Background
- Implements a collection of machine learning algorithms for solving data mining problems
- Provides tools for data pre-processing, classification, regression, clustering, association rules
- Offers visualization of data and two GUIs named as WEKA Explorer and WEKA Experimenter

Selected Algorithms
- C4.5 decision tree learner
- Naïve Bayesian learner
- Multi-response linear regression, logistic regression
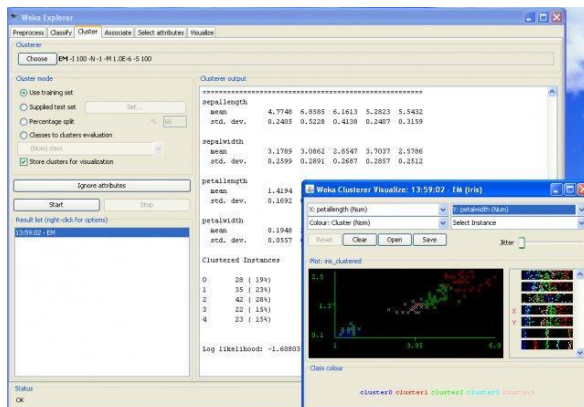- Support Vector Machine (linear, polynomial, radial basis function kernels)

Datasets & Data Files
- Implemented as data set 'Instance' class with attributes of different types: nominal (one of a predefined list of values), numeric (a real/integer number), or string (characters as "text")
- Handling data files with ARFF files (external representation of instance class): header with attribute types and data as comma-separated list
- Use WEKA.FILTERS package in order to transform datasets in the pre-processing phase: e.g. removing/adding attributes, resampling dataset, removing examples

Technical Short Description
- Written in Java (runs on most platforms), GNU General Public License
- Apply directly to a dataset as a tool (command-line client, GUIs)
- Embed as library called from Java code

Tool Snapshot (from [4])

Download and availability
- WEKA - Version 3.6.10
- Download @ Sourceforge [4]

References

[1] WEKA Home page with Documentation, Online: http://www.cs.waikato.ac.nz/ml/weka/

[2] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.

[3] WEKA Primer on Wiki, Online: http://weka.wikispaces.com/Primer

[4] WEKA @ SourceForge, Online: http://sourceforge.net/projects/weka/

Tags: **Machine Learning, Data Mining, Java, SVM, C4.5, Bayesian, Regression**