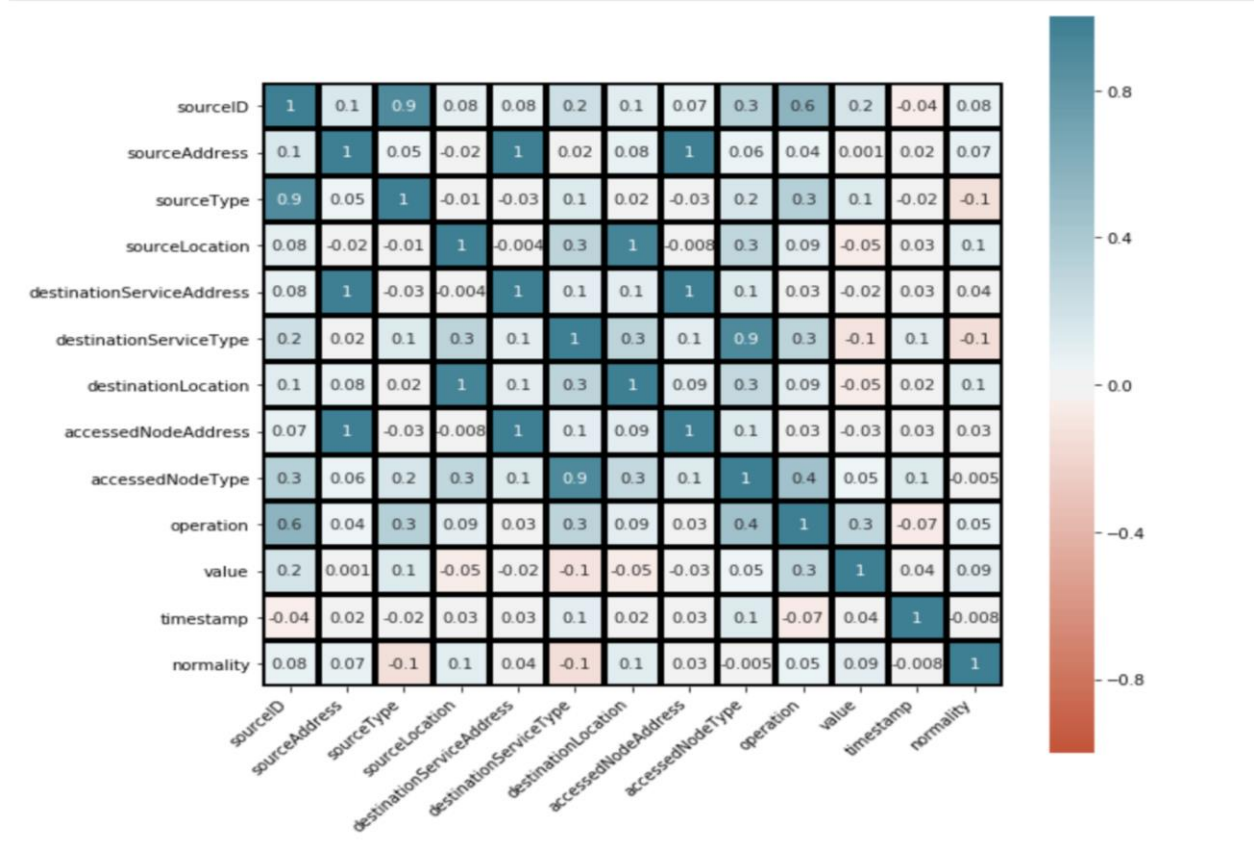


1. MULTICOLLINEARITY HEAT MAP



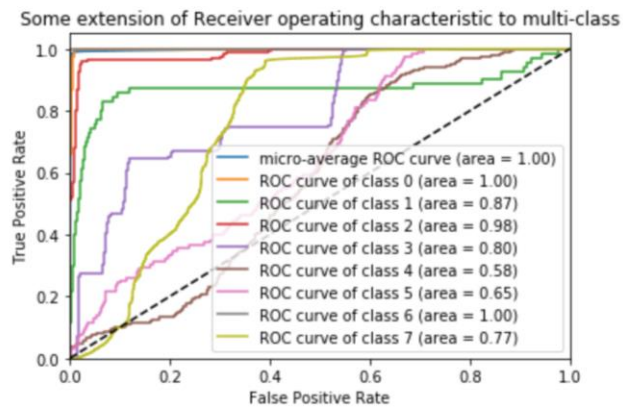
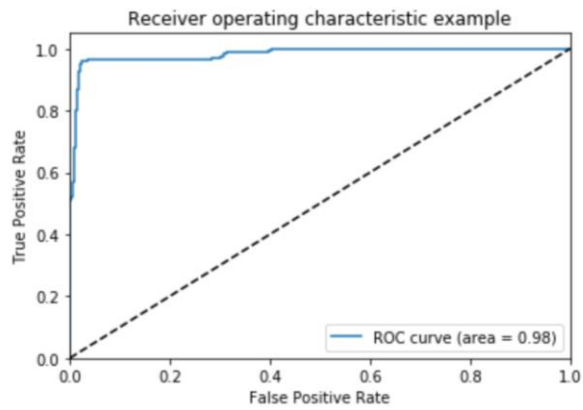
CLASSIFICATION REPORT AND CONFUSION MATRIX FOR LOGISTIC REGRESSION

	precision	recall	f1-score	support
0	0.00	0.00	0.00	1159
1	0.00	0.00	0.00	75
2	0.00	0.00	0.00	173
3	0.00	0.00	0.00	166
4	0.00	0.00	0.00	328
5	0.00	0.00	0.00	95
6	0.00	0.00	0.00	23
7	0.97	1.00	0.99	69572
accuracy			0.97	71591
macro avg	0.12	0.12	0.12	71591
weighted avg	0.95	0.97	0.96	71591

C:\Users\ranka\Anaconda3\lib\site-packages\sklearn\metrics\classification\_report.py:10: FutureWarning: The 'precision' parameter is deprecated and will be removed in a future version. Please use 'precision' instead.

```
cm = confusion_matrix(y_test,y_preds)
cm
array([[ 0,  0,  0,  0,  0,  0,  0, 1159],
       [ 0,  0,  0,  0,  0,  0,  0,  75],
       [ 0,  0,  0,  0,  0,  0,  0, 173],
       [28,  0,  0,  0,  0,  0,  0, 138],
       [ 3,  0,  0,  0,  0,  0,  0, 324],
       [18,  0,  0,  0,  0,  0,  0,  77],
       [ 0,  0,  0,  0,  0,  0,  0,  23],
       [19,  0,  0,  0,  0,  0,  0, 69553]],
      dtype=int64)
```

## ROC CURVE FOR LOGISITC REGRESSION:



## LEARNING CURVE FOR LOGISTIC REGRESSION



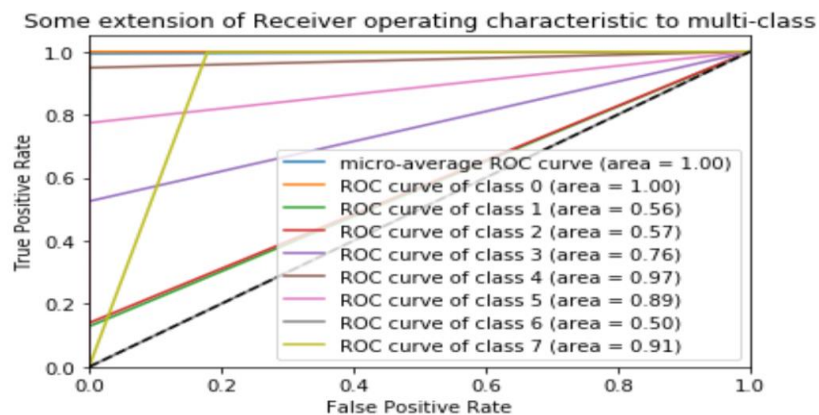
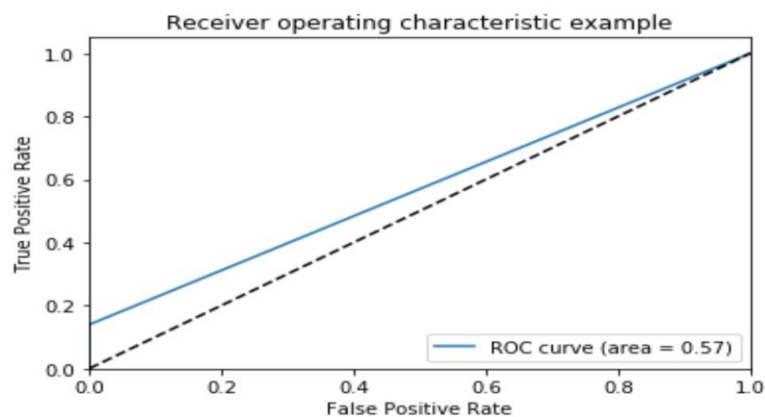
## CLASSIFICATION REPORT AND CONFUSION MATRIX FOR KNEIGHBORS CLASSIFIER

	precision	recall	f1-score	support
0	0.99	1.00	1.00	1098
1	0.39	0.14	0.21	76
2	0.86	0.13	0.23	193
3	0.88	0.47	0.61	159
4	0.91	0.91	0.91	325
5	0.90	0.74	0.81	115
6	0.00	0.00	0.00	31
7	0.99	1.00	1.00	69594
accuracy			0.99	71591
macro avg	0.74	0.55	0.60	71591
weighted avg	0.99	0.99	0.99	71591

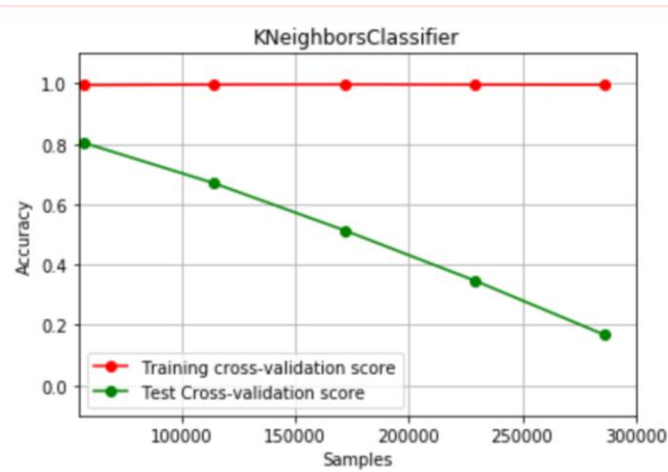
C:\Users\ranka\Anaconda3\lib\site-packages\sklearn\metrics\classification\_report.py:131: UserWarning: Some labels are ill-defined and being set to 0.0 in labels with no predicted samples. Use 'precision', 'predicted', average, warn\_for) to control the behavior.

```
cm1 = confusion_matrix(y_test,y_preds1)
cm1
array([[ 1098,    0,    0,    0,    0,    0,    0,    0],
       [    0,   11,    0,    0,    0,    0,    0,    0],
       [    0,    0,   25,    0,    0,    0,    0,   65],
       [    0,    0,    0,   74,    0,    0,    0,   85],
       [    0,    0,    0,    0,  296,    0,    0,   29],
       [    0,    0,    0,    0,    0,   85,    0,   30],
       [    0,    0,    0,    0,    0,    0,    0,   31],
       [    8,   17,    4,   10,   30,    9,    0, 69516]],
      dtype=int64)
```

## ROC CURVE FOR KNEIGHBORS CLASSIFIER



LEARNING CURVE OF KNEIGHBORS CLASSIFIER RESULTING IN OVERFITTING



LEARNING CURVE OF KNEIGHBORS CLASSIFIER AFTER SOLVING OVERFITTING BY ELIMINATING FEATURE VARIABLE TIMESTAMP

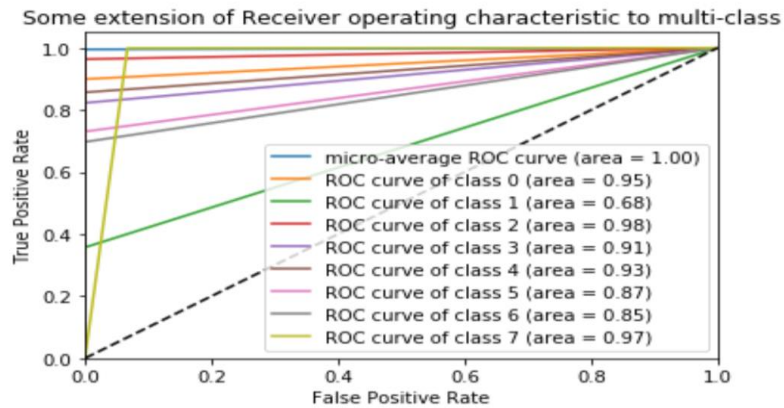
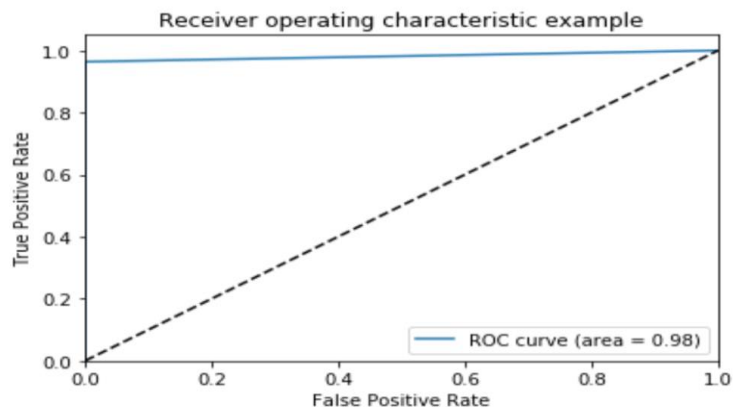


CLASSIFICATION REPORT AND CONFUSION MATRIX FOR RANDOM FOREST CLASSIFIER

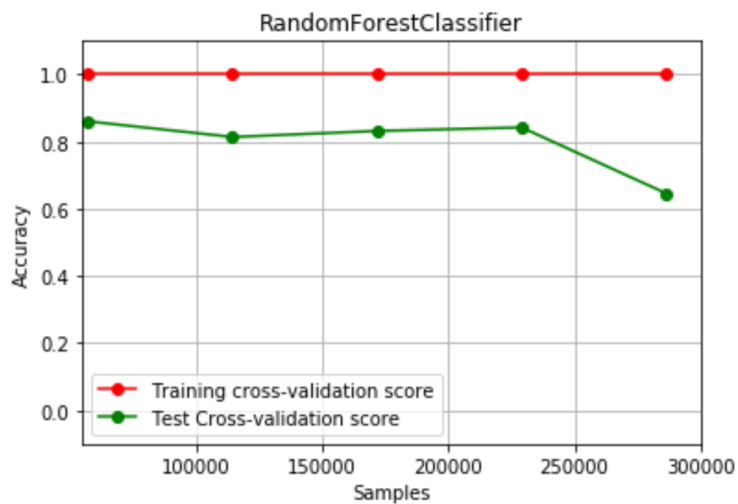
	precision	recall	f1-score	support
0	1.00	1.00	1.00	1109
1	1.00	1.00	1.00	64
2	1.00	1.00	1.00	173
3	1.00	1.00	1.00	155
4	1.00	1.00	1.00	308
5	1.00	1.00	1.00	85
6	1.00	1.00	1.00	24
7	1.00	1.00	1.00	69673
accuracy			1.00	71591
macro avg	1.00	1.00	1.00	71591
weighted avg	1.00	1.00	1.00	71591

```
#Confusion Matrix
cm2 = confusion_matrix(y_test,y_preds2)
cm2
array([[ 1109,    0,    0,    0,    0,    0,    0,    0],
       [    0,    64,    0,    0,    0,    0,    0,    0],
       [    0,    0,   173,    0,    0,    0,    0,    0],
       [    0,    0,    0,   155,    0,    0,    0,    0],
       [    0,    0,    0,    0,   308,    0,    0,    0],
       [    0,    0,    0,    0,    0,   85,    0,    0],
       [    0,    0,    0,    0,    0,    0,   24,    0],
       [    0,    0,    0,    0,    0,    0,    0, 69673]],
      dtype=int64)
```

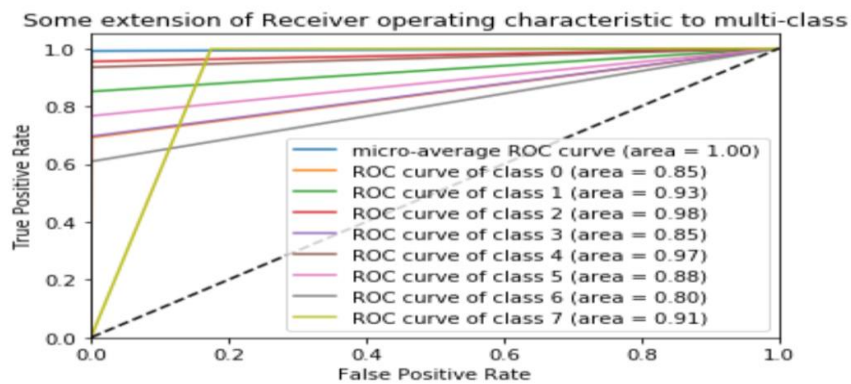
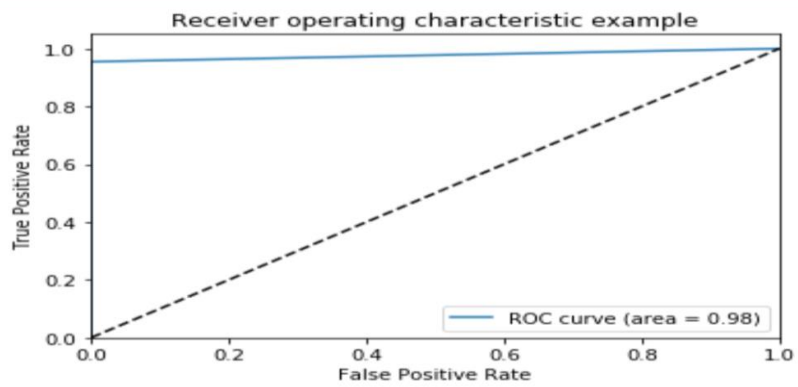
## ROC CURVE FOR RANDOM FOREST CLASSIFIER



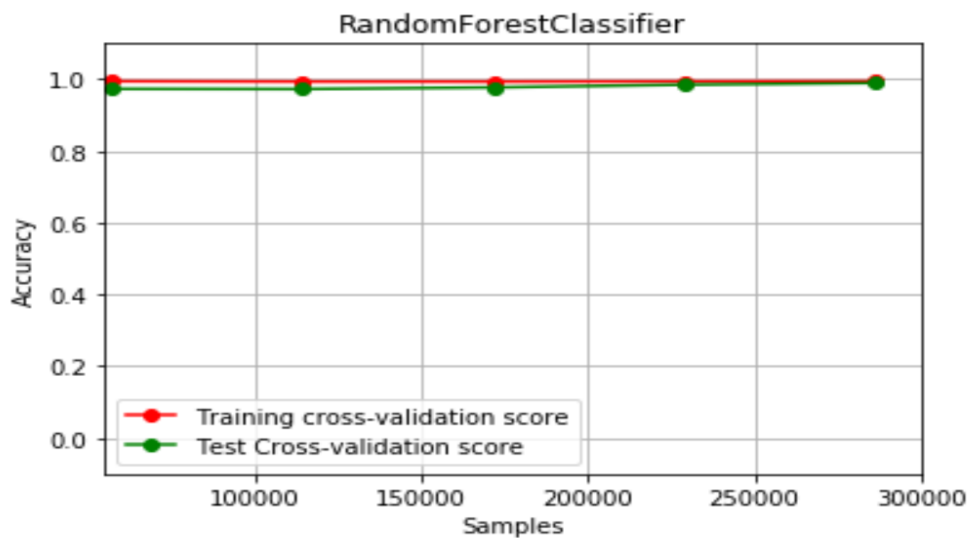
LEARNING CURVE FOR RANDOM FOREST CLASSIFIER WHICH RESULTS IN OVERFITTING AS TEST-CROSS VALIDATION SCORE IS VERY LESS AS COMPARED TO TRAINING CROSS VALIDATION SCORE



## IMPROVED ROC CURVE AFTER ELIMINATING FEATURE VARIABLE TIMESTAMP



## LEARNING CURVE OF RANDOM FOREST CLASSIFIER AFTER SOLVING OVERFITTING BY ELIMINATING TIMESTAMP FEATURE VARIABLE



## CLASSIFICATION REPORT AND CONFUSION MATRIX OF DECISION TREE CLASSIFIER

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1162
1	1.00	1.00	1.00	73
2	1.00	0.99	0.99	168
3	0.99	0.98	0.99	165
4	1.00	1.00	1.00	294
5	1.00	1.00	1.00	96
6	1.00	1.00	1.00	21
7	1.00	1.00	1.00	69612
accuracy			1.00	71591
macro avg	1.00	1.00	1.00	71591
weighted avg	1.00	1.00	1.00	71591

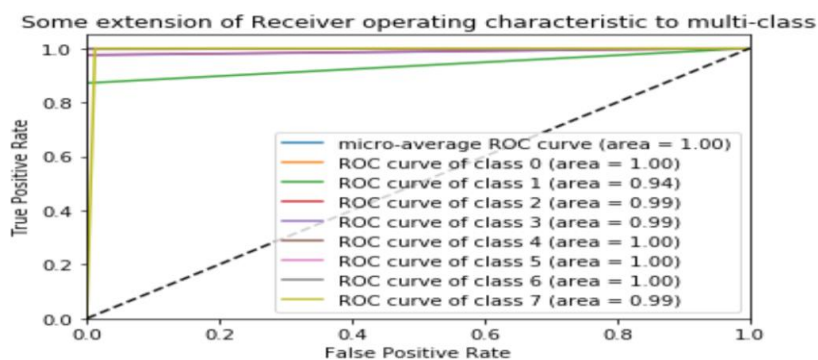
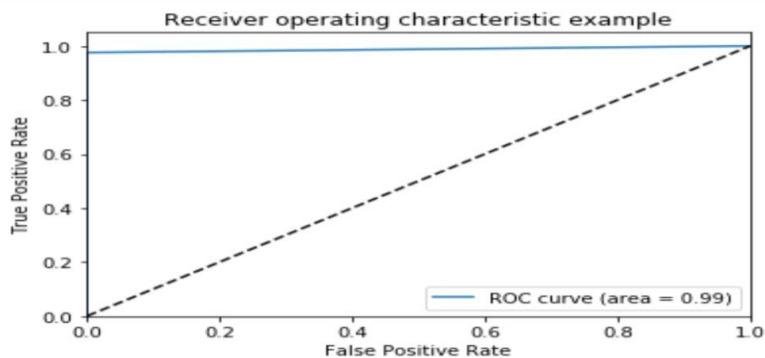
```

: #Confusion Matrix
cm3 = confusion_matrix(y_test,y_preds3)
cm3

: array([[ 1162,    0,    0,    0,    0,    0,    0,    0],
       [    0,    73,    0,    0,    0,    0,    0,    0],
       [    0,    0,   166,    0,    0,    0,    0,    2],
       [    0,    0,    0,   162,    0,    0,    0,    3],
       [    0,    0,    0,    0,   293,    0,    0,    1],
       [    0,    0,    0,    0,    0,   96,    0,    0],
       [    0,    0,    0,    0,    0,    0,   21,    0],
       [    0,    0,    0,    1,    0,    0,    0, 69611]],
      dtype=int64)

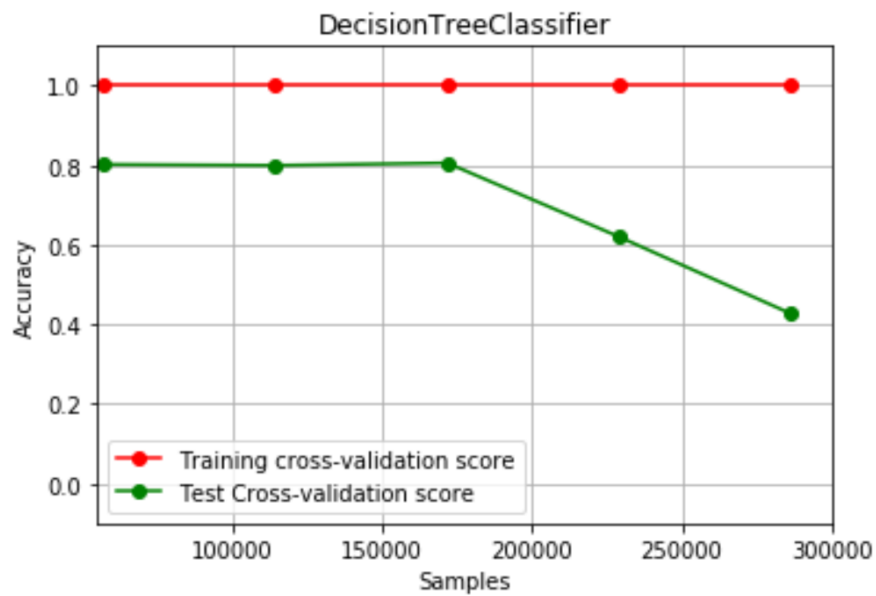
```

## ROC CURVE FOR DECISION TREE CLASSIFIER





LEARNING CURVE ORF DECISION TREE CLASSIFIER RESULTING IN OVERFITTING AS AGAIN TEST CROSS VALIDATION SCORE IS MORE LOWER THAN TRAINING CROSS VALIDATION SCORE



LEARNING CURVE OF DECISION TREE CLASSIFIER AFTER SOLVING OVERFITTING BY ELIMINATING FEATURE VARIABLE TIMESTAMP

