

Title: Weka (software)

URL: [https://en.wikipedia.org/wiki/Weka\\_\(software\)](https://en.wikipedia.org/wiki/Weka_(software))

PageID: 3829034

Categories: Category:Data mining and machine learning software, Category:Free artificial intelligence applications, Category:Free data analysis software, Category:Free science software, Category:Free software programmed in Java (programming language), Category:Software using the GNU General Public License

Source: Wikipedia (CC BY-SA 4.0). Content may require attribution.

-----

Waikato Environment for Knowledge Analysis ( Weka ) is a collection of machine learning and data analysis free software licensed under the GNU General Public License . It was developed at the University of Waikato , New Zealand and is the companion software to the book "Data Mining: Practical Machine Learning Tools and Techniques".

#### Description

Weka contains a collection of visualization tools and algorithms for data analysis and predictive modeling , together with graphical user interfaces for easy access to these functions. The original non-Java version of Weka was a Tcl / Tk front-end to (mostly third-party) modeling algorithms implemented in other programming languages, plus data preprocessing utilities in C , and a makefile -based system for running machine learning experiments. This original version was primarily designed as a tool for analyzing data from agricultural domains, but the more recent fully Java -based version (Weka 3), for which development started in 1997, is now used in many different application areas, in particular for educational purposes and research. Advantages of Weka include:

Free availability under the GNU General Public License .

Portability, since it is fully implemented in the Java programming language and thus runs on almost any modern computing platform.

A comprehensive collection of data preprocessing and modeling techniques.

Ease of use due to its graphical user interfaces.

Weka supports several standard data mining tasks, more specifically, data preprocessing, clustering , classification , regression , visualization , and feature selection . Input to Weka is expected to be formatted according the Attribute-Relational File Format and with the filename bearing the .arff extension. All of Weka's techniques are predicated on the assumption that the data is available as one flat file or relation, where each data point is described by a fixed number of attributes (normally, numeric or nominal attributes, but some other attribute types are also supported). Weka provides access to SQL databases using Java Database Connectivity and can process the result returned by a database query. Weka provides access to deep learning with Deeplearning4j . It is not capable of multi-relational data mining, but there is separate software for converting a collection of linked database tables into a single table that is suitable for processing using Weka. Another important area that is currently not covered by the algorithms included in the Weka distribution is sequence modeling.

#### Extension packages

In version 3.7.2, a package manager was added to allow the easier installation of extension packages. Some functionality that used to be included with Weka prior to this version has since been moved into such extension packages, but this change also makes it easier for others to contribute extensions to Weka and to maintain the software, as this modular architecture allows independent updates of the Weka core and individual extensions.

#### History

In 1993, the University of Waikato in New Zealand began development of the original version of Weka, which became a mix of Tcl/Tk, C, and makefiles.

In 1997, the decision was made to redevelop Weka from scratch in Java, including implementations of modeling algorithms.

In 2005, Weka received the SIGKDD Data Mining and Knowledge Discovery Service Award.

In 2006, Pentaho Corporation acquired an exclusive licence to use Weka for business intelligence . It forms the data mining and predictive analytics component of the Pentaho business intelligence suite. Pentaho has since been acquired by Hitachi Vantara, and Weka now underpins the PMI (Plugin for Machine Intelligence) open source component.

#### Related tools

Auto-WEKA is an automated machine learning system for Weka.

Environment for DeveLoping KDD-Applications Supported by Index-Structures ( ELKI ) is a similar project to Weka with a focus on cluster analysis , i.e., unsupervised methods.

H2O.ai is an open-source data science and machine learning platform

KNIME is a machine learning and data mining software implemented in Java .

Massive Online Analysis (MOA) is an open-source project for large scale mining of data streams, also developed at the University of Waikato in New Zealand .

Neural Designer is a data mining software based on deep learning techniques written in C++ .

Orange is a similar open-source project for data mining, machine learning and visualization based on scikit-learn .

RapidMiner is a commercial machine learning framework implemented in Java which integrates Weka.

scikit-learn is a popular machine learning library in Python.

See also

Free and open-source software portal

List of numerical-analysis software

#### References

#### External links

Official website at University of Waikato in New Zealand

v

t

e

Advanced Simulation Library

ADMB

Chapel

Euler Mathematical Toolbox

FreeFem++

FreeMat

Genius

Gmsh

GNU Octave

gretl

Julia

Jupyter ( Ju lia, Pyt hon, R ; IPython )

MFEM

OpenFOAM

Python

R

SageMath

Salome

ScicosLab

Scilab

X10

Weka

Fortress

DADiSP

FEATool Multiphysics

GAUSS

LabVIEW

Maple

Mathcad

Mathematica

MATLAB

MWorks

SAS ( SAS Viya )

Speakeasy

VisSim

Comparison