EXPERIMENT NO. 6

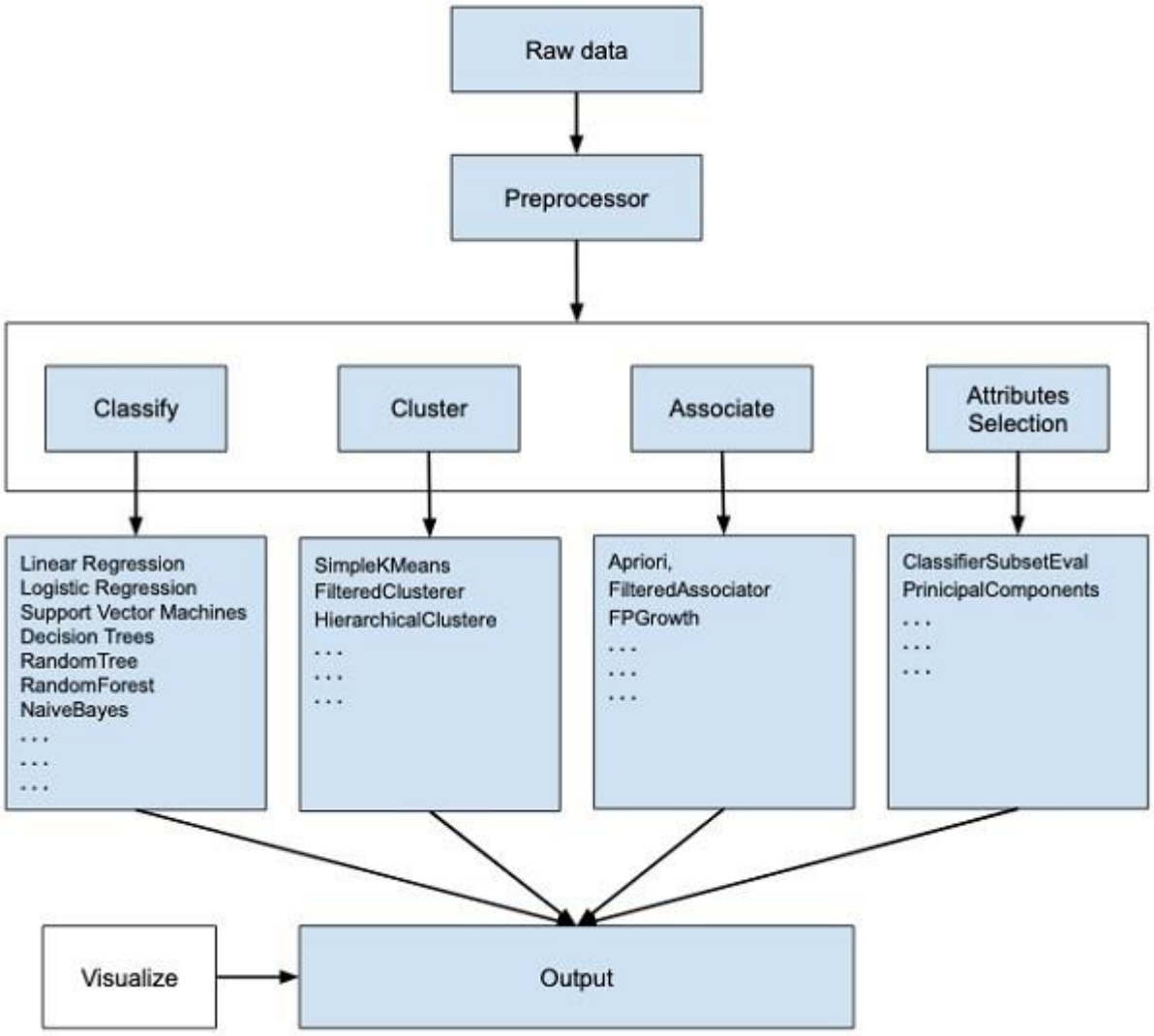
Aim: To perform data Pre-processing and implement Classification, Clustering and Association algorithms on data set.

Requirement: Windows O.S and Weka Tool.

Theory:

Weka Tool:

WEKA - an open source software provides tools for data pre-processing, implementation of several Machine Learning algorithms, and visualization tools so that you can develop machine learning techniques and apply them to real-world data mining problems. What WEKA offers is summarized in the following diagram −



If you observe the beginning of the flow of the image, you will understand that there are many stages in dealing with Big Data to make it suitable for machine learning −

First, you will start with the raw data collected from the field. This data may contain several null values and irrelevant fields. You use the data pre-processing tools provided in WEKA to cleanse the data.

Then, you would save the pre-processed data in your local storage for applying ML algorithms.

Next, depending on the kind of ML model that you are trying to develop you would select one of the options such as **Classify, Cluster**, or **Associate**. The **Attributes Selection** allows the automatic selection of features to create a reduced dataset.

Note that under each category, WEKA provides the implementation of several algorithms. You would select an algorithm of your choice, set the desired parameters and run it on the dataset.

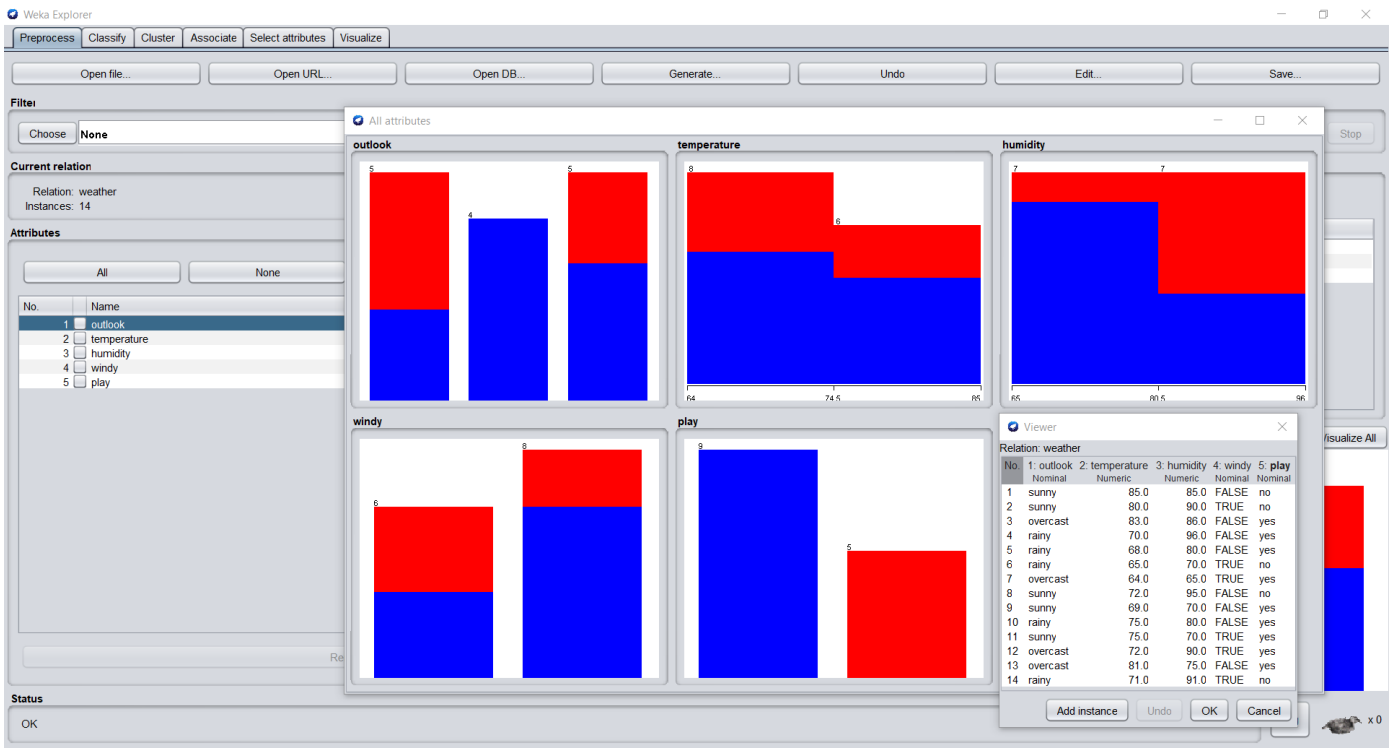
Then, WEKA would give you the statistical output of the model processing. It provides you a visualization tool to inspect the data.

The various models can be applied on the same dataset. You can then compare the outputs of different models and select the best that meets your purpose.

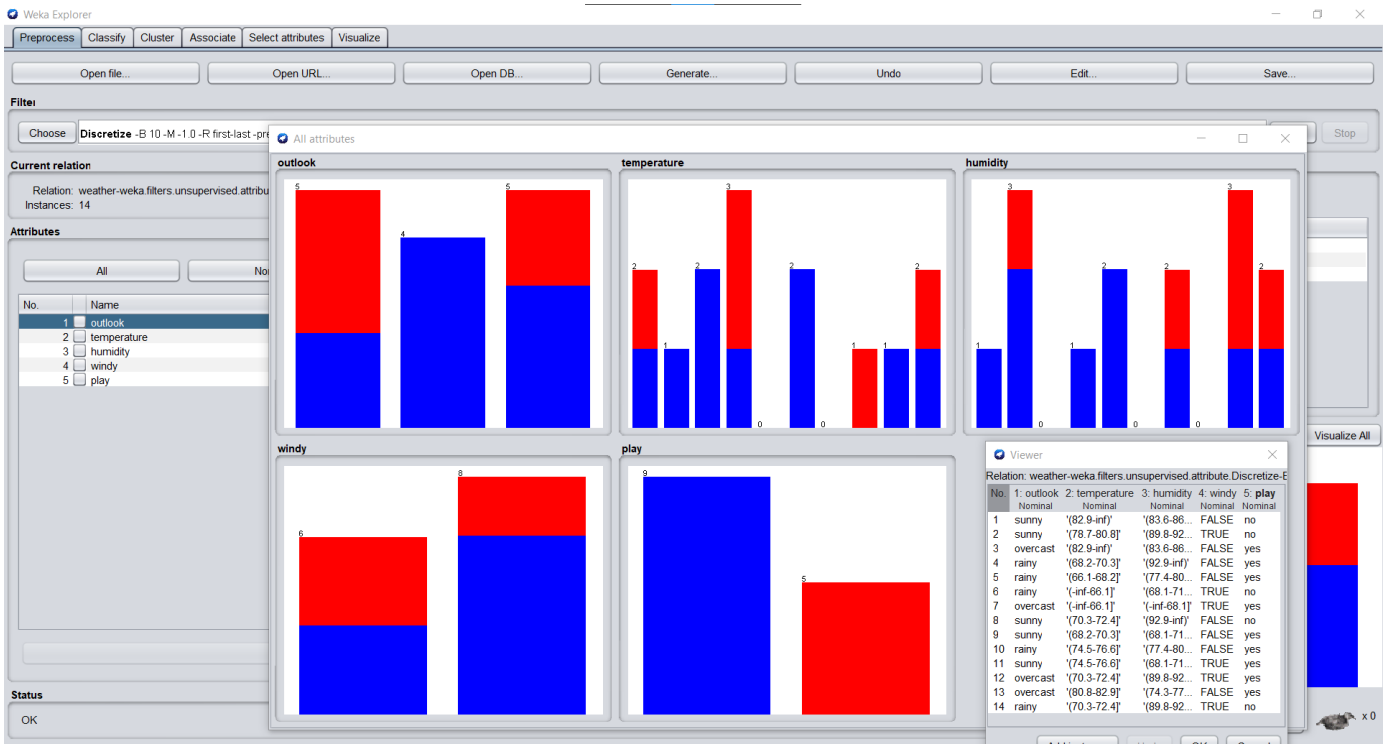
Thus, the use of WEKA results in a quicker development of machine learning models on the whole.

Output:

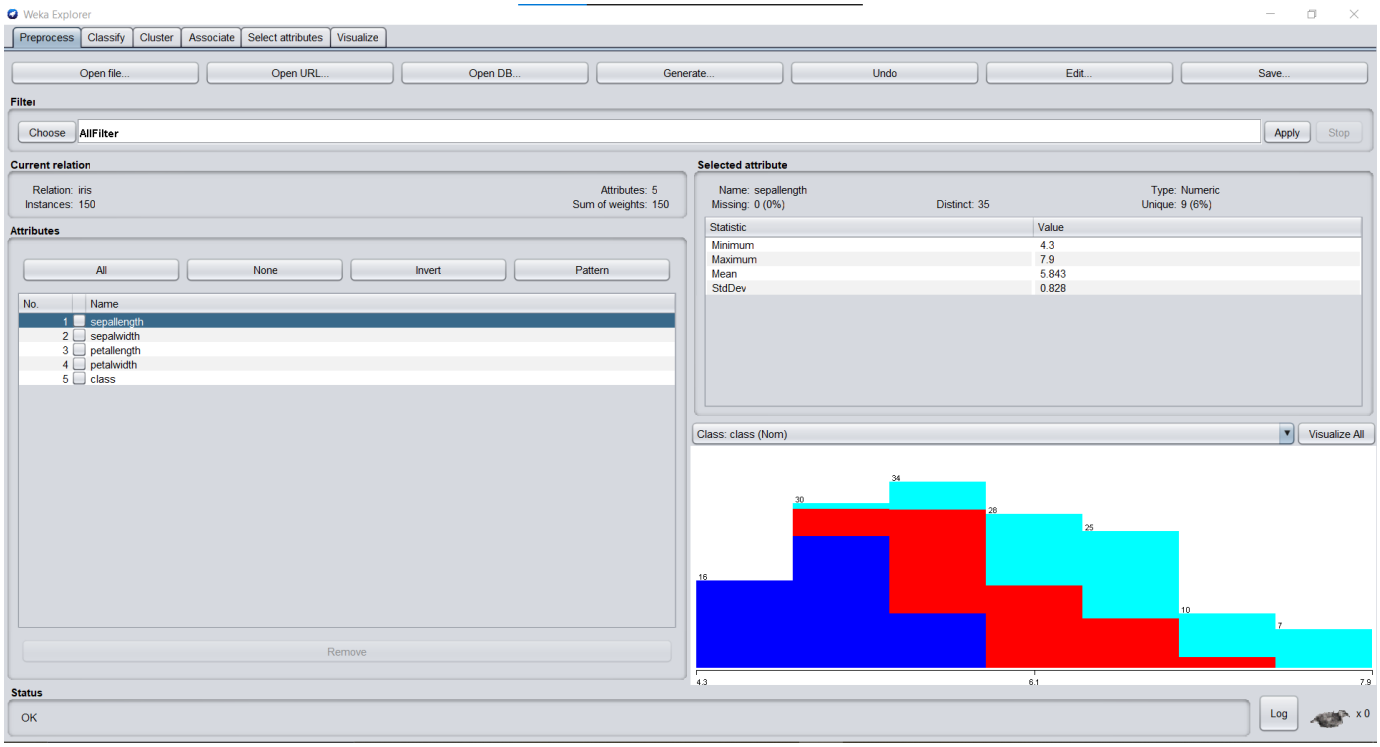
Data Set before Data Pre-processing (Data Discretization):



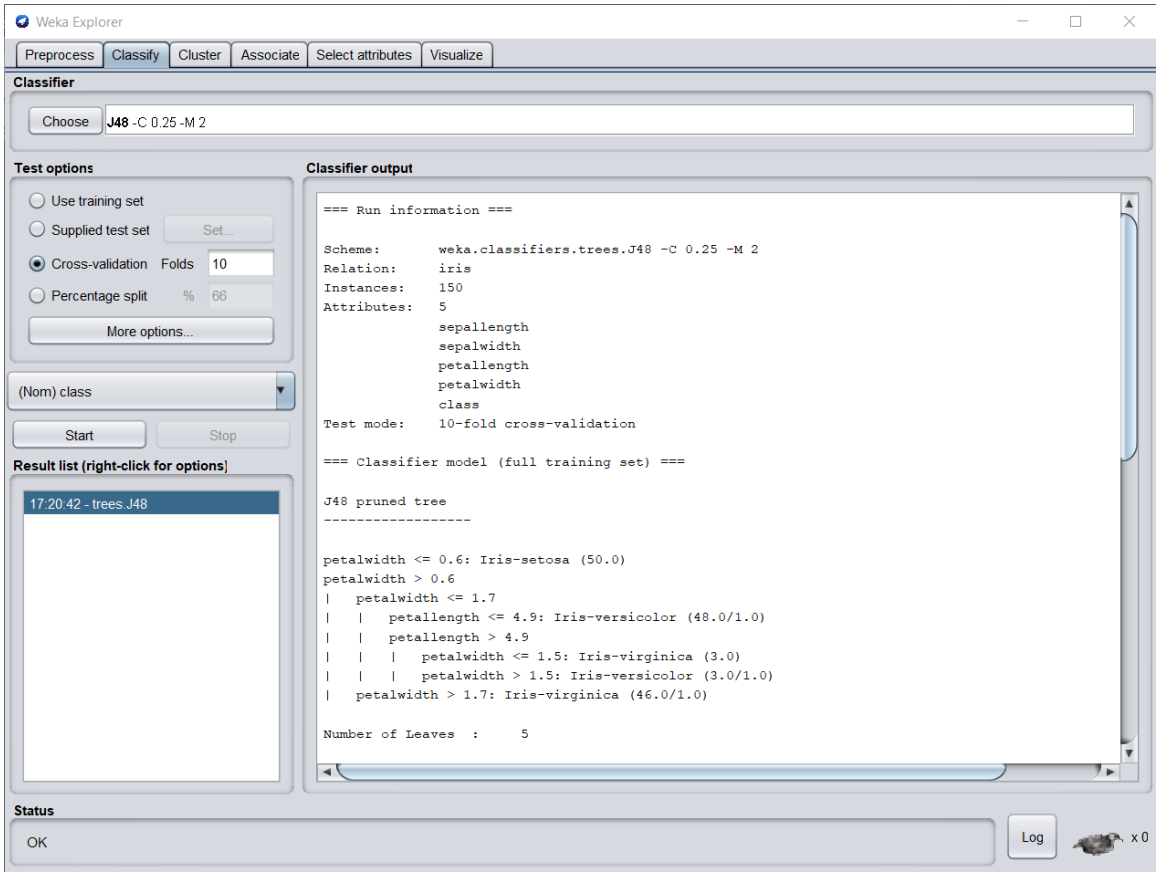
Data Set after Data Pre-processing (Discretization):

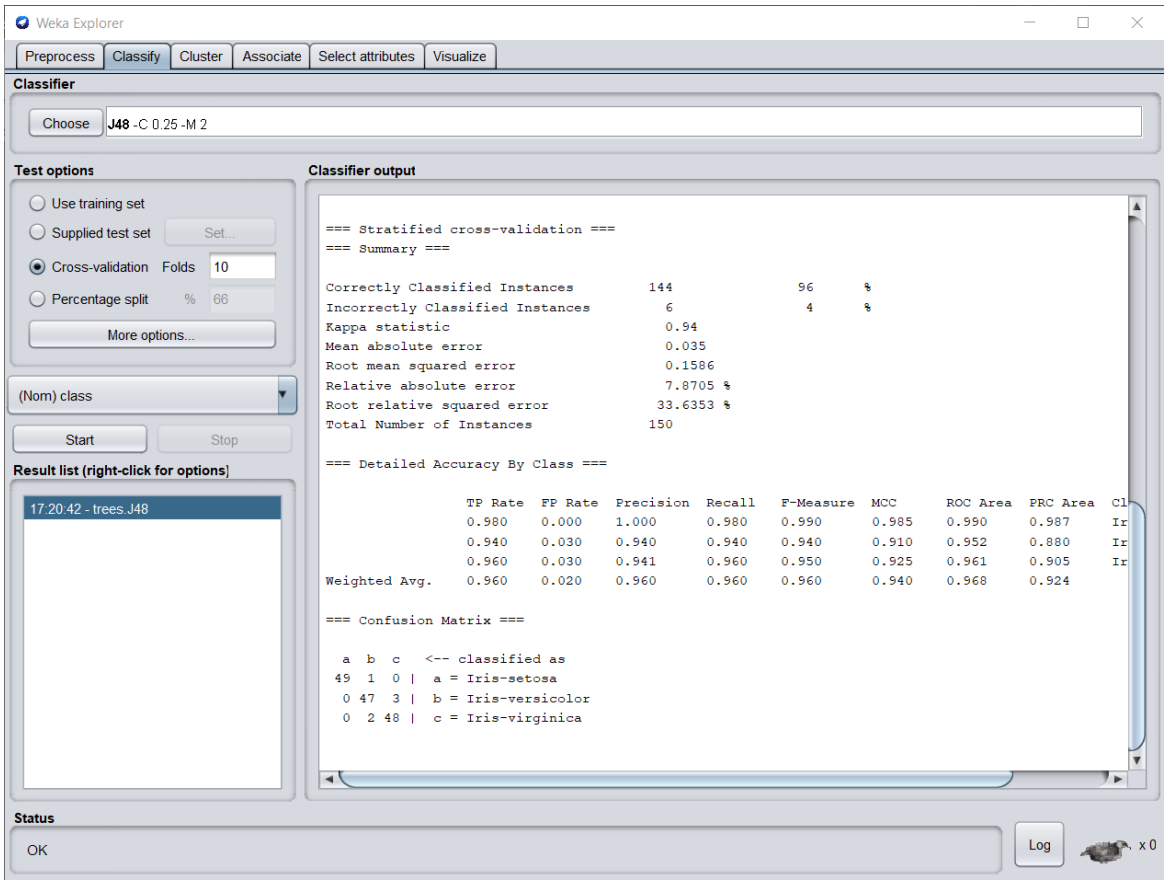


Iris Data Set:

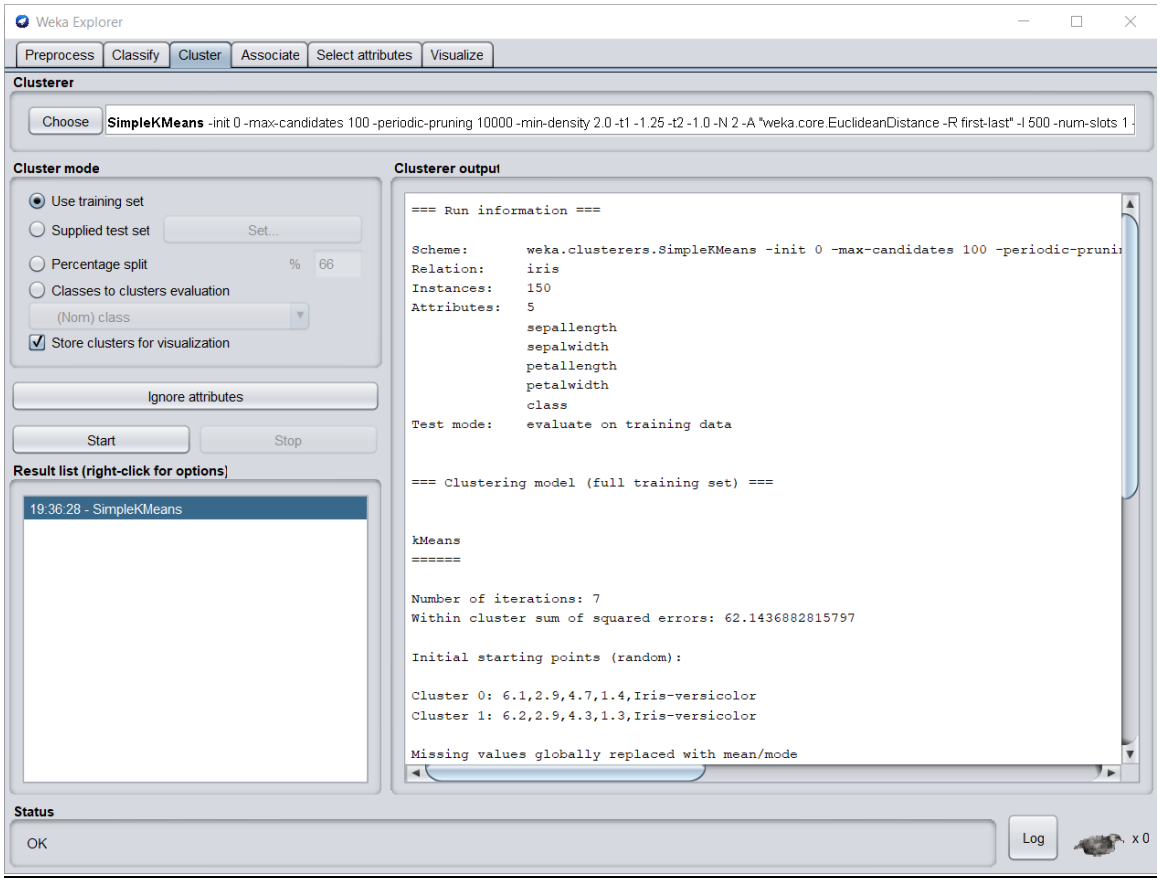


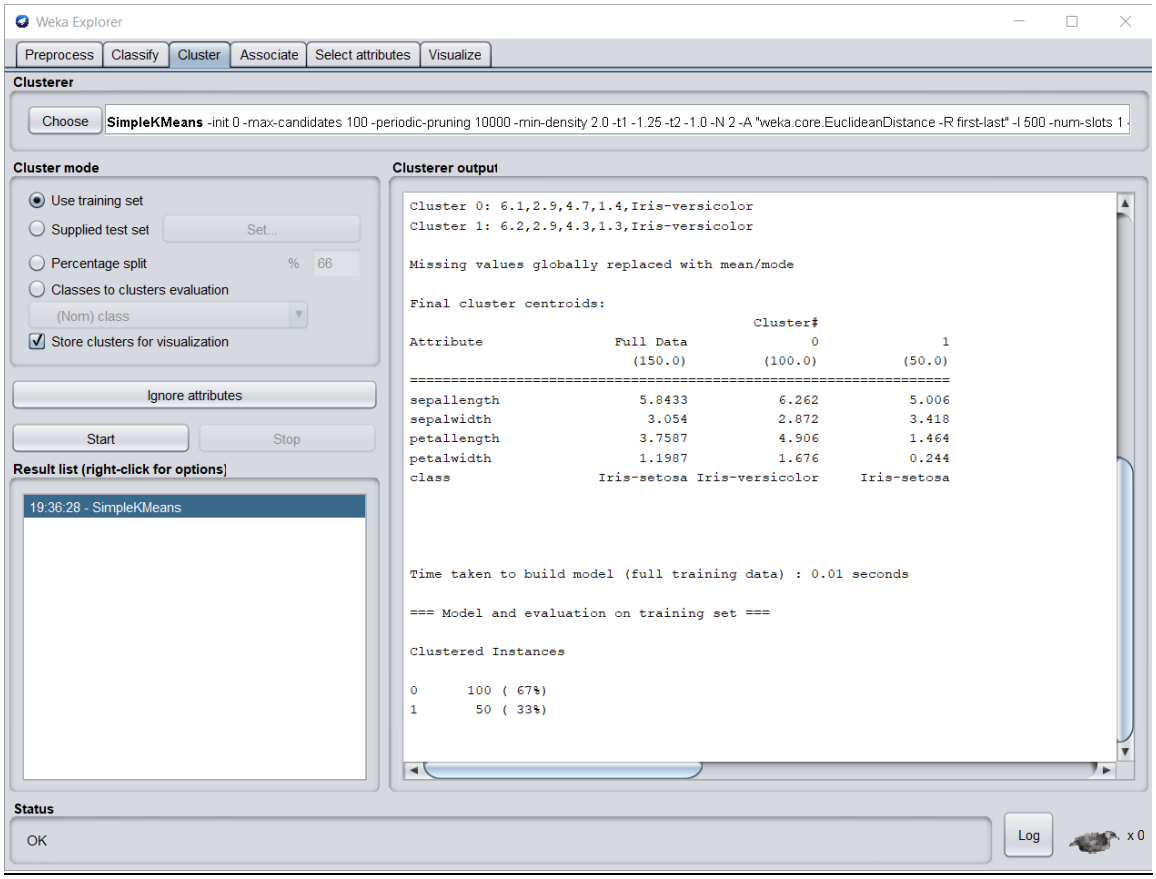
J48 Tree classification on iris data set:

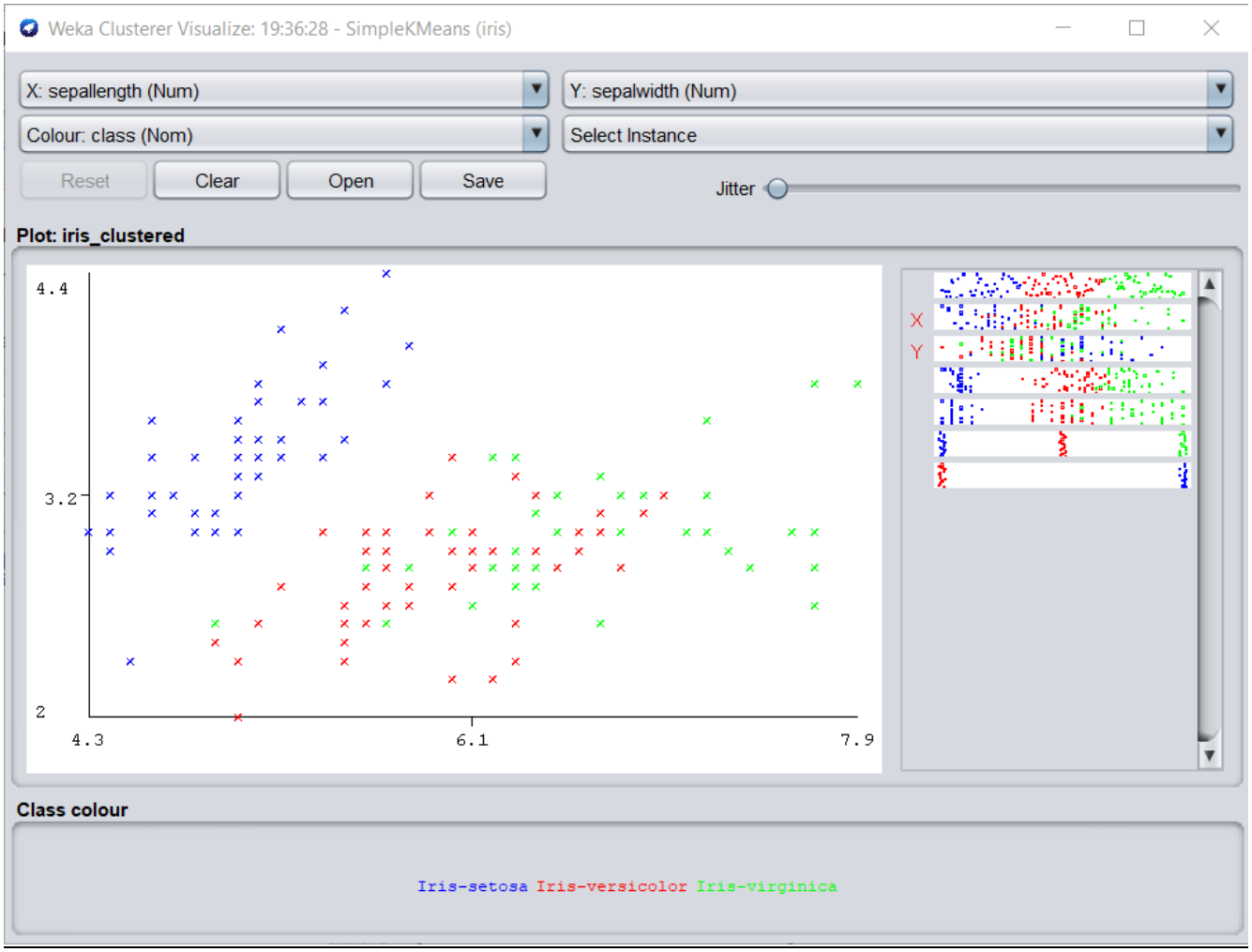




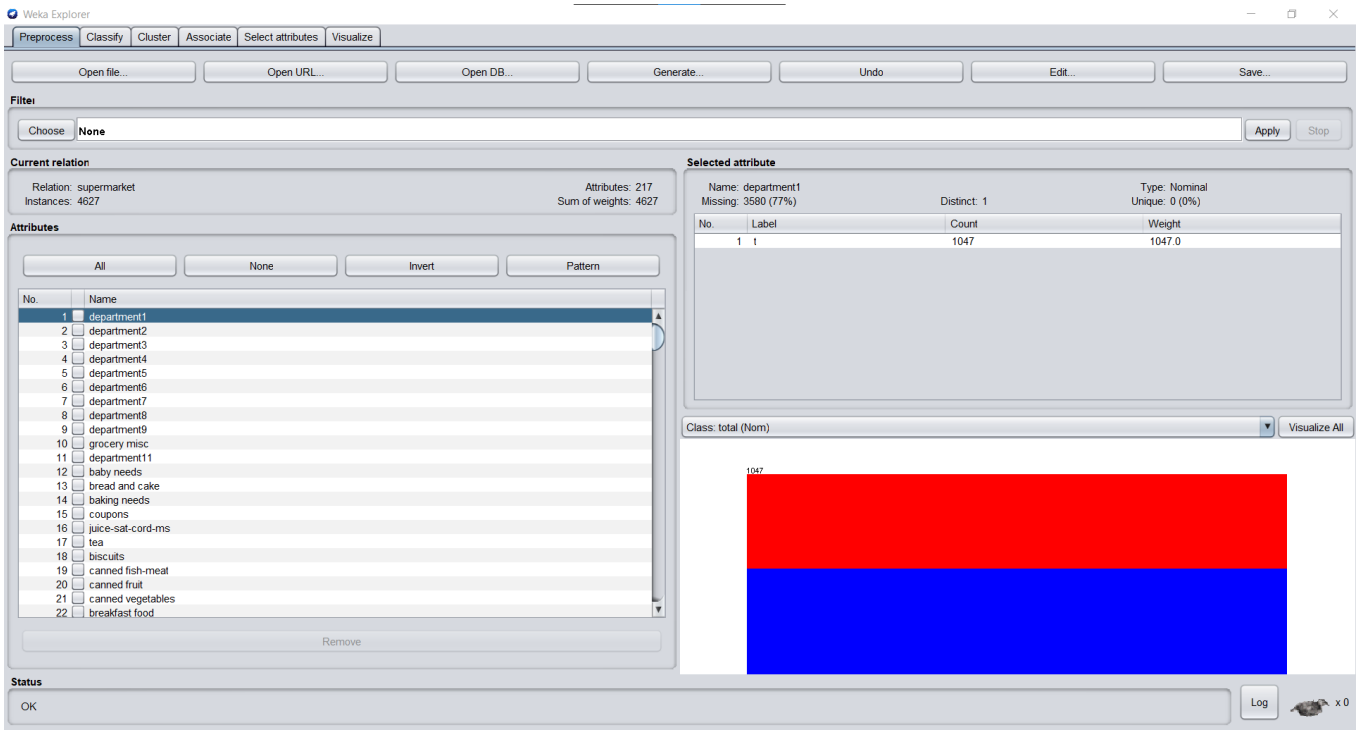
Data Clustering (K-Means/ K-Medoids) on iris Data Set:

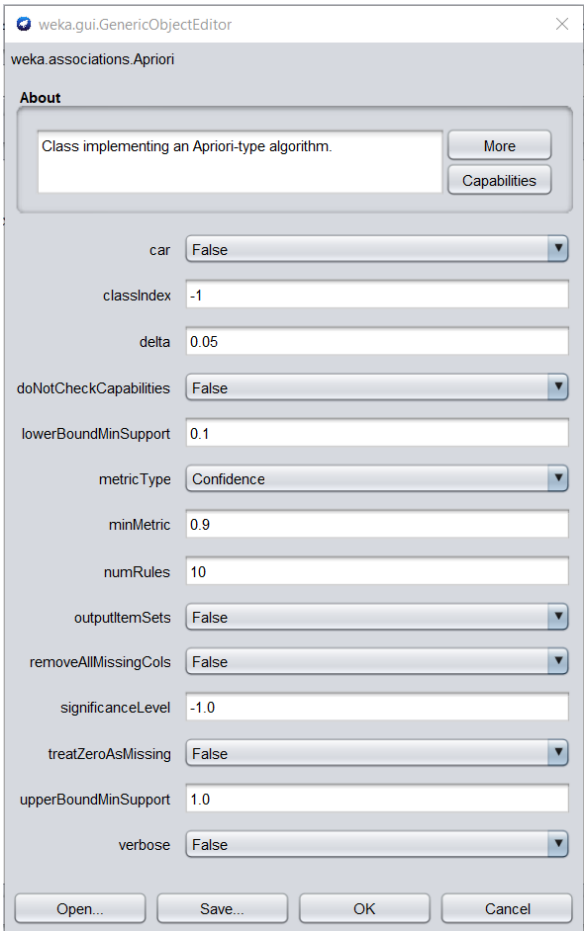


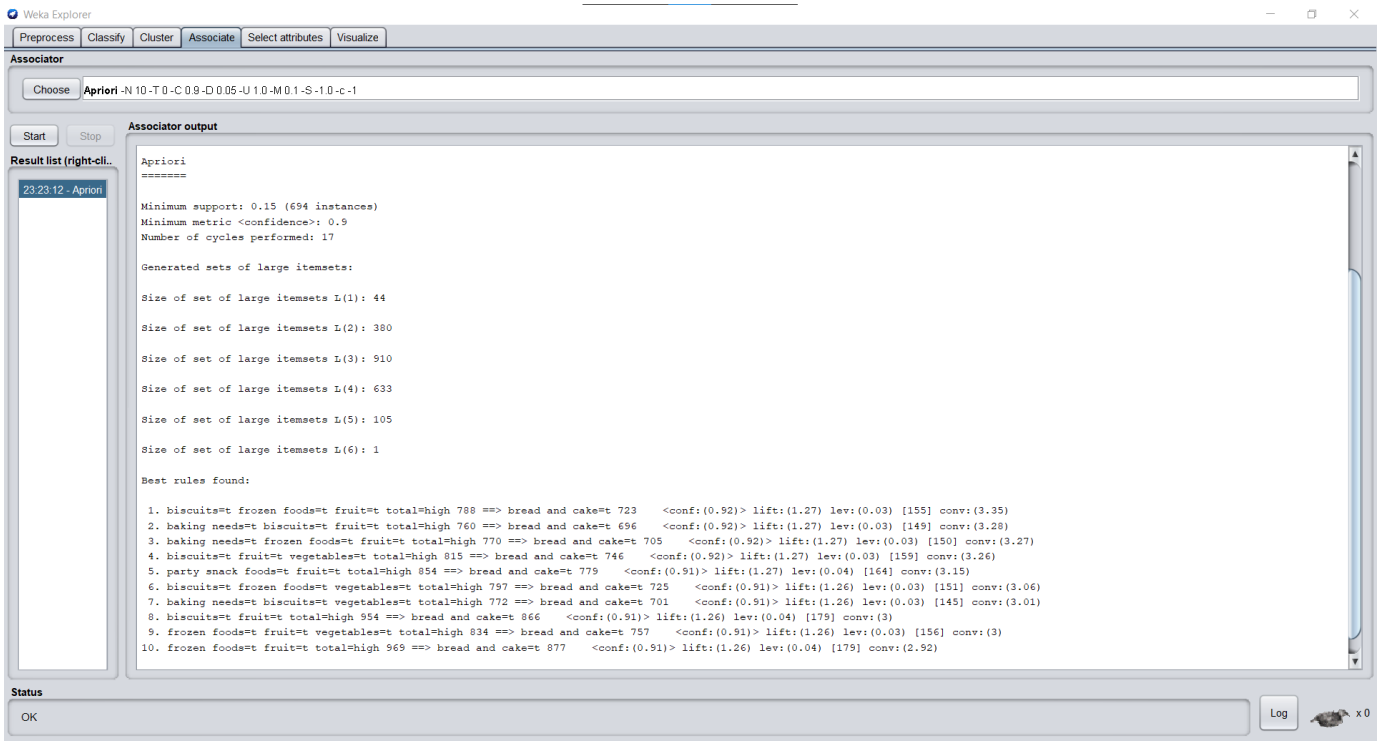




Association algorithm (Apriori) on Supermarket Data Set:







Conclusion: We have successfully Pre-processed data and implemented Classification, Clustering and Association algorithms on data sets using Weka Tool.