosis of cervical cancer. The features cover demographic information, habits, and historic medical records. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 858 </td>
<td> 36 </td>
<td> 2017 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td> </td><td> Quality Assessment of Digital Colposcopies </td></tr></table></td>
<!-- <td> This dataset explores the subjective quality assessment of digital colposcopies. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 287 </td>
<td> 69 </td>
<td> 2017 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> KASANDR </td></tr></table></td>
<!-- <td> KASANDR is a novel, publicly available collection for recommendation systems that records the behavior of customers of the European leader in e-Commerce advertising, Kelkoo.   </td> -->
<td> Multivariate </td>
<td> Causal-Discovery </td>
<td> Integer </td>
<td> 17764280 </td>
<td> 2158859 </td>
<td> 2017 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td> </td><td> FMA: A Dataset For Music Analysis </td></tr></table></td>
<!-- <td> FMA features 106,574 tracks and includes song title, album, artist, genres; play counts, favorites, comments; description, biography, tags; together with audio (343 days, 917 GiB) and features. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 106574 </td>
<td> 518 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> Air quality </td></tr></table></td>
<!-- <td> Contains the responses of a gas multisensor device deployed on the field in an Italian city.   </td> -->
<td> Multivariate, Time-Series </td>
<td> Regression </td>
<td> Real </td>
<td> 9358 </td>
<td> 15 </td>
<td> 2016 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td> </td><td> Epileptic Seizure Recognition </td></tr></table></td>
<!-- <td> This dataset is a pre-processed and re-structured/reshaped version of a very commonly used dataset featuring epileptic seizure detection.   </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Clustering </td>
<td> Integer, Real </td>

<td> Integer, Real </td>
<td> 11500 </td>
<td> 179 </td>
<td> 2017 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Devanagari+Handwritten+Character+Dataset">Devanagari Handwritten Character
Dataset </td></tr></table></td>
<!-- <td> This is an image database of Handwritten Devanagari characters.
There are 46 classes of characters with 2000 examples each. The dataset is split into training set
(85%) and testing set(15%).   </td> -->
<td>   </td>
<td> Classification </td>
<td> Integer </td>
<td> 92000 </td>
<td>   </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Stock+portfolio+performance">Stock portfolio performance </td></tr></tab
le></td>
<!-- <td> The data set of performances of weighted scoring stock portfolios
are obtained with mixture design from the US stock market historical database. </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 315 </td>
<td> 12 </td>
<td> 2016 </td>
<!-- <td> Business </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/MoCap+Hand+Postures">MoCap Hand Postures </td></tr></table></td>
<!-- <td> 5 types of hand postures from 12 users were recorded using unlabeled
markers attached to fingers of a glove in a motion capture environment. Due to resolution and occl
usion, missing values are common. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> Integer, Real </td>
<td> 78095 </td>
<td> 38 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><a
href="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech"><img s
rc="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a h
ref="datasets/Early+biomarkers+of+Parkinson%92s+disease+based+on+natural+connected+speech">Early b
iomarkers of Parkinson’s disease based on natural connected speech </td></tr></table></
td>
<!-- <td> Predict a pattern of neurodegeneration in the dataset of speech
features obtained from patients with early untreated Parkinson’s disease and patients at high risk
developing Parkinson’s disease. </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Integer, Real </td>
<td> 130 </td>
<td> 65 </td>
<td> 2017 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a
href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting">Data for So
ftware Engineering Teamwork Assessment in Education Setting </td></tr></table></td>
<!-- <td> Data include over 100 Team Activity Measures and outcomes (ML classe
s) obtained from activities of 74 student teams during the creation of final class project in SW E
ng. classes at SFSU, Fulda, FAU </td> -->
<td> Sequential, Time-Series </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 74 </td>

<td> 274 </td>
<td> 102 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/PM2.5+Data+of+Five+Chinese+Cities">PM2.5 Data of Five Chinese Cities </t
d></tr></table></td>
<!-- <td> This hourly data set contains the PM2.5 data in Beijing, Shanghai, G
uangzhou, Chengdu and Shenyang. Meanwhile, meteorological data for each city are also
included. </td> -->
<td> Multivariate, Time-Series </td>
<td> Regression </td>
<td> Integer, Real </td>
<td> 52854 </td>
<td> 86 </td>
<td> 2017 </td>
<!-- <td> Physical </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a
href="datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet">Parkinson
Disease Spiral Drawings Using Digitized Graphics Tablet </td></tr></table></td>
<!-- <td> Handwriting database consists of 62 PWP(People with Parkinson) and 1
5 healthy individuals. Three types of recordings (Static Spiral Test, Dynamic Spiral Test and
Stability Test) are taken. </td> -->
<td> Multivariate </td>
<td> Classification, Regression, Clustering </td>
<td> Integer </td>
<td> 77 </td>
<td> 7 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Sales_Transactions_Dataset_Weekly">Sales_Transactions_Dataset_Weekly </t
d></tr></table></td>
<!-- <td> Contains weekly purchased quantities of 800 over products over 52 we
eks. Normalised values are provided too. </td> -->
<td> Multivariate, Time-Series </td>
<td> Clustering </td>
<td> Integer, Real </td>
<td> 811 </td>
<td> 53 </td>
<td> 2017 </td>
<!-- <td>   </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Las+Vegas+Strip">Las Vegas Strip </td></tr></table></td>
<!-- <td> This dataset includes quantitative and categorical features from onl
ine reviews from 21 hotels located in Las Vegas Strip, extracted from TripAdvisor
(http://www.tripadvisor.com). </td> -->
<td>   </td>
<td> Classification, Regression </td>
<td> Integer </td>
<td> 504 </td>
<td> 20 </td>
<td> 2017 </td>
<!-- <td> Business </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Eco-hotel">Eco-hotel </td></tr></table></td>
<!-- <td> This dataset includes Online Textual Reviews from both online (e.g.,
TripAdvisor) and offline (e.g., Guests' book) sources from the Areias do Seixo Eco-Resort. </
p></td> -->
<td> Text </td>
<td>   </td>
<td>   </td>
<td> 401 </td>
<td> 1 </td>
<td> 2017 </td>
<!-- <td> Business </td> -->
</tr><tr bgcolor="DDEEFF">

```
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/MEU-Mobile+KSD"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 m
obile device (Nexus 7). The dataset contains 2856 records, 51 records per subject for 56 subjects.
 </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal">Integer, Real </p></td>
<td><p class="normal">2856 </p></td>
<td><p class="normal">71 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Computer </p></td> -->
</tr><tr>
<td><table><tr><td><a href="datasets/Crowdsourced+Mapping"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Crowdsourced+Mapping">Crowdsourced Mapping</a></b></p></td></tr></table></td>
<!-- <td><p class="normal">Crowdsourced data from OpenStreetMap is used to automate the
classification of satellite images into different land cover classes (impervious, farm, forest,
grass, orchard, water).  </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification </p></td>
<td><p class="normal"> </p></td>
<td><p class="normal">10546 </p></td>
<td><p class="normal">29 </p></td>
<td><p class="normal">2016 </p></td>
<!-- <td><p class="normal">Physical </p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/gene+expression+cancer+RNA-Seq"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 s
et, it is a random extraction of gene expressions of patients having different types of tumor: BRC
A, KIRC, COAD, LUAD and PRAD. </p></td> -->
<td><p class="normal">Multivariate </p></td>
<td><p class="normal">Classification, Clustering </p></td>
<td><p class="normal">Real </p></td>
<td><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 </p></td> -->
</tr><tr>
<td><table><tr><td><a
href="datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer"><im
g src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><
a
href="datasets/Hybrid+Indoor+Positioning+Dataset+from+WiFi+RSSI%2C+Bluetooth+and+magnetometer">Hybr
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.  </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 </p></td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/chestnut+%E2%80%93+LARVIC"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 che
stnuts 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><td><p
class="normal"><a
href="datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%28OBS%29+N
rk">Burst Header Packet (BHP) flooding attack on Optical Burst Switching (OBS) Network
</td></tr></table></td>
<!-- <td> One of the primary challenges in identifying the risks of the Burst
Header Packet (BHP) flood attacks in Optical Burst Switching networks (OBS) is the scarcity of
reliable historical data.   </td> -->
<td> Text </td>
<td> Classification </td>
<td> Integer </td>
<td> 1075 </td>
<td> 22 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Motion+Capture+Hand+Postures">Motion Capture Hand Postures </td></tr></t
able></td>
<!-- <td> 5 types of hand postures from 12 users were recorded using unlabeled
markers on fingers of a glove in a motion capture environment. Due to resolution and occlusion, mi
ssing values are common. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 78095 </td>
<td> 38 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Anuran+Calls+%28MFCCs%29">Anuran Calls (MFCCs) </td></tr></table></td>
<!-- <td> Acoustic features extracted from syllables of anuran (frogs) calls,
including the family, the genus, and the species labels (multilabel).   </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td> Real </td>
<td> 7195 </td>
<td> 22 </td>
<td> 2017 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a href="datasets/TTC-
3600%3A+Benchmark+dataset+for+Turkish+text+categorization"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization">TTC-3600: Benchmark
dataset for Turkish text categorization </td></tr></table></td>
<!-- <td> The TTC-3600 data set is a collection of Turkish news and articles i
ncluding categorized 3,600 documents from 6 well-known portals in Turkey. It has 4 different forms
in ARFF Weka format. </td> -->
<td> Text </td>
<td> Classification, Clustering </td>
<td> Integer </td>
<td> 3600 </td>
<td> 4814 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img src=
"assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a href
="datasets/Gastrointestinal+Lesions+in+Regular+Colonoscopy">Gastrointestinal Lesions in Regular Co
lonoscopy </td></tr></table></td>
<!-- <td> This dataset contains features extracted from colonoscopy videos use
d to detect gastrointestinal lesions. It contains 76 lesions: 15 serrated adenomas, 21
hyperplastic lesions and 40 adenoma.   </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 76 </td>
<td> 698 </td>
<td> 2016 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a

href="datasets/Daily+Demand+Forecasting+Orders">Daily Demand Forecasting Orders </td></tr></table></td>
<!-- <td> The dataset was collected during 60 days, this is a real database of a brazilian logistics company. </td> -->
<td> Time-Series </td>
<td> Regression </td>
<td> Integer </td>
<td> 60 </td>
<td> 13 </td>
<td> 2017 </td>
<!-- <td> Business </td> -->
</tr><tr>
<td><table><tr><td> </td><td> Paper Reviews </td></tr></table></td>
<!-- <td> This sentiment analysis data set contains scientific paper reviews from an international conference on computing and informatics. The task is to predict the orientation or the evaluation of a review. </td> -->
<td> Text </td>
<td> Classification, Regression </td>
<td> Integer </td>
<td> 405 </td>
<td> 10 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> extention of Z-Alizadeh sani dataset </td></tr></table></td>
<!-- <td> It was collected for CAD diagnosis. </td> -->
<td>   </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 303 </td>
<td> 59 </td>
<td> 2017 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td> </td><td> Z-Alizadeh Sani </td></tr></table></td>
<!-- <td> It was collected for CAD diagnosis. </td> -->
<td>   </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 303 </td>
<td> 56 </td>
<td> 2017 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> Dynamic Features of VirusShare Executables </td></tr></table></td>
<!-- <td> This dataset contains the dynamic features of 107,888 executables, collected by VirusShare from Nov/2010 to Jul/2014. </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Regression </td>
<td> Integer </td>
<td> 107888 </td>
<td> 482 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td> </td><td> IDA2016Challenge </td></tr></table></td>
<!-- <td> The dataset consists of data collected from heavy Scania trucks in everyday usage.   </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 76000 </td>
<td> 171 </td>
<td> 2017 </td>
<!-- <td> Computer </td> -->

```
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/DSRC+Vehicle+Communications"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/DSRC+Vehicle+Communications">DSRC Vehicle Communications</a></b></p></td></tr></tab
le></td>
   <!-- <td><p class="normal">This set Provides data regarding wireless communications between veh
icles and road side units. two separate data sets are provided (normal scenario) and in the presen
ce of attacker (jammer). </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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Mturk+User-Perceived+Clusters+over+Images"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 NT
HU, Taiwan. There're 325 user-perceived clusters from 100 users and their corresponding
descriptions. </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 </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Character+Font+Images"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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"><img src="assets/MLimages/SmallLargedefault.jpg" border=1
/></a> </td><td><p class="normal"><b><a href="datasets/DeliciousMIL%3A+A+Data+Set+for+Multi-L
abel+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. </p></td> -->
   <td><p class="normal">Text </p></td>
   <td><p class="normal">Classification </p></td>
   <td><p class="normal">Integer </p></td>
   <td><p class="normal">12234 </p></td>
   <td><p class="normal">8519 </p></td>
   <td><p class="normal">2016 </p></td>
   <!-- <td><p class="normal">Computer </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++">
<img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><
b><a href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++">Autistic Spectrum Di
sorder Screening Data for Children  </a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">Children screening data for autism suitable for classification and p
redictive 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><a
href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
```

```
href="datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Adolescent+++">Autistic Spectrum Diso
rder Screening Data for Adolescent   </a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">Autistic Spectrum Disorder Screening Data for Adolescent. This datas
et 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><a href="datasets/APS+Failure+at+Scania+Trucks"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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></t
able></td>
   <!-- <td><p class="normal">The datasets' positive class consists of component failures for a sp
ecific component of the APS system. The negative class consists of trucks with failures for compon
ents 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"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Wireless+Indoor+Localization">Wireless Indoor Localization</a></b></p></td></tr></t
able></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><a href="datasets/HCC+Survival"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 Un
iversity Hospital in Portugal. It contains real clinical data of 165 patients diagnosed with HCC.&
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">165 </p></td>
   <td><p class="normal">49 </p></td>
   <td><p class="normal">2017 </p></td>
   <!-- <td><p class="normal">Life </p></td> -->
   </tr><tr>
   <td><table><tr><td><a
href="datasets/CSM+%28Conventional+and+Social+Media+Movies%29+Dataset+2014+and+2015"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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. B
oth conventional features, collected from movies databases on Web as well as social media
features(YouTube,Twitter). </p></td> -->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal">Classification, Regression </p></td>
   <td><p class="normal">Integer </p></td>
   <td><p class="normal">217 </p></td>
   <td><p class="normal">12 </p></td>
   <td><p class="normal">2017 </p></td>
   <!-- <td><p class="normal">Computer </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a
href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/University+of+Tehran+Question+Dataset+2016+%28UTQD.2016%29">University of Tehran Qu
estion Dataset 2016 (UTQD.2016)</a></b></p></td></tr></table></td>
```

<!-- <td> Persian questions gathered from a jeopardy game broadcasted on Irani
an national television.   </td> -->
<td> Text </td>
<td> Classification </td>
<td>   </td>
<td> 1175 </td>
<td> 3 </td>
<td> 2017 </td>
<!-- <td> Other </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Autism+Screening+Adult">Autism Screening Adult </td></tr></table></td>
<!-- <td> Autistic Spectrum Disorder Screening Data for Adult. This dataset is
related to classification and predictive tasks. </td> -->
<td>   </td>
<td> Classification </td>
<td> Integer </td>
<td> 704 </td>
<td> 21 </td>
<td> 2017 </td>
<!-- <td> Social </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><a
href="datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor">
 </td><td> <
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 </td></t
r></table></td>
<!-- <td> Sequential motion data from 14 healthy older people aged 66 to 86
years old using a batteryless, wearable sensor on top of their clothing for the recognition of act
ivities in clinical environments. </td> -->
<td> Sequential </td>
<td> Classification </td>
<td> Real </td>
<td> 75128 </td>
<td> 9 </td>
<td> 2016 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Immunotherapy+Dataset">Immunotherapy Dataset </td></tr></table></td>
<!-- <td> This dataset contains information about wart treatment results of 90
patients using immunotherapy. </td> -->
<td> Univariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 90 </td>
<td> 8 </td>
<td> 2018 </td>
<!-- <td> Life </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Cryotherapy+Dataset+">Cryotherapy Dataset </td></tr></table></td>
<!-- <td> This dataset contains information about wart treatment results of 90
patients using cryotherapy. </td> -->
<td> Univariate </td>
<td> Classification </td>
<td> Integer, Real </td>
<td> 90 </td>
<td> 7 </td>
<td> 2018 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><
img src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <b
>OCT data & Color
Fundus Images of Left & Right Eyes </td></tr></table></td>
<!-- <td> This dataset contains OCT data (in mat format) and color fundus data
(in jpg format) of left & right eyes of 50 healthy persons. </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Real </td>
<td> 50 </td>

```
   <td><p class="normal">2 </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/Discrete+Tone+Image+Dataset"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Discrete+Tone+Image+Dataset">Discrete Tone Image Dataset</a></b></p></td></tr></tab
le></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.
 </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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/News+Popularity+in+Multiple+Social+Media+Platforms"><img s
rc="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a h
ref="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. </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"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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"><
img src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b
><a href="datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation">BLE RSSI Dataset for I
ndoor 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. </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> </td><td> Container Crane Controller Data Set </td></tr></table></td>
<!-- <td> A container crane has the function of transporting containers from one point to another point. </td> -->
<td> Univariate, Domain-Theory </td>
<td> Classification, Regression </td>
<td> Real </td>
<td> 15 </td>
<td> 3 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> Residential Building Data Set </td></tr></table></td>
<!-- <td> Data set includes construction cost, sale prices, project variables, and economic variables corresponding to real estate single-family residential apartments in Tehran, Iran.   </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Real </td>
<td> 372 </td>
<td> 105 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td> </td><td> Health News in Twitter </td></tr></table></td>
<!-- <td> The data was collected in 2015 using Twitter API. This dataset contains health news from more than 15 major health news agencies such as BBC, CNN, and NYT.   </td> -->
<td> Text </td>
<td> Clustering </td>
<td> Real </td>
<td> 58000 </td>
<td> 25000 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> chipseq </td></tr></table></td>
<!-- <td> ChIP-seq experiments characterize protein modifications or binding at
specific genomic locations in specific samples. The machine learning
problem in these data is structured binary classification. </td> -->
<td> Sequential </td>
<td> Classification </td>
<td> Integer </td>
<td> 4960 </td>
<td>   </td>
<td> 2018 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td> </td><td> SGEMM GPU kernel performance </td></tr></table></td>
<!-- <td> Running times for multiplying two 2048 x 2048 matrices using a GPU OpenCL SGEMM kernel with varying parameters (using the library 'CLTune'). </td> -->
<td> Multivariate </td>
<td> Regression </td>
<td> Integer </td>
<td> 241600 </td>
<td> 18 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td> </td><td> Repeat Consumption Matrices </td></tr></table></td>
<!-- <td> The dataset contains 7 datasets of User - Item matrices, where each

```
entry represents how many times a user consumed an item. Item is used as an umbrella term for vari
ous categories. </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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/detection_of_IoT_botnet_attacks_N_BaIoT"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><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. </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">1000000 </p></td>
   <td><p class="normal">115 </p></td>
   <td><p class="normal">2018 </p></td>
   <!-- <td><p class="normal">Computer </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Absenteeism+at+work"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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 Ju
ly 2007 to July 2010 at a courier company in Brazil. </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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/SCADI"><img src="assets/MLimages/SmallLargedefault.jpg" bo
rder=1 /></a> </td><td><p class="normal"><b><a href="datasets/SCADI">SCADI</a></b></p></td></
tr></table></td>
   <!-- <td><p class="normal">First self-care activities dataset based on ICF-CY. </p></td> -
->
   <td><p class="normal">Multivariate </p></td>
   <td><p class="normal">Classification, Clustering </p></td>
   <td><p class="normal"> </p></td>
   <td><p class="normal">70 </p></td>
   <td><p class="normal">206 </p></td>
   <td><p class="normal">2018 </p></td>
   <!-- <td><p class="normal">Life </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Condition+monitoring+of+hydraulic+systems"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Condition+monitoring+of+hydraulic+systems">Condition monitoring of hydraulic
systems</a></b></p></td></tr></table></td>
   <!-- <td><p class="normal">The data set addresses the condition assessment of a hydraulic test
rig based on multi sensor data. Four fault types are superimposed with several severity grades
impeding selective quantification. </p></td> -->
   <td><p class="normal">Multivariate, Time-Series </p></td>
   <td><p class="normal">Classification, Regression </p></td>
   <td><p class="normal">Real </p></td>
   <td><p class="normal">2205 </p></td>
   <td><p class="normal">43680 </p></td>
   <td><p class="normal">2018 </p></td>
   <!-- <td><p class="normal">Computer </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Carbon+Nanotubes"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Optical+Interconnection+Network+">Optical Interconnection Network </td>
</tr></table></td>
<!-- <td> This dataset contains 640 performance measurements from a simulation
of 2-Dimensional Multiprocessor Optical Interconnection Network.   </td> -->
<td> Multivariate </td>
<td> Classification, Regression </td>
<td> Integer, Real </td>
<td> 640 </td>
<td> 10 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Sports+articles+for+objectivity+analysis">Sports articles for objectivity
analysis </td></tr></table></td>
<!-- <td> 1000 sports articles were labeled using Amazon Mechanical Turk as ob
jective or subjective. The raw texts, extracted features, and the URLs from which the articles wer
e retrieved are provided. </td> -->
<td> Multivariate, Text </td>
<td> Classification </td>
<td> Integer </td>
<td> 1000 </td>
<td> 59 </td>
<td> 2018 </td>
<!-- <td> Social </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Breast+Cancer+Coimbra">Breast Cancer Coimbra </td></tr></table></td>
<!-- <td> Clinical features were observed or measured for 64 patients with bre
ast cancer and 52 healthy controls.   </td> -->
<td> Multivariate </td>
<td> Classification </td>
<td> Integer </td>
<td> 116 </td>
<td> 10 </td>
<td> 2018 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/GNFUV+Unmanned+Surface+Vehicles+Sensor+Data">GNFUV Unmanned Surface Vehicles Sensor
Data </td></tr></table></td>
<!-- <td> The data-set contains four (4) sets of mobile sensor readings data (
humidity, temperature) corresponding to a swarm of four (4) Unmanned Surface Vehicles (USVs) in a
test-bed in Athens (Greece).   </td> -->
<td> Multivariate, Time-Series </td>
<td> Regression </td>
<td> Real </td>
<td> 1672 </td>
<td> 5 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Dishonest+Internet+users+Dataset">Dishonest Internet users Dataset </td>
</tr></table></td>
<!-- <td> The dataset was used to test an architecture based on a trust model
capable to cope with the evaluation of the trustworthiness of users interacting in pervasive
environments. </td> -->
<td> Multivariate </td>
<td> Classification, Clustering </td>
<td>   </td>
<td> 322 </td>
<td> 5 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Victorian+Era+Authorship+Attribution">Victorian Era Authorship Attribution<

/p></td></tr></table></td>
<!-- <td> To create the largest authorship attribution dataset, we extracted w
orks of 50 well-known authors. To have a non-exhaustive learning, in training there are 45 authors
whereas, in the testing, it's 50 </td> -->
<td> Text </td>
<td> Classification </td>
<td>   </td>
<td> 93600 </td>
<td> 1000 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Simulated+Falls+and+Daily+Living+Activities+Data+Set">Simulated Falls and Daily
Living Activities Data Set </td></tr></table></td>
<!-- <td> 20 falls and 16 daily living activities were performed by 17
volunteers with 5 repetitions while wearing 6 sensors (3.060 instances) that attached to their hea
d, chest, waist, wrist, thigh and ankle. </td> -->
<td> Time-Series </td>
<td> Classification </td>
<td> Integer </td>
<td> 3060 </td>
<td> 138 </td>
<td> 2018 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><a
href="datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Multimodal+Damage+Identification+for+Humanitarian+Computing">Multimodal Damage Iden
tification for Humanitarian Computing </td></tr></table></td>
<!-- <td> 5879 captioned images (image and text) from social media related to
damage during natural disasters/wars, and belong to 6 classes: Fires, Floods, Natural landscape, I
nfrastructural, Human, Non-damage. </td> -->
<td> Multivariate, Text </td>
<td> Classification </td>
<td> Integer </td>
<td> 5879 </td>
<td>   </td>
<td> 2018 </td>
<!-- <td> Social </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img src
="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a hre
f="datasets/EEG+Steady-State+Visual+Evoked+Potential+Signals">EEG Steady-State Visual Evoked
Potential Signals </td></tr></table></td>
<!-- <td> This database consists on 30 subjects performing Brain Computer Inte
rface for Steady State Visual Evoked Potentials (BCI-SSVEP).   </td> -->
<td> Multivariate, Time-Series </td>
<td> Classification, Regression </td>
<td> Integer </td>
<td> 9200 </td>
<td> 16 </td>
<td> 2018 </td>
<!-- <td> Life </td> -->
</tr><tr>
<td><table><tr><td><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /> </td><td> <a
href="datasets/Roman+Urdu+Data+Set">Roman Urdu Data Set </td></tr></table></td>
<!-- <td> Roman Urdu (the scripting style for Urdu language) is one of the lim
ited resource languages.A data corpus comprising of more than 20000 records was collected. </
p></td> -->
<td> Text </td>
<td> Classification </td>
<td>   </td>
<td> 20000 </td>
<td> 2 </td>
<td> 2018 </td>
<!-- <td> Computer </td> -->
</tr><tr bgcolor="DDEEFF">
<td><table><tr><td><img src="assets/MLimages/SmallLargedefault.jpg" bo
rder=1 /> </td><td> Avila </td></
tr></table></td>
<!-- <td> The Avila data set has been extracted from 800 images of the 'Avila
Bible', an XII century giant Latin copy of the Bible. The prediction task consists in associating
each pattern to a copyist. </td> -->

```
   <td><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 </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/PANDOR"><img src="assets/MLimages/SmallLargedefault.jpg" b
order=1 /></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/).  </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 </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Drug+Review+Dataset+%28Druglib.com%29"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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. </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"> </p></td> -->
   </tr><tr>
   <td><table><tr><td><a href="datasets/Drug+Review+Dataset+%28Drugs.com%29"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></a> </td><td><p class="normal"><b><a
href="datasets/Drug+Review+Dataset+%28Drugs.com%29">Drug Review Dataset (Drugs.com)</a></b></p></t
d></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. </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 </p></td> -->
   </tr><tr bgcolor="DDEEFF">
   <td><table><tr><td><a href="datasets/Physical+Unclonable+Functions"><img
src="assets/MLimages/SmallLargedefault.jpg" border=1 /></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) s
imulation, specifically XOR Arbiter PUFs. PUFs are used for authentication purposes. For more info
, refer to our paper below. </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 </p></td> -->
   </tr></table>
 </td>
</tr>
</table>

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



<table cellpadding=5 align=center><tr valign=center>
```

```
  <td><p class="normal">Supported By:</p></td>
        <td><img src="assets/nsfe.gif" height=60 /> </td>
        <td><p class="normal"> In Collaboration With:</p></td>
        <td><img src="assets/rexaSmall.jpg" /></td>
</tr></table>

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

</body>
</html>
```

In [113]:

```
nombre = re.findall("<a href=\"datasets/(.+)</a>", response.text)
tipo = re.findall("<p class=\"normal(.+)>;</p>", response.text)
```

In [114]:

```
print("Número de nombres de bases de datos",str(len(nombre)))
print("Número de nombres de bases de datos",str(len(Tipo_datos)))
```

```
Número de nombres de bases de datos 446
Número de nombres de bases de datos 4033
```

In [93]:

```
db = {"Nombre":nombre}
```

In [94]:

```
database = pd.DataFrame(db)
database
```

Out[94]:

	Nombre
0	Abalone"><img src="assets/MLimages/SmallLarge1...
1	Adult"><img src="assets/MLimages/SmallLarge2.j...
2	Annealing"><img src="assets/MLimages/SmallLarg...
3	Anonymous+Microsoft+Web+Data"><img src="assets...
4	Arrhythmia"><img src="assets/MLimages/SmallLar...
5	Artificial+Characters"><img src="assets/MLimag...
6	Audiology+%28Original%29"><img src="assets/MLi...
7	Audiology+%28Standardized%29"><img src="assets...
8	Auto+MPG"><img src="assets/MLimages/SmallLarge...
9	Automobile"><img src="assets/MLimages/SmallLar...
10	Badges"><img src="assets/MLimages/SmallLargede...
11	Balance+Scale"><img src="assets/MLimages/Small...
12	Balloons"><img src="assets/MLimages/SmallLarge...
13	Breast+Cancer"><img src="assets/MLimages/Small...
14	Breast+Cancer+Wisconsin+%28Original%29"><img s...

	Nombre
15	Breast+Cancer+Wisconsin+%28Prognostic%29"><img...
16	Breast+Cancer+Wisconsin+%28Diagnostic%29"><img...
17	Pittsburgh+Bridges"><img src="assets/MLimages/...
18	Car+Evaluation"><img src="assets/MLimages/Smal...
19	Census+Income"><img src="assets/MLimages/Small...
20	Chess+%28King-Rook+vs.+King-Knight%29"><img sr...
21	Chess+%28King-Rook+vs.+King-Pawn%29"><img src=...
22	Chess+%28King-Rook+vs.+King%29"><img src="asse...
23	Chess+%28Domain+Theories%29"><img src="assets/...
24	Bach+Chorales"><img src="assets/MLimages/Small...
25	Connect-4"><img src="assets/MLimages/SmallLarg...
26	Credit+Approval"><img src="assets/MLimages/Sma...
27	Japanese+Credit+Screening"><img src="assets/ML...
28	Computer+Hardware"><img src="assets/MLimages/S...
29	Contraceptive+Method+Choice"><img src="assets/...
...	...
416	News+Popularity+in+Multiple+Social+Media+Platf...
417	Ultrasonic+flowmeter+diagnostics"><img src="as...
418	ICMLA+2014+Accepted+Papers+Data+Set"><img src=...
419	BLE+RSSI+Dataset+for+Indoor+localization+and+N...
420	Container+Crane+Controller+Data+Set"><img src=...
421	Residential+Building+Data+Set"><img src="asset...
422	Health+News+in+Twitter"><img src="assets/MLima...
423	chipseq"><img src="assets/MLimages/SmallLarged...
424	SGEMM+GPU+kernel+performance"><img src="assets...
425	Repeat+Consumption+Matrices"><img src="assets/...
426	detection_of_IoT_botnet_attacks_N_BaIoT"><img ...
427	Absenteeism+at+work"><img src="assets/MLimages...
428	SCADI"><img src="assets/MLimages/SmallLargedef...
429	Condition+monitoring+of+hydraulic+systems"><im...
430	Carbon+Nanotubes"><img src="assets/MLimages/Sm...
431	Optical+Interconnection+Network+"><img src="as...
432	Sports+articles+for+objectivity+analysis"><img...
433	Breast+Cancer+Coimbra"><img src="assets/MLimag...
434	GNFUV+Unmanned+Surface+Vehicles+Sensor+Data"><...
435	Dishonest+Internet+users+Dataset"><img src="as...
436	Victorian+Era+Authorship+Attribution"><img src...
437	Simulated+Falls+and+Daily+Living+Activities+Da...
438	Multimodal+Damage+Identification+for+Humanitar...
439	EEG+Steady-State+Visual+Evoked+Potential+Signa...
440	Roman+Urdu+Data+Set"><img src="assets/MLimages...
441	Avila"><img src="assets/MLimages/SmallLargedef...
442	PANDOR"><img src="assets/MLimages/SmallLargede...
443	Drug+Review+Dataset+%28Druglib.com%29"><img sr...
444	Drug+Review+Dataset+%28Drugs.com%29"><img src=...

	Nombre
445	Physical+Unclonable+Functions"><img src="asset

446 rows × 1 columns