

Weka is a free open source tool for data analysis , machine learning and data mining tasks

The tool is written in Java

Features:

1. User-friendly interface that is easy to explore different features
2. Not much programming skills are required to use the tool
3. Supports different data formats like
  - a. ARFF - attribute relation file format
  - b. CSV- comma separated values
  - c. Json
  - d. Excel files
  - e. C4.5 used in decision trees
  - f. LibSVM for support vector machines
  - g. Database connectivity
4. Number of libraries for data preprocessing
  - a. Handle missing data
  - b. Remove duplicates & outliers
  - c. Attribute selection
  - d. Normalization- attribute values are brought to similar ranges
  - e. Attribute transformations- convert categorical and strings to numeric
  - f. Data integration- data can be pulled from multiple sources and combine
5. Classification algorithms- decision trees, Random forests, Support vector machines, Logistic regression, naive bayes, multi layer perceptrons
6. Ensembling methods like Adaboost, gradient boost, bagging algorithms etc
7. Clustering algorithms to group the datapoints.
  - a. Algorithms are like k-means, hierarchical clustering, aglomerative clustering, DBScan, Gaussian mixture models .
  - b. these algorithms are used for
    - i. document classification to group articles like education articles, political, entertainment, sports etc
    - ii. customer segmantation group them based on preferences

- iii. Anomaly detection- identify the one that is unusual,
  - iv. image segmentation- identify different types of objects in a image
8. All visualizations- graphs & charts like line plot, bar, scatter, box, histogram , pie etc
  9. Integrate weka with python using weka.core library
  10. Association mining

#### Use cases

1. In education to teach machine learning concepts
2. Used by researchers to develop proto type models and evaluate models
3. Scientists for primary data analysis and model selection