

Breast Cancer Wisconsin

[http://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+\(diagnostic\)](http://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic))

Data Set Characteristics:	Multivariate	Number of Instances:	569	Area:	Life
Attribute Characteristics:	Real	Number of Attributes:	32	Date Donated	1995-11-01
Associated Tasks:	Classification	Missing Values?	No	Number of Web Hits:	1231977

DESCRIPCIÓN DEL PROBLEMA: Pacientes con tumores mamarios benignos y malignos.

CLASES Número: 2 clases (benigno (2), malign (4)). Desbalanceo: Sí, menos instancias clasificadas “maligno” (proporción 1.9)

VALORES PERDIDOS: Sí, 16 instancias, todas en la variable séptima.

:

Communities and Crime

<http://archive.ics.uci.edu/ml/datasets/communities+and+crime>

Data Set Characteristics:	Multivariate	Number of Instances:	1994	Area:	Social
Attribute Characteristics:	Real	Number of Attributes:	128	Date Donated	2009-07-13
Associated Tasks:	Regression	Missing Values?	Yes	Number of Web Hits:	278072

DESCRIPCIÓN DEL PROBLEMA: Se busca predecir el número total de crímenes violentos por cada 100.000 habitantes

VALORES PERDIDOS: Sí, 39202 repartidos en 25 columnas. 1871 tienen al menos un valor perdido en algún atributo.

Housing

<http://archive.ics.uci.edu/ml/machine-learning-databases/housing/>

Data Set Characteristics:	Multivariate	Number of Instances:	506	Area:	--
Attribute Characteristics:	Integer, Real	Number of Attributes:	14	Date Donated	1993
Associated Tasks:	Regression	Missing Values?	0	Number of Web Hits:	

DESCRIPCIÓN DEL PROBLEMA: Se puede afrontar el problema de dos modos distintos: predecir el valor de óxido de nitrógeno (variable NOX) o el precio medio de una casa.

VALORES PERDIDOS: No

DOCUMENTACIÓN: En otro documento. Completa y de fácil comprensión.

COMENTARIOS:

Cardiotocography

<http://archive.ics.uci.edu/ml/datasets/Cardiotocography>

Data Set Characteristics:	Multivariate	Number of Instances:	2126	Area:	Life
Attribute Characteristics:	Real	Number of Attributes:	23	Date Donated	2010-09-07
Associated Tasks:	Classification	Missing Values?	N/A	Number of Web Hits:	163281

DESCRIPCIÓN DEL PROBLEMA: Se estudian cardiotocogramas en fetos e intenta clasificar. Dos opciones de clasificación: CLASS-FHR (código de patrón de clase, son números del 1 al 10), NSP (código de estado de clase del feto, puede ser normal, sospechoso o patológico).

VALORES PERDIDOS: No

DOCUMENTACIÓN: En otro documento además de en el propio excel.

Completa y de fácil comprensión.

COMENTARIOS: Los datos se encuentran en un Excel dividido en varias hojas.

Occupancy detection

URL: <http://archive.ics.uci.edu/ml/datasets/Occupancy+Detection+>

Abstract: 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.

Data Set Characteristics:	Multivariate, Time-Series	Number of Instances:	20560	Area:	Computer
Attribute Characteristics:	Real	Number of Attributes:	7	Date Donated	2016-02-29
Associated Tasks:	Classification	Missing Values?	N/A	Number of Web Hits:	135433

Series temporales multivariable. Cada instancia contiene la medida de cada variable para un instante de tiempo dado.

Características:

Series temporales multivariable. 8143 instancias

7 Atributos por instancia.. 2 Clases.

Atributos: Reales.

Default of Credit Card Clients (clasificación)

URL: <http://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients>

Abstract: This research aimed at the case of customers' default payments in Taiwan and compares the predictive accuracy of probability of default among six data mining methods.

Data Set Characteristics:	Multivariate	Number of Instances:	30000	Area:	Business
Attribute Characteristics:	Integer, Real	Number of Attributes:	24	Date Donated	2016-01-26
Associated Tasks:	Classification	Missing Values?	N/A	Number of Web Hits:	463065

Problema de clasificación binaria Con un único fichero de datos. default of credit card clients.xls (Train)

Características:

30000 instancias en total., 24 Atributos por instancia., 2 Clases.

Atributos: Reales y enteros.

No existen valores perdidos.

Internet Advertisements (clasificación)

URL: <http://archive.ics.uci.edu/ml/datasets/internet+advertisements>

Abstract: This dataset represents a set of possible advertisements on Internet pages.

Data Set Characteristics:	Multivariate	Number of Instances:	3279	Area:	Computer
Attribute Characteristics:	Categorical, Integer, Real	Number of Attributes:	1558	Date Donated	1998-07-01
Associated Tasks:	Classification	Missing Values?	Yes	Number of Web Hits:	350410

Clasificación

Características:

3279 instancias en total., 1558 Atributos por instancia., 2 Clases.

Atributos: Reales, enteros y categóricos.

Human Activity Recognition Using Smartphones

URL:

<http://archive.ics.uci.edu/ml/datasets/Smartphone+Based+Recognition+of+Human+Activities+and+Postural+Transitions>

Abstract: Activity recognition data set built from the recordings of 30 subjects performing basic activities and postural transitions while carrying a waist-mounted smartphone with embedded inertial sensors.

Data Set Characteristics:	Multivariate, Time-Series	Number of Instances:	10929	Area:	Life
Attribute Characteristics:	Real	Number of Attributes:	561	Date Donated	2015-07-29
Associated Tasks:	Classification	Missing Values?	N/A	Number of Web Hits:	167105

Características:

Series temporales multivariable.

7767 instancias train, 3162 instancias de test.

561 Atributos por instancia.

12 Clases.

Image Segmentation

URL: <http://archive.ics.uci.edu/ml/datasets/image+segmentation>

Abstract: Image data described by high-level numeric-valued attributes, 7 classes

Data Set Characteristics:	Multivariate	Number of Instances:	2310	Area:	N/A
Attribute Characteristics:	Real	Number of Attributes:	19	Date Donated	1990-11-01
Associated Tasks:	Classification	Missing Values?	No	Number of Web Hits:	193966

Clasificación. Es necesario eliminar las 3 primeras líneas de los ficheros, junto a la quinta, para poder realizar la lectura de los datos.

Características:

2100 instancias train, 210 instancias de test. 19 Atributos por instancia. 7

Clases.

Atributos: Reales. No existen valores perdidos.

Student Performance Data Set (???)

URL: <http://archive.ics.uci.edu/ml/datasets/Student+Performance>

Abstract: Predict student performance in secondary education (high school).

Data Set Characteristics:	Multivariate	Number of Instances:	649	Area:	Social
Attribute Characteristics:	Integer	Number of Attributes:	33	Date Donated	2014-11-27
Associated Tasks:	Classification, Regression	Missing Values?	N/A	Number of Web Hits:	640325

Características:

649 instancias en el curso de portugués y 395 en el curso de matemáticas.

33 Atributos por instancia.

18 Clases. (Posibles valores numéricos del 0 al 20)

Atributos: Enteros y categóricos.

Ausencia de valores perdidos.

Clases prácticamente desbalanceadas.

(Matemáticas: 38/1/7/15/9/31/28/56/47/31/31/27/33/16/6/12/5/1)

(Portugués: 15/1/1/3/10/35/35/97/104/72/82/63/49/36/29/15/2)

APS Failure at Scania Trucks Data Set

Abstract: The datasets' positive class consists of component failures for a specific component of the APS system. The negative class consists of trucks with failures for components not related to the APS.

Data Set Characteristics:	Multivariate	Number of Instances	60000	Area:	Computer
Attribute Characteristics:	Integer, Real	Number of Attributes	171	Date Donated	2017-12-08
Associated Tasks:	Classification	Missing Values?	Yes	Number of Web Hits:	52387

Bank Marketing Data Set

Abstract: 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).

Data Set Characteristics:	Multivariate	Number of Instances:	45211	Area:	Business
Attribute Characteristics:	Real	Number of Attributes:	17	Date Donated	2012-02-14
Associated Tasks:	Classification	Missing Values?	N/A	Number of Web Hits:	1175692

Devanagari Handwritten Character Dataset Data Set

Download: [Data Folder](#), [Data Set Description](#)

<http://archive.ics.uci.edu/ml/datasets/Devanagari+Handwritten+Character+Dataset>

Abstract: 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%).

Data Set Characteristics:	N/A	Number of Instances:	92000	Area:	Computer
Attribute Characteristics:	Integer	Number of Attributes:	N/A	Date Donated	2016-09-01
Associated Tasks:	Classification	Missing Values?	N/A	Number of Web Hits:	26264

IDA2016Challenge Data Set

Download: [Data Folder](#), [Data Set Description](#)

<http://archive.ics.uci.edu/ml/datasets/IDA2016Challenge>

Abstract: The dataset consists of data collected from heavy Scania trucks in everyday usage.

Data Set Characteristics:	Multivariate	Number of Instances:	76000	Area:	Computer
Attribute Characteristics:	Integer	Number of Attributes:	171	Date Donated	2017-01-17
Associated Tasks:	Classification	Missing Values?	Yes	Number of Web Hits:	15209

Interesante data set con costes....

BlogFeedback Data Set

Download: [Data Folder](#), [Data Set Description](#)

<http://archive.ics.uci.edu/ml/datasets/BlogFeedback>

Abstract: 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.

Data Set Characteristics:	Multivariate	Number of Instances:	60021	Area:	Social
Attribute Characteristics:	Integer, Real	Number of Attributes:	281	Date Donated	2014-05-29
Associated Tasks:	Regression	Missing Values?	N/A	Number of Web Hits:	94031

Relative location of CT slices on axial axis Data Set

Download: [Data Folder](#), [Data Set Description](#)

<http://archive.ics.uci.edu/ml/datasets/Relative+location+of+CT+slashes+on+axial+axis>

Abstract: The dataset consists of 384 features extracted from CT images. The class variable is numeric and denotes the relative location of the CT slice on the axial axis of the human body.

Data Set Characteristics:	Domain-Theory	Number of Instances:	53500	Area:	Computer
Attribute Characteristics:	Real	Number of Attributes:	386	Date Donated	2011-07-07
Associated Tasks:	Regression	Missing Values?	N/A	Number of Web Hits:	55568

Online News Popularity Data Set

Download: [Data Folder](#), [Data Set Description](#)

<http://archive.ics.uci.edu/ml/datasets/Online+News+Popularity>

Abstract: This dataset summarizes a heterogeneous set of features about articles published by Mashable in a period of two years. The goal is to predict the number of shares in social networks (popularity).

Data Set Characteristics:	Multivariate	Number of Instances:	39797	Area:	Business
Attribute Characteristics:	Integer, Real	Number of Attributes:	61	Date Donated	2015-05-31
Associated Tasks:	Classification, Regression	Missing Values?	N/A	Number of Web Hits:	284127

Nomao Data Set

Download: [Data Folder](#), [Data Set Description](#)

<http://archive.ics.uci.edu/ml/datasets/Nomao>

Abstract: 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.

Data Set Characteristics:	Univariate	Number of Instances:	34465	Area:	Computer
Attribute Characteristics:	Real	Number of Attributes:	120	Date Donated	2012-07-04
Associated Tasks:	Classification	Missing Values?	Yes	Number of Web Hits:	50949

Letter Recognition Data Set

Download: [Data Folder](#), [Data Set Description](#)

<http://archive.ics.uci.edu/ml/datasets/Letter+Recognition>

Abstract: Database of character image features; try to identify the letter

Data Set Characteristics:	Multivariate	Number of Instances:	20000	Area:	Computer
Attribute Characteristics:	Integer	Number of Attributes:	16	Date Donated	1991-01-01
Associated Tasks:	Classification	Missing Values?	No	Number of Web Hits:	382385

<https://archive.ics.uci.edu/ml/datasets/Geographical+Original+of+Music>

Geographical Original of Music Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: 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.

Data Set Characteristics:	Multivariate	Number of Instances:	1059	Area:	N/A
Attribute Characteristics:	Real	Number of Attributes:	68	Date Donated	2014-10-18
Associated Tasks:	Classification, Regression	Missing Values?	N/A	Number of Web Hits:	95753

<https://archive.ics.uci.edu/ml/datasets/Weight+Lifting+Exercises+monitored+with+Inertial+Measurement+Units>

Weight Lifting Exercises monitored with Inertial Measurement Units Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Six young health subjects were asked to perform 5 variations of the biceps curl weight lifting exercise. One of the variations is the one predicted by the health professional.

Data Set Characteristics:	Multivariate	Number of Instances:	39242	Area:	Physical
Attribute Characteristics:	Real	Number of Attributes:	152	Date Donated	2013-11-24
Associated Tasks:	Classification	Missing Values?	Yes	Number of Web Hits:	49398

<https://archive.ics.uci.edu/ml/datasets/Grammatical+Facial+Expressions>

Grammatical Facial Expressions Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: This dataset supports the development of models that make possible to interpret Grammatical Facial Expressions from Brazilian Sign Language (Libras).

Data Set Characteristics:	Multivariate, Sequential	Number of Instances:	27965	Area:	Computer
Attribute Characteristics:	Real	Number of Attributes:	100	Date Donated	2014-10-06
Associated Tasks:	Classification, Clustering	Missing Values?	N/A	Number of Web Hits:	78402

<https://archive.ics.uci.edu/ml/datasets/Unmanned+Aerial+Vehicle+%28UAV%29+Intrusion+Detection>

Unmanned Aerial Vehicle (UAV) Intrusion Detection Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: For UAV identification, each input is an encrypted WiFi traffic record while the output is whether the current traffic is from a UAV or not. Meta-info on attribute relationship is also provided.

Data Set Characteristics:	Multivariate	Number of Instances:	17256	Area:	Computer
Attribute Characteristics:	Real	Number of Attributes:	55	Date Donated	2020-04-12
Associated Tasks:	Classification	Missing Values?	N/A	Number of Web Hits:	8854

<https://archive.ics.uci.edu/ml/datasets/Swarm+Behaviour>

Swarm Behaviour Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: This dataset achieved from an online survey, which is run by UNSW, Australia. It contains three data of ' Flocking - Not Flocking', 'Aligned - Not Aligned', and 'Grouped - Not Grouped'.

Data Set Characteristics:	Multivariate	Number of Instances:	24017	Area:	Computer
Attribute Characteristics:	Real	Number of Attributes:	2400	Date Donated	2020-06-16
Associated Tasks:	Classification	Missing Values?	N/A	Number of Web Hits:	23042

<https://archive.ics.uci.edu/ml/datasets/Seoul+Bike+Sharing+Demand>

Seoul Bike Sharing Demand Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: The dataset contains count of public bikes rented at each hour in Seoul Bike haring System with the corresponding Weather data and Holidays information

Data Set Characteristics:	Multivariate	Number of Instances:	8760	Area:	Computer
Attribute Characteristics:	Integer, Real	Number of Attributes:	14	Date Donated	2020-03-01
Associated Tasks:	Regression	Missing Values?	N/A	Number of Web Hits:	27350

<https://archive.ics.uci.edu/ml/datasets/Bias+correction+of+numerical+prediction+model+temperature+forecast>

Bias correction of numerical prediction model temperature forecast Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: It contains fourteen numerical weather prediction (NWP)'s meteorological forecast data, two in-situ observations, and five geographical auxiliary variables over Seoul, South Korea in the summer.

Data Set Characteristics:	Multivariate	Number of Instances:	7750	Area:	Physical
Attribute Characteristics:	Real	Number of Attributes:	25	Date Donated	2020-02-18
Associated Tasks:	Regression	Missing Values?	Yes	Number of Web Hits:	24651

<https://archive.ics.uci.edu/ml/datasets/Wave+Energy+Converters>

Wave Energy Converters Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: This data set consists of positions and absorbed power outputs of wave energy converters (WECs) in four real wave scenarios from the southern coast of Australia.

Data Set Characteristics:	Multivariate	Number of Instances:	288000	Area:	Computer
Attribute Characteristics:	Real	Number of Attributes:	49	Date Donated	2019-06-30
Associated Tasks:	Regression	Missing Values?	Yes	Number of Web Hits:	26548

<https://archive.ics.uci.edu/ml/datasets/Facebook+Comment+Volume+Dataset>

Facebook Comment Volume Dataset Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Instances in this dataset contain features extracted from facebook posts. The task associated with the data is to predict how many comments the post will receive.

Data Set Characteristics:	Multivariate	Number of Instances:	40949	Area:	N/A
Attribute Characteristics:	Integer, Real	Number of Attributes:	54	Date Donated	2016-03-11
Associated Tasks:	Regression	Missing Values?	N/A	Number of Web Hits:	108042

<https://archive.ics.uci.edu/ml/datasets/optical+recognition+of+handwritten+digits>

Optical Recognition of Handwritten Digits Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Two versions of this database available; see folder

Data Set Characteristics:	Multivariate	Number of Instances:	5620	Area:	Computer
Attribute Characteristics:	Integer	Number of Attributes:	64	Date Donated	1998-07-01
Associated Tasks:	Classification	Missing Values?	No	Number of Web Hits:	306734

<https://archive.ics.uci.edu/ml/datasets/airfoil+self-noise>

Airfoil Self-Noise Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: NASA data set, obtained from a series of aerodynamic and acoustic tests of two and three-dimensional airfoil blade sections conducted in an anechoic wind tunnel.

Data Set Characteristics:	Multivariate	Number of Instances:	1503	Area:	Physical
Attribute Characteristics:	Real	Number of Attributes:	6	Date Donated	2014-03-04
Associated Tasks:	Regression	Missing Values?	N/A	Number of Web Hits:	153731

<http://archive.ics.uci.edu/ml/datasets/Communities+and+Crime>

Communities and Crime Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Communities within the United States. The data combines socio-economic data from the 1990 US Census, law enforcement data from the 1990 US LEMAS survey, and crime data from the 1995 FBI UCR.

Data Set Characteristics:	Multivariate	Number of Instances:	1994	Area:	Social
----------------------------------	--------------	-----------------------------	------	--------------	--------

Attribute Characteristics:	Real	Number of Attributes:	128	Date Donated	2009-07-13
Associated Tasks:	Regression	Missing Values?	Yes	Number of Web Hits:	319228