Question 6-

a)

Performance of Random Forest made from scratch

Parameters :

- num_of_trees=30
- num_of_features=15
- **Accuracy - 93.33 %**
- **Time taken for execution- 28.874 sec**

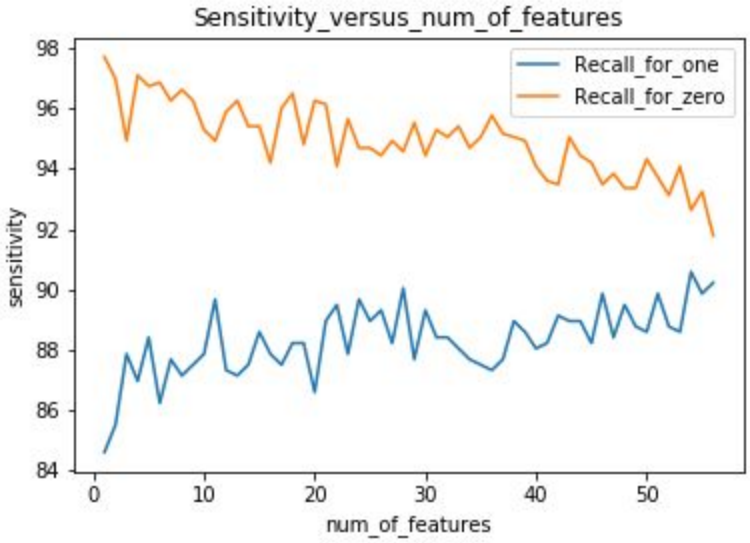
Performance of Random Forest made by scikit learn

Parameters :

- num_of_trees=30
- num_of_features=15
- **Accuracy - 94.92 %**
- **Time taken for execution- 0.904 sec**

b)

Sensitivity of Random forest to the parameter m (the number of features).



- X-axis has num_of_features and Y-axis has sensitivity (Recall)

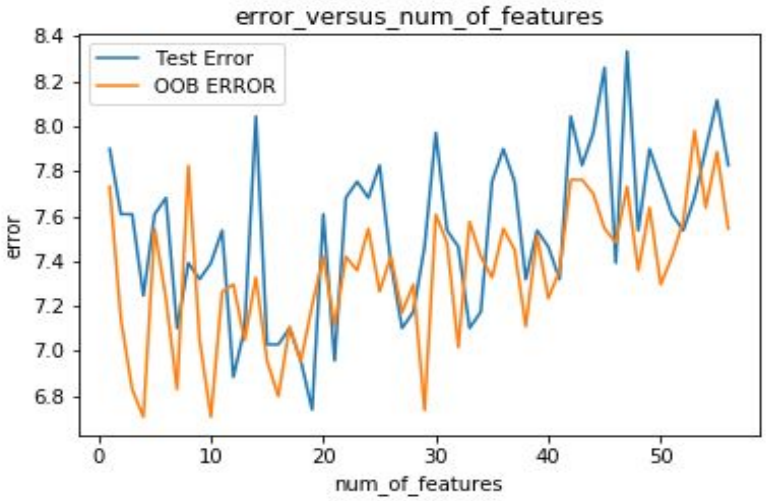
$$\text{Recall} = \frac{tp}{tp + fn}$$

- This graph is for 35 trees at each features from (1 to 56), blue line is for recall for one and orange line shows recall for zero
- Recall for zero decreases as the number of features increases

- As the data is biased towards zero i.e large number of zeroes and less number of one in label so, it may be the case that it has less number of rows to train the model for label 1,so that is why it is giving less recall for 1 as compared to 0.

c)

Plot for OOB error and test error against num_of_features



- X-axis has num_of_features and Y-axis contain error in percent
- Blue line is showing test error and orange line is OOB error
- As we know that the difference between OOB error and test error must be less and my graph is also showing the same behaviour they both are almost same