Experiment Number 06

Aim: Implementation of Naïve Bayes classifier using WEKA

Problem Statement: To implement Naïve Bayes classifier using WEKA

Theory:

It is a simple probabilistic classifier based on applying Bayesian' theorem with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model".

In simple terms, a naive Bayesian classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. For example, a fruit may be considered to be an apple if it is red, round, and about 4" in diameter. Even if these features depend on each other or upon the existence of the other features, a naive Bayesian classifier considers all of these properties to independently contribute to the probability that this fruit is an apple.

Depending on the precise nature of the probability model, naive Bayesian classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayesian models uses the method of maximum likelihood; in other words, one can work with the naive Bayesian model without believing in Bayesian probability or using any Bayesian methods.

An advantage of the naive Bayesian classifier is that it only requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix.
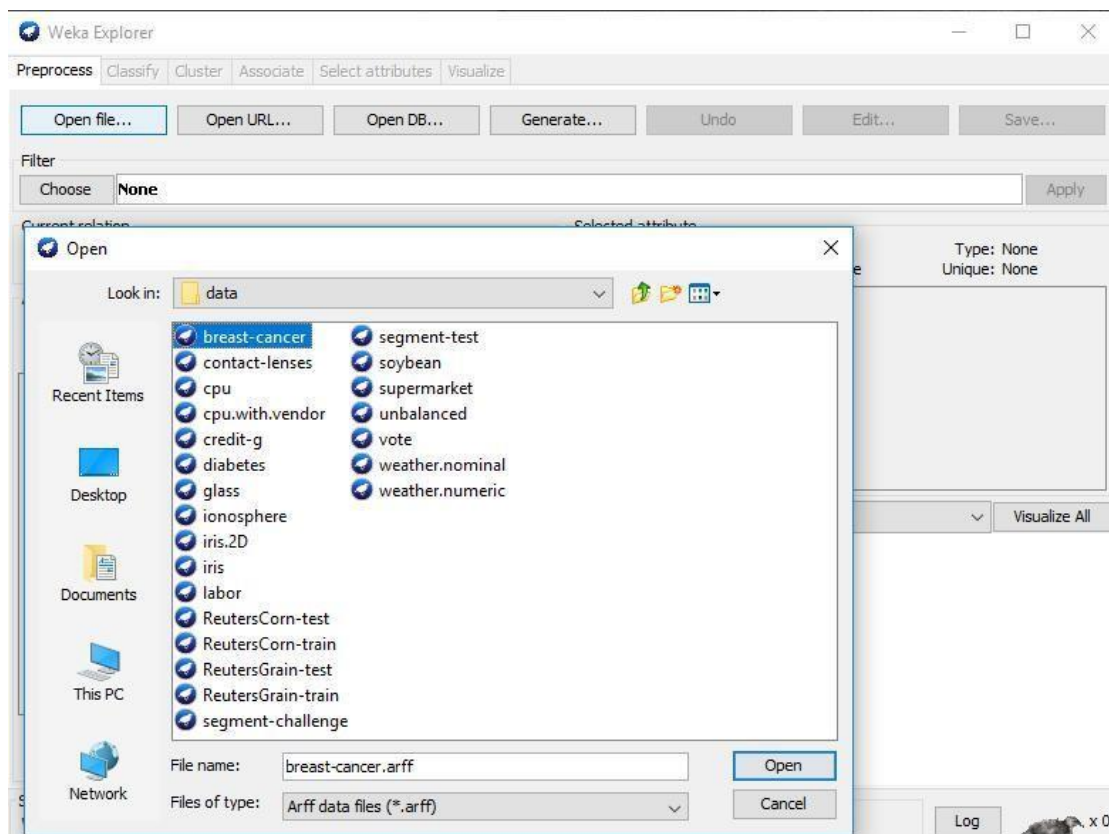
Procedure:

Open a data file

In "Preprocess" panel, click "Open file" button, and choose an ARFF file from "data" folder. Now the other tabs are active.
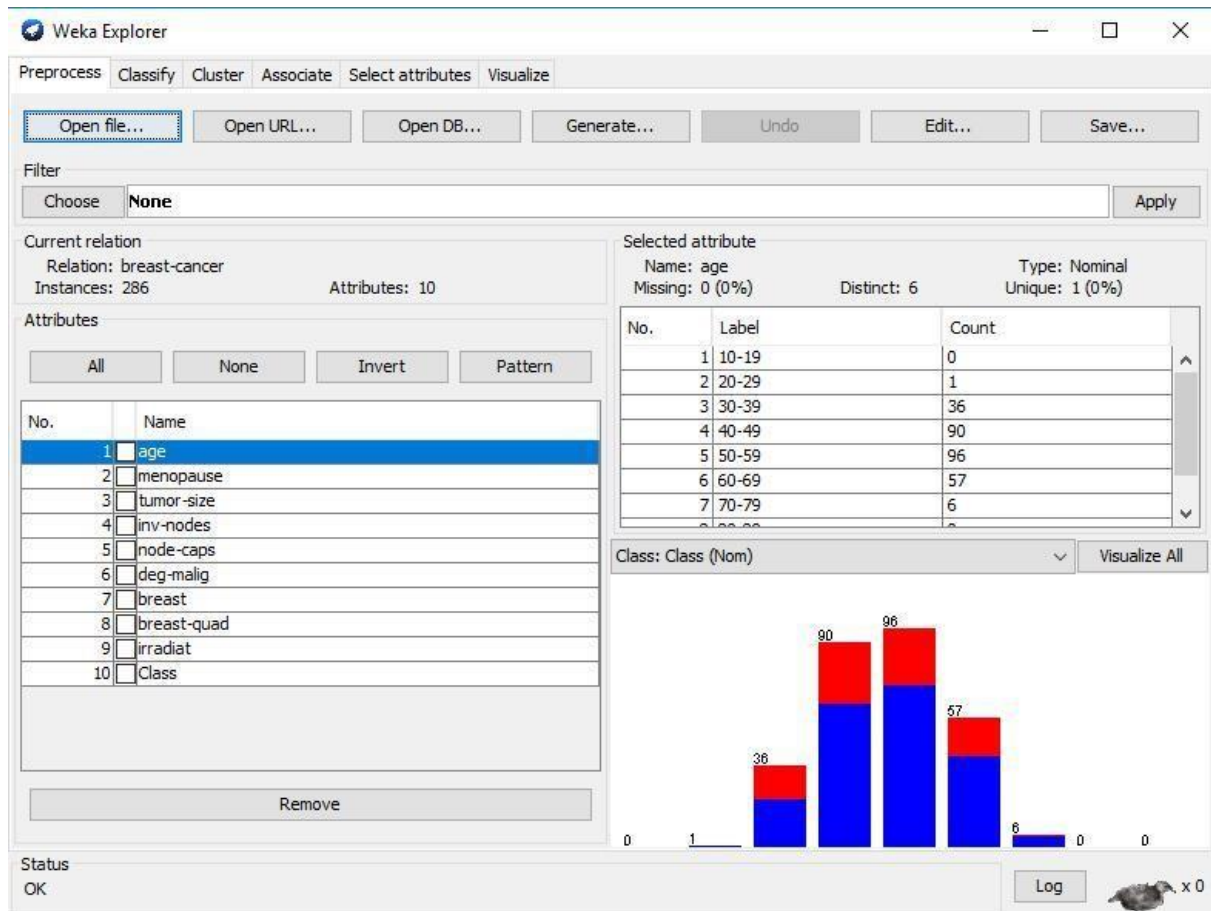
If you specify a "CSV" file, it will be automatically converted into ARFF file.

Select for example breast-cancer.arff .

Pre-process Panel

Pre-process Panel

- Current relation pane: specification of data file
- Attributes and Remove button to possibly remove some attributes from the experiment.
- Statistics about selected attribute (highlighted attribute.) – If you select a numeric attribute, If you select a numeric attribute, Min, Max, Mean, and standard deviation are shown.
- Histogram shows the distribution of the class as the function of the selected attribute.
- Edit button to edit input data on a separate window – In the Edit window, right click on the caption of column or values opens a list of various available editions on data. – Undo, and Save buttons have the known function
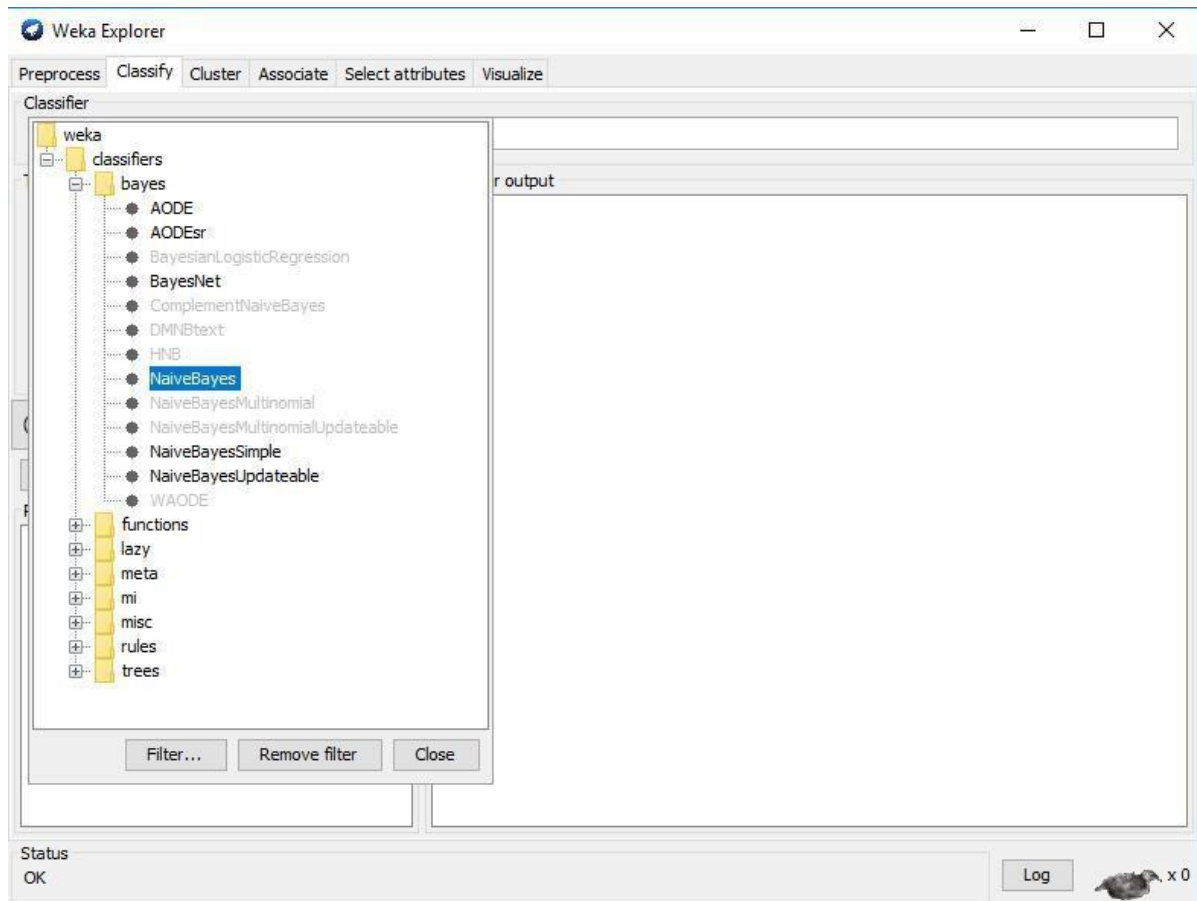
Build a Naïve bayes classifier

- Switch to "Classify" tab
- Select "Bayes"– by Clicking "Choose" button – Selecting classifiers >> Bayes >> NaiveBayes

- Invoke classifier by clicking "start" button
- Weka keeps the results of different classifiers in the "Result List" pane.

Classifier

Naïve Bayes classifier Result

Classifier Output

Classifier output includes:

- Summary of the data set
- Correctly Classified Instances
- Incorrectly Classified Instances
- Total Number of Instances
- Detailed Accuracy By Class
- Confusion Matrix
- Some other statistics

Conclusion:

Thus we have studied implementation of Naïve Bayes classification in weka.