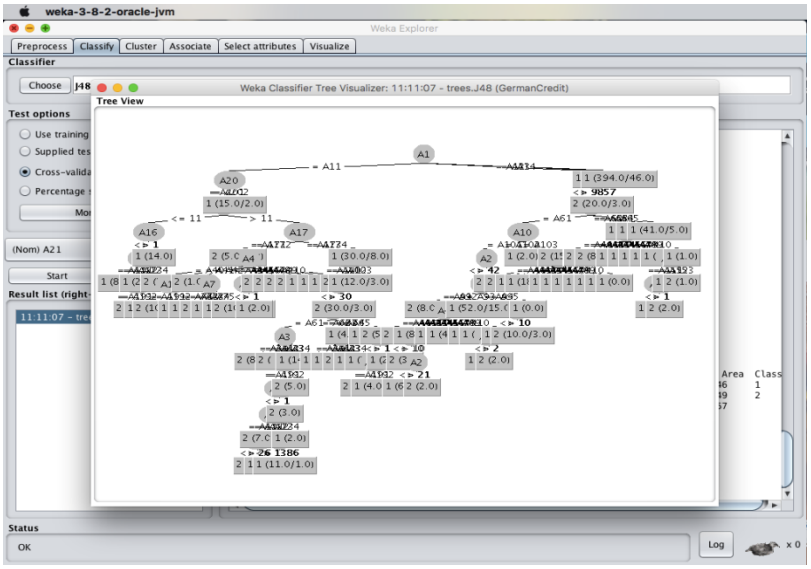
EX.No: 15
Date :

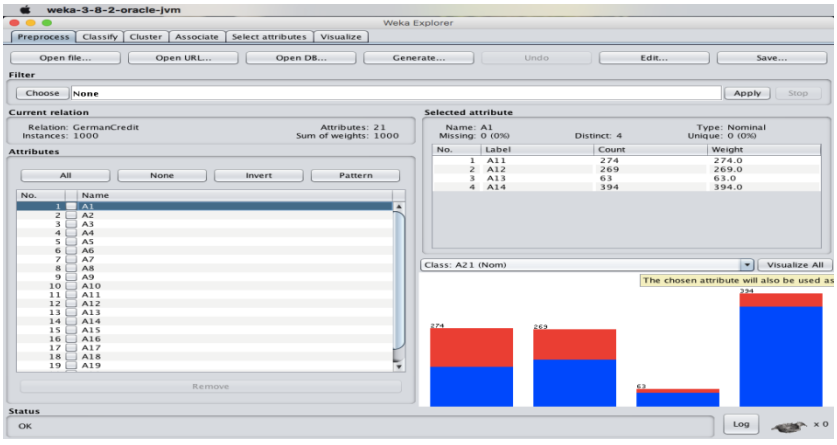
PREDICTION OF CATEGORICAL DATA USING DECISION TREE ALGORTIHM THROUGH WEKA

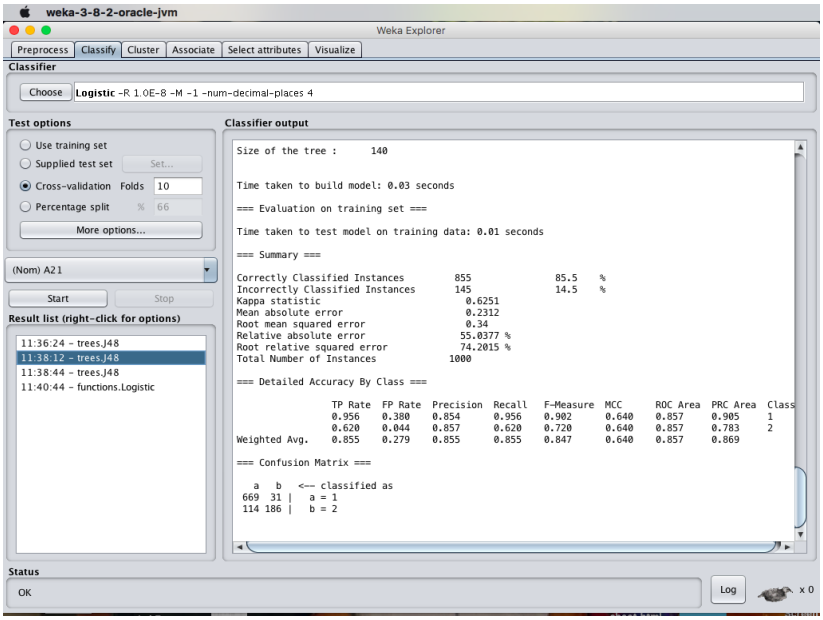
❖ Decision Tree :

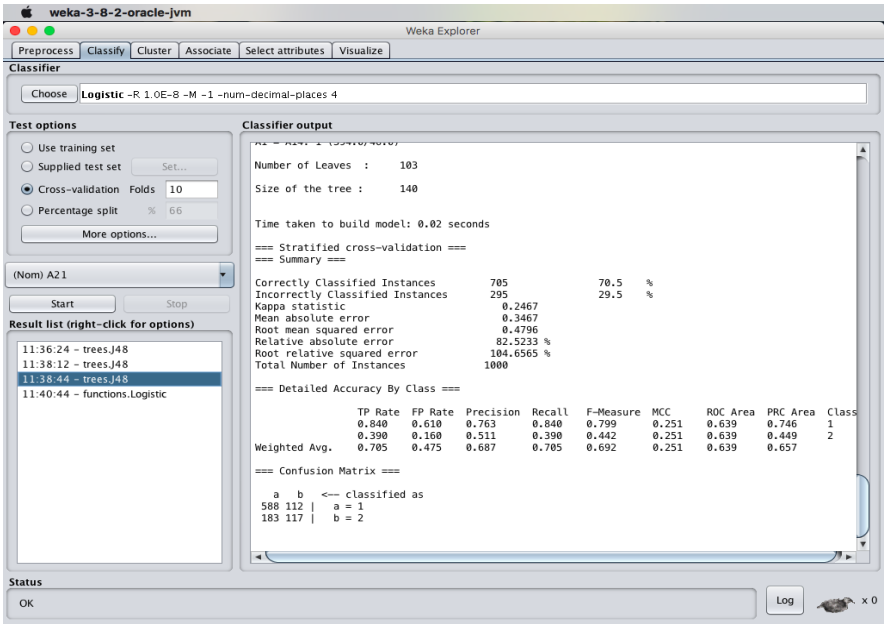
Visualize the decision tree for the given dataset.



or.

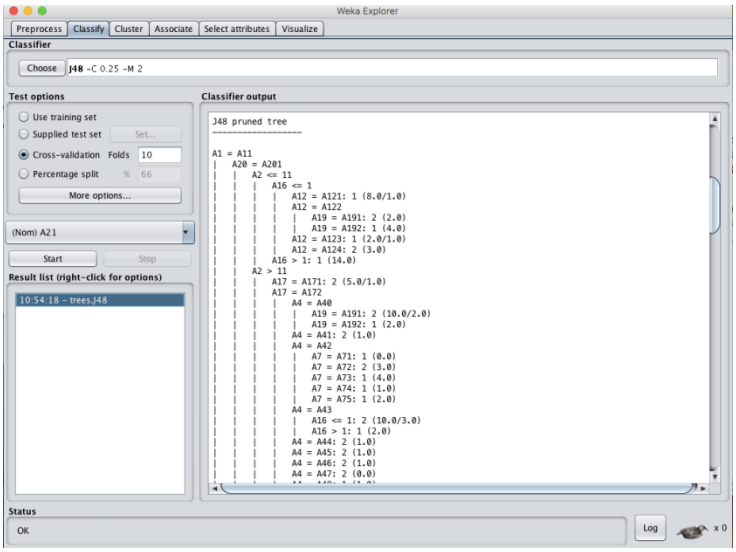


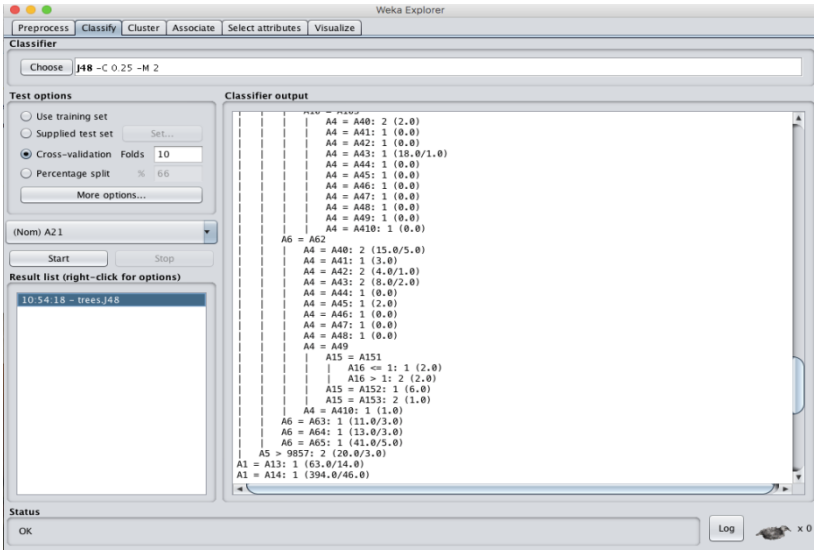


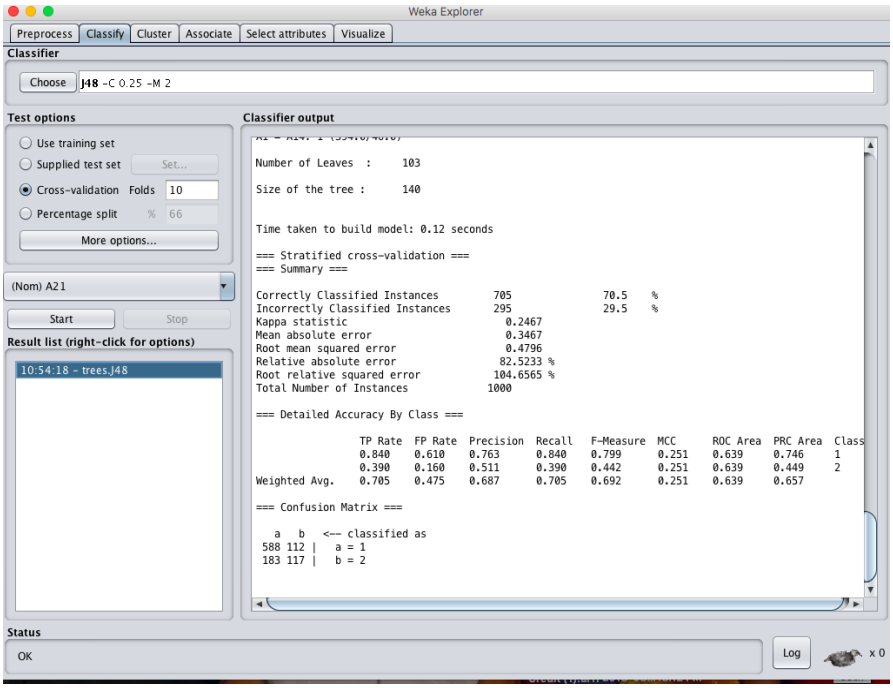


- **CROSS VALIDATION ANALYSIS :**

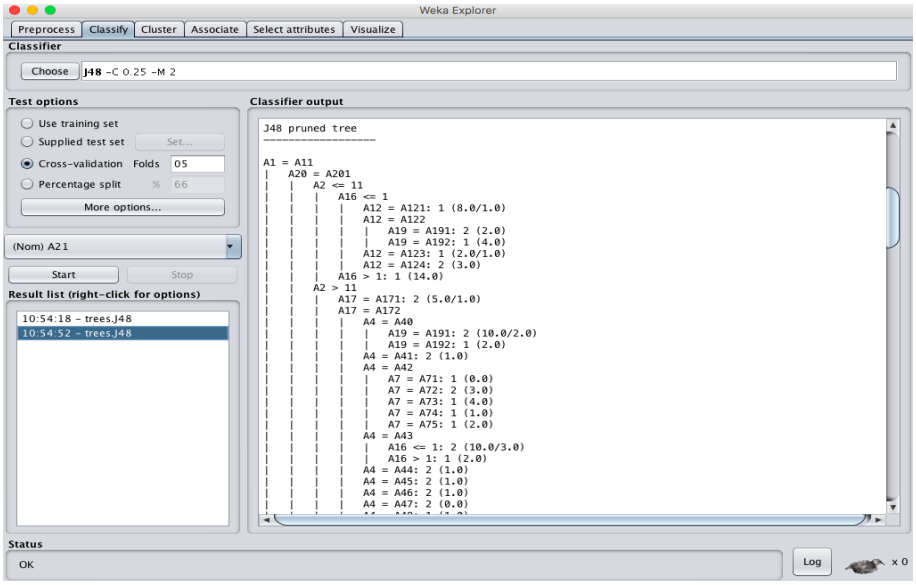
  - **When cross validation folds are 10 :**

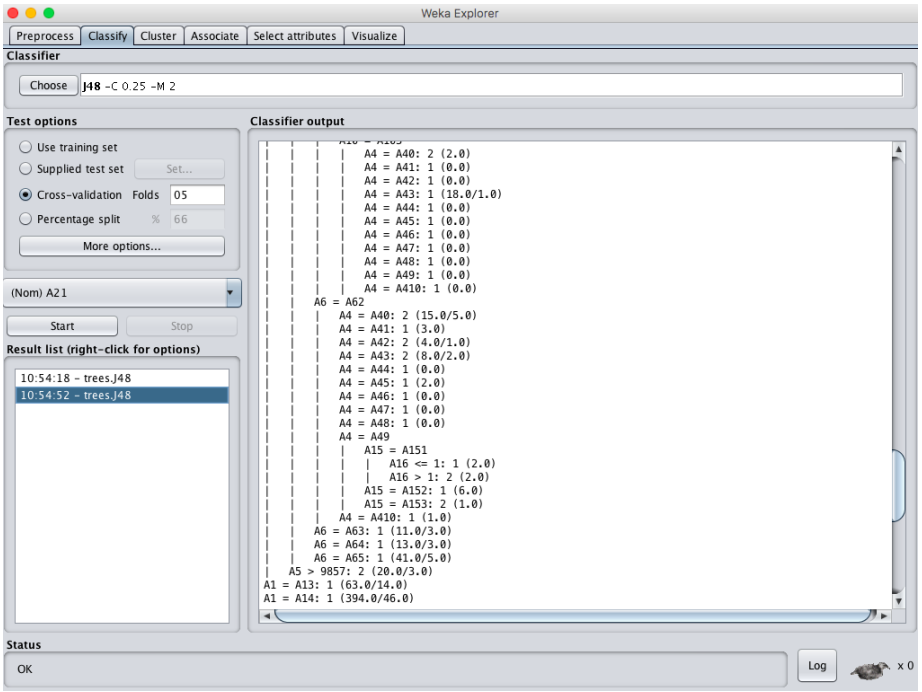


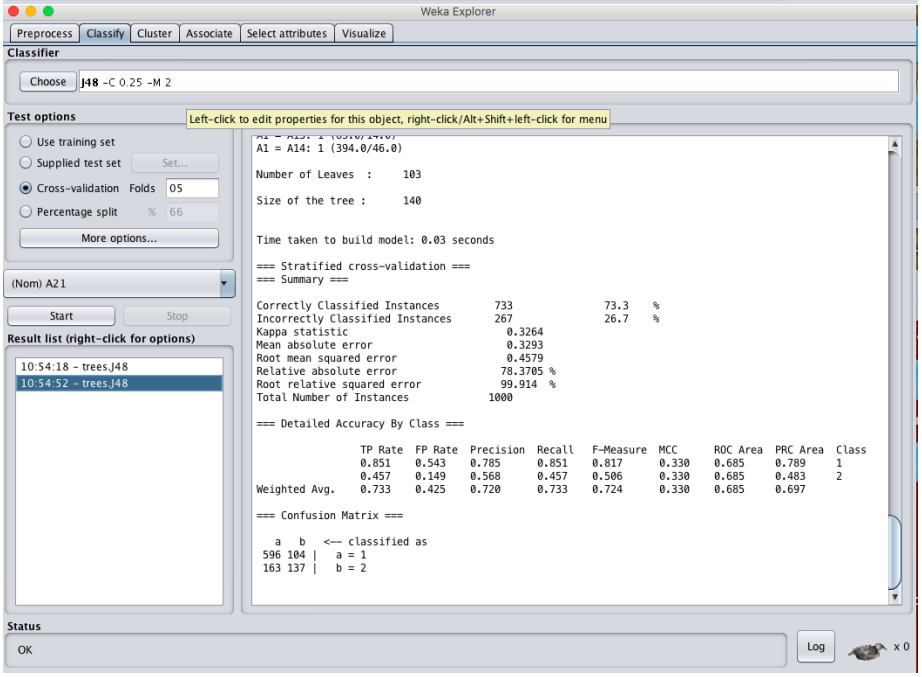




- **When cross validation folds are : 05 :-**







RESULT :

Thus, the observations and evaluations done on the german_credit dataset are analyzed. The decision tree has been successfully visualized. Various evaluations and comparisons done through the cross validation folds change. Which lead to the change of values in confusion matrix.

EX.No: 16
Date :

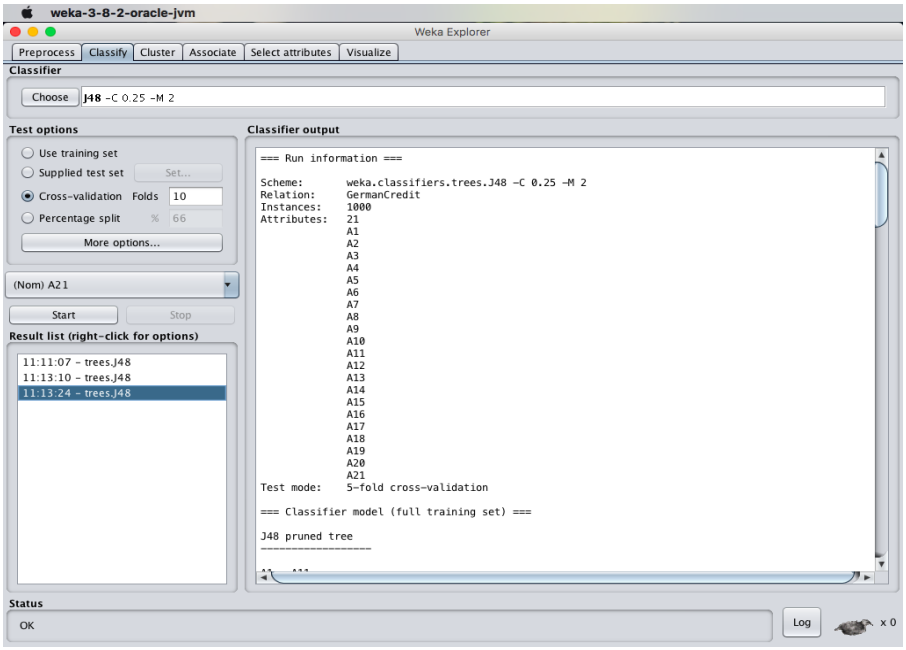
PREDICTION OF CATEGORICAL DATA USING SMO ALGORTIHM THROUGH WEKA

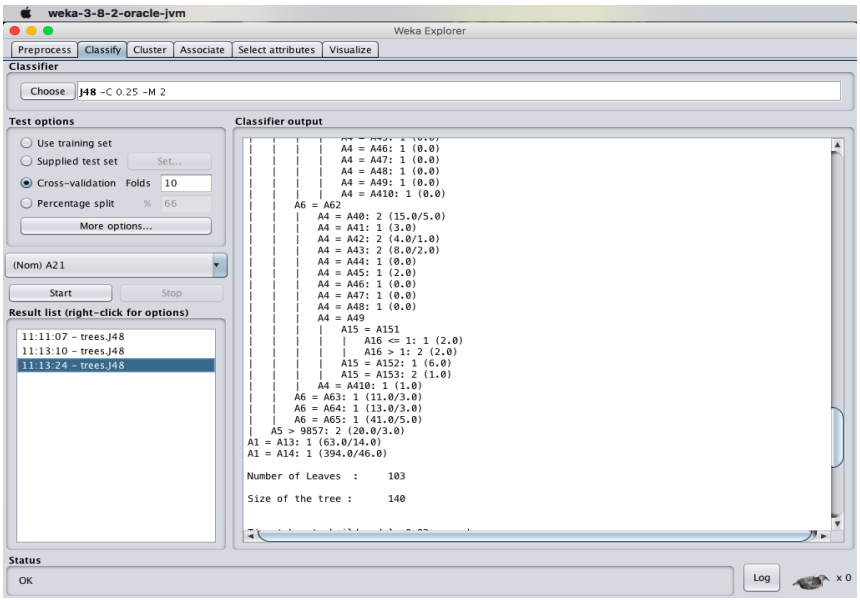
DESCRIPTION :

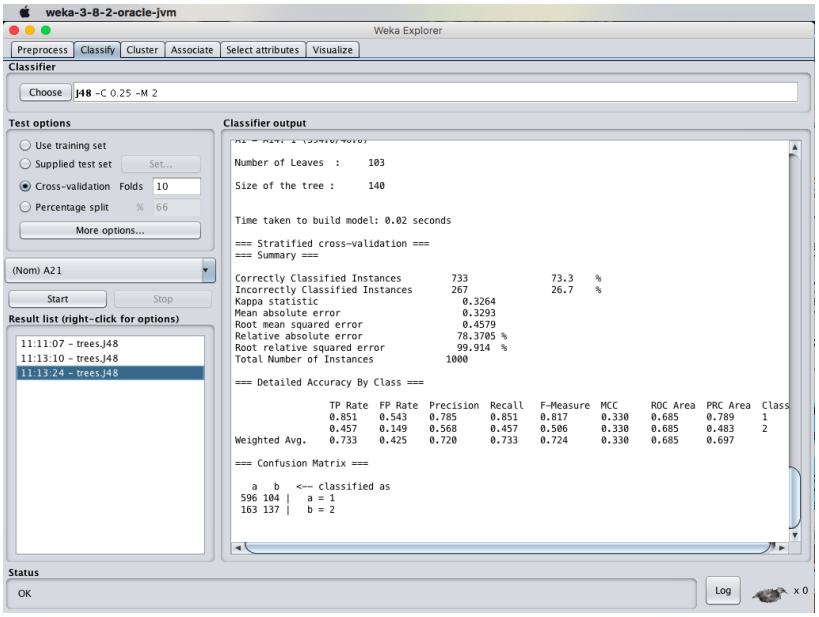
Consider the german credit dataset which can be downloaded from the UCI repository.

❖ DECISION TREE :

A tree has many analogies in real life, and turns out that it has influenced a wide area of machine learning, covering both classification and regression. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. As the name goes, it uses a tree-like model of decisions. Though a commonly used tool in data mining for deriving a strategy to reach a particular goal, its also widely used in machine learning, which will be the main focus of this article.

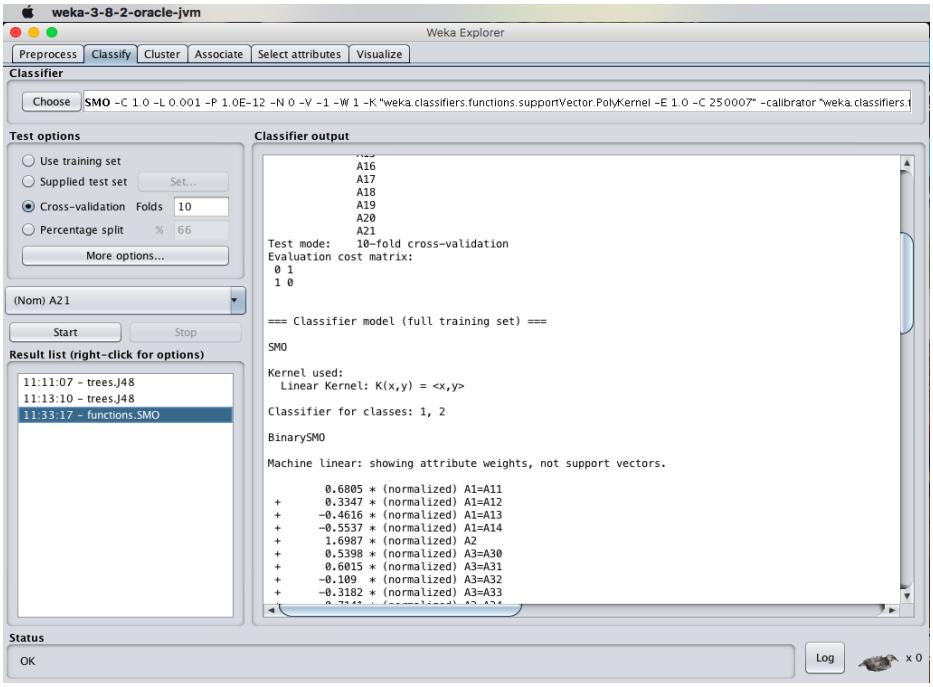


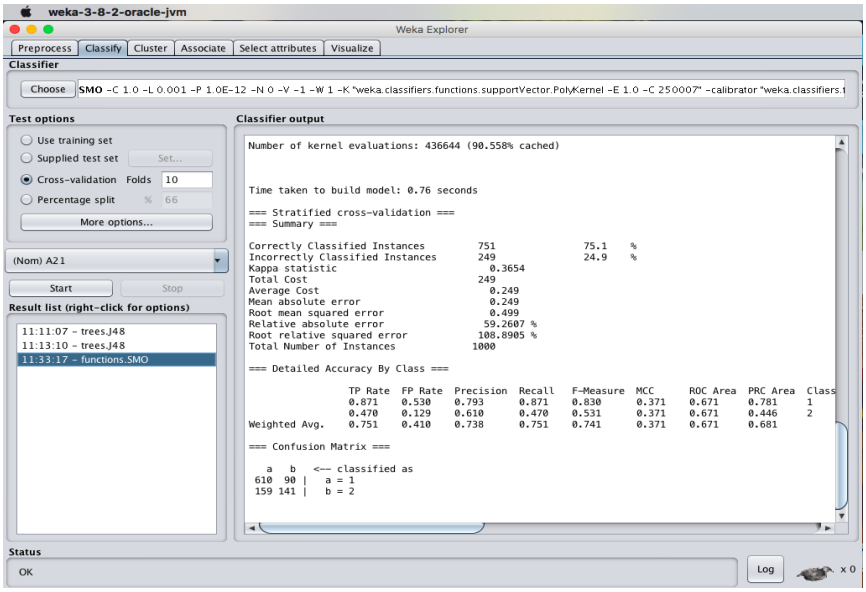




❖ SMO ALGORITHM:

The iterative algorithm Sequential Minimal Optimization (SMO) is used for solving quadratic programming (QP) problems. One example where QP problems are relevant is during the training process of support vector machines (SVM). The SMO algorithm is used to solve in this example a constraint optimization problem. John Platt proposed this algorithm in 1998 and it was successfully used since then. We describe here the basics of the algorithm in the light of big data.





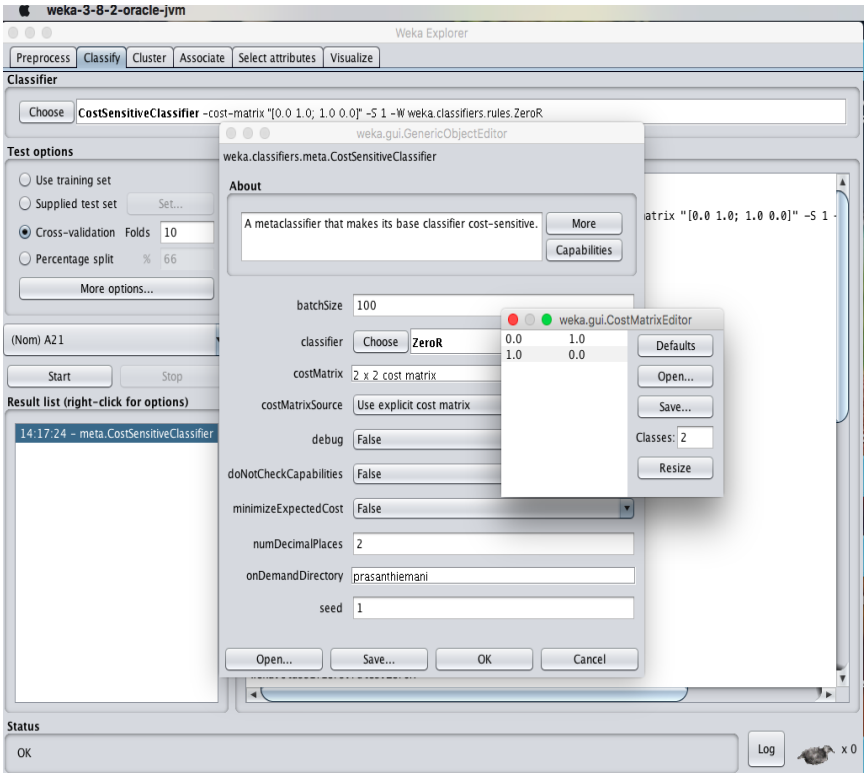
1. Set the cost sensitive evaluation and compare the obtained results.

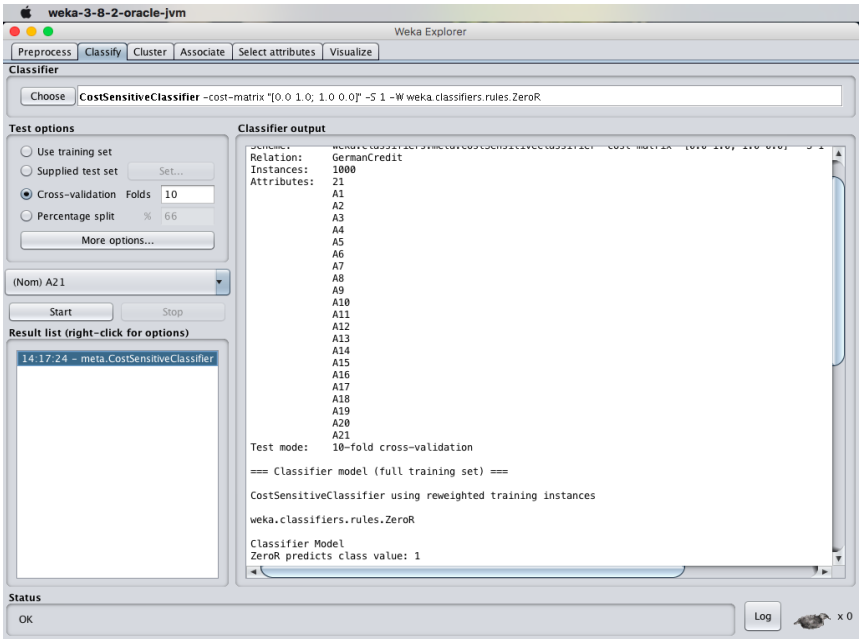
Cost-Sensitive Learning is a type of learning in data mining that takes the misclassification costs (and possibly other types of cost) into consideration. The goal of this type of learning is to minimize the total cost. The key difference between cost-sensitive learning and cost-insensitive learning is that cost-sensitive learning treats the different misclassifications differently. Costinsensitive learning does not take the misclassification costs into consideration. The goal of this type of learning is to pursue a high accuracy of classifying examples into a set of known classes.

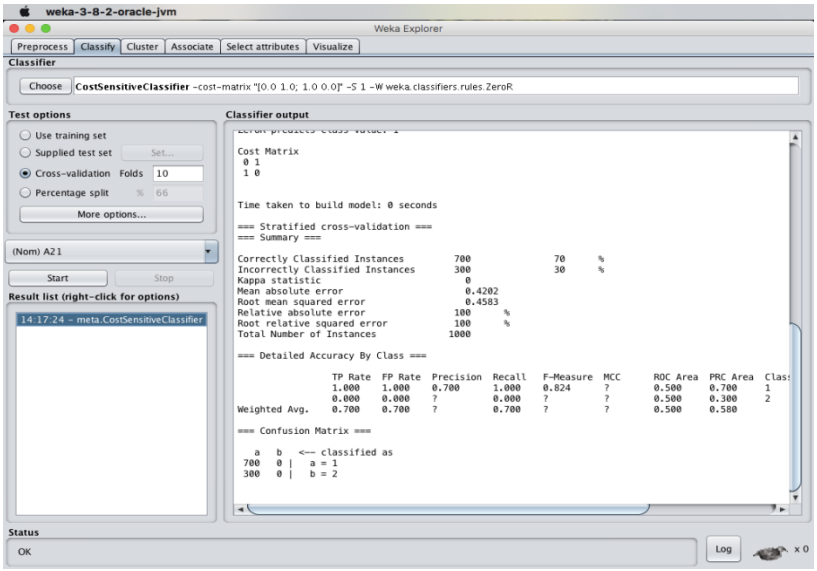
STEPS :

- Classifythe dataset with the cost sensitive classifier technique.
- Change the cost matrix to 2*2 matrix and execute.

ANALYSIS :







2. What is the significance of the following parameters :

a) Mean Absolute Error :

Mean Absolute Error (MAE) is similar to the Mean Squared Error, but it uses absolute values instead of squaring. This measure is not as popular as MSE, though its meaning is more intuitive (the "average error").

b) Total Number of Instances :

The data present consists of various instances of the class. In the case of german_credit dataset, the total number of instances present in the german credit dataset are 1000 instances.

RESULT :

Thus, the observations and evaluations done on the german_credit dataset are analyzed. The comparison between decision tree and Sequential Minimal Optimization (SMO) has been successfully visualized. In addition to that cost sensitive classifier is been used to analyze few things.