Credit Card Fraud Detection

Mina Jambajantsan
&
Matthew Newton

Fraud Numbers

Frauds

- Credit Card
- Payment cards
- Remote banking
- Cheques

Remote purchase (card not present or CNP), counterfeit, lost and stolen, card not received and card ID theft

Numbers

Credit Cards: £671.4 M
Remote banking and cheques: £844.8 M

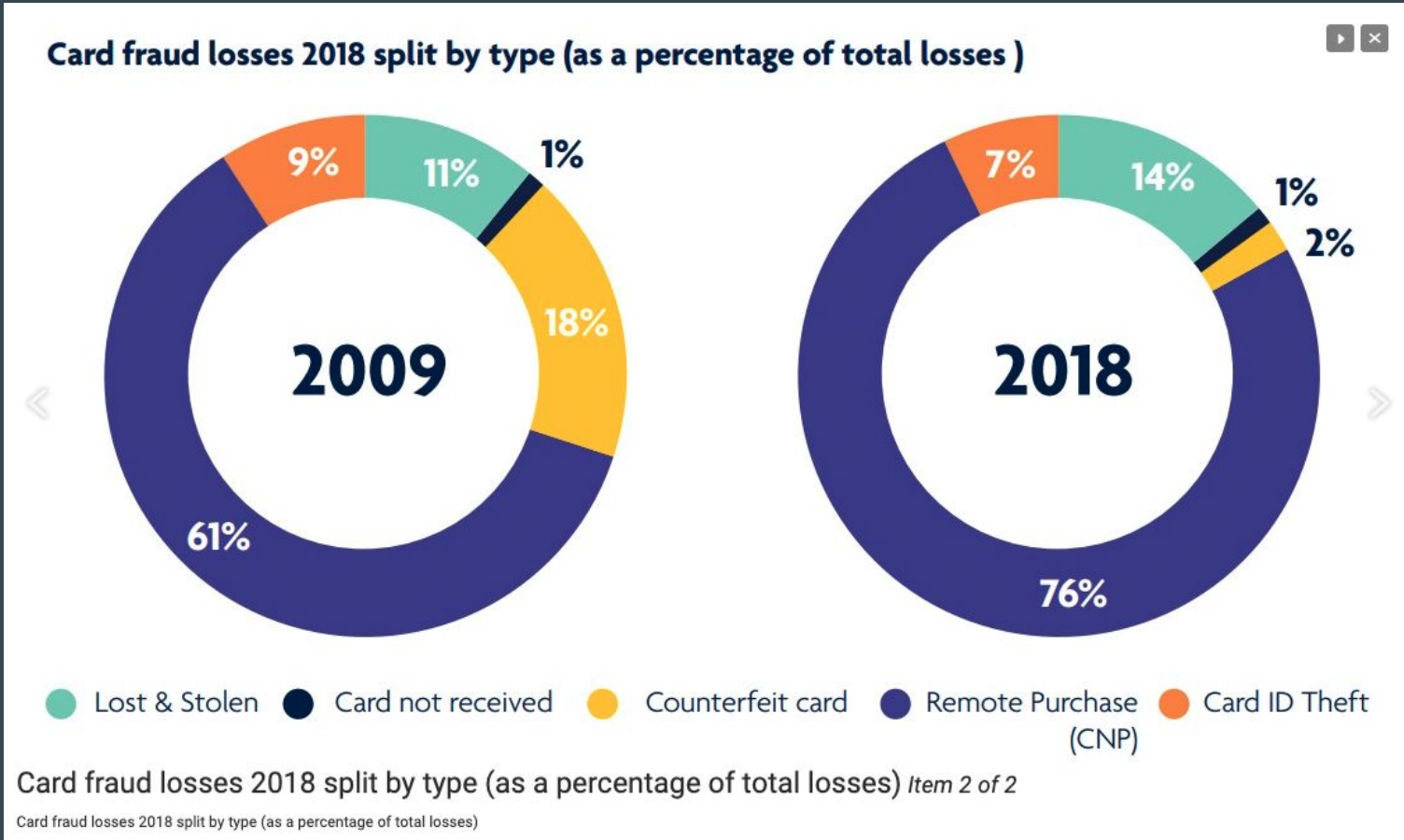
Prevented £1.66 billion

£6.27 in every £10 of attempted card fraud being prevented

Punishment

From low level community orders and fines to up to the equivalent of 150 % of your weekly income

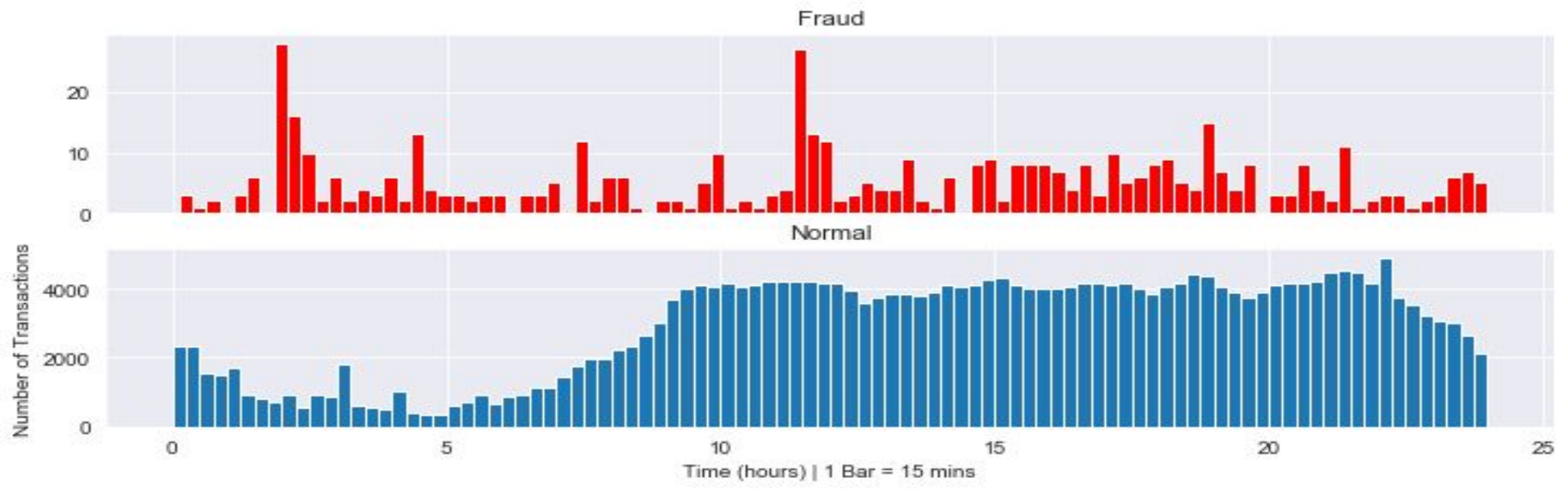
Lengthy prison sentences



Source: https://www.paymentscardsandmobile.com/uk-fraud-report-2019/

Our Data

- 2 days of September 2013
- 284 315 Legitimate transactions
- 492 Fraud transactions (0.172%)



Project Workflow:

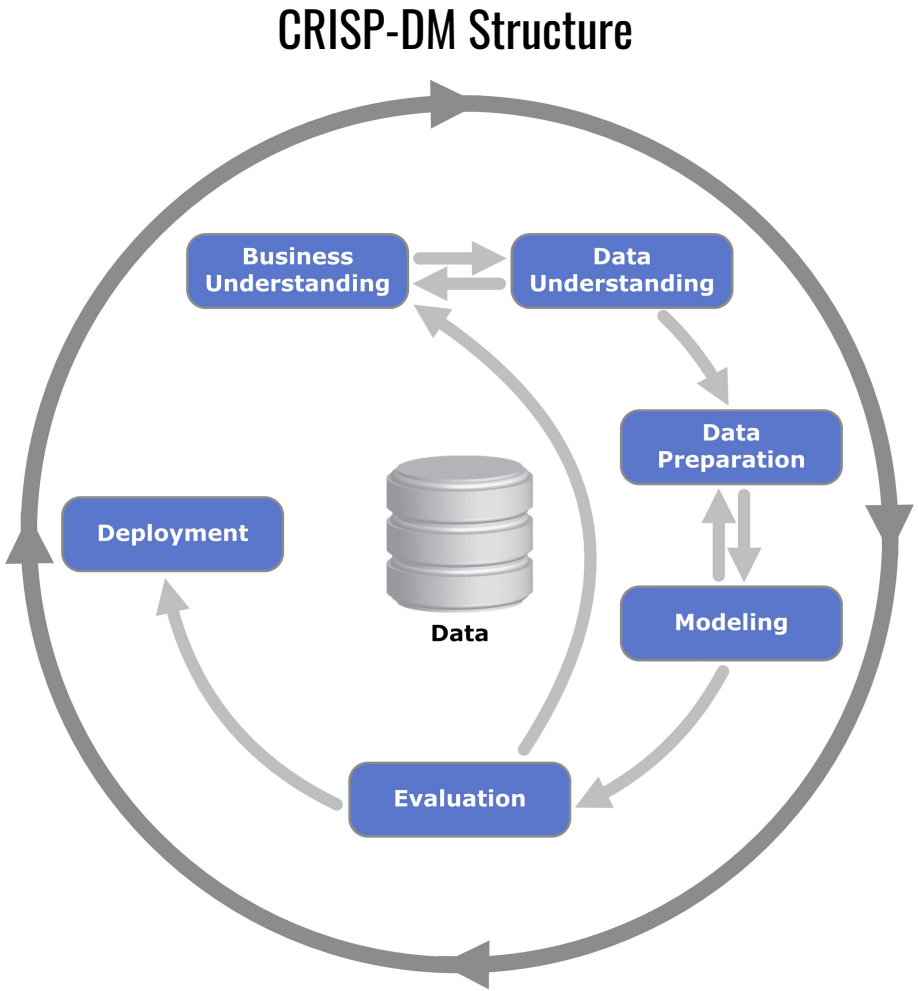
PCA Already Conducted

Resample Data (SMOTE)

Broad Search ML Models

Refined Search ML Models

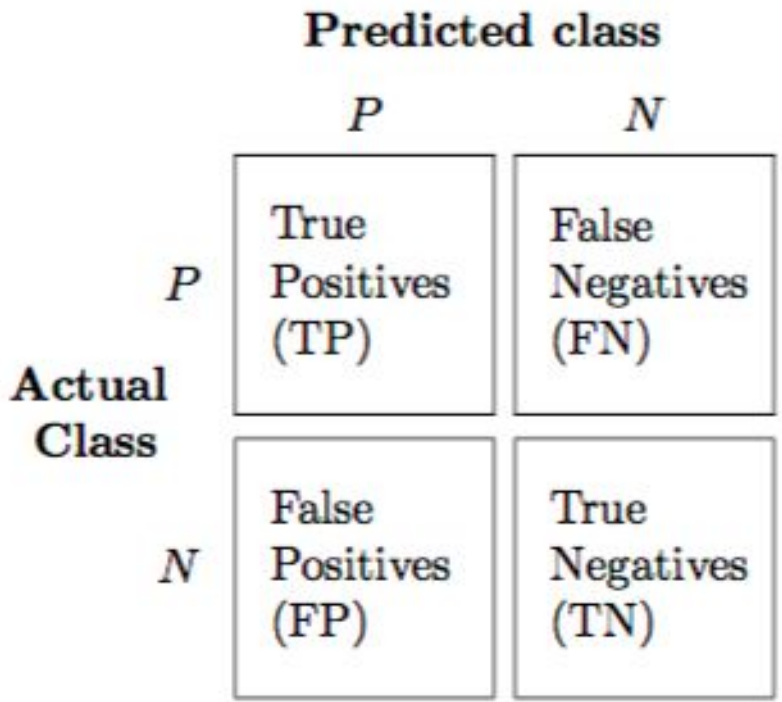
Final Model Evaluation Using Test Data



Model Selection Criteria:

1. Minimise Non-Detected Fraudulent Transaction - Type II error (FN)

2. Acceptable Level of Falsely Flagged Non-Fraudulent Transaction - Type I error (FP)

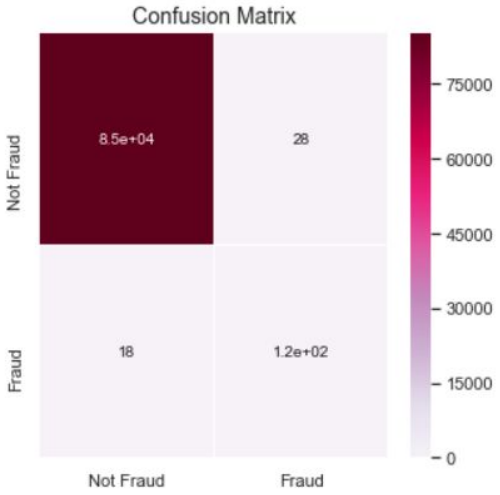


Random Forest

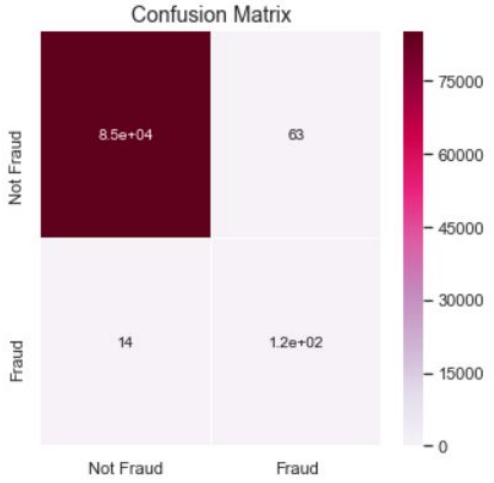
← Final Model

Adjusted Model →

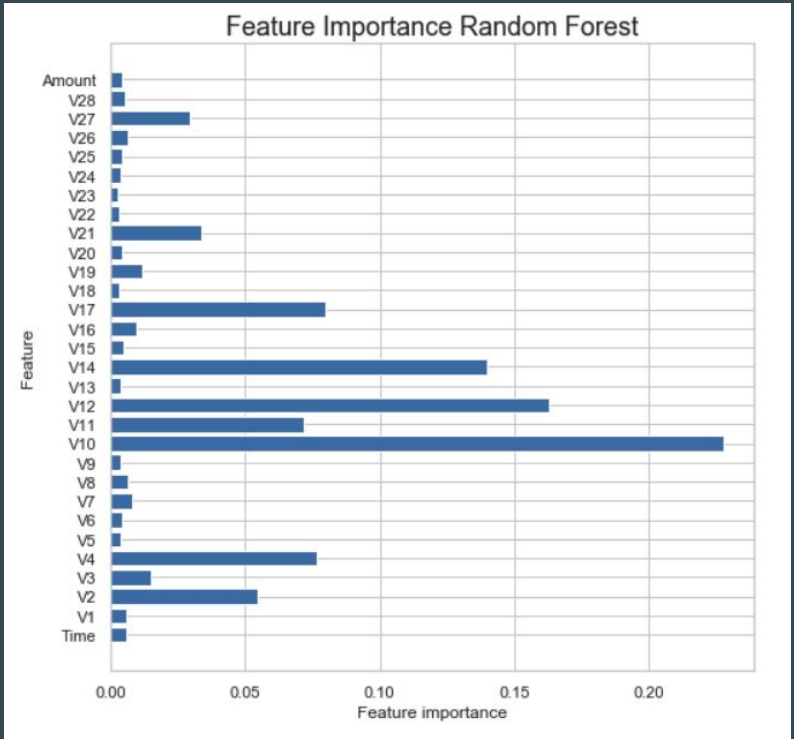
```
Recall: [0.99967177 0.86764706]
Precision: [0.99978897 0.80821918]

 clasification report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     85307
           1       0.81      0.87      0.84       136

avg / total       1.00      1.00      1.00     85443


 confussion matrix:
 [[85279    28]
 [   18   118]]

 Accuracy Percentage  is : 99.94616293903537%
```

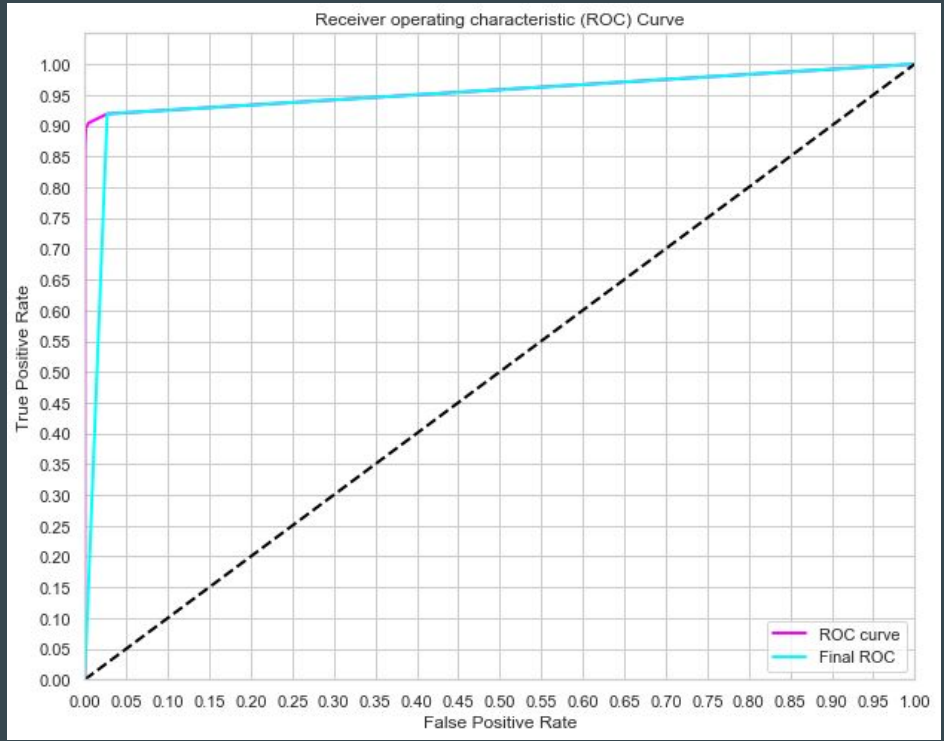



```
Recall: [0.99926149 0.89705882]
Precision: [0.99983579 0.65945946]

 clasification report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     85307
           1       0.66      0.90      0.76       136

avg / total       1.00      1.00      1.00     85443


 confussion matrix:
 [[85244    63]
 [   14   122]]

 Accuracy Percentage  is : 99.9098814414288%
```



Final Model - Recommendations





Thank you for your attention

Q & A

Models:

Undersampling	• Majority under-sampling • Faster, more balanced, but data loss
Oversampling	• Minority over-sampling • Duplicates minority, but may do overfitting
SMOTE	• Synthetic Minority Over-sampling • Creates points closer in feature space
Logistic Regression	• Baseline Model • Third Best

Models:

K Nearest Neighbors	• Simplistic and quick • Not accurate enough
XGBoost	• Second best • Fairly quick
Random Forest	• Best Model / Final Model • Used threshold of 0.7 at the Final Model
SVM	• Very slow • Works better on small samples

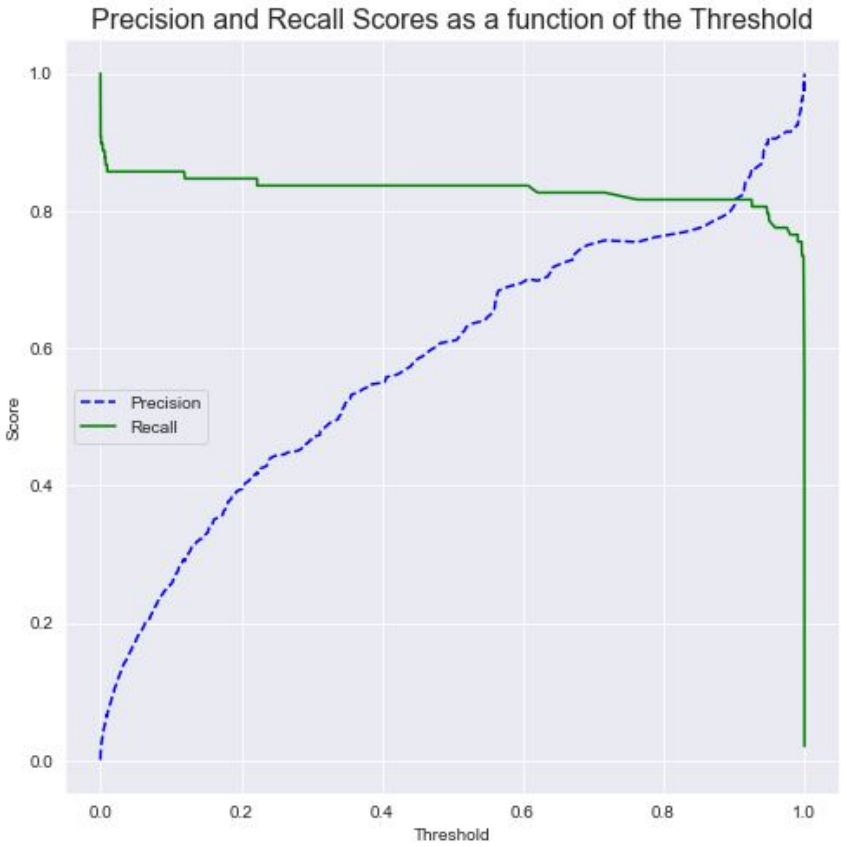
XGBoost Model - Threshold Selection - Validation Set

$$m = \frac{1 - Prevalence}{Prevalence} * \frac{Fraud\_Coef(FP - TN)}{Normal\_Coef(FN - TP)}$$

$$Prevalence = 0.00172$$
$$Fraud\_coef = Mean\ Fraud\ Transaction = 122.21$$
$$Normal\_coef = Mean\ Normal\ Transaction = 88.29$$

$$f_m = TPR - m(FPR)$$

Threshold	fm	FPR	TPR
0.996023	0.714915	0.000050	0.755102
0.998756	0.711094	0.000017	0.724490
0.998571	0.707902	0.000033	0.734694
0.998657	0.697698	0.000033	0.724490
0.996626	0.694507	0.000050	0.734694

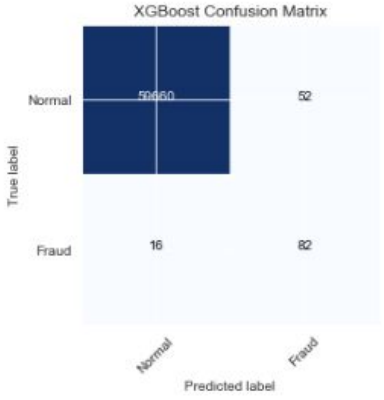


Threshold = 0.5

clasification report:

	precision	recall
0	1.00	1.00
1	0.61	0.84
accuracy		
macro avg	0.81	0.92
weighted avg	1.00	1.00

Frauds: 98 | Missed: 16.3%
Total Flagged: 134 | Real: 837.5%

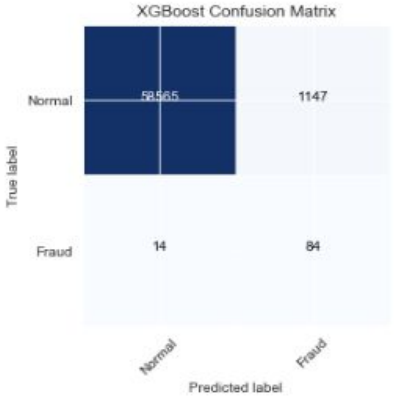


Threshold = 0.99

clasification report:

	precision	recall
0	1.00	0.98
1	0.07	0.86
accuracy		
macro avg	0.53	0.92
weighted avg	1.00	0.98

Frauds: 98 | Missed: 14.3%
Total Flagged: 1231 | Real: 1.1%

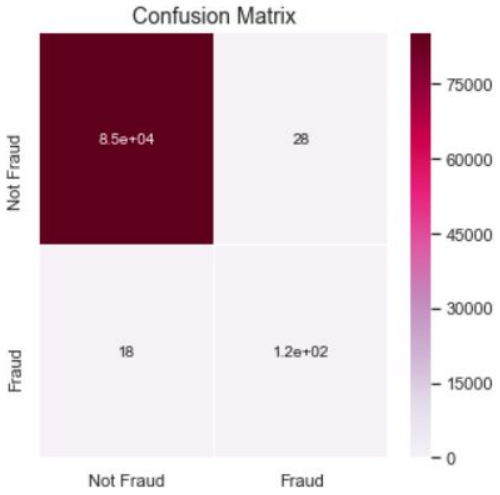


Random Forest

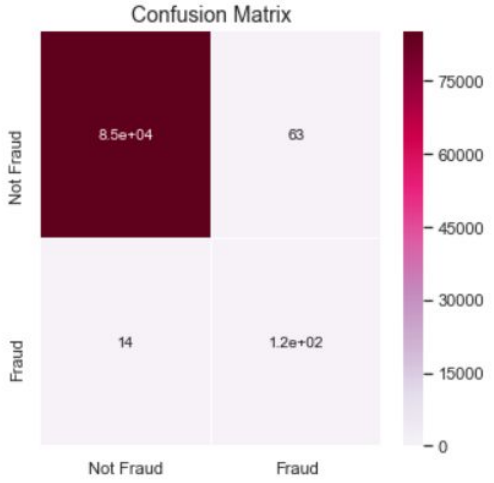
← Final Model

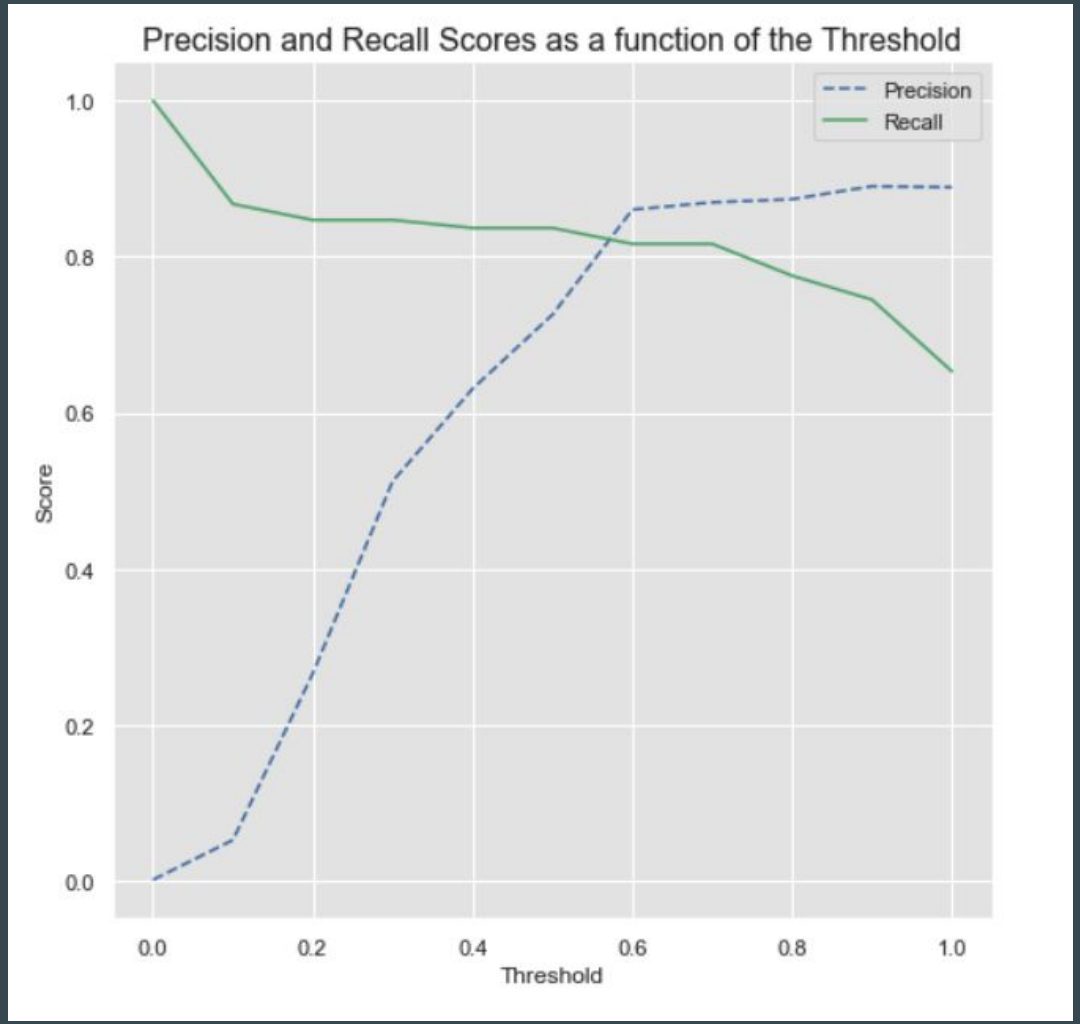
Adjusted Model →

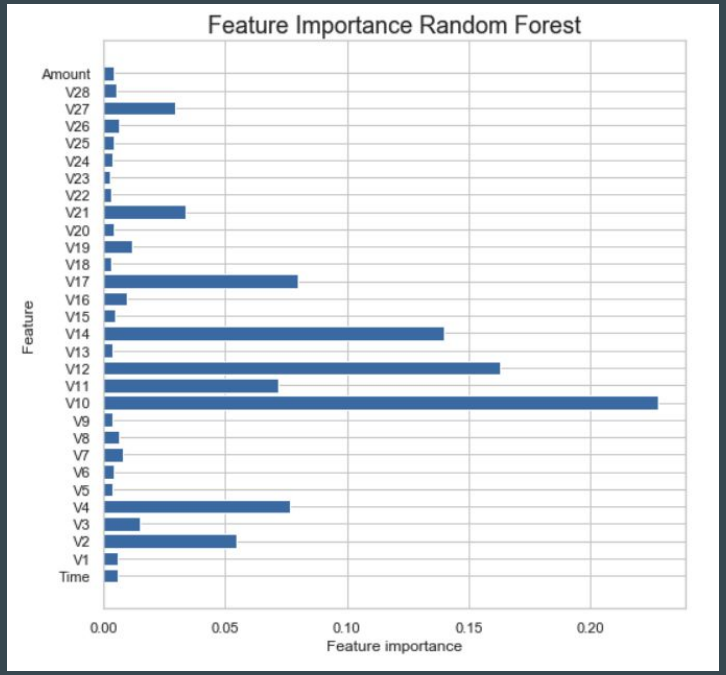
```
Recall: [0.99967177 0.86764706]
Precision: [0.99978897 0.80821918]

 clasification report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     85307
           1       0.81      0.87      0.84       136

avg / total       1.00      1.00      1.00     85443


 confussion matrix:
 [[85279    28]
 [   18   118]]

 Accuracy Percentage  is : 99.94616293903537%
```

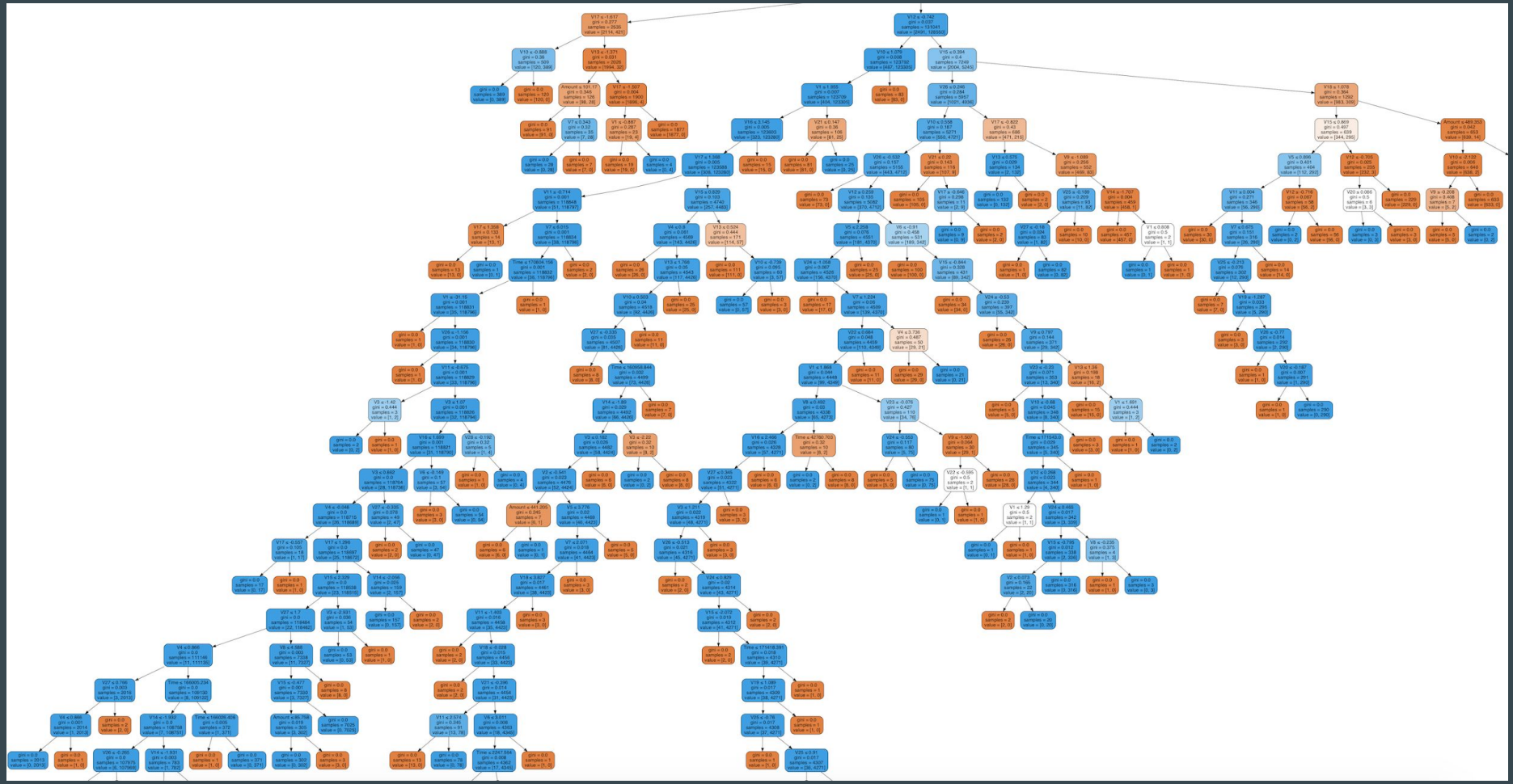


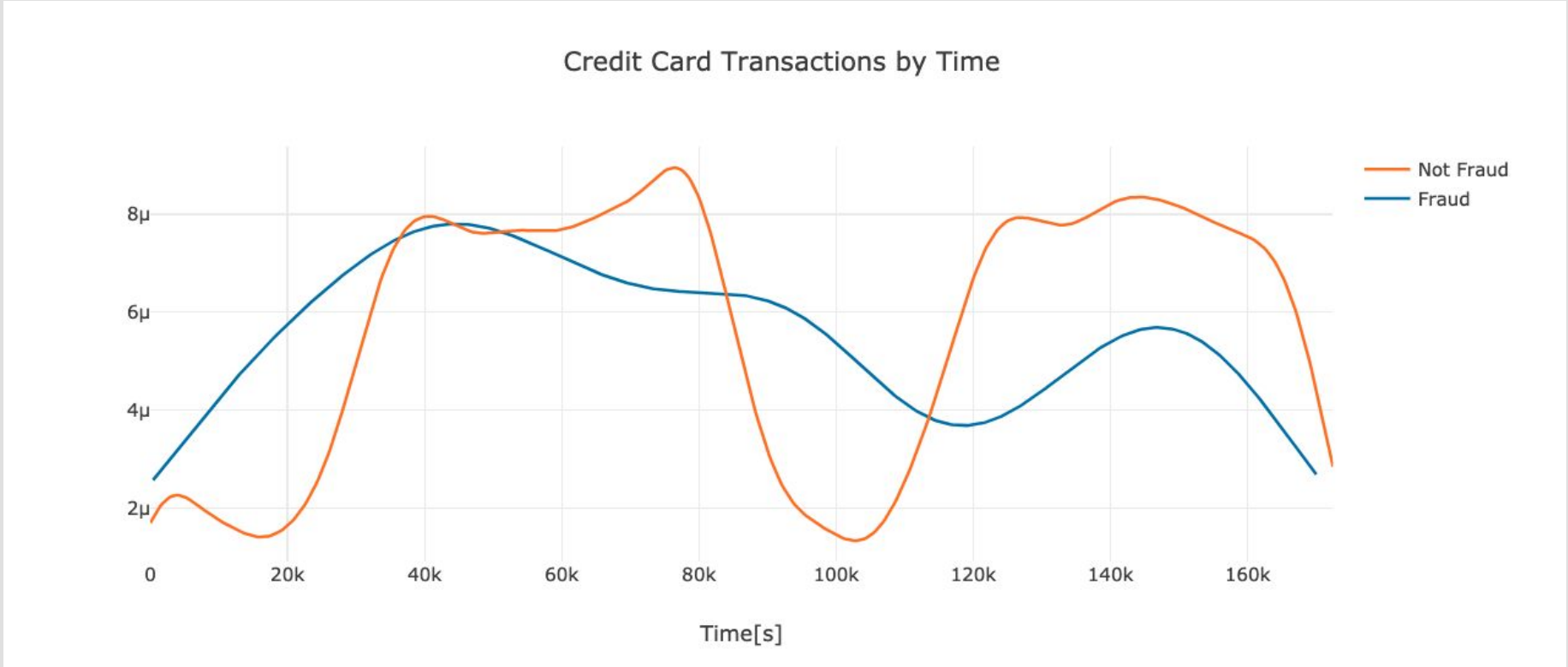

```
Recall: [0.99926149 0.89705882]
Precision: [0.99983579 0.65945946]

 clasification report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     85307
           1       0.66      0.90      0.76       136

avg / total       1.00      1.00      1.00     85443


 confussion matrix:
 [[85244    63]
 [   14   122]]

 Accuracy Percentage  is : 99.9098814414288%
```




Precision and Recall Scores as a function of the Threshold
Precision
Recall
Score
Threshold
0.0
0.2
0.4
0.6
0.8
1.0


Feature Importance Random Forest
Feature
Amount
V28
V27
V26
V25
V24
V23
V22
V21
V20
V19
V18
V17
V16
V15
V14
V13
V12
V11
V10
V9
V8
V7
V6
V5
V4
V3
V2
V1
Time
0.00
0.05
0.10
0.15
0.20
Feature importance




Credit Card Transactions by Time
Not Fraud
Fraud
8μ
6μ
4μ
2μ
0
20k
40k
60k
80k
100k
120k
140k
160k
Time[s]