Assignment -3
Cs19btech11020

Q4.

a) Random forest :
hyperparameters:(`n_trees=8,n_bootstrap=800,n_features=10, max_depth=10`)

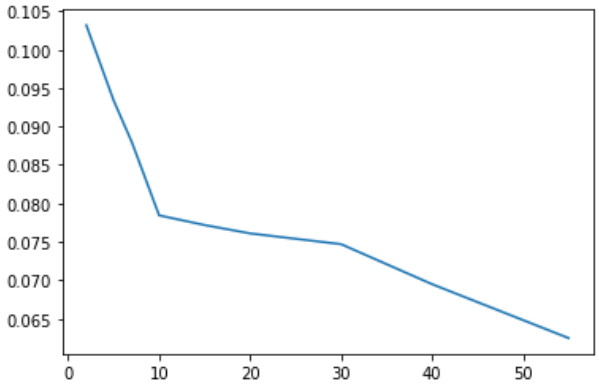
Accuracy of my random forest: 92.35667752442996
Accuracy of inbuilt classifier: 92.72529858849077

Time taken on my random forest: 9s
Time taken on inbuilt classifier : 1s

b) Sensitivity of random forest to parameter m(number of features of best split):
Accuracy increases as we increase the number of features.

| No of features | Accuracy |
|---|---|
| 2 | 89.68512486427795 |
| 5 | 90.66232356134635 |
| 7 | 91.20521172638436 |
| 10 | 92.15667752442996 |
| 15 | 92.28241042345277 |
| 20 | 92.39098805646036 |
| 30 | 92.53094462540716 |
| 40 | 93.05103148751357 |

c) Plot: m vs test error, test error decreases as we increase m(number of features)



Q5.

a) Preprocessing steps:
   - Dropped column with all null values and single values.
   - Removed rows with loan status as current
   - Dropped column with useless information such as id,member id, address, url, description
   - Dropped columns with more than 20% missing values
   - Removed rows with null values
   - Dropped column containing less correlation with target variable
   - Dropped columns with least feature importance
   - Transformed categorical data into binary features, and any other relevant columns to suitable data types.

b) Best accuracy=0.994214748534403
   - Hyperparameters
```
(n_estimators=400,
            learning_rate=0.2,
            criterion='friedman_mse',
            max_depth=4,
            random_state=10,
            max_features='auto')
```

| Model | precision | recall |
|---|---|---|
| n_estimators=200, learning_rate=0.2, criterion='mse',max_depth=4,max_features='auto' | 0.9909781152300134 | 0.9997296566639632 |
| (n_estimators=300, learning_rate=0.2, criterion='mse',max_depth=4, max_features='auto' | 0.9924859110832811 | 0.9998197711093089 |
| n_estimators=300,learning_rate=0.2, criterion='friedman_mse', max_depth=4, max_features='auto' | 0.99394090499445 | 0.9990567439848 |
| n_estimators=400,learning_rate=0.2, criterion='friedman_mse', max_depth=4, max_features='auto' | 0.9933751119068934 | 0.9999098855546544 |

   - Increasing number of trees in the classifier accuracy increases

| | Accuracy | Precision | Recall |
|---|---|---|---|
| My best model | 0.994 | 0.993 | 0.999 |
| Inbuilt classifier | 0.928 | 0.923 | 0.999 |