**DATA MINING LAB -3**

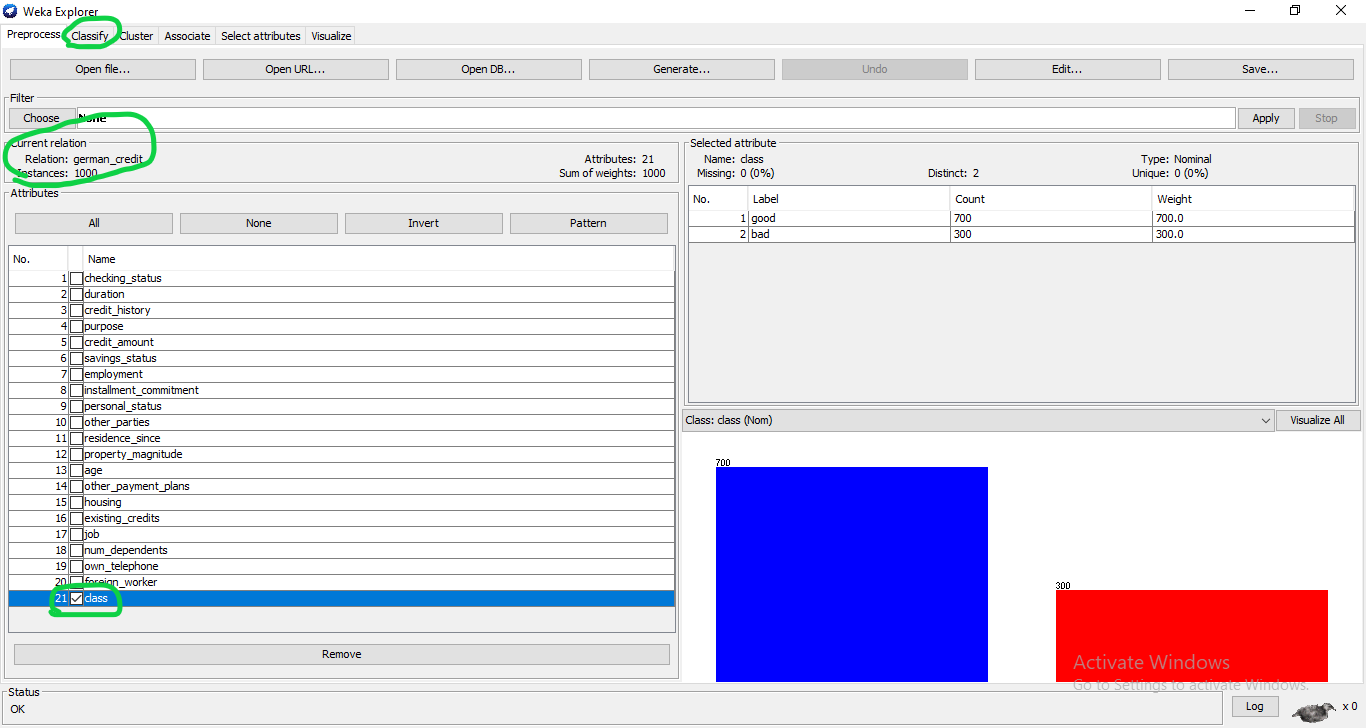
**Q.) TO DERIVE A DECISION TREE FROM A GIVEN DATASET:**

STEP 1: open Weka.

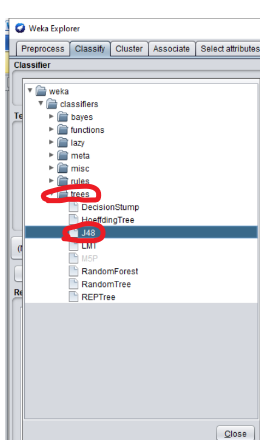
STEP 2: open explorer -> choose required dataset.

* Here choose dataset is “German Credit Dataset.”

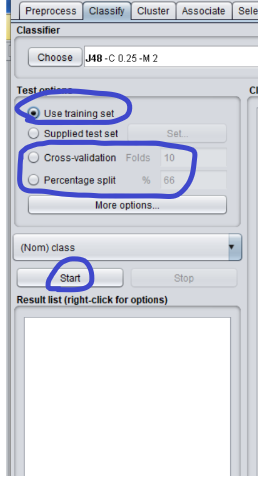
STEP 3: select class attribute from the given dataset and On the top click the classify tab.

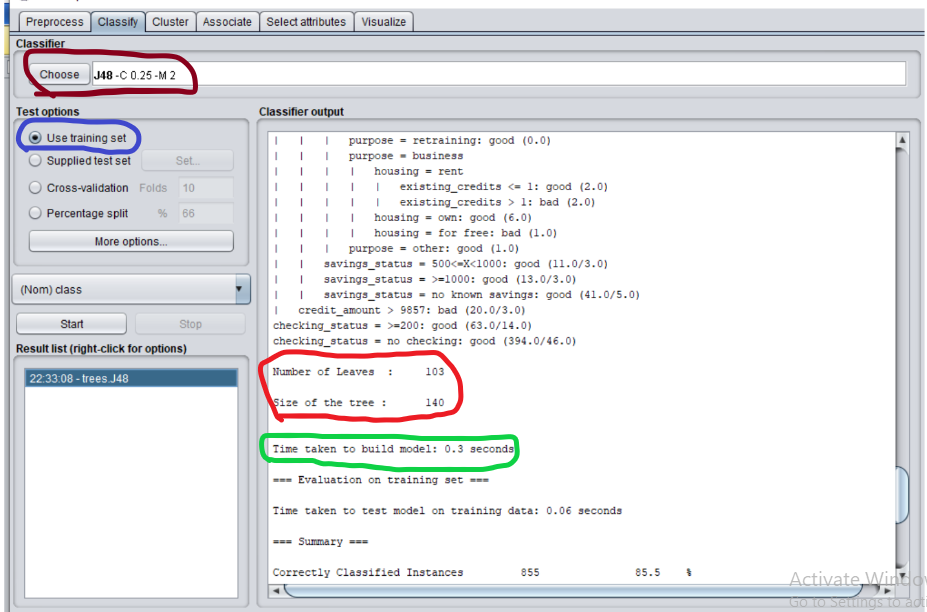


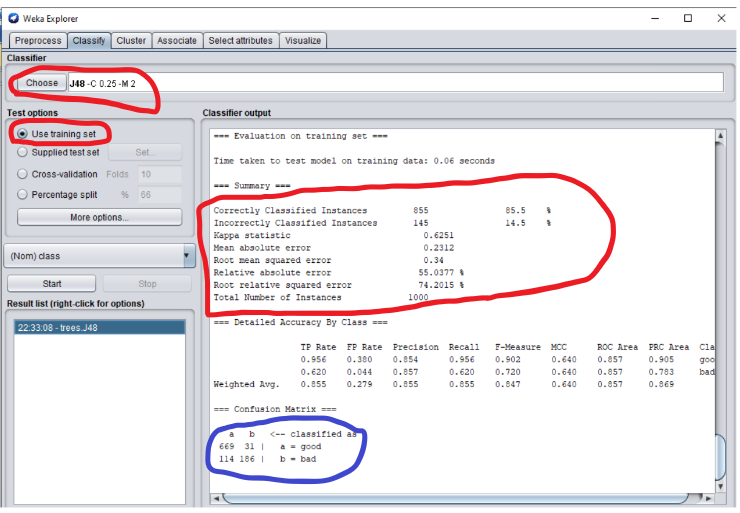
STEP 4: from classify select tree in that select J48.



STEP 5: Choose the training set from the test option and then start.

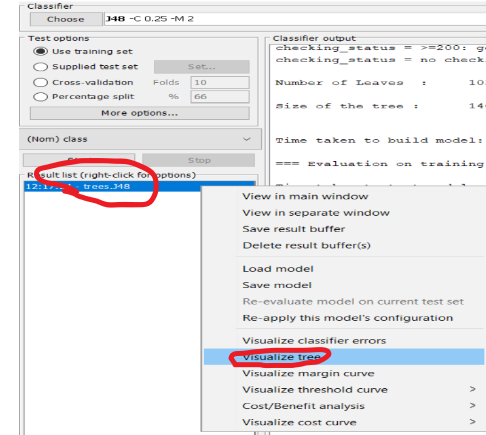


STEP 6: 

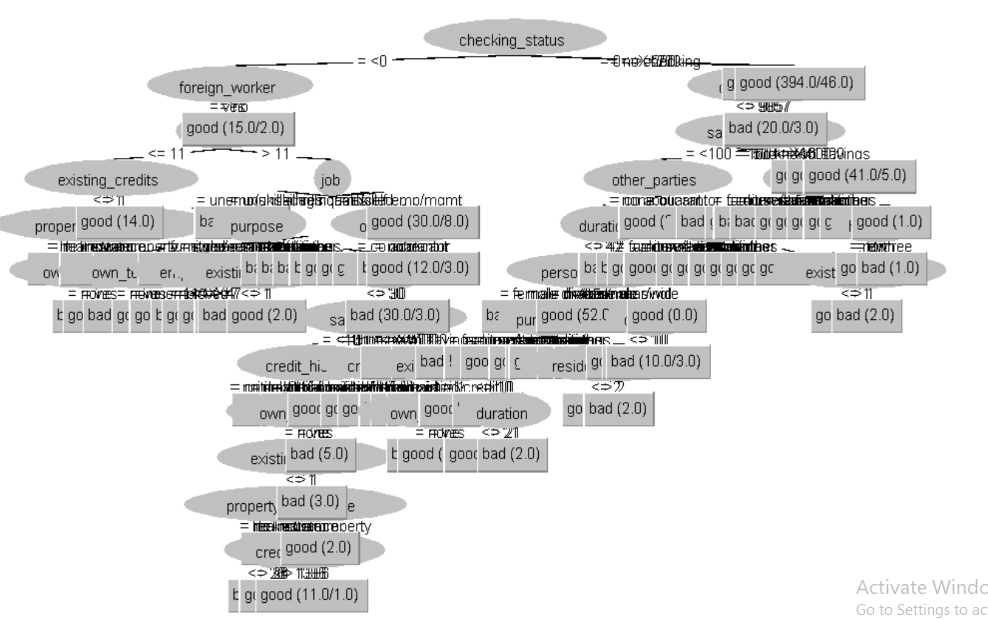


* Here the number of leaves: 103.
* Size of tree: 140
* Time taken to build a model: 0.3sec
* We get “Confusion Matrix” as the output.
* When **percentage split = 66%** the accuracy will be 85.5%.

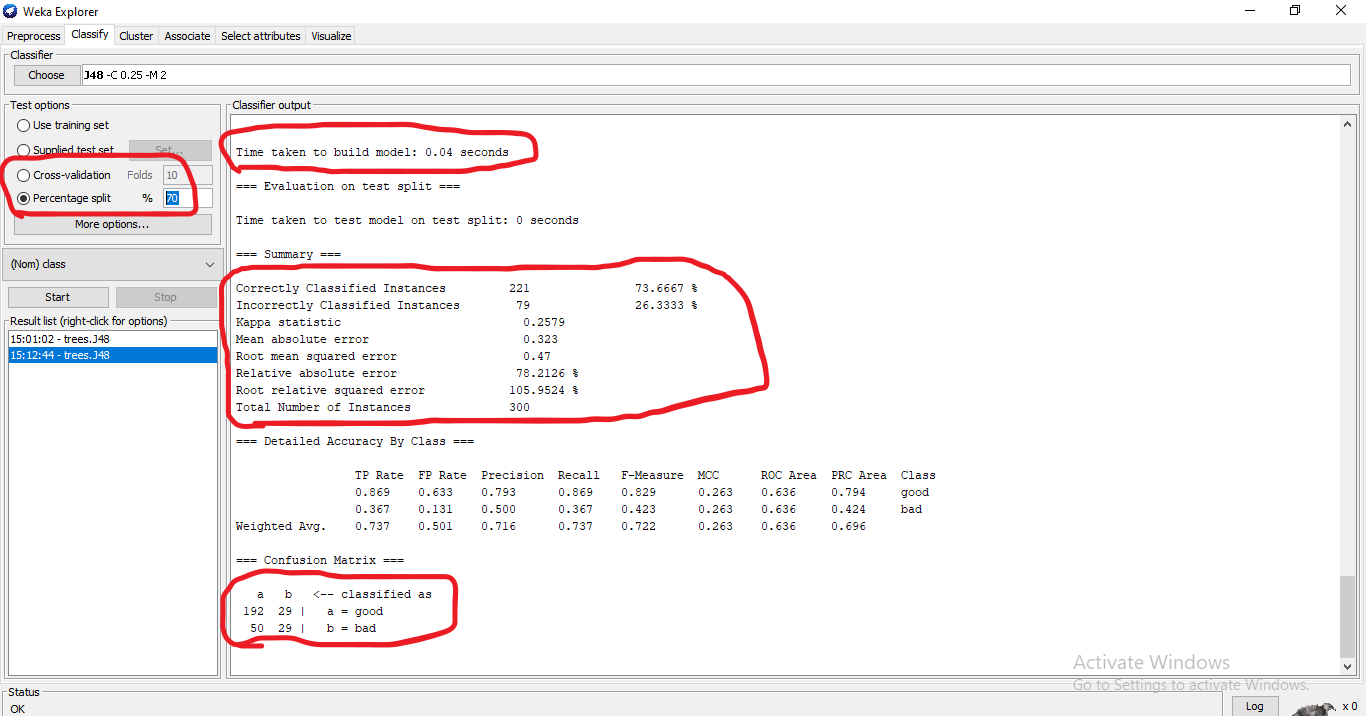
STEP 7: To get a tree. Right click on the option in the result list and select the visualize tree option.



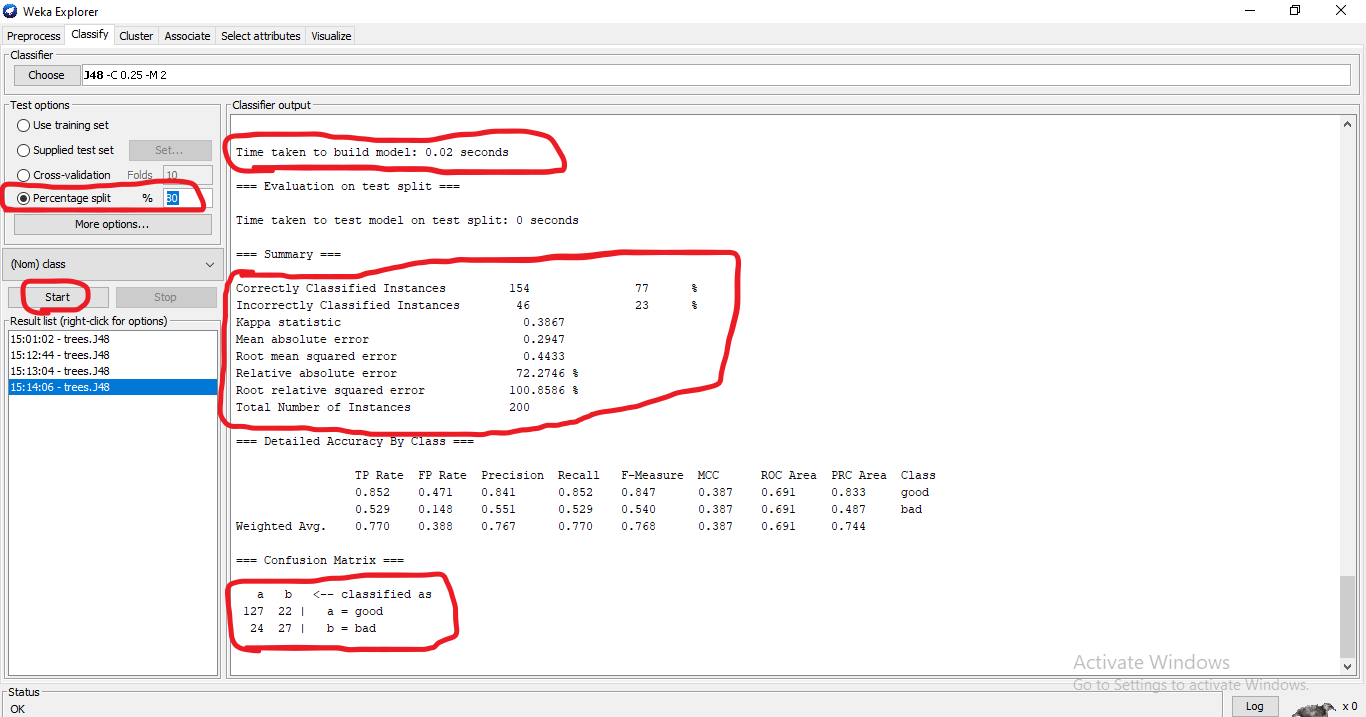
STEP 8: the output of the tree will be like this:



* When Percentage split = 70% the accuracy will be 73.6%.
* Time to build a model = 0.04 sec.

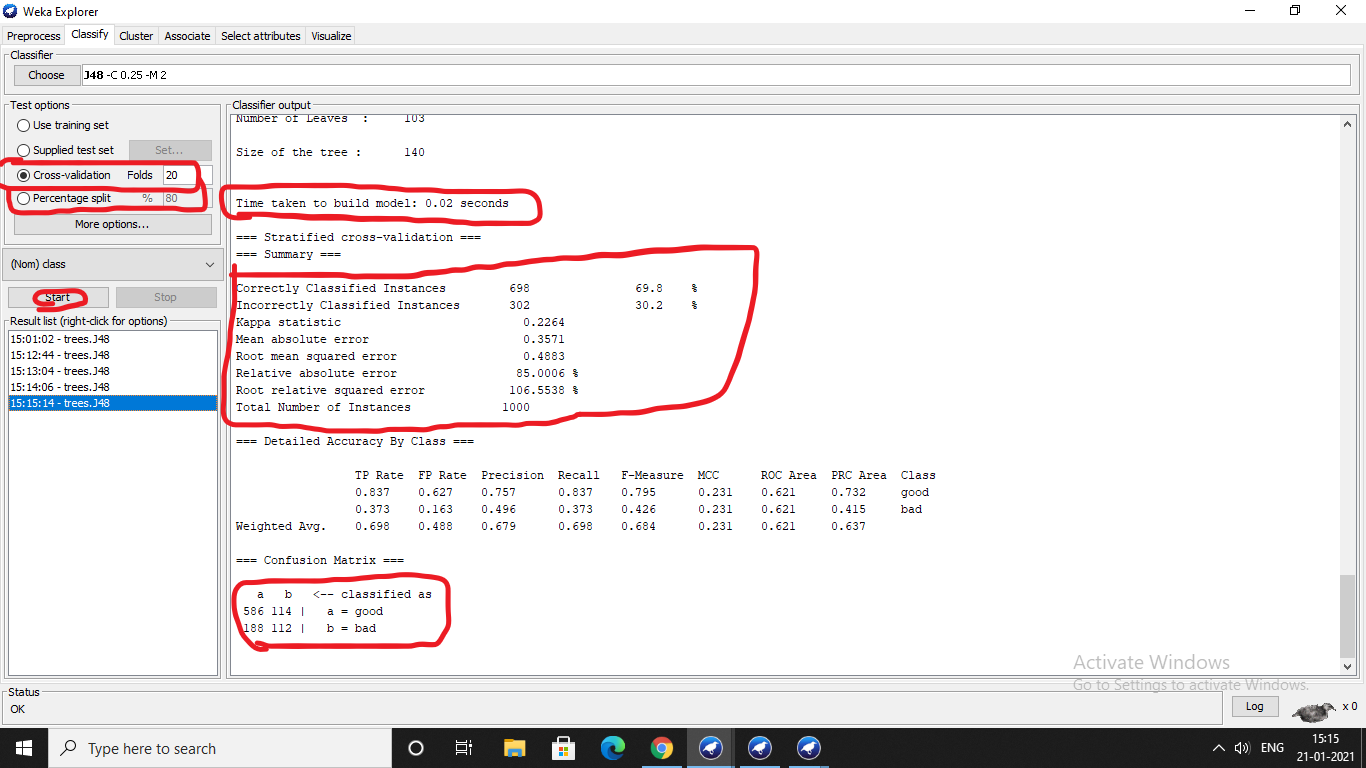


* When **Percentage split = 80%** the accuracy will be 77%
* Time to build a model = 0.02 sec



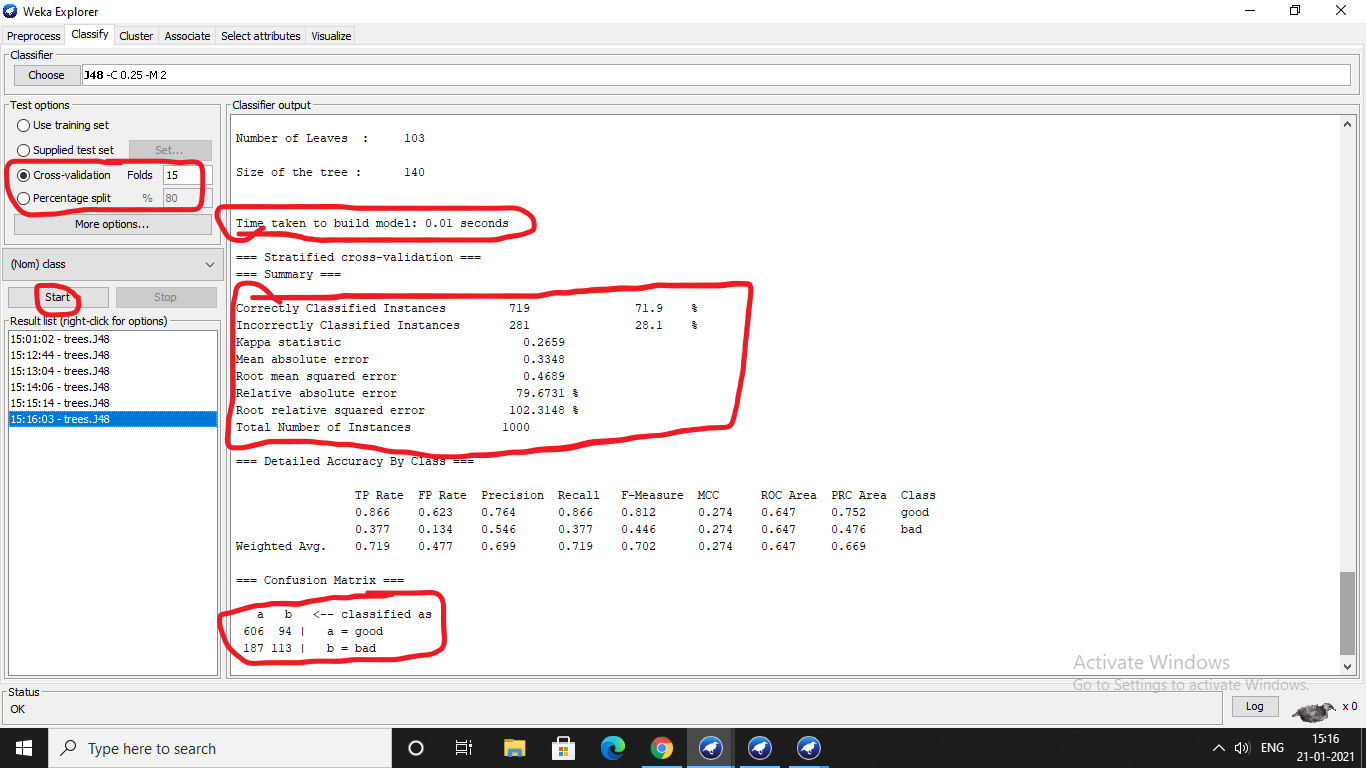
When cross-validation folds is increased by 20 and percentage split = 80%.

* Time taken to build a model = 0.02 sec
* Accuracy will be = 69.8%.



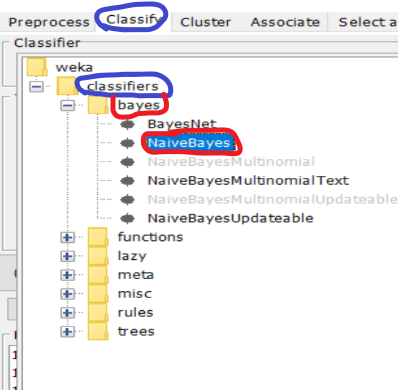
When cross validation folds = 15 and the percentage split = 80%.

* Time to build a model = 0.01 sec.



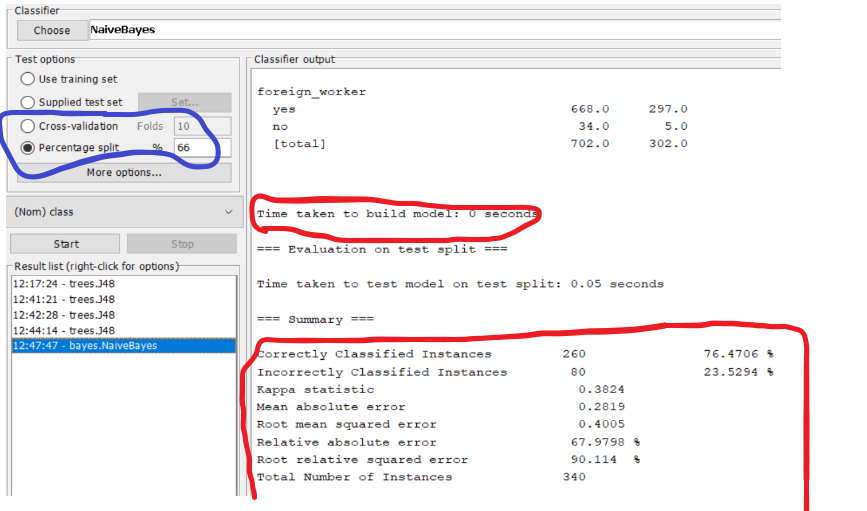
**NAÏVE BAYES CLASSIFIER**:

* To select naïve baye click choose option in the classify tab, select bayes in that select naïve bayes.



When **percentage split =66%**, cross-validation folds = 10 then

* Time to build a model = 0 sec
* Accuracy = 76.4%.



When **percentage split= 80%**, cross-validation = 10.

* Time taken to build a model = 0sec
* Accuracy = 74.5%.

