## DATA MINING ASSIGNMENT 2

Decision tree classification

TASK 1: One type of model that you can create is a decision tree. Train a decision tree 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 the Classify tab.

5) Here the c4.5 algorithm has been chosen which is entitled as j48 in Java and can be selected by clicking the button choose and select tree j48.

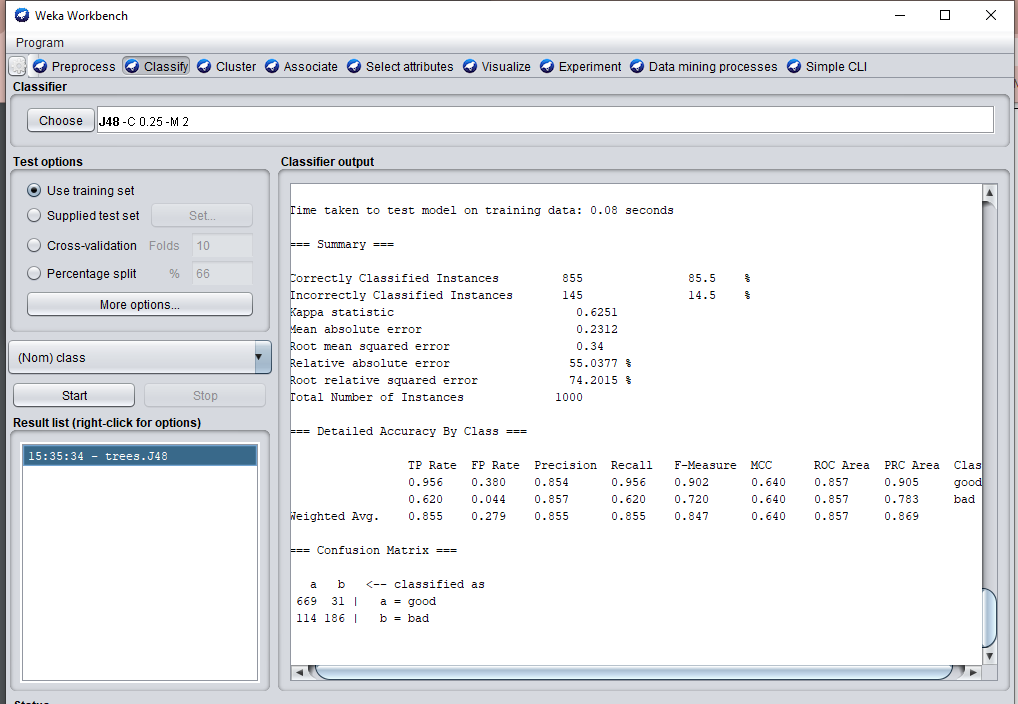
6) Select Test options “Use training set”.

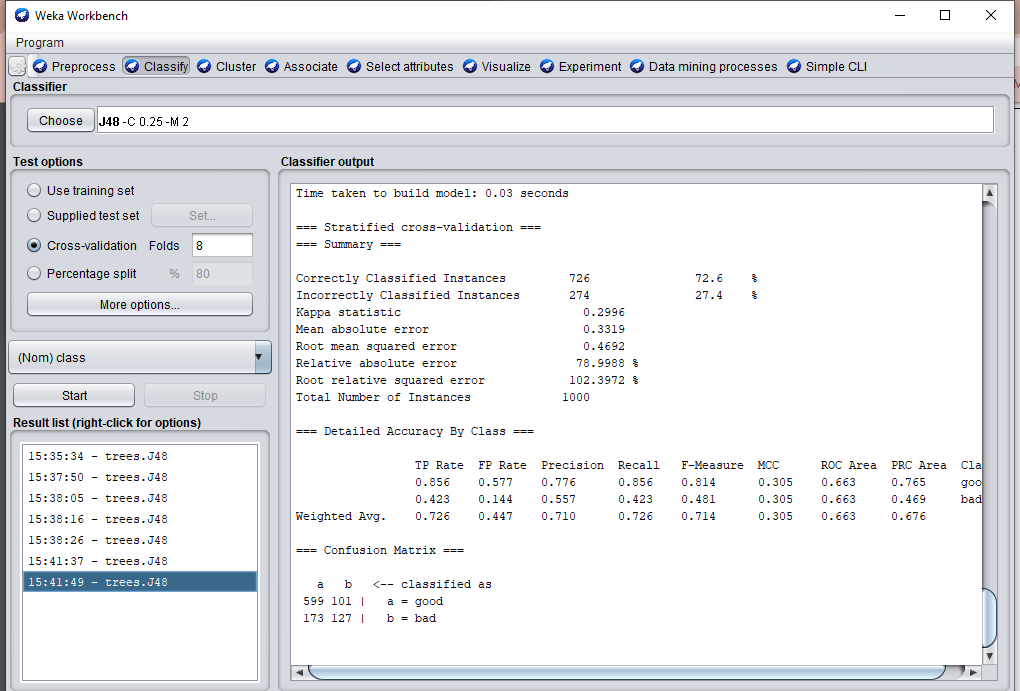
7) Select class attribute.

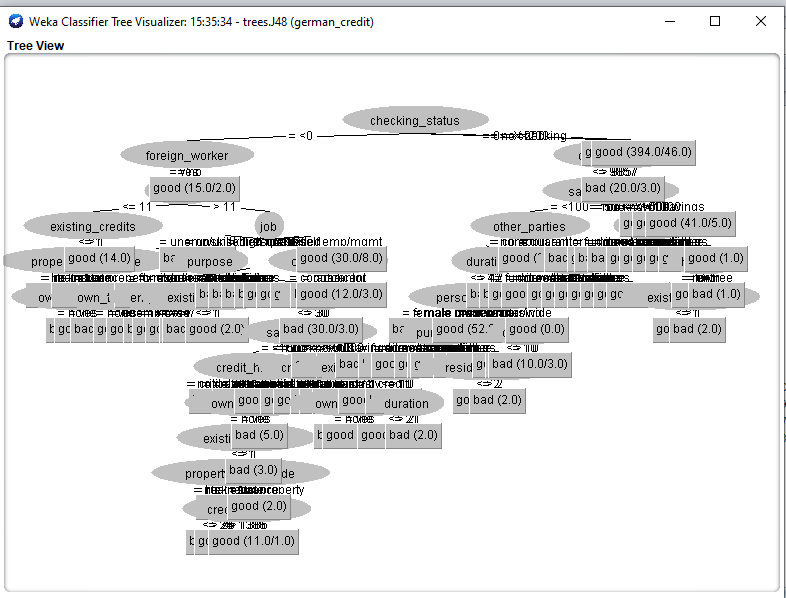
8) Click Start.

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

10) Right click on the result list and select “visualize tree” option.

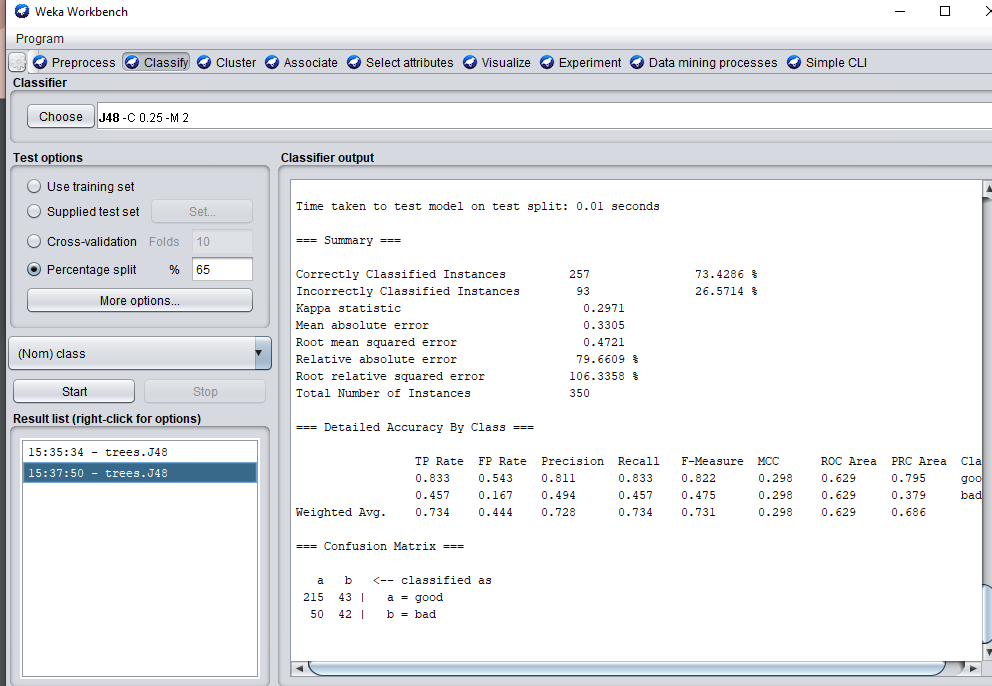




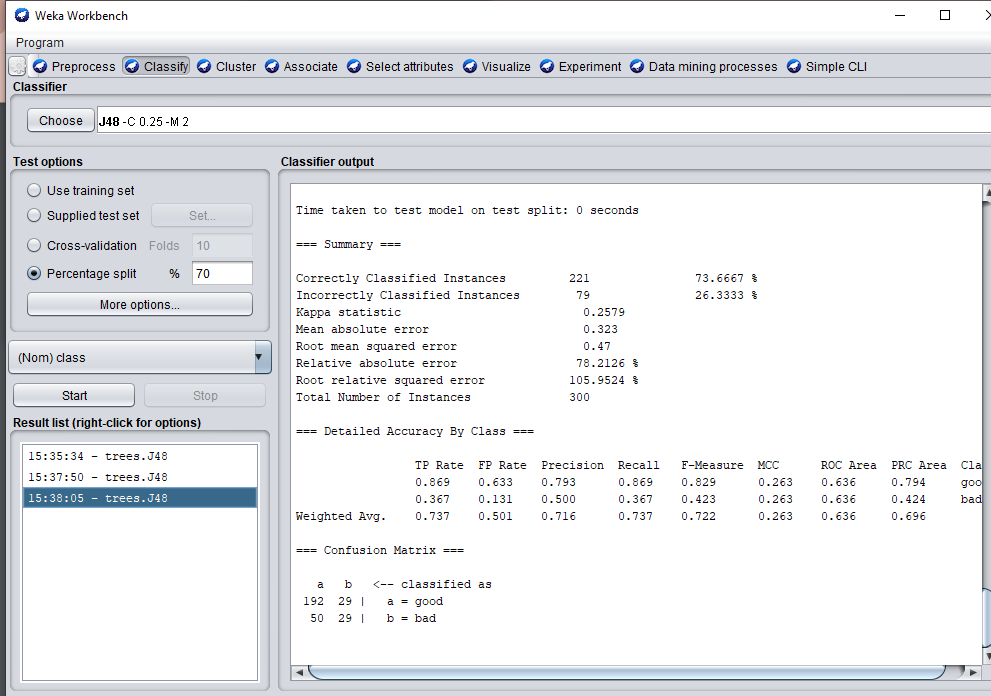


TASK 2: Train a Decision Tree using percentage split and report your results. Increase the percentage split by 5% upto 80% starting from 65% and check at which percentage split we are getting the best accuracy.

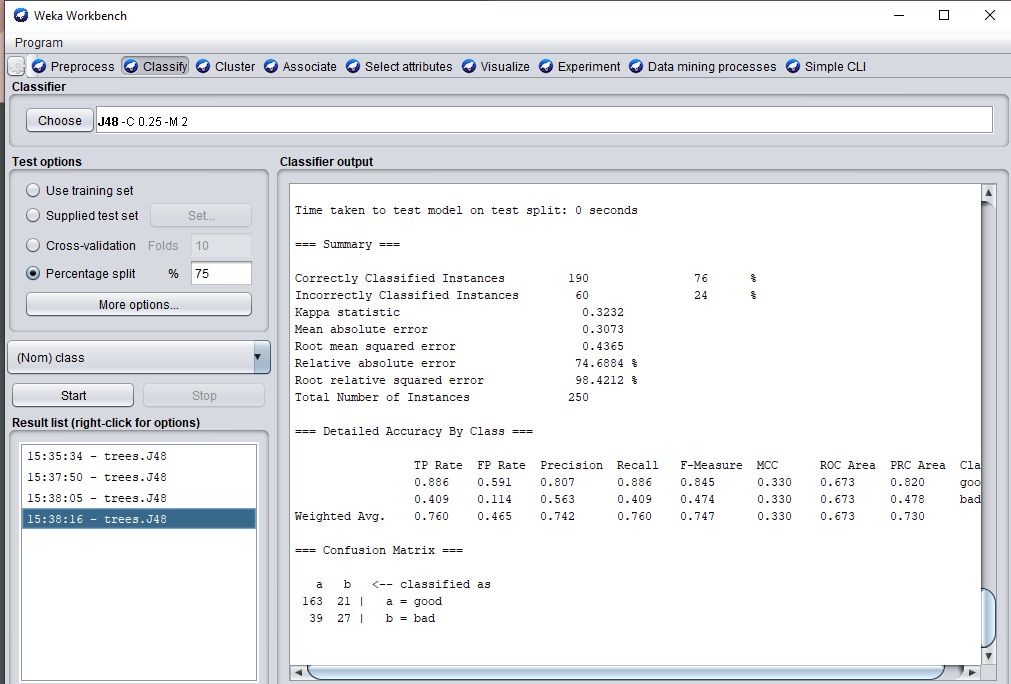
1. When percentage split is 65%, the accuracy is 73.4286%



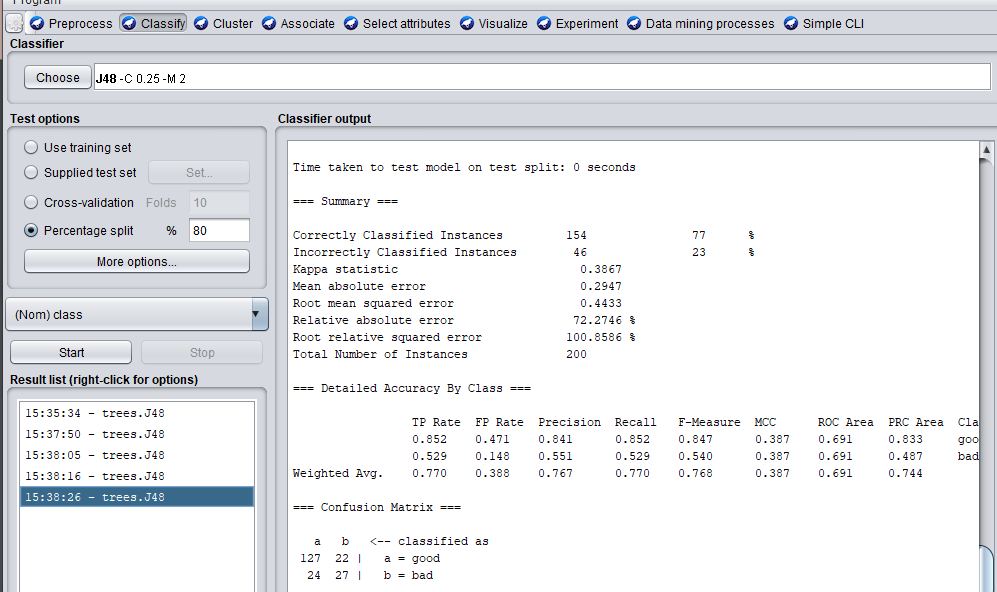
1. When percentage split is 70%, the accuracy is 73.6667%



1. When percentage split is 75%, the accuracy is 76%



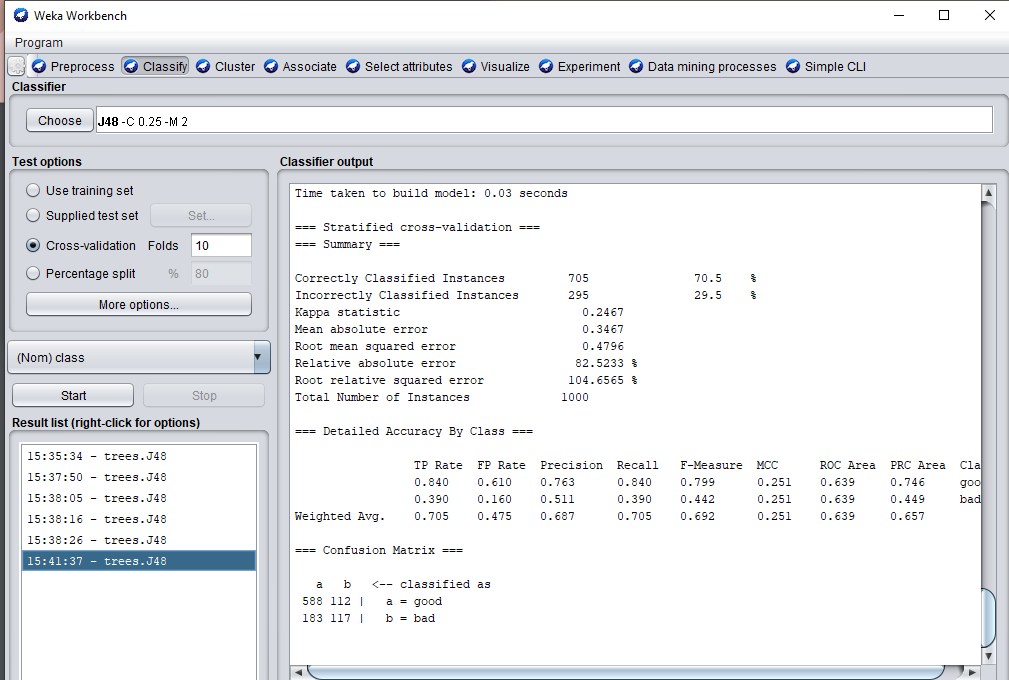
1. When percentage split is 80%, the accuracy is 77%



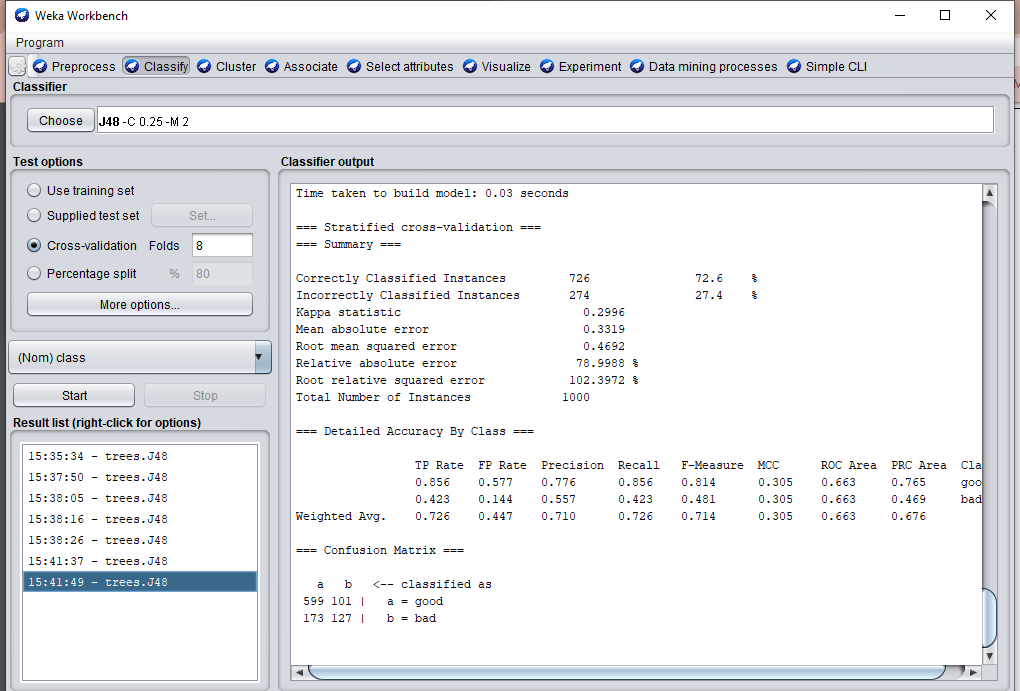
CONCLUSION: When the percentage split is 80%, the accuracy is high(77%).

TASK 3: Train a Decision Tree using cross validation and report your results.

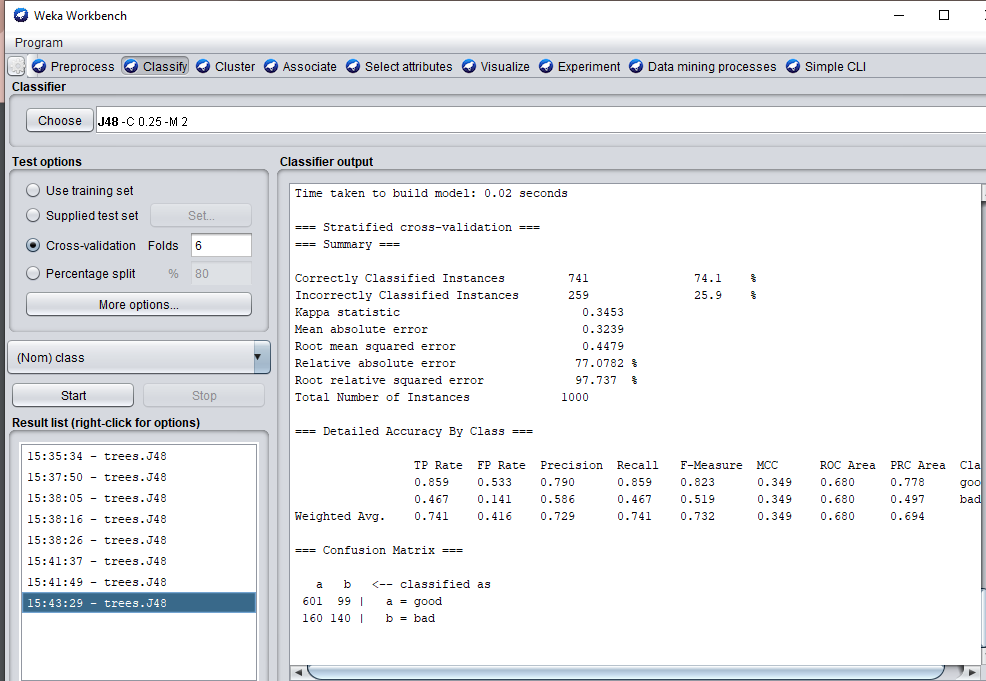
1. When cross validation folds : 10, accuracy is 70.5%



1. When cross validation folds : 8, accuracy is 72.6%



1. When cross validation folds : 6, accuracy is 74.1%



CONCLUSION: The accuracy is high(74.1%) when cross validation folds: 6