DATA MINING ASSIGNMENT 2

NaiveBayes Classification

TASK 1: One type of model that you can create is a Naivebayes. Train a Naivebayes using the complete dataset as the training data. Report the model obtained after training.

PROCEDURE:

1) Open Weka GUI Chooser.

2) Select WORKBENCH present in Applications.

3) Go to OPEN file and browse the file that is already stored in the system “credit-g.arff”.

4) Go to Classify tab.

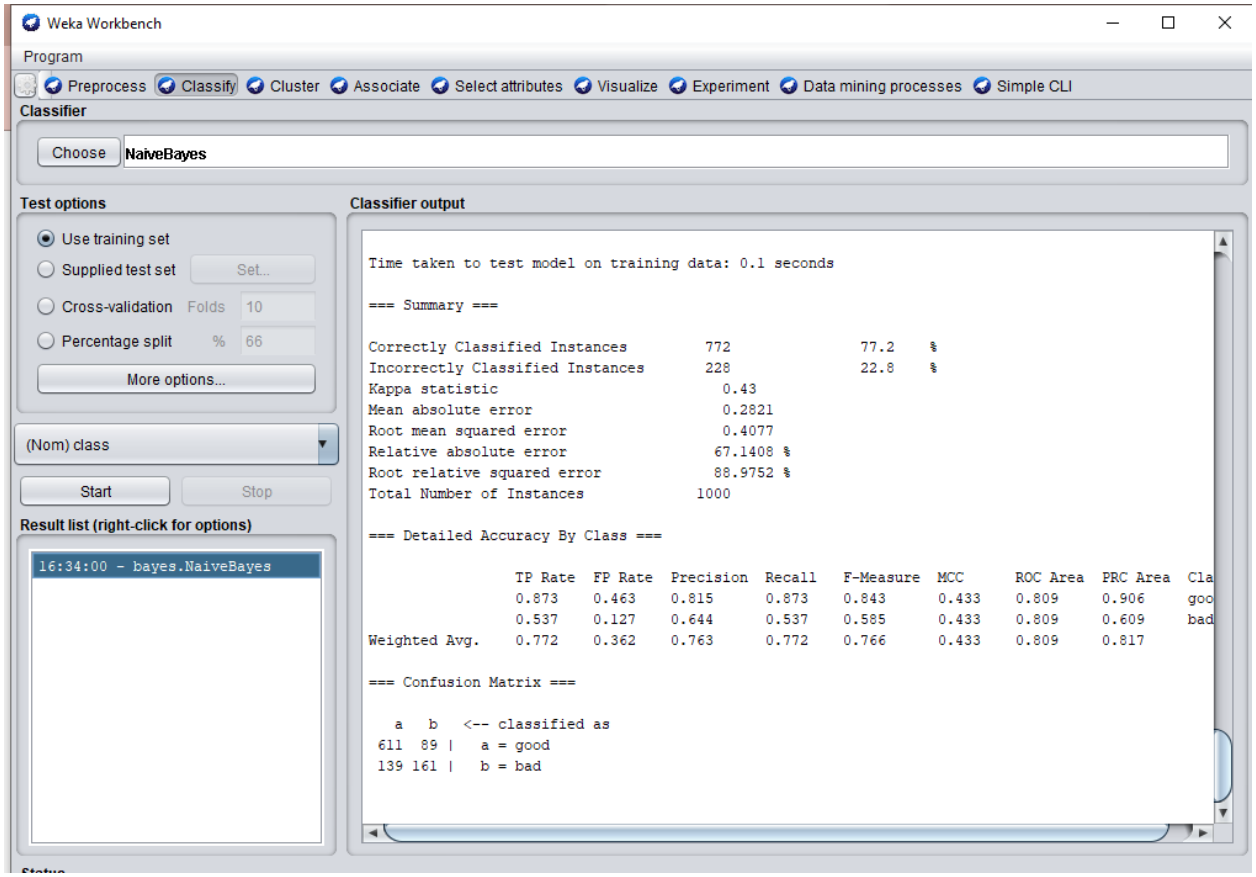
5) Click on choose button then select NaiveBayes in Bayes dropdown list.

6) Select Test options “Use training set”.

7) Select class attribute.

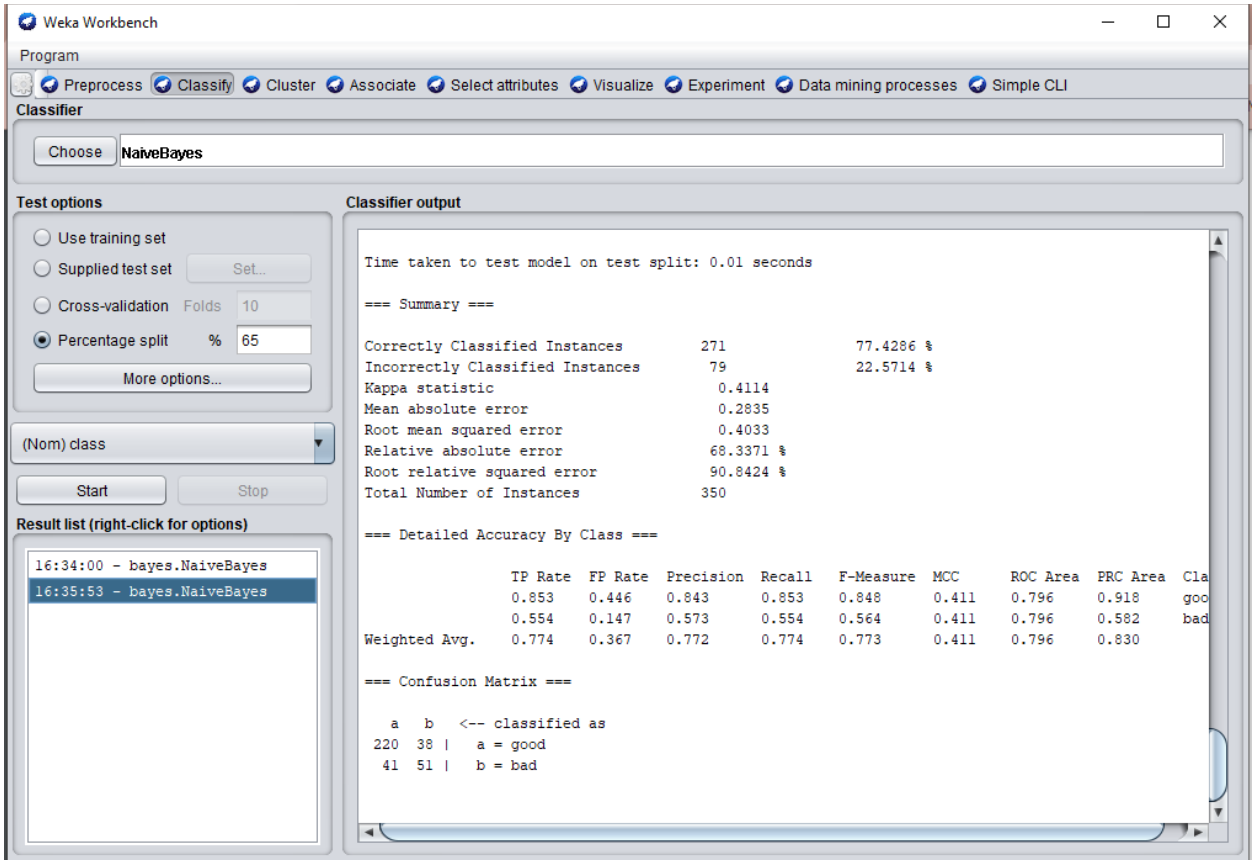
8) Click Start.

9) Now we can see the output details in the Classifier output.

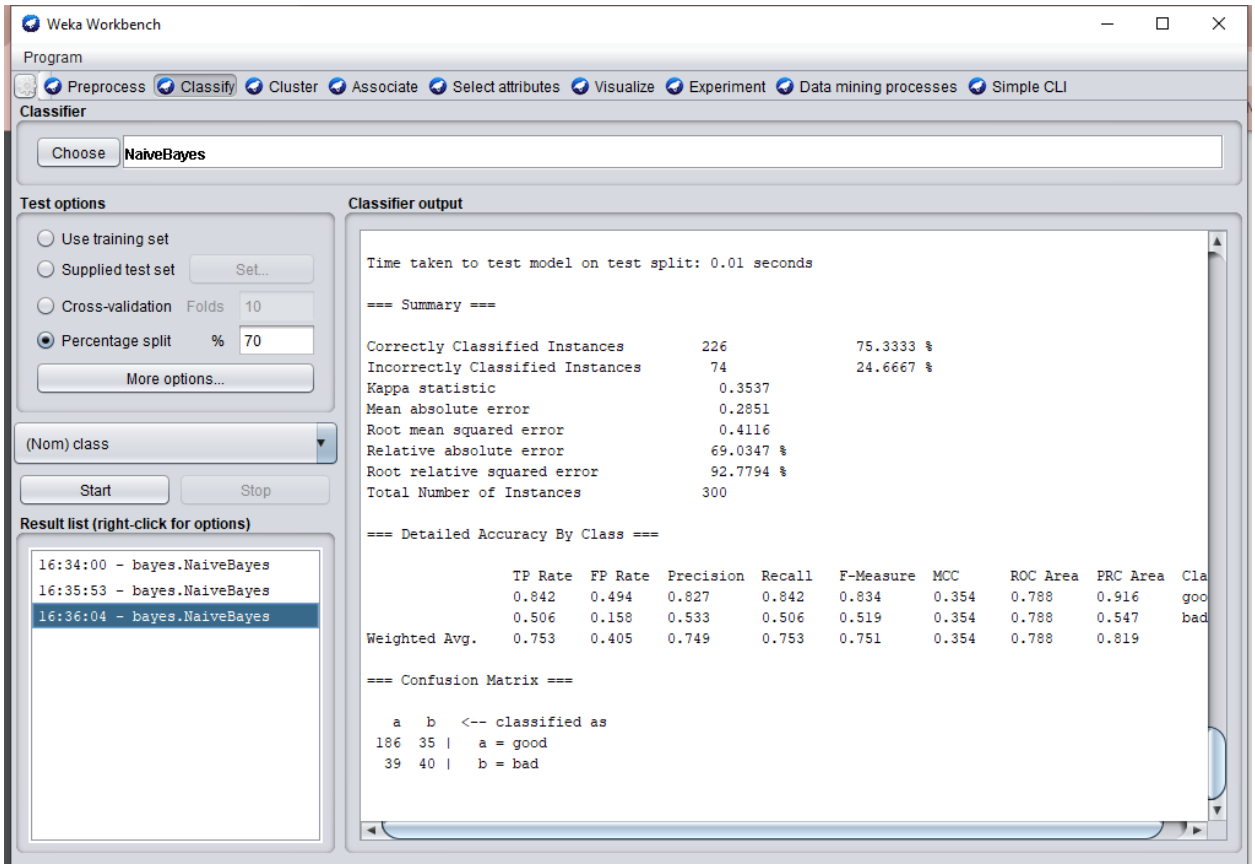

Weka Workbench
Program
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Experiment
Data mining processes
Simple CLI
Classifier
Choose
NaiveBayes
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
16:34:00 - bayes.NaiveBayes
Classifier output
Time taken to test model on training data: 0.1 seconds
=== Summary ===
Correctly Classified Instances 772 77.2 %
Incorrectly Classified Instances 228 22.8 %
Kappa statistic 0.43
Mean absolute error 0.2821
Root mean squared error 0.4077
Relative absolute error 67.1408 %
Root relative squared error 88.9752 %
Total Number of Instances 1000
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Cla
0.873 0.463 0.815 0.873 0.843 0.433 0.809 0.906 goo
0.537 0.127 0.644 0.537 0.585 0.433 0.809 0.609 bad
Weighted Avg. 0.772 0.362 0.763 0.772 0.766 0.433 0.809 0.817
=== Confusion Matrix ===
a b <-- classified as
611 89 | a = good
139 161 | b = bad

TASK 2:Train a NaiveBayes using percentage split and report your results.Increase percentage split by 5% upto 80% starting from 65% and check at which percentage split we are getting the best accuracy.

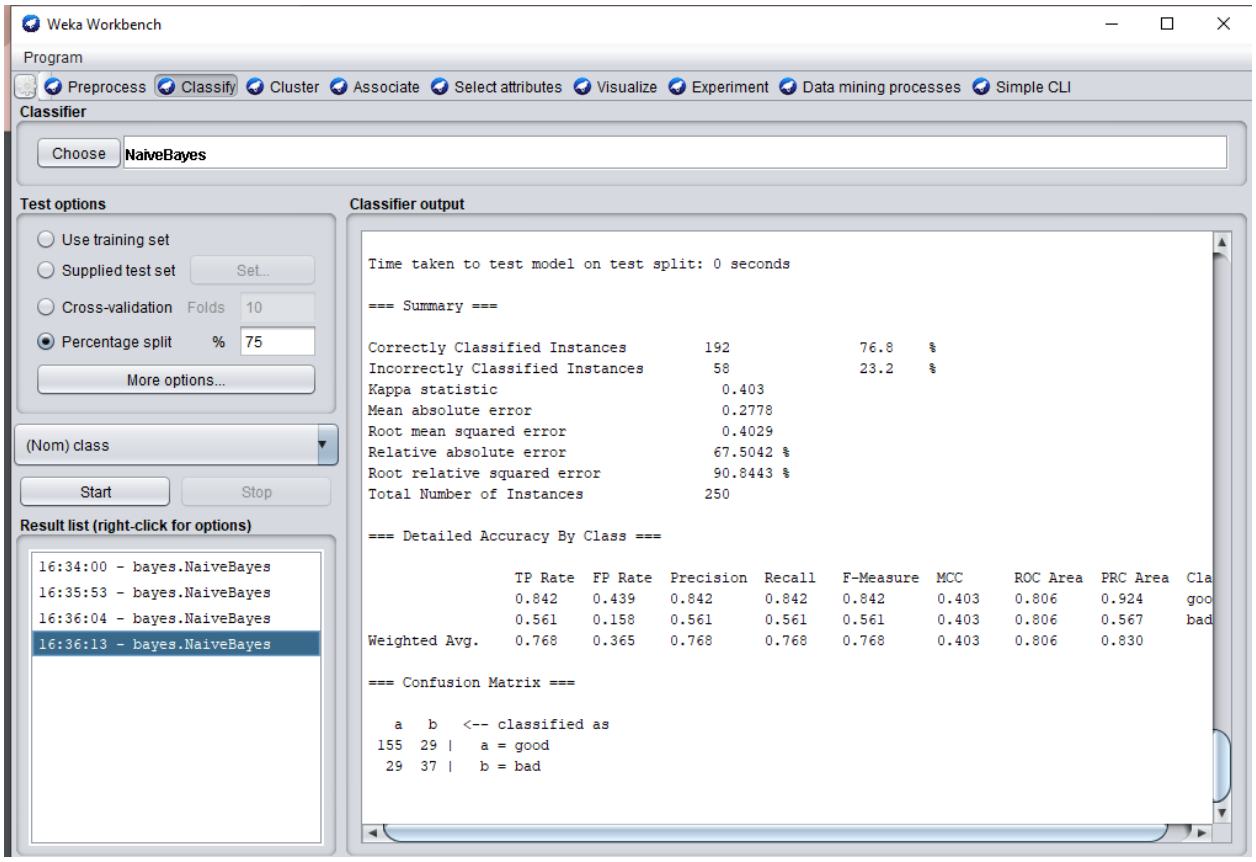
1.When the percentage split is 65%,the accuracy is 77.4286%.



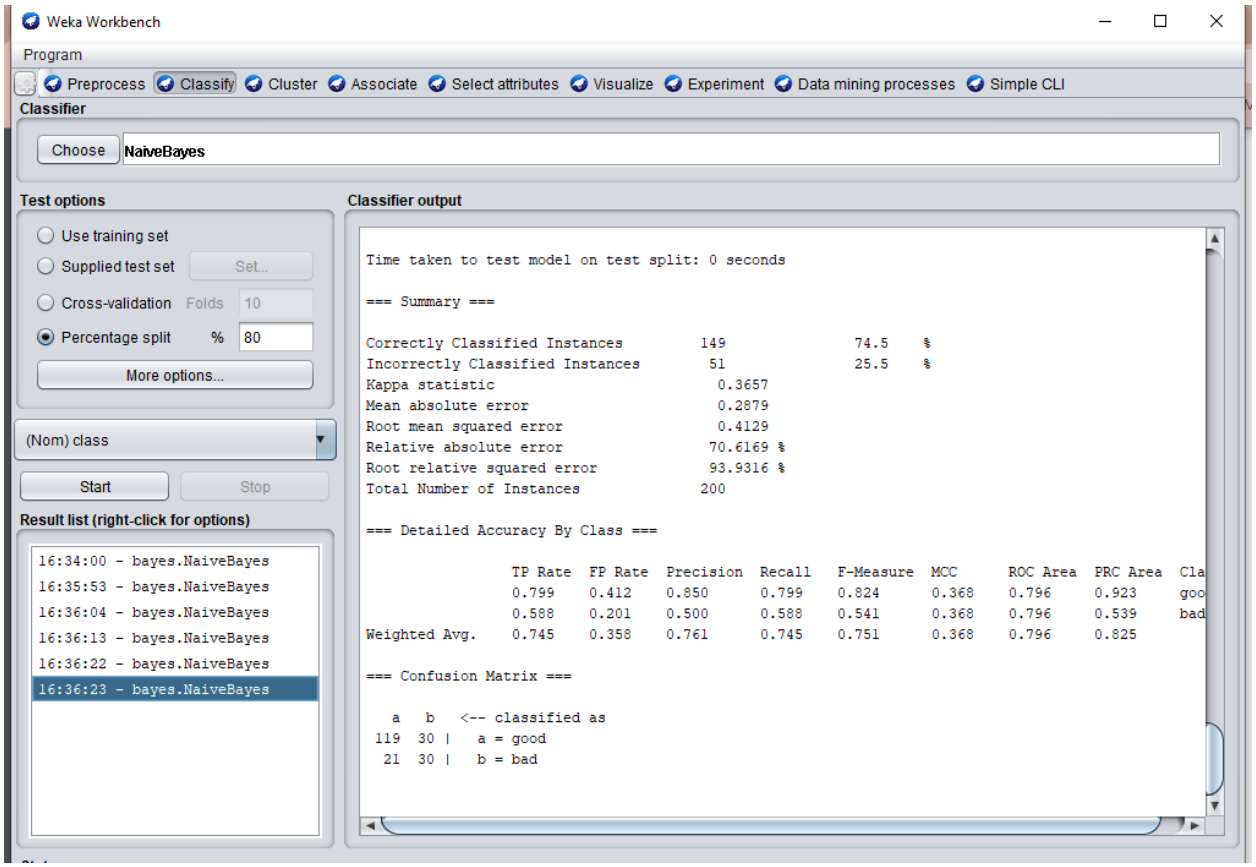
2.When the percentage is 70% the accuracy is 75.3333%



3.When the percentage is 75% the accuracy is 76.8%



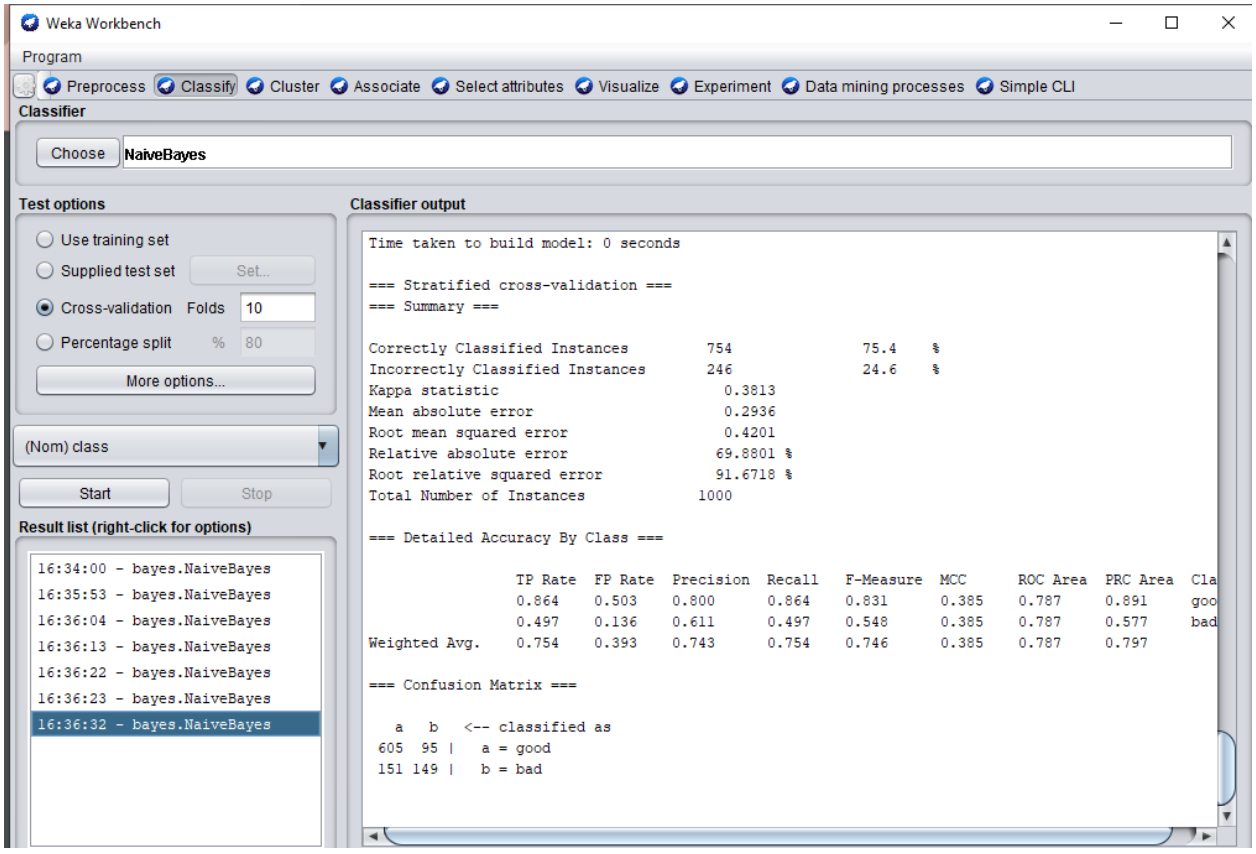
4.When the percentage is 80 the accuracy is 74.5%



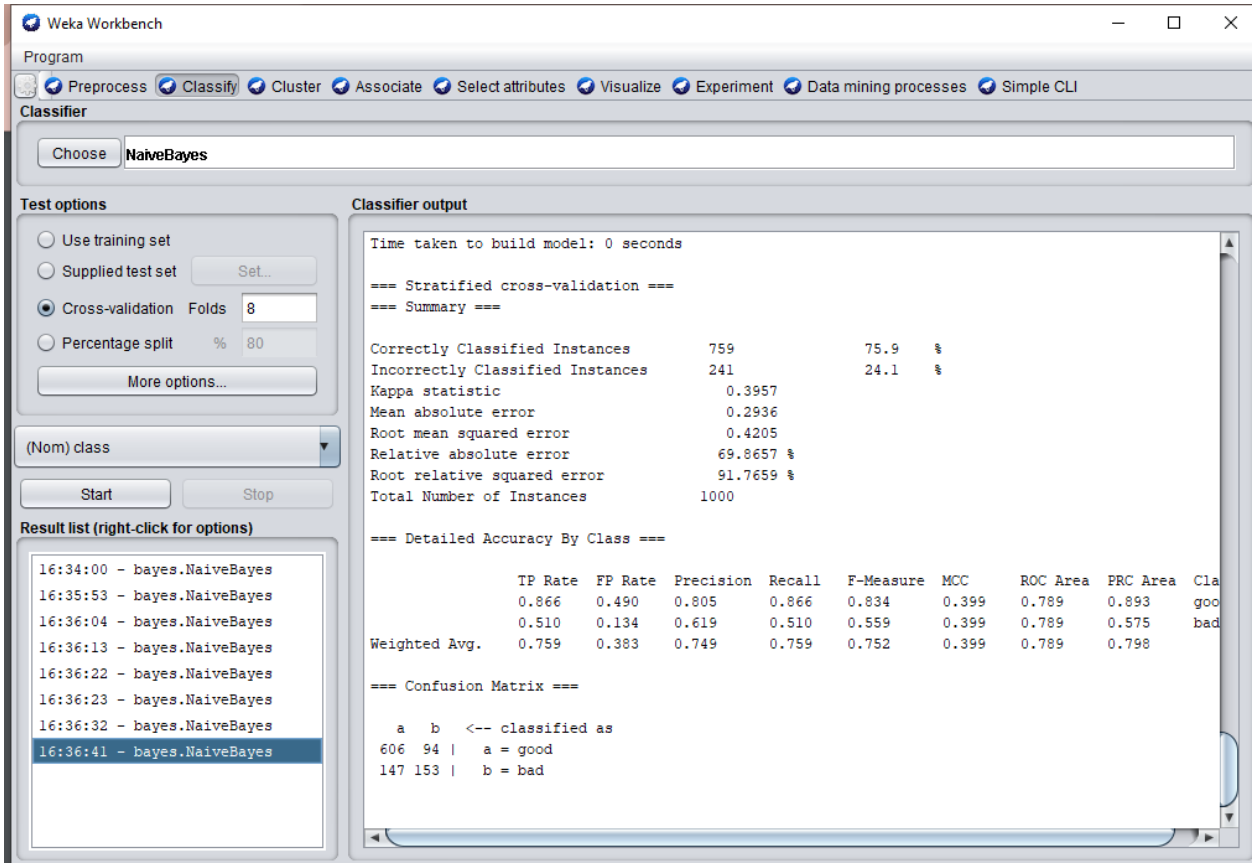
CONCLUSION : When the percentage split is 65%,the accuracy is high which is 77%.

TASK 3:Train the NaiveBayes using cross validation and report the results.

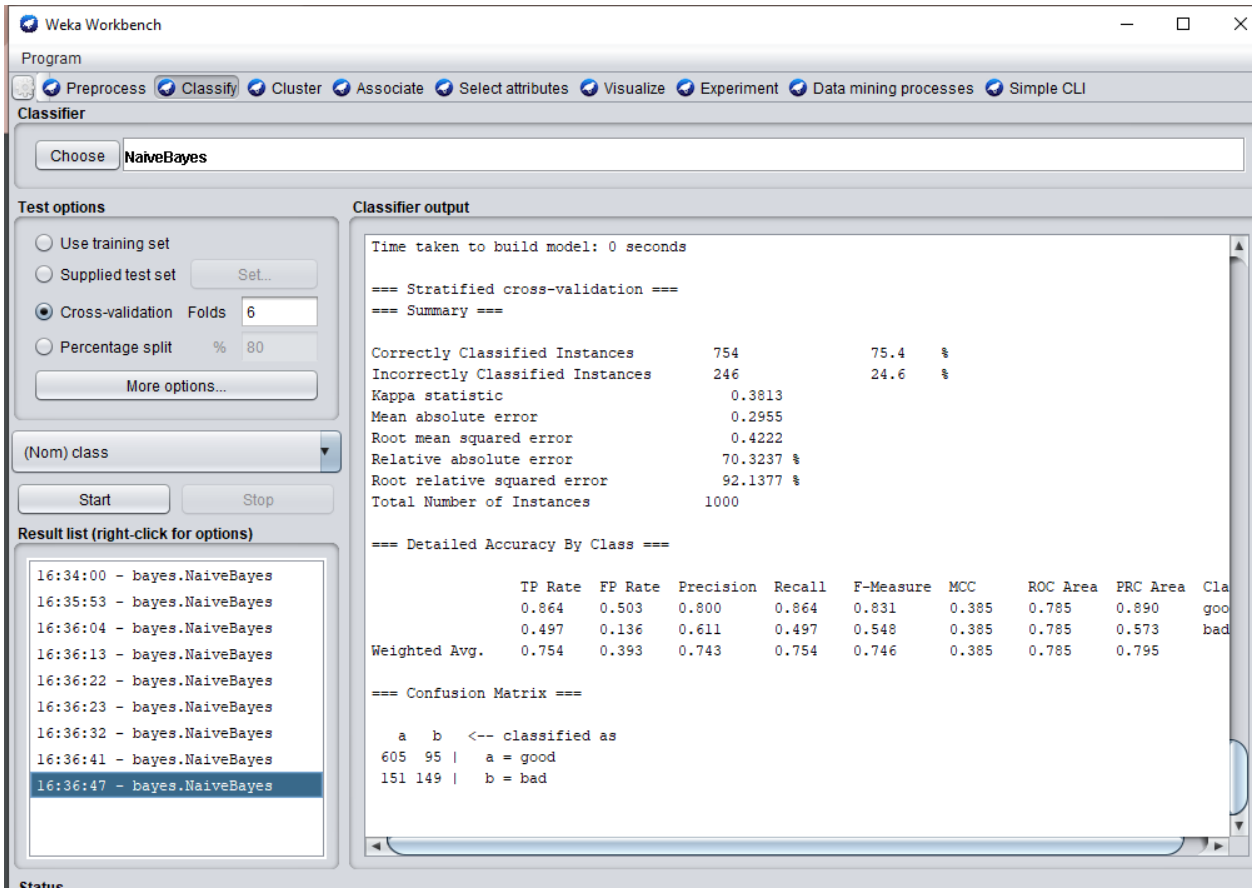
1.When cross validation folds:10,accuracy is 75.4%.



2.When cross validation folds : 8,accuracy is 75.9%.



3.When cross Validation folds:6, accuracy is 75.4%.



CONCLUSION : The accuracy is high when the number of cross validation folds are 8 which is 75.9%.