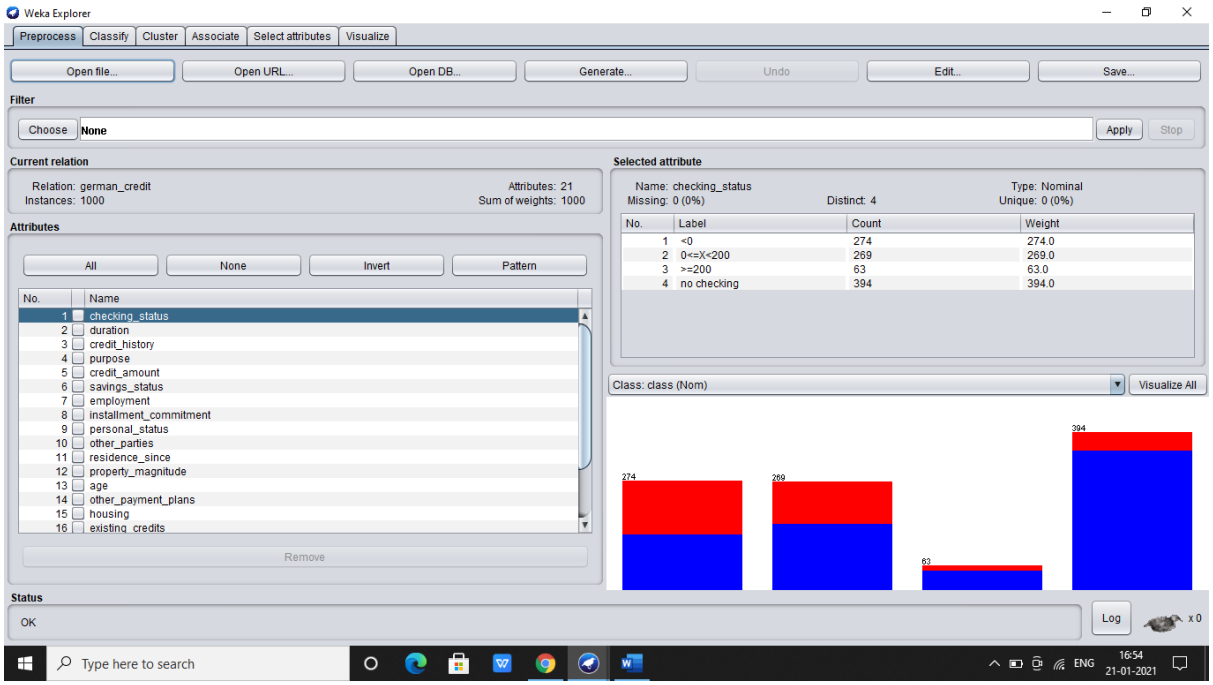
Performing Classification in German Credit Data set

Classification: It is one of the data mining tasks which is used to find a model that distinguishes data classes . There are several techniques to perform classification. Some of them are decision trees, K-Nearest neighbour, neural networks etc.,

Steps to perform classification in weka using decision trees:

Decision Tree algorithm is one of the supervised learning algorithms. It can be used for solving regression and classification problems too. The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules derived from training data.

- Load the german credit dataset into weka platform.



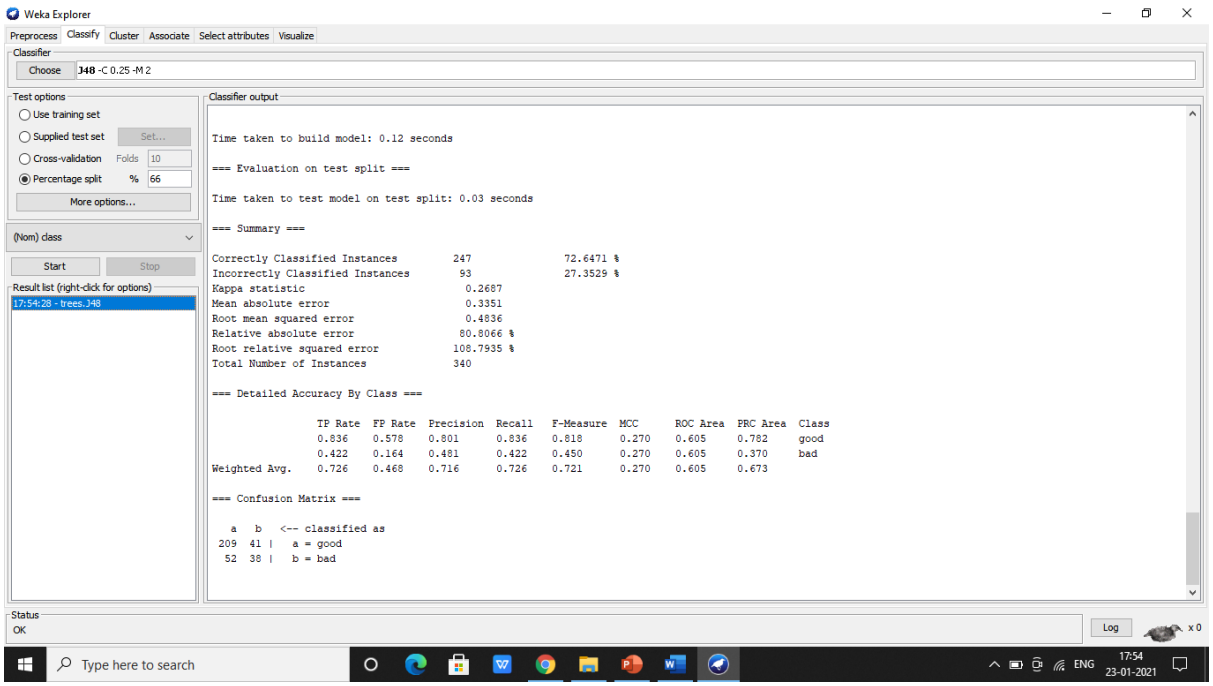
- To classify the dataset opt the classify option and choose the required algorithm. In present case we choose J48 to use decision tree algorithm. Now perform the classification by choose cross validation and percentage split.

Observations By changing percentage splits and cross validation folds:

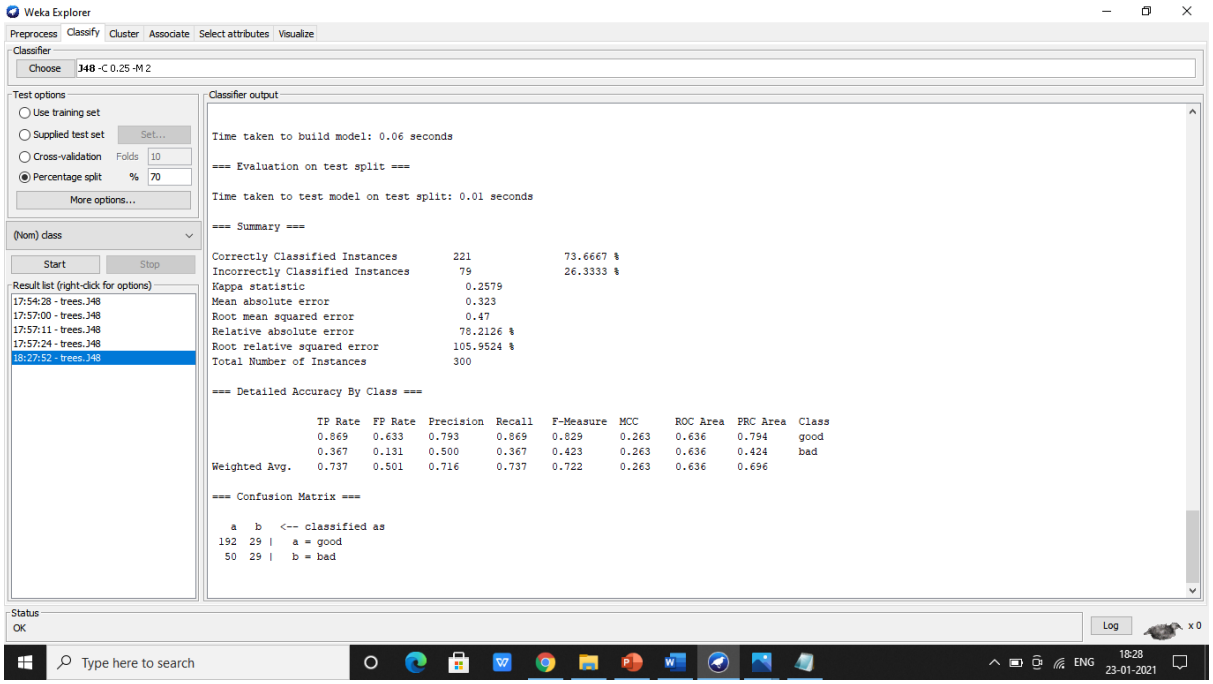
Percentage split: It is a percentage which describe the percent of data used for training and testing purpose.

Cross Validation: Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample. The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into. The procedure is often called k-fold cross-validation. When a specific value for k is chosen, it may be used in place of k in the reference to the model, such as k=10 becoming 10-fold cross-validation.

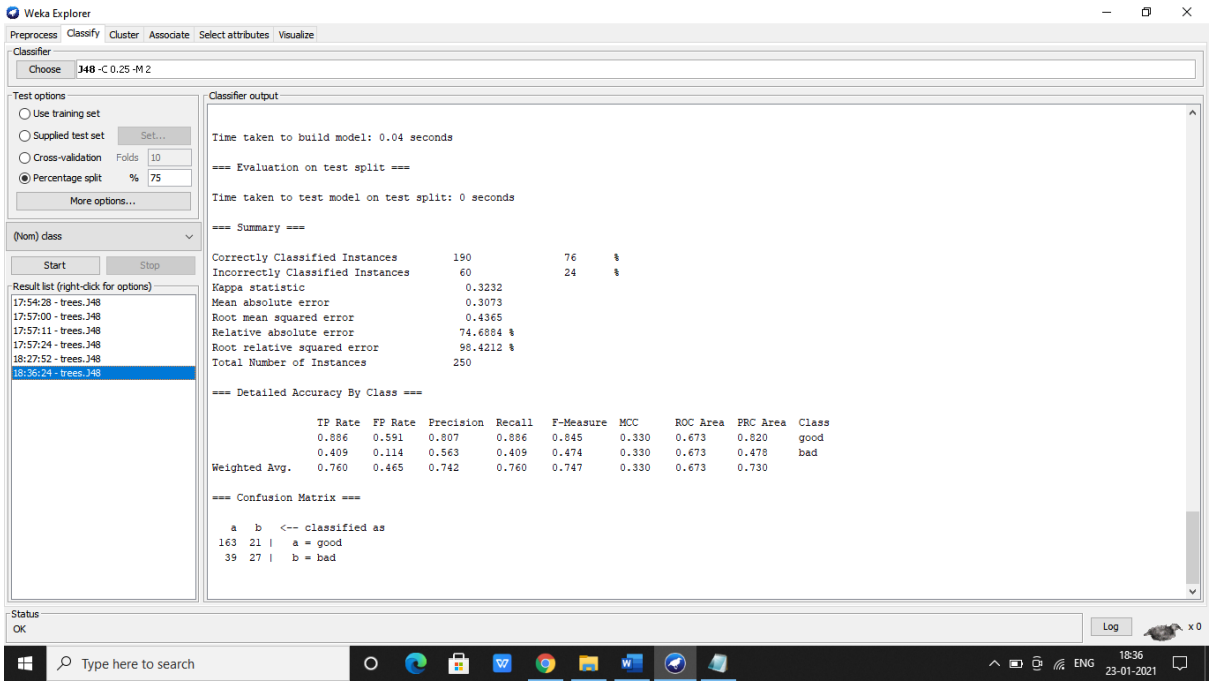
- By classifying with default percentage split i.e 66% and with cross validation folds 10 the accuracy observed is 72.64% .Time taken to build model is 0.12 sec.



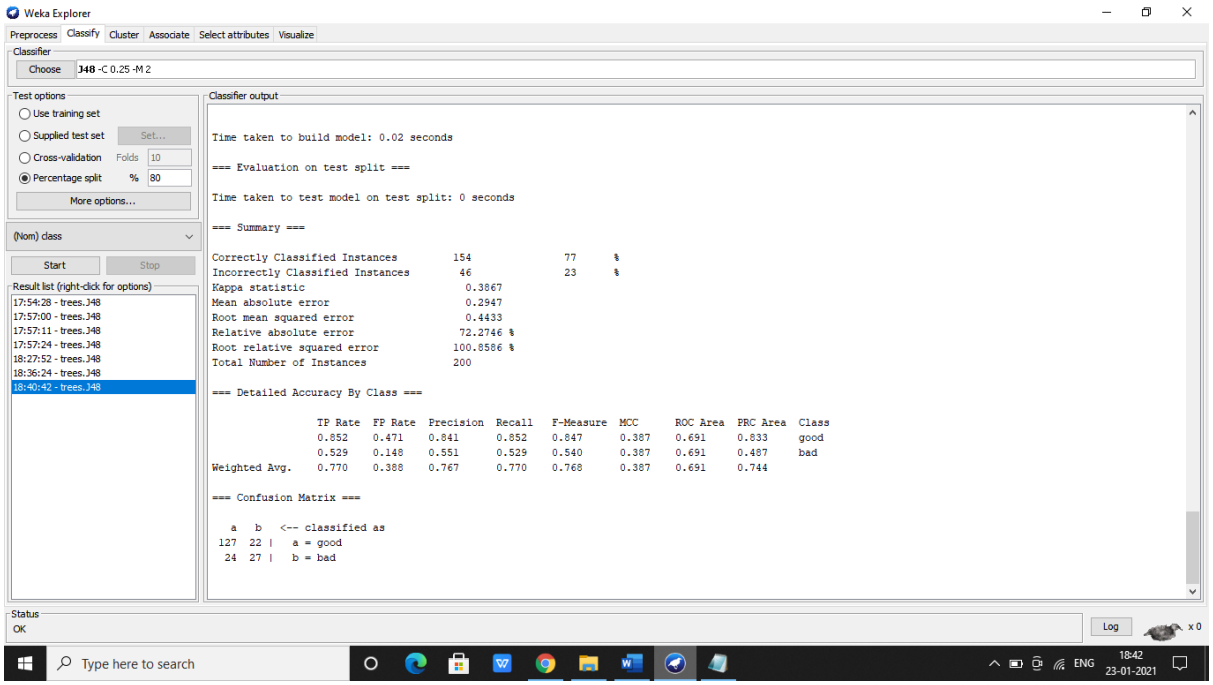
- When the percentage split is 70% and with cross validation folds 10 the accuracy observed is 73.67% .Time taken to build model is 0.06 sec.



- When the percentage split is 75% and with cross validation folds 10 the accuracy observed is 76% .Time taken to build model is 0.04 sec.

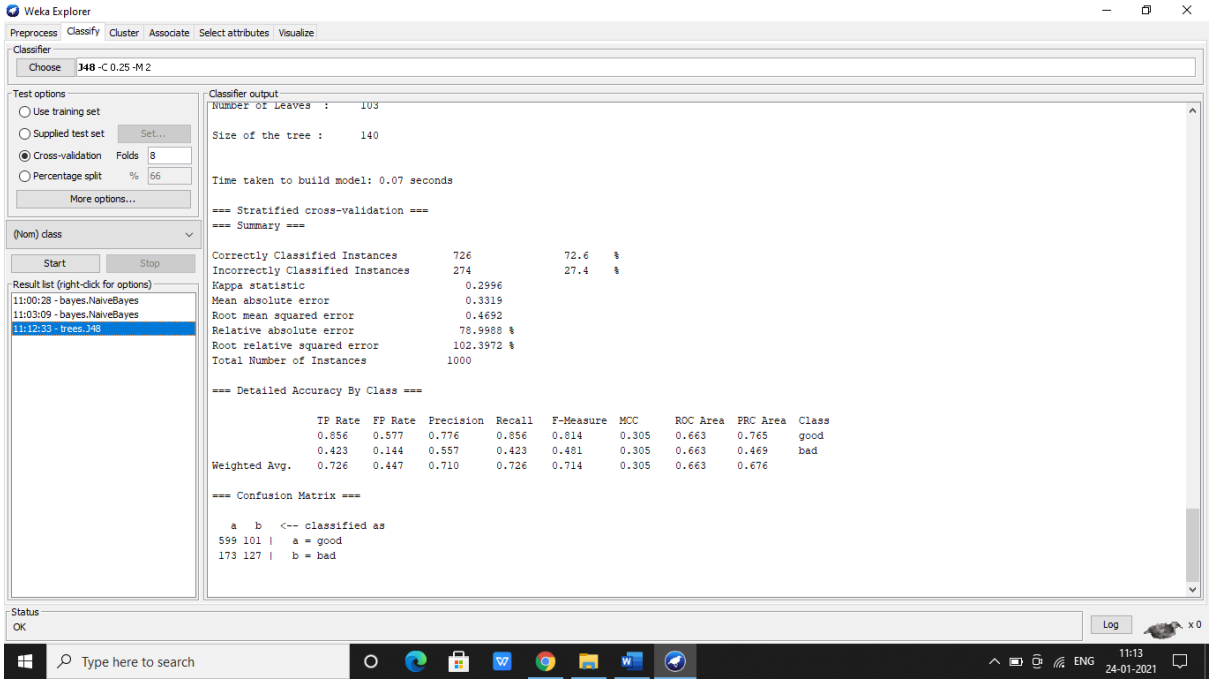


- When the percentage split is 80% and with cross validation folds 10 the accuracy observed is 77% .Time taken to build model is 0.02 sec.

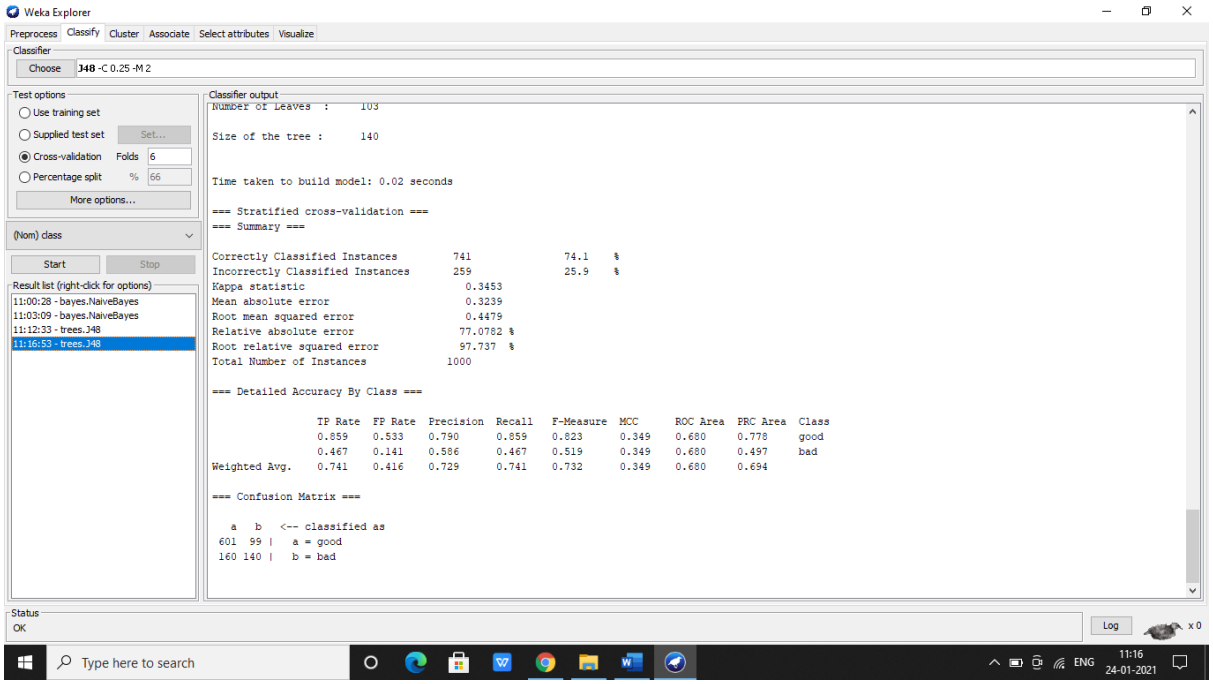


- Number of leaves : 103
- Size of the tree: 140

- When the percentage split is 66% and with cross validation folds 8 the accuracy observed is 72.6% .

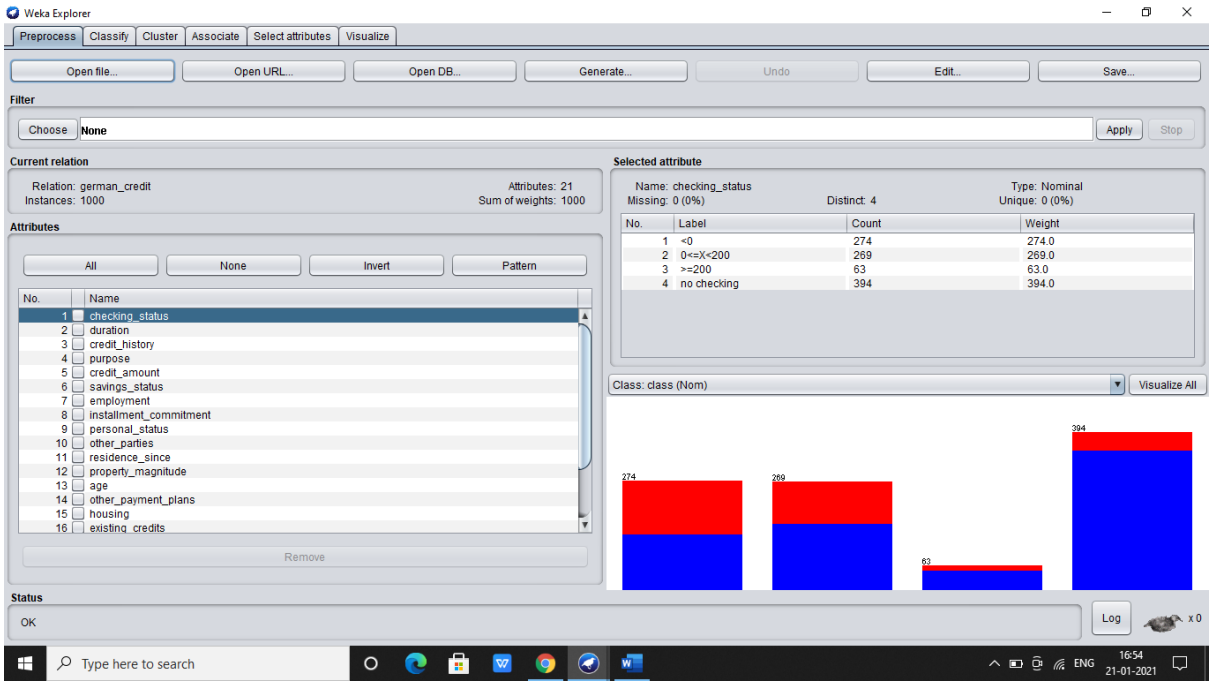


- When the percentage split is 66% and with cross validation folds 6 the accuracy observed is 74.1% .



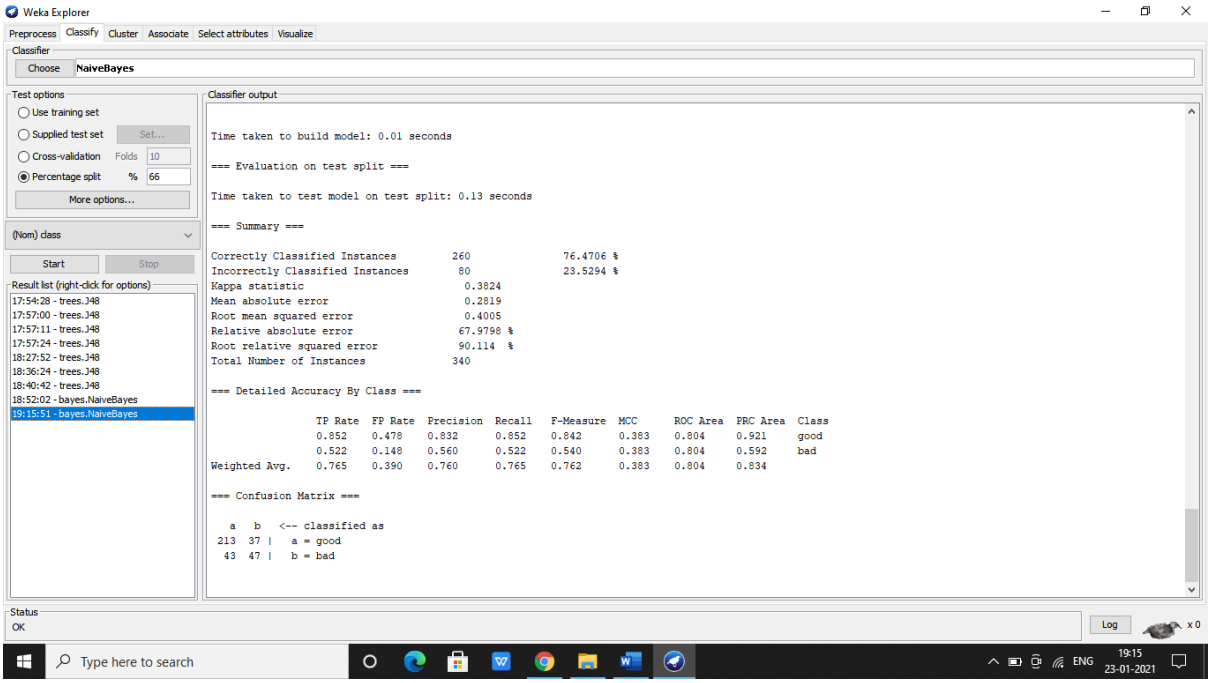
Classification using Naivebayes algorithm:

- Load the german credit dataset into weka platform

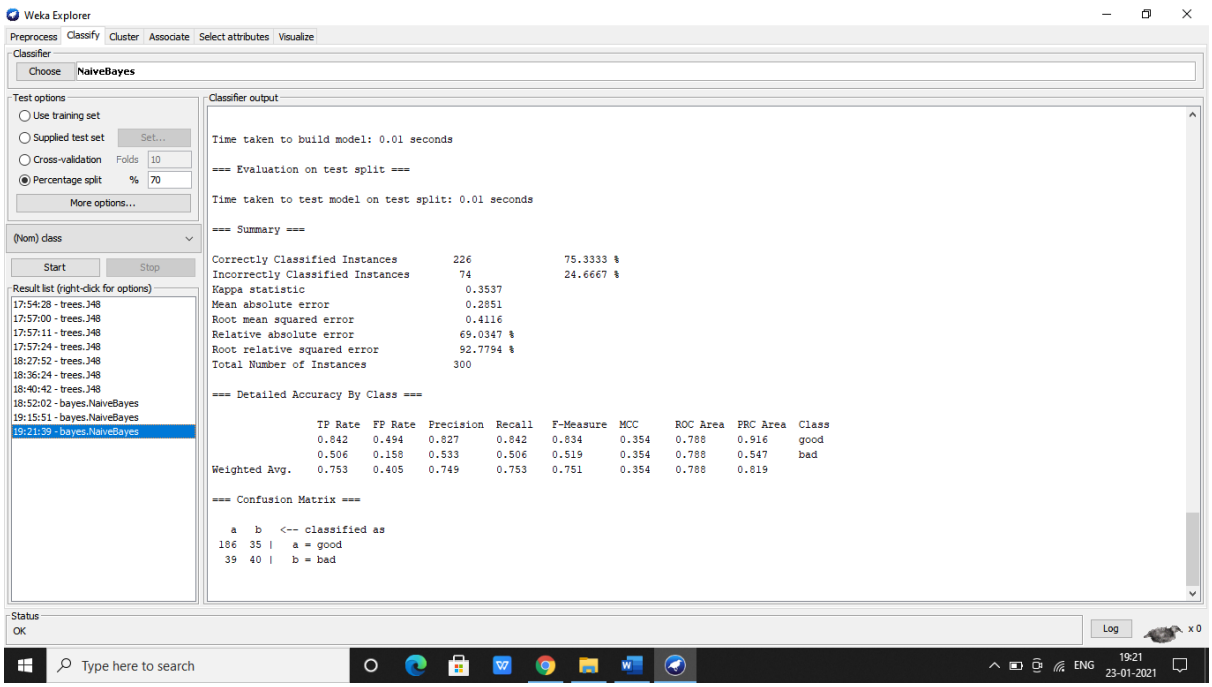


- To classify the dataset opt the classify option and choose the required algorithm. In present case we choose Naivebayes. Now perform the classification by choose cross validation and percentage split.

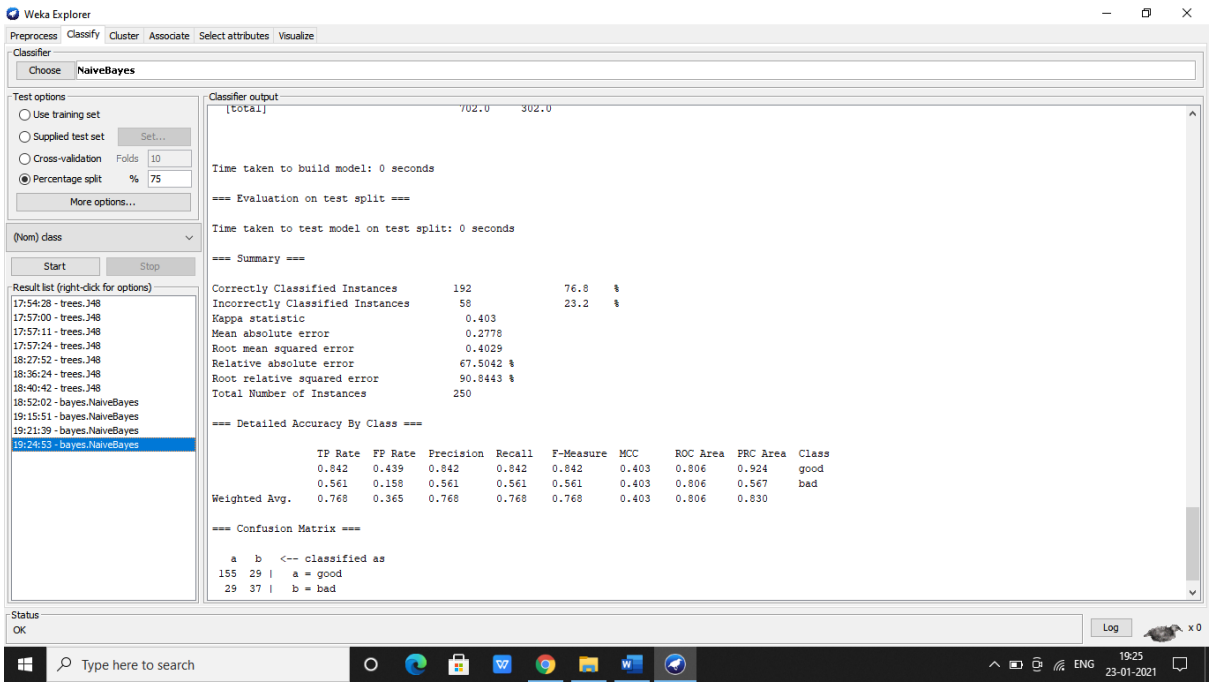
- When the percentage split is 66% and with cross validation folds 10 the accuracy observed is 76.47% .Time taken to build model is 0.01 sec.



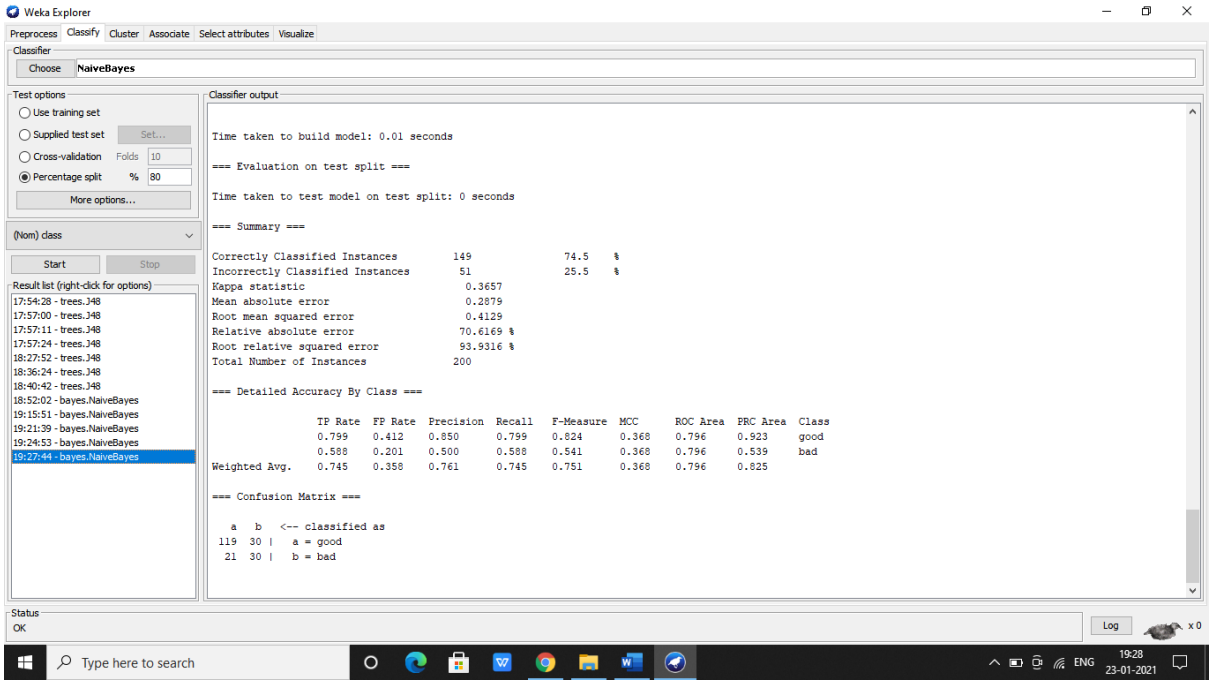
- When the percentage split is 70% and with cross validation folds 10 the accuracy observed is 75.33% .Time taken to build model is 0.01 sec.



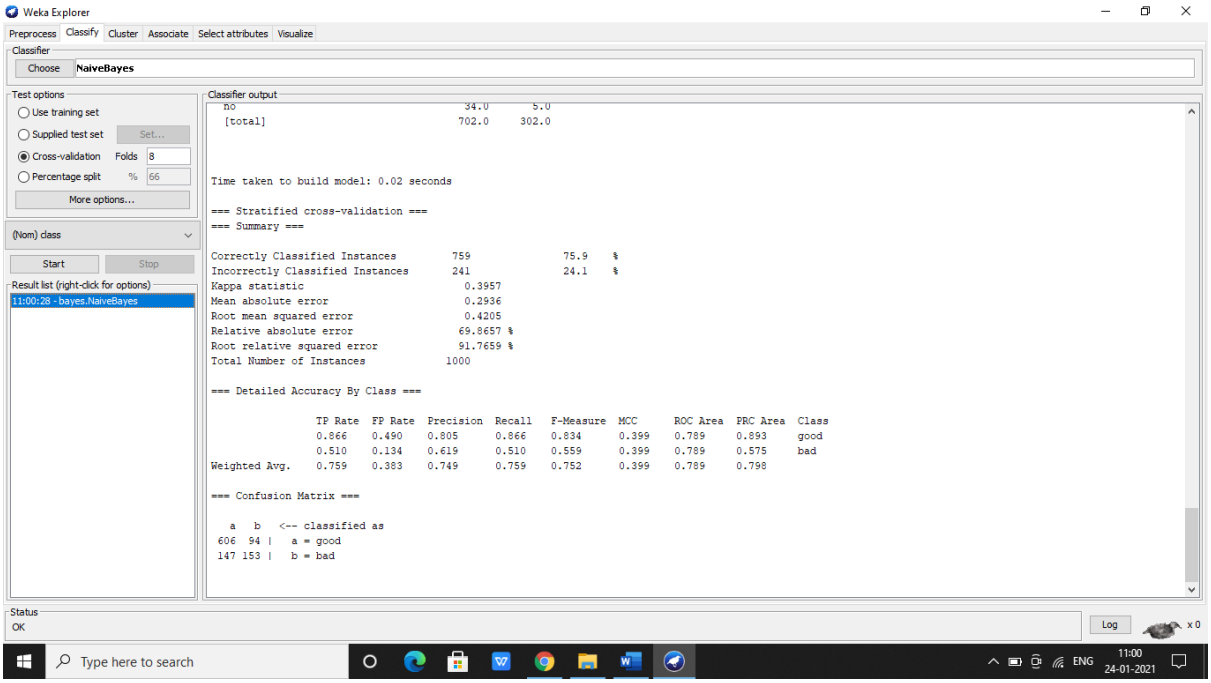
- When the percentage split is 75% and with cross validation folds 10 the accuracy observed is 76.8% .Time taken to build model is 0.02 sec.



- When the percentage split is 80% and with cross validation folds 10 the accuracy observed is 74.5% .Time taken to build model is 0.01 sec.



- When the percentage split is 66% and with cross validation folds 8 the accuracy observed is 75.9% .



- When the percentage split is 66% and with cross validation folds 6 the accuracy observed is 75.4%.