ABOUT WEKA TERMS:

What are the features of Weka?

**Weka features** include machine learning, data mining, preprocessing, classification, regression, clustering, association rules, attribute selection, experiments, workflow and visualization.

What do you mean by data mining?

**Definition** of '**Data Mining**' **Definition**: In simple words, **data mining** is defined as a process used to extract usable **data** from a larger set of any raw **data**. It implies analysing **data** patterns in large batches of **data** using one or more software. ... **Data mining** is also known as Knowledge Discovery in **Data** (KDD).

Introduction to **Data Mining**

We can simply define **data mining** as a process that involves searching, collecting, filtering and analyzing the **data**. ... The **data** collected can be stored for future use. Storage of information is quite **important** whenever it is required.

What is data mining and why it is required?

**Data mining** is the process of analyzing hidden patterns of **data** according to different perspectives for categorization into useful information, which is collected and assembled in common areas, such as **data** warehouses, for efficient analysis, **data mining** algorithms, facilitating business decision making and other ...

What are the four major types of data mining tools?

**Data Mining** is an important analytic process designed to explore **data**.  
...  
**4 Data Mining Techniques for Businesses (That Everyone Should Know)**

* Regression (predictive)
* Association Rule Discovery (descriptive)
* Classification (predictive)
* Clustering (descriptive)

What are the types of data mining?

Five **Data Mining** Techniques That Help Create Business Value. There are many **different types** of analysis to retrieve information from big **data**. ... **Data mining** is a buzzword that often is used to describe the entire range of big **data** analytics, including collection, extraction, analysis and statistics.

Why is datamining used?

For businesses, **data mining is used** to discover patterns and relationships in the data in order to help make better business decisions. **Data mining** can help spot sales trends, develop smarter marketing campaigns, and accurately predict customer loyalty.

What are the benefits of data mining?

In finance and banking, **data mining** is used to create accurate risk models for loans and mortgages. They are also very helpful when detecting fraudulent transactions. In marketing, **data mining** techniques are used to improve conversions, increase customer satisfaction and created targeted advertising campaigns.

What are the four data mining techniques?

**The 7 Most Important Data Mining Techniques**

* Data Mining Techniques.
* Tracking patterns. One of the most basic techniques in data mining is learning to recognize patterns in your data sets. ...
* Classification. ...
* Association. ...
* Outlier detection. ...
* Clustering. ...
* Regression. ...
* Prediction

What are the tools for data mining?

**Top 15 Best Free Data Mining Tools: The Most Comprehensive List**

* #1) Rapid Miner.
* #2) Orange.
* #3) Weka.
* #4) KNIME.
* #4) Sisense.
* #5) SSDT (SQL Server Data Tools)
* #6) Apache Mahout.
* #7) Oracle Data Mining.
* #8) Rattle
* #9) DataMelt
* #10) IBM Cognos
* #11) IBM SPSS Modeler
* #12) SAS Data Mining
* #13) Teradata
* #14) Board
* #15) Dundas BI

What is Apriori algorithm in data mining?

• The **Apriori Algorithm** is an influential **algorithm** for **mining** frequent itemsets for boolean association rules. • **Apriori** uses a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as candidate generation, and groups of candidates are tested against the **data**.

What is decision tree in data mining?

**Data Mining** - **Decision Tree** Induction. Advertisements. A **decision tree** is a structure that includes a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. The topmost node in the **tree** is the root node.

What is decision tree in weka?

Package **weka**.classifiers.**trees**

A Hoeffding **tree** (VFDT) is an incremental, anytime **decision tree** induction algorithm that is capable of learning from massive data streams, assuming that the distribution generating examples does not change over time. J48. Class for generating a pruned or unpruned C4.5 **decision tree**. LMT.

What is decision tree in machine learning?

**Decision Trees** are a type of Supervised **Machine Learning** (that is you explain what the input is and what the corresponding output is in the training data) where the data is continuously split according to a certain parameter. The **tree** can be explained by two entities, namely **decision** nodes and leaves.

**Weka** Knowledge **Explorer**. ... Each of the major **weka** packages Filters, Classifiers, Clusterers, Associations, and Attribute Selection is represented in the **Explorer** along with a Visualization tool which allows datasets and the predictions of Classifiers and Clusterers to be visualized in two dimensions.

Why is Weka used?

**Weka** contains tools for data pre-processing, classification, regression, clustering, association rules, and visualisation. ... Data mining uses machine language to find valuable information from large volumes of data. **Weka**. **Weka** is data mining software that uses a collection of machine learning algorithms.

What is Weka API?

**Weka** is a standard Java tool for performing both machine learning experiments and for embedding trained models in Java applications. It can be used for supervised and unsupervised learning. There are three ways to use **Weka** first using command line, second using **Weka** GUI, and third through its **API** with Java.

What is an ARFF file?

**ARFF** stands for Attribute-Relation **File** Format. It is an ASCII text **file** that describes a list of instances sharing a set of attributes. **ARFF files** were developed by the Machine Learning Project at the Department of Computer Science of The University of Waikato for use with the Weka machine learning software.

What is Weka regression?

What is decision tree analysis?

**Decision Tree Analysis**. Definition: The **Decision Tree Analysis** is a schematic representation of several **decisions** followed by different chances of the occurrence. ... These **decisions** are followed by the chance points, represented by circles, are the uncertain points, where the outcomes are dependent on the chance process.

Linear **regression** only supports **regression** type problems. It works by estimating coefficients for a line or hyperplane that best fits the training data. It is a very simple **regression** algorithm, fast to train and can have great performance if the output variable for your data is a linear combination of your inputs.

What is Weka workbench?

The **WEKA workbench** is a collection of machine learning algorithms and data preprocessing tools that includes virtually all the algorithms described in our book. It is designed so that you can quickly try out existing methods on new datasets in flexible ways.

What is classification in Weka?

**Weka** makes a large number of **classification** algorithms available. The large number of machine learning algorithms available is one of the benefits of using the **Weka** platform to work through your machine learning problems. ... The key configuration parameters for 5 top **classification** algorithms.

What is classification in data mining?

**Classification** is a **data mining** function that assigns items in a collection to target categories or classes. The goal of **classification** is to accurately predict the target class for each case in the **data**. For example, a **classification** model could be used to identify loan applicants as low, medium, or high credit risks.

What is the use of classification?

**Classification** is a data mining technique that assigns categories to a collection of data in order to aid in more accurate predictions and analysis. Also called sometimes called a Decision Tree, **classification** is one of several methods intended to make the analysis of very large datasets effective.

How do you classify data?

**Data classification** is the process of sorting and categorizing **data** into various types, forms or any other distinct class. **Data classification** enables the separation and **classification** of **data** according to **data** set requirements for various business or personal objectives. It is mainly a **data** management process.

What is the goal of classification?

The **goal of classification** is to take input data, and predict a category for that data from a discrete set of possible values. For example: Classifying emails as spam or not spam. Giving a diagnosis for a patient, given a set of symptoms.

What is j48 in Weka?

The C4.5 algorithm for building decision trees is implemented in **Weka** as a classifier called **J48**. Classifiers, like filters, are organized in a hierarchy: **J48** has the full name **weka**.classifiers.trees.**J48**. The classifier is shown in the text box next to the Choose button: It reads **J48** –C 0.25 –M 2

What is j48?

**J48** is an extension of ID3. The additional features of **J48** are accounting for missing values, decision trees pruning, continuous attribute value ranges, derivation of rules, etc. In the WEKA data mining tool, **J48** is an open source Java implementation of the C4.5 algorithm

What is j48 in data mining?

It involves systematic analysis of large **data** sets. The classification is used to manage **data**, sometimes tree modeling of **data** helps to make predictions about new **data**. This paper is focused on **J48** algorithm which is used to create univariate decision trees.

What is Kappa statistics in Weka?

In **weka**, **Kappa** is a chance-‐corrected measure of agreement between the classifications and the true classes. It's calculated by taking the agreement expected by chance away from the observed agreement and dividing by the maximum possible agreement

What is the use of kappa statistics?

Cohen's **kappa coefficient** (κ) is a **statistic** which measures inter-rater agreement for qualitative (categorical) items. It is generally thought to be a more robust measure than simple percent agreement calculation, as κ takes into account the possibility of the agreement occurring by chance.

What does negative Kappa mean?

**Kappa** is a chance corrected measure of agreement, and can be **negative**. A **negative Kappa means** that there is less agreement than **would** be expected by chance given the marginal distributions of. ratings.

What are the applications of Weka?

**Weka** is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. **Weka** contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization.

How does Weka tool work?

algorithms can either be applied directly to a dataset or called from your own Java code. **Weka** contains **tools** for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes. More info in **Weka** Manual p. 15 & 16.

How do we implement decision tree in weka?

Is Weka an open source?

**Weka** is a collection of machine learning algorithms for data mining tasks. It contains tools for data preparation, classification, regression, clustering, association rules mining, and visualization. ... **Weka** is **open source** software issued under the GNU General Public License.

How does Weka tool work?

algorithms can either be applied directly to a dataset or called from your own Java code. **Weka** contains **tools** for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes. More info in **Weka** Manual p. 15 & 16

How do I save a file as Arff?

**Save** your dataset in **ARFF** format by clicking the “**File**” menu and selecting “**Save** as…”. Enter a filename with a .**arff** extension and click the “**Save**” button. You can now load your **saved** .**arff file** directly into Weka.

How do I convert a CSV file to Arff?

**Steps**

1. Open Weka. ...
2. Click the Tools menu. ...
3. Click ArffViewer. ...
4. Click the File menu. ...
5. Click Open. ...
6. Navigate to the folder that contains the .CSV file.
7. Select CSV data files (\*.csv) from the ″Files of Type″ menu. ...
8. Select the .CSV and click Open.

What is correctly classified instances in Weka?

Similarly, incorrectly **classified instances** means the sum of FP and FN. The total number of **correctly instances** divided by total number of **instances** gives the accuracy. In **weka**, % of **correctly classified instances** give the accuracy of the model. Cheers! 2 Recommendations.

What is SMO in Weka?

Sequential minimal optimization (**SMO**) is an algorithm for solving the quadratic programming (QP) problem that arises during the training of support-vector machines (SVM). ... **SMO** is widely used for training support vector machines and is implemented by the popular LIBSVM tool.

What is JRIP algorithm?

It is based in association rules with reduced error pruning (REP), a very common and effective technique found in decision tree **algorithms**. ... At each stage of simplification, the pruning operator chosen is the one that yields the greatest reduction of error on the pruning set.

What is naive Bayes algorithm in data mining?

The **Naive Bayes** classification **algorithm** is a probabilistic **classifier**. It is based on probability models that incorporate strong independence assumptions. ... **Data mining** in DB2® Warehouse is based on the maximum likelihood for parameter estimation for **Naive Bayes** models.

What is naive Bayes classifier algorithm?

**Naive Bayes classifiers** are a collection of classification **algorithms** based on **Bayes**' Theorem. It is not a single **algorithm** but a family of **algorithms** where all of them share a common principle, i.e. every pair of features being classified is independent of each other. To start with, let us consider a dataset.

How does decision stump work?

A **decision stump is** a machine learning model consisting of a one-level **decision** tree. That **is**, it **is** a **decision** tree with one internal node (the root) which **is** immediately connected to the terminal nodes (its leaves). A **decision stump** makes a prediction based on the value of just a single input feature.

What does confusion matrix mean?

A **confusion matrix is** a table that **is** often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values **are** known. The **confusion matrix** itself **is** relatively simple to understand, but the related terminology **can** be **confusing**.

**Converting CSV File to ARFF file:**

Either load the CSV file in the Explorer or use the CSV converter on the commandline as follows:

java weka.core.converters.CSVLoader filename.csv > filename.arff