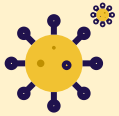
CATCH ME IF YOU CAN

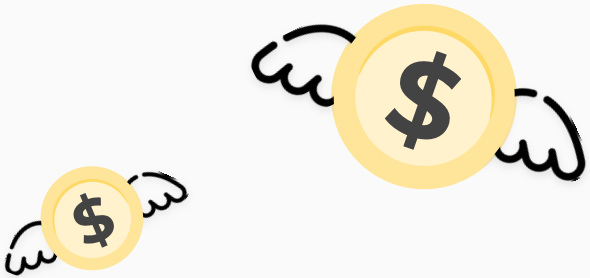
Jenny Wang | May 5, 2020



COVID-19

NO CASH

DEFAULT



Can we predict potential default accounts?

APPROACH

01
DATA COLLECTION

UCI Machine Learning Datasets

02
DATA SCRUBBING

03
DATA EXPLORING

04
DATA MODELING

DATA

CREDIT CARD CLIENTS IN TAIWAN

30,000 OBSERVATIONS

24 FEATURES

ORIGINAL FEATURE CATEGORIES

CREDIT INFO

Credit Line

DEMOGRAPHICS

Gender, Education, Age, Marital Status

PAYMENT HISTORY

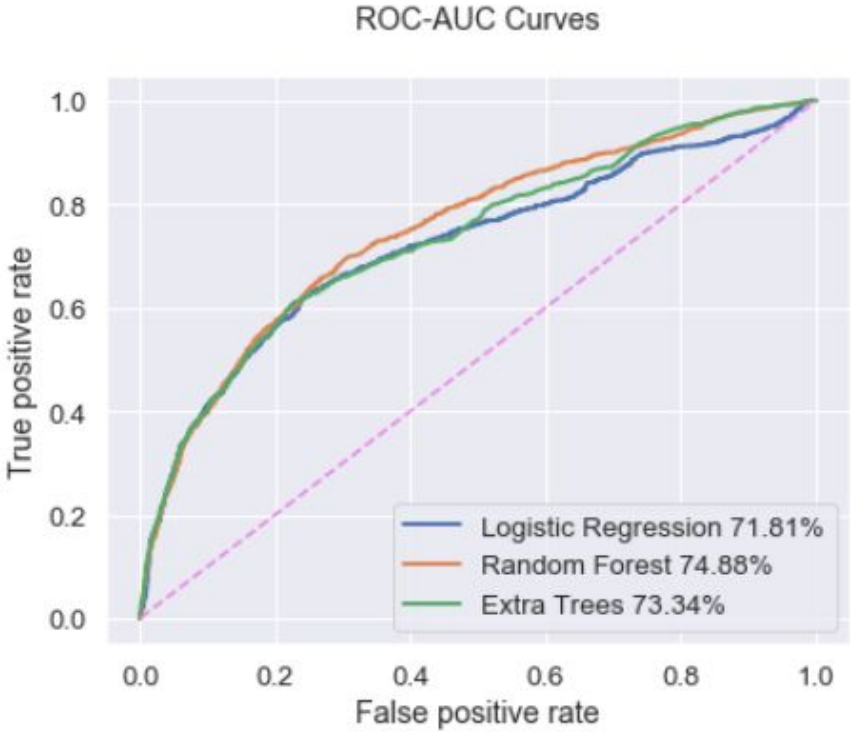
Repayment Status, Payments, & Bill Statements from Apr ~ Sep 2005

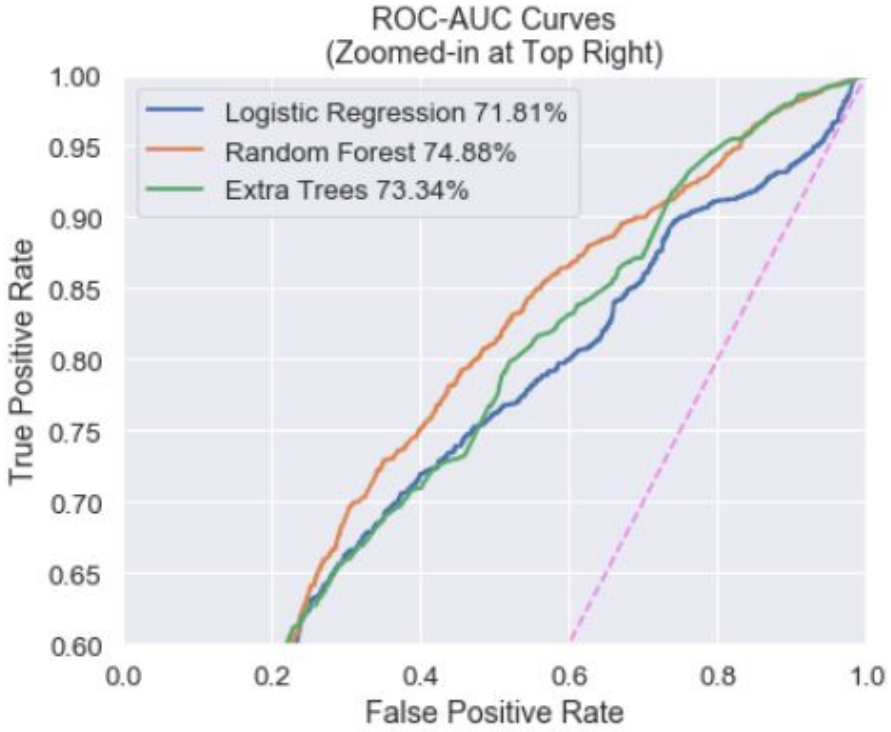
TARGET

For next month, will the credit card client...

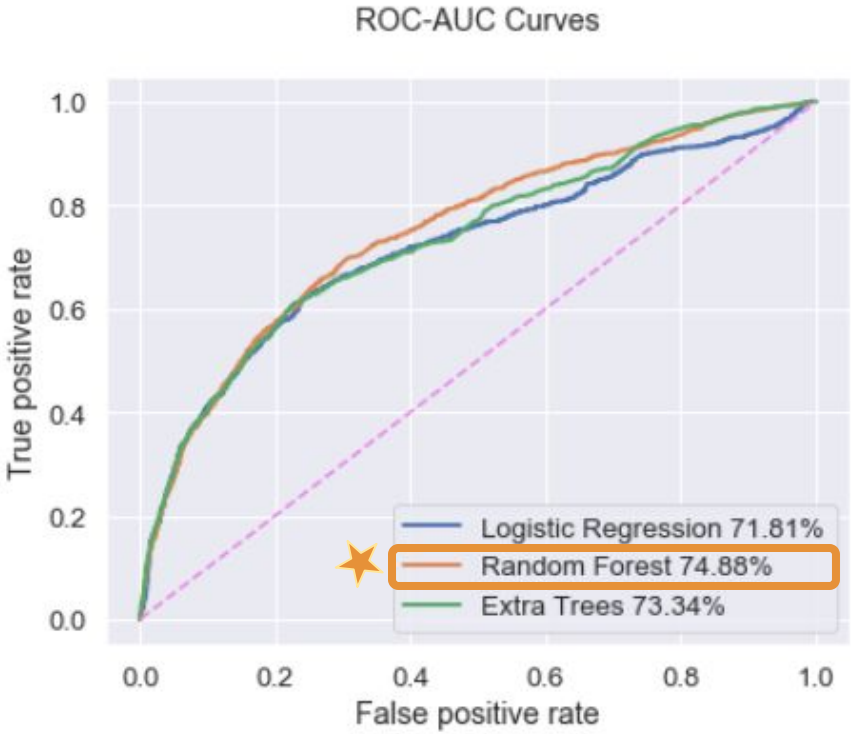
DEFAULT vs. NOT DEFAULT

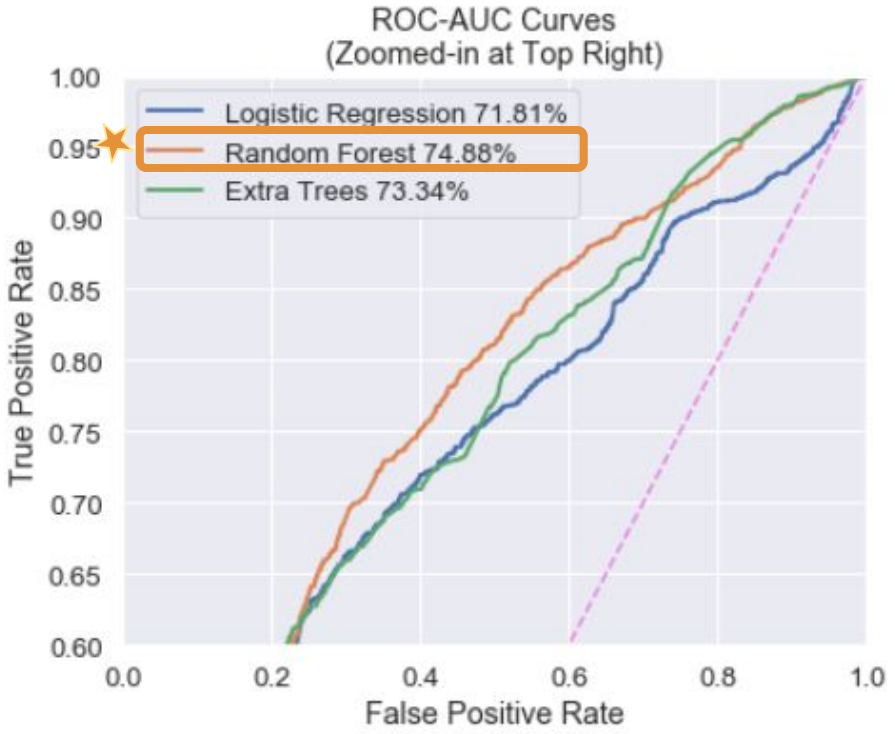
ROC - AUC CURVES FROM TOP 3 MODELS



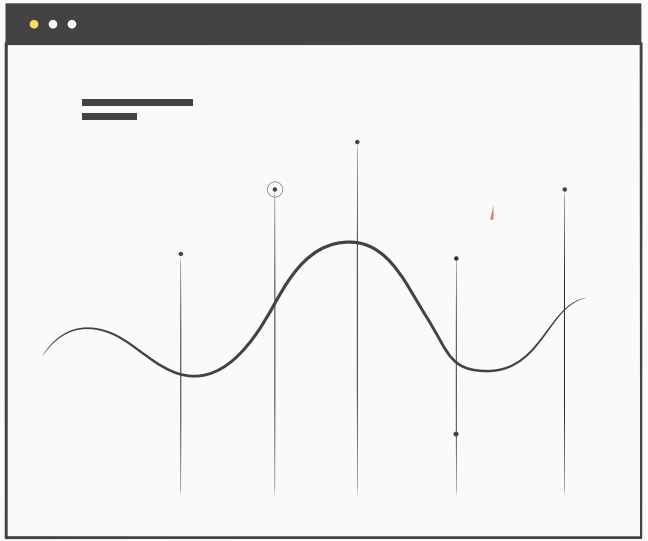


ROC - AUC CURVES FROM TOP 3 MODELS

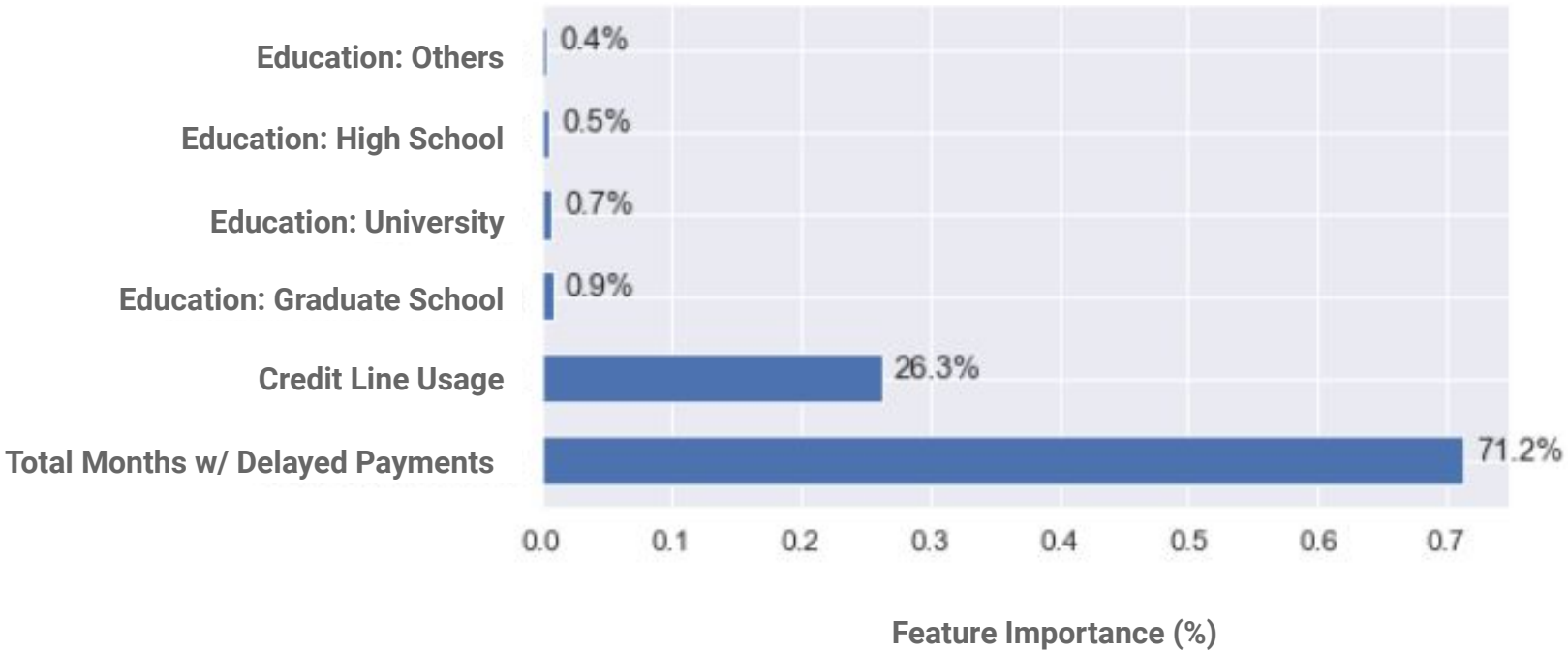




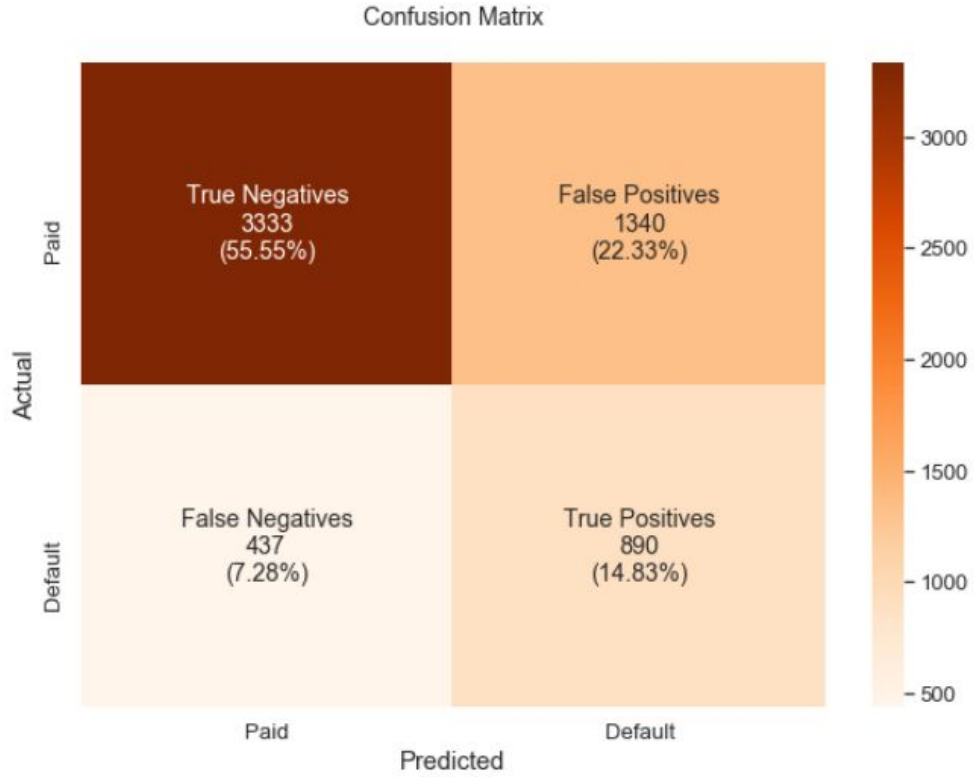
RANDOM FOREST



RANDOM FOREST: FEATURE IMPORTANCE



RANDOM FOREST: CONFUSION MATRIX



67.07%

Recall Score

FUTURE WORK

1) Obtain different datasets
EX: length of credit history, history of new credits, types of credit in use, etc.

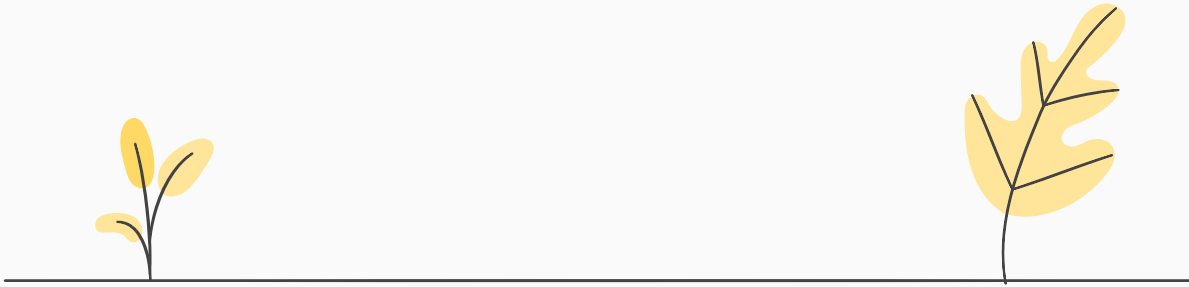
2) Gather more recent data
The current data is from 2005

THANK YOU

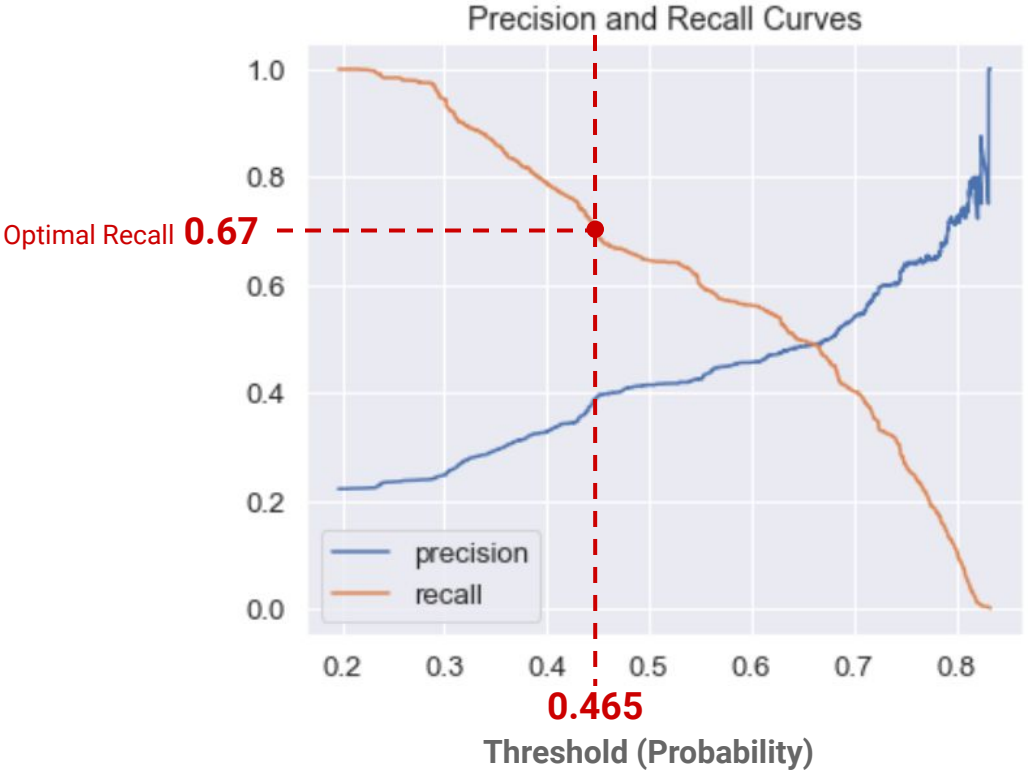
ms.jcwang@gmail.com
alohajenny.github.io

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

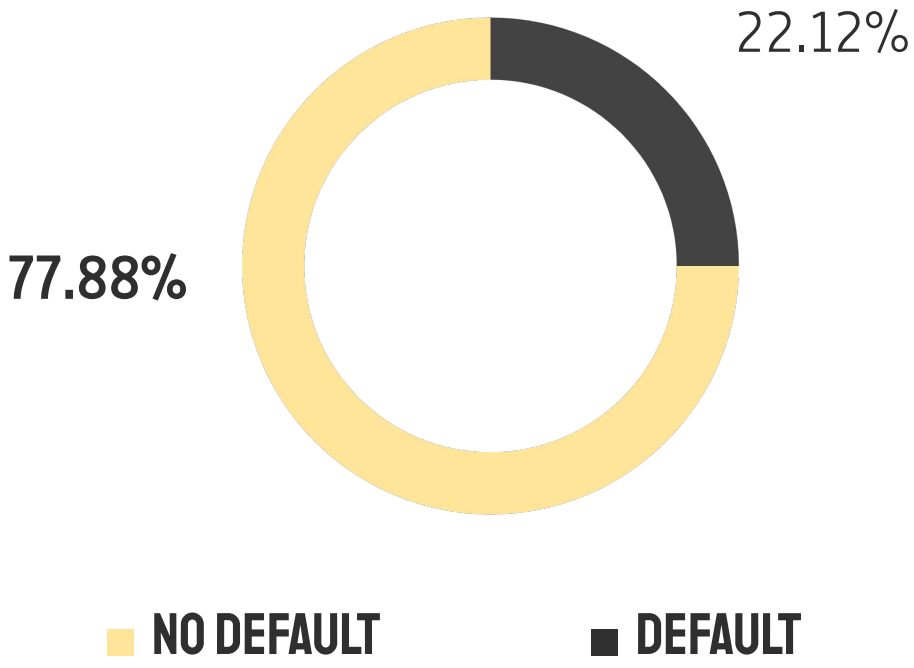
APPENDIX



RANDOM FOREST: PRECISION & RECALL CURVES

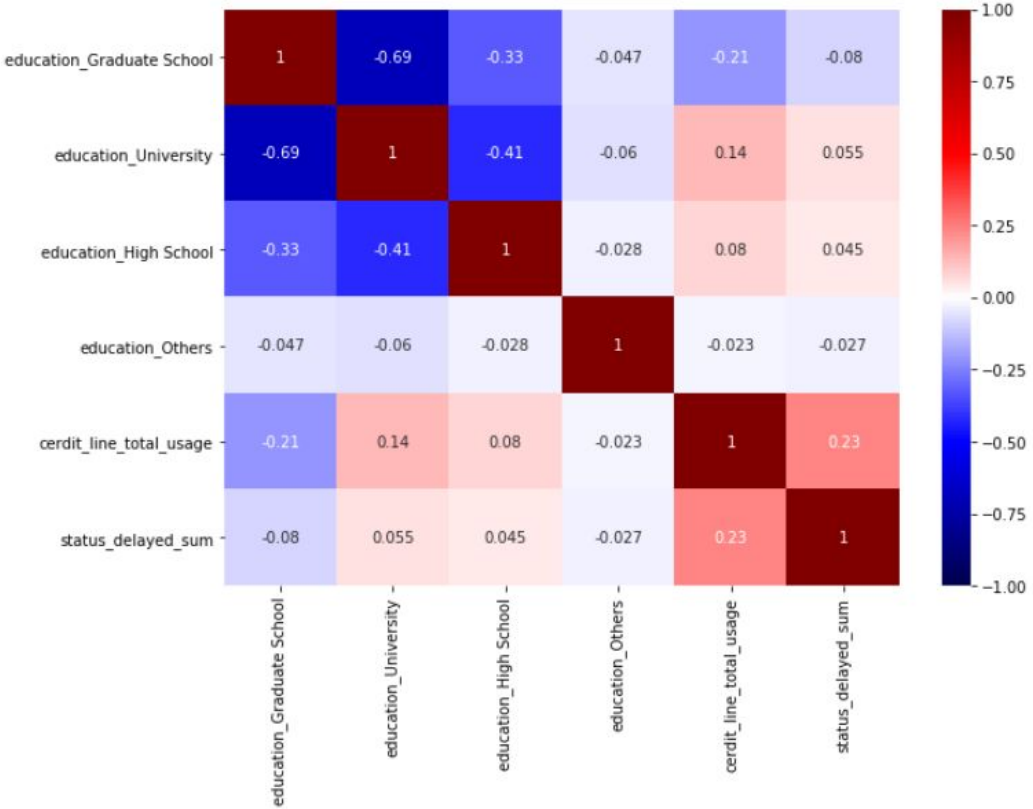


TARGET DISTRIBUTION



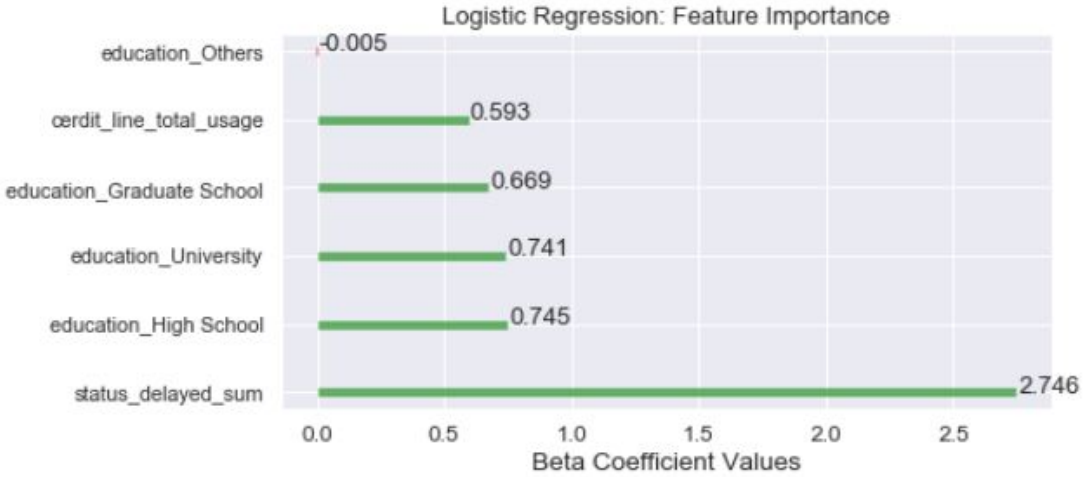
APPENDIX

CORRELATION HEATMAP (FEATURES)



APPENDIX

LOGISTIC REGRESSION

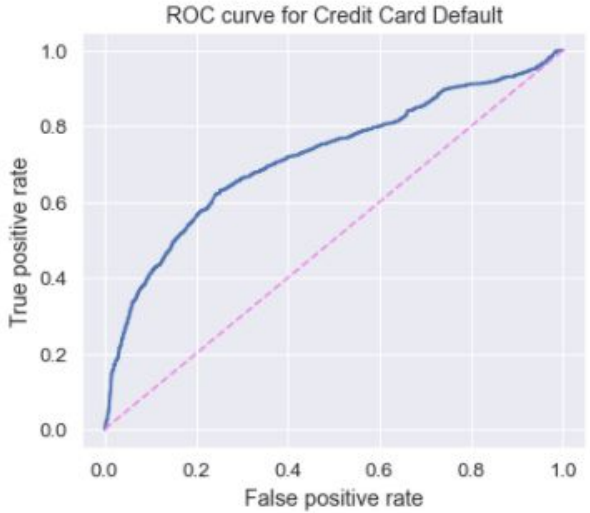


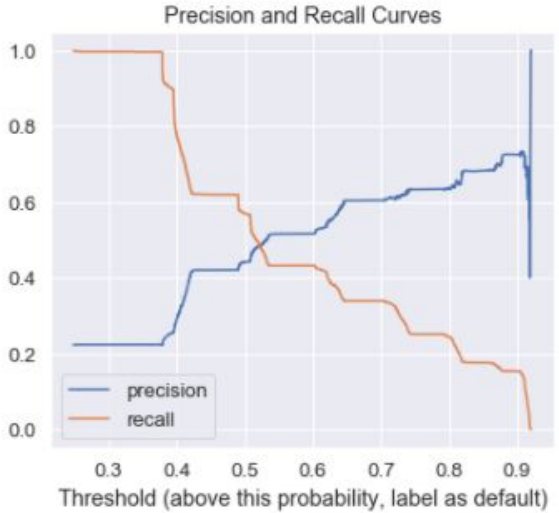

```
{'test_recall_mean': '65.53%',
 'test_recall_std': '1.19%',
 'test_precision_mean': '42.95%',
 'test_precision_std': '1.22%',
 'test_accuracy_mean': '73.1%',
 'test_accuracy_std': '0.92%',
 'test_balanced_accuracy_mean': '70.39%',
 'test_balanced_accuracy_std': '0.92%',
 'test_f1_mean': '51.88%',
 'test_f1_std': '1.18%',
 'test_roc_auc_mean': '72.94%',
 'test_roc_auc_std': '0.85%'}
```

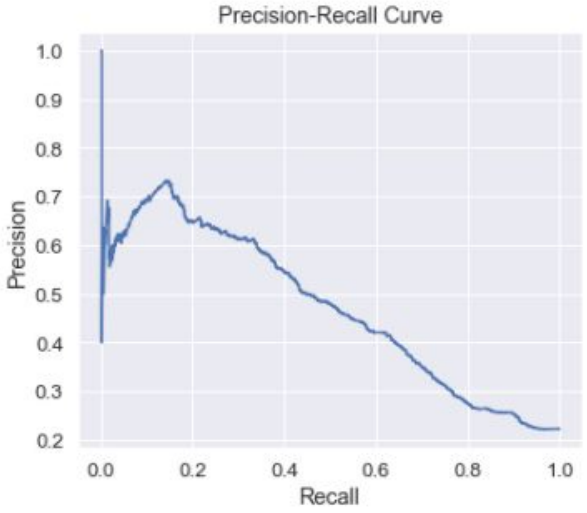
APPENDIX

LOGISTIC REGRESSION

```
ROC AUC score =  0.7180977286020431
```

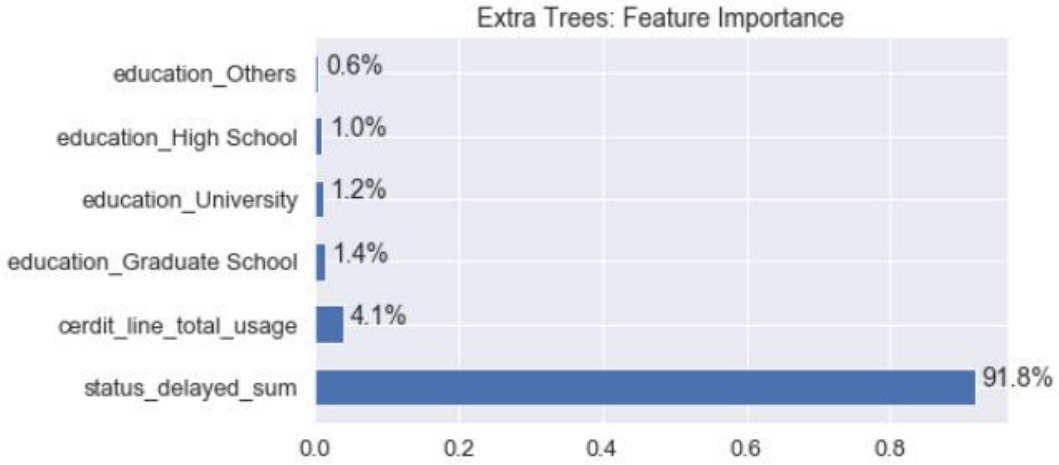






APPENDIX

EXTRA TREES

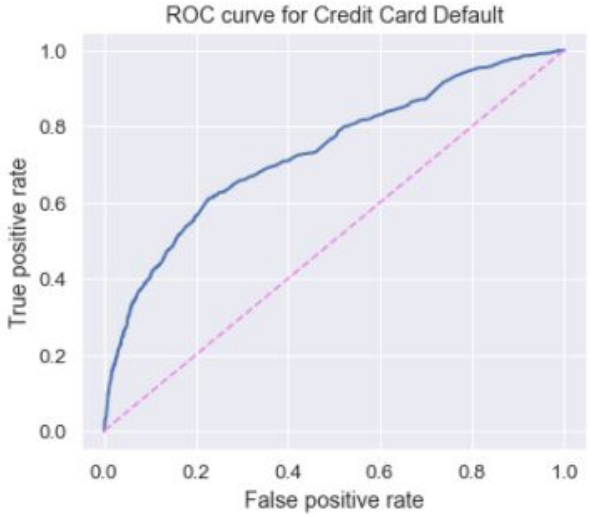


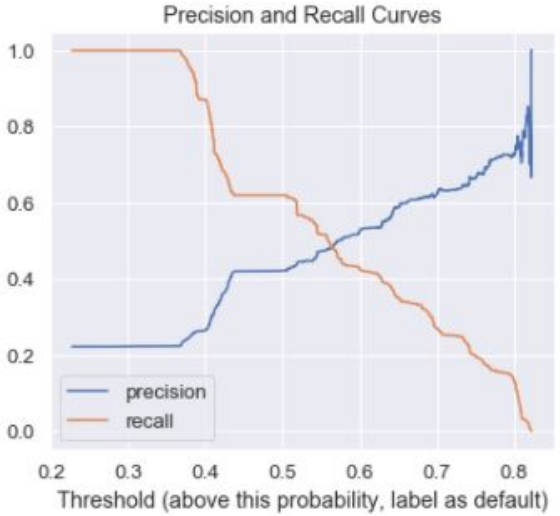

```
{'test_recall_mean': '65.46%',
 'test_recall_std': '1.2%',
 'test_precision_mean': '43.17%',
 'test_precision_std': '1.23%',
 'test_accuracy_mean': '73.28%',
 'test_accuracy_std': '0.92%',
 'test_balanced_accuracy_mean': '70.48%',
 'test_balanced_accuracy_std': '0.92%',
 'test_f1_mean': '52.02%',
 'test_f1_std': '1.19%',
 'test_roc_auc_mean': '74.89%',
 'test_roc_auc_std': '1.12%'}
```

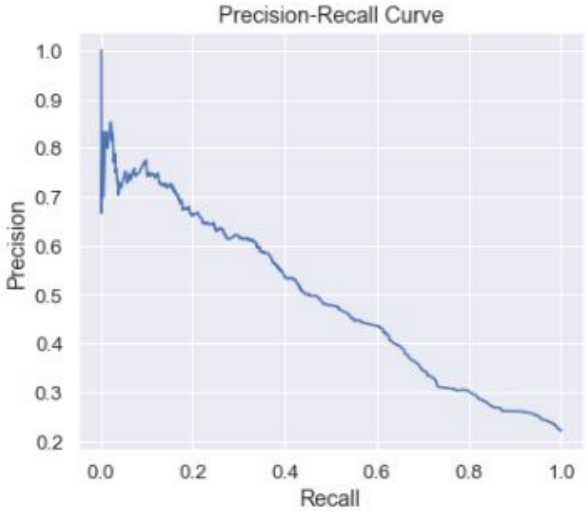
APPENDIX

EXTRA TREES

```
ROC AUC score =  0.7333726706241551
```







K-NEAREST NEIGHBOR

```
{'test_recall_mean': '60.48%',
 'test_recall_std': '0.71%',
 'test_precision_mean': '33.35%',
 'test_precision_std': '0.93%',
 'test_accuracy_mean': '64.49%',
 'test_accuracy_std': '1.07%',
 'test_balanced_accuracy_mean': '63.06%',
 'test_balanced_accuracy_std': '0.74%',
 'test_f1_mean': '42.99%',
 'test_f1_std': '0.81%',
 'test_roc_auc_mean': '66.77%',
 'test_roc_auc_std': '0.43%'}
```

LINEAR SVC

```
{'test_recall_mean': '53.1%',
 'test_recall_std': '2.06%',
 'test_precision_mean': '49.54%',
 'test_precision_std': '1.27%',
 'test_accuracy_mean': '77.65%',
 'test_accuracy_std': '0.62%',
 'test_balanced_accuracy_mean': '68.86%',
 'test_balanced_accuracy_std': '0.91%',
 'test_f1_mean': '51.24%',
 'test_f1_std': '1.28%',
 'test_roc_auc_mean': '73.74%',
 'test_roc_auc_std': '1.24%'}
```