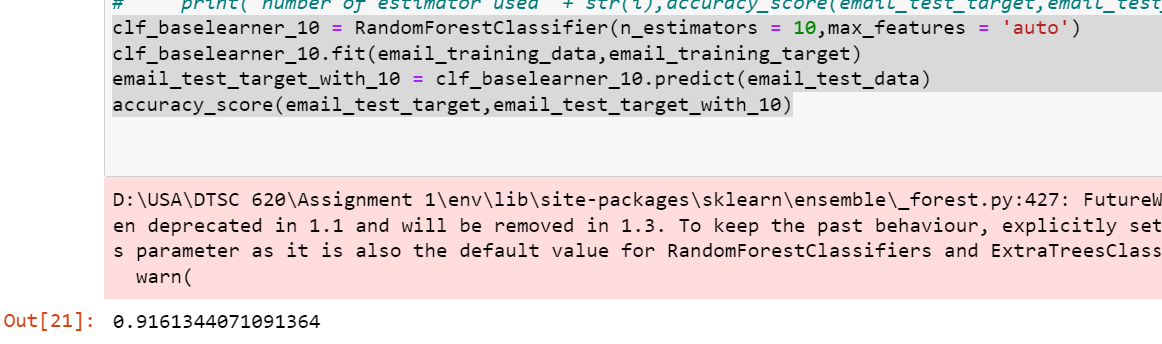
**QUESTION 1**

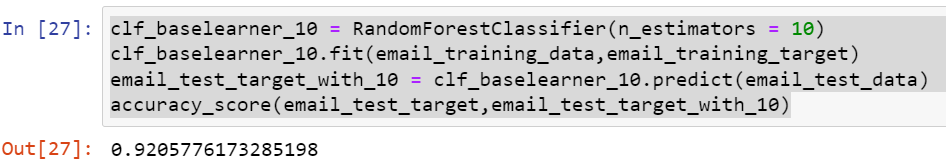
|  |  |  |
| --- | --- | --- |
| Number of Base Learners | Features split = ‘auto’ | Features split = ‘sqrt’ |
| 10 | 0.9161344071091364 | 0.9205776173285198 |
| 50 | 0.9280755345737295 | 0.9322410441544016 |
| 100 | 0.9305748403221328 | 0.9355734518189391 |
| 500 | 0.9344626492640933 | 0.9344626492640933 |
| 1000 | 0.9350180505415162 | 0.933074146070536 |
| 5000 | 0.9347403499028047 | 0.9339072479866704 |

The classification accuracy with 10 base learners irrespective to feature split has the lowest accuracy which has average of 91.8%.

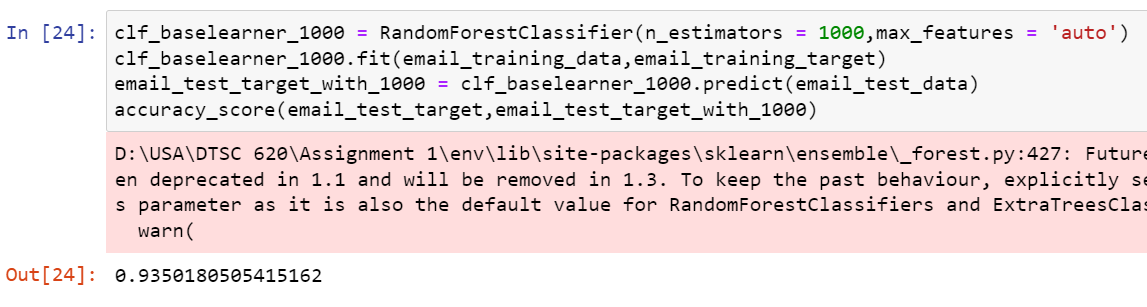
The below Screenshot uses Random Forest Classifier with base learner 10 and max\_features of auto has the accuracy 91.6%.

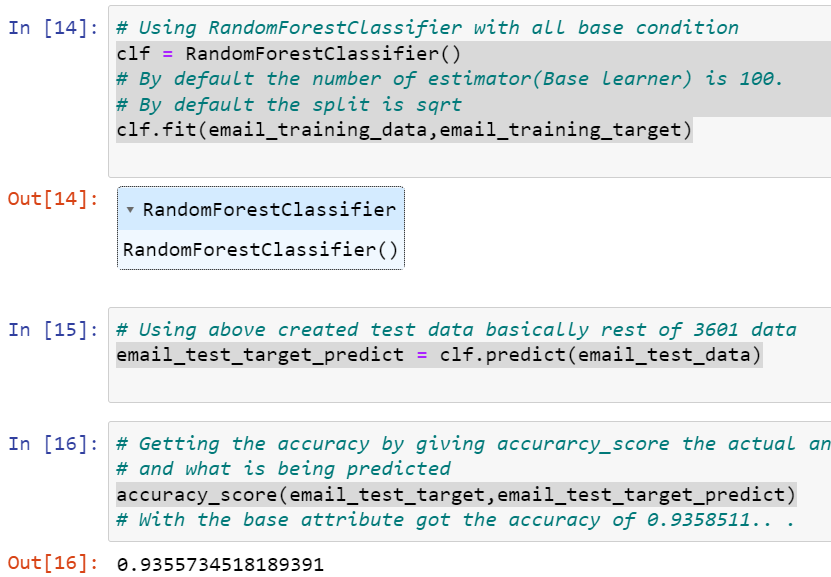


Below Classifier has 10 base learners and max\_features of sqrt which has slightly high accuracy.



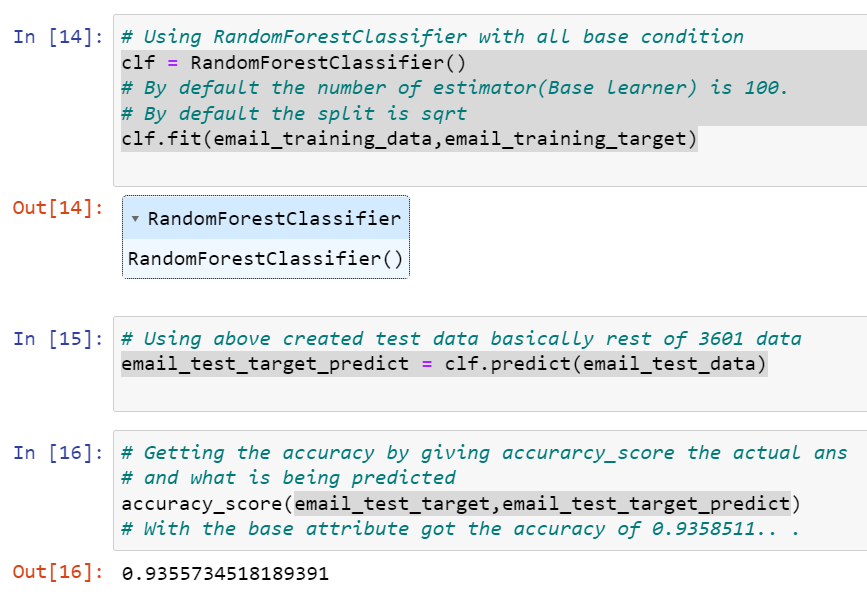
In summary, the classification accuracy increases as number of base learners increase no matter how we split the features ‘auto’ or ‘sqrt’. According to me it will be efficient to use feature split of ‘sqrt’ because with a smaller number of base learners it has achieved the highest accuracy. The best parameter for random forest classification is the default one which is 100 base learner and feature split ‘sqrt’. We can also get similar accuracy with 1000 base learners and auto feature split but then the execution time will increase.

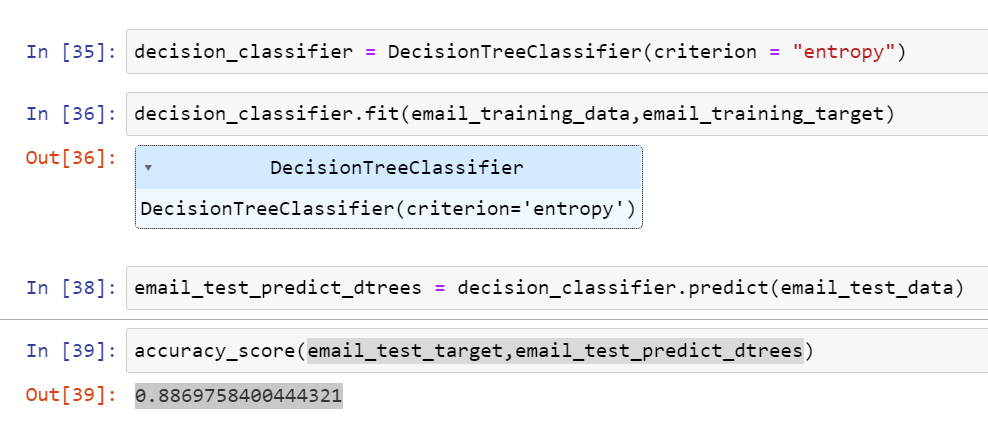




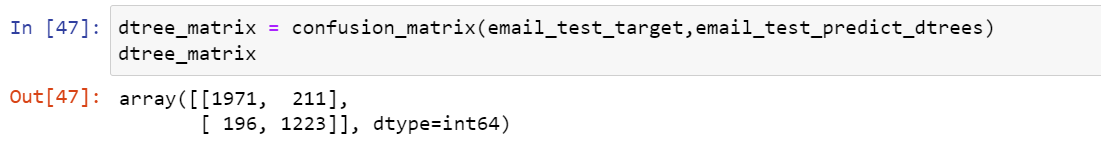
**Question 2**

The decision Tree classifier has classification accuracy of 0.8869758400444321 Random Forest classifier has accuracy has 0.9355734518189391 with the best parameters set for it.



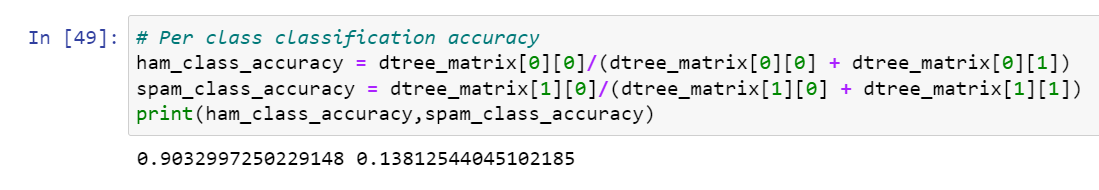


Confusion matrix for Decision Tree classifier is:

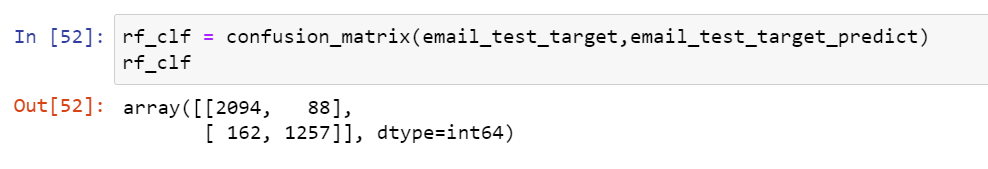


Per class classification accuracy is

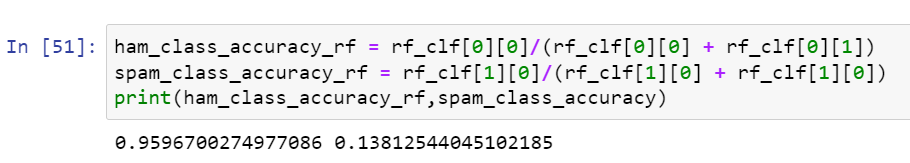
(Class ham ,Class spam)



Confusion matrix for Random Forest classifier is:



Per class classification accuracy



In summary if we see the per class classification accuracy then random forest classifier is best classifier because its classification accuracy for correctly classifying ham class is 95% where as in decision Tree classification the accuracy is 90% which is 5% less.