

Data Science Bootcamp

Project 2 Transaction Fraud Detection

Marwa Salah





Project Overview



Identify potential fraudulent activity (transactions) using ML based on existing card transaction data



Anonymized credit card transactions labeled as fraudulent or genuine
Transformed using PCA
2 Features were not transformed using PCA which are Time & Amount

k
source



Due to nature classes are highly unbalanced.

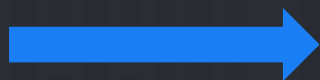
0.17% vs 99.8%
Class Balance



Transaction Fraud Detection

WHY?

1B+



11K+

Transactions per Day

Transactions per Second
~19 Fraudulent/sec

\$2-3

Cost of each dollar lost
to fraud

\$ 29B → \$ 35B

Fraud Losses 2019

Expected Fraud Losses 2022

Impacts

Individuals | Businesses |
Financial Institutions

<https://fortunly.com/statistics/cash-versus-credit-card-spending-statistics/#gref>

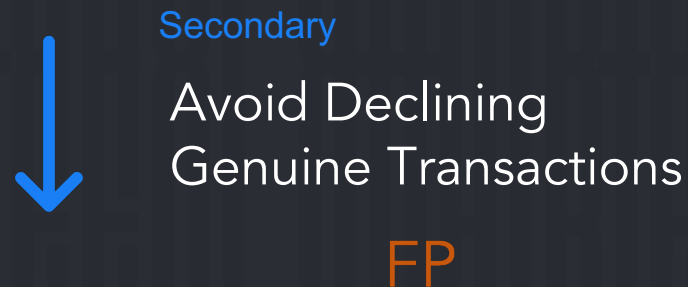
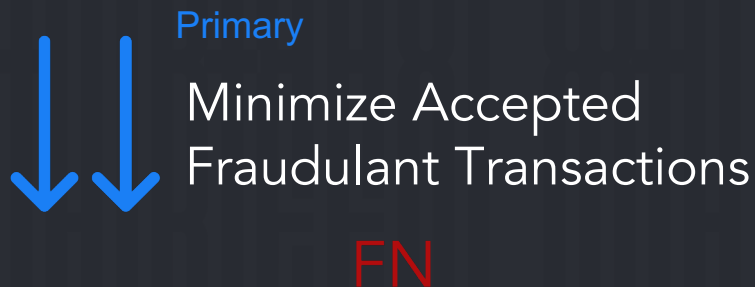
<https://www.ncr.com/blogs/payments/credit-card-fraud-detection>



Objectives

Identify potentially fraudulent activity

Detecting fraudulent activity before those transactions are even completed



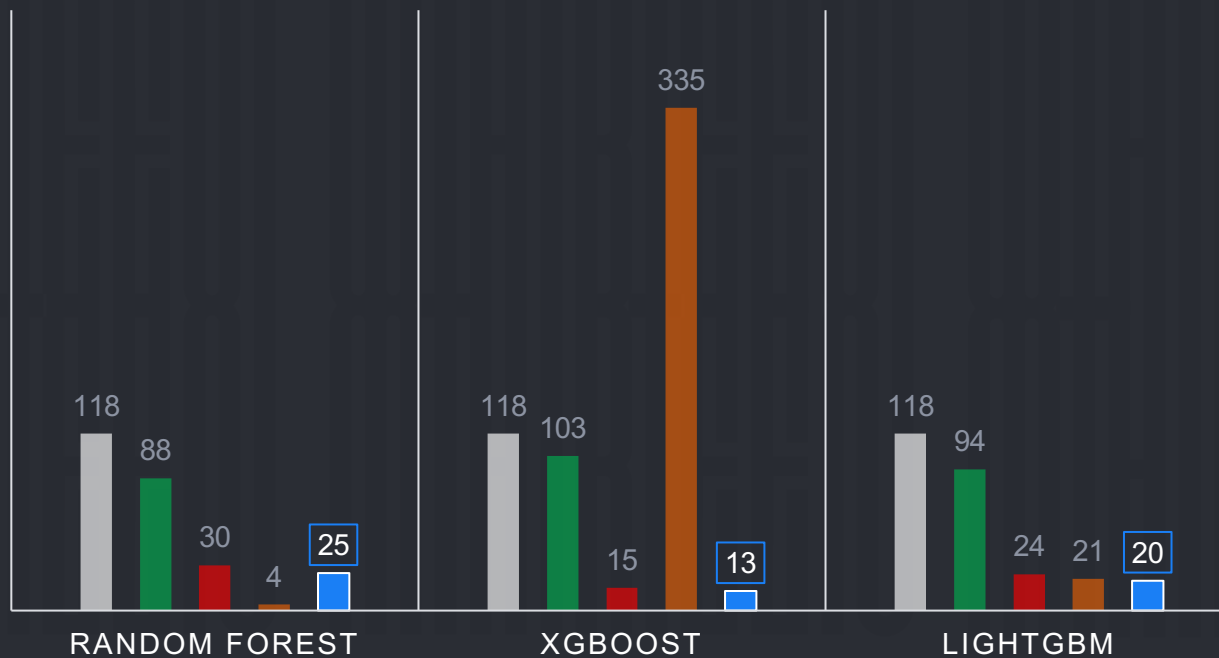


Outcome

MODEL COMPARISON

■ Actual Fraud ■ Caught ■ Missed ■ Genuine Blocked ■ % Fraud Missed

~71k
#Transactions
Tested



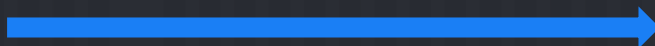


Conclusion & Next Step

Full Dataset

Fraud: \$ 60K
Genuine: \$ 25M

@87% Detection Rate



Fraud: \$ 9K
Genuine: \$ 25M (-Genuine Blocked if not validated)

Example based cost-sensitive learning & Cost-dependant Classification

	Actual Fraud $y_{\text{true}} = 1$	Actual Legitimate $y_{\text{true}} = 0$
Predicted Fraud $y_{\text{pred}} = 1$	True Positive $\text{cost}_{\text{TP}} = \text{Admin}$	False Positive $\text{cost}_{\text{FP}} = \text{Admin}$
Predicted Legitimate $y_{\text{pred}} = 0$	False Negative $\text{cost}_{\text{FN}} = \text{Transaction}$	True Negative $\text{cost}_{\text{TN}} = \0

<https://towardsdatascience.com/fraud-detection-with-cost-sensitive-machine-learning-24b8760d35d9>



Challenges

- Anonymization of Data
- Diversity of Data
- Time
- Compute Resources



Technical Project Details



Project Description

- In this project we were initially tasked with selecting a couple of datasets for analysis that meet the following rules:
 - Available to use
 - Suitable for use in a professional environment
 - Does not contain any personal information
 - Not used for any assignment, lecture or task from this environment
- My Proposed Sets: Credit Card Transaction Data & Financial Sales Data. Former was approved.
- So far, the project covers the following progress:
 - Data Cleaning & EDA
 - Evaluating & Tuning 3 Different Models (RandomForest, XGBoost, LightGBM)
 - Testing Impact of example-based cost sensitive learning
 - Model performance comparison



Data Description



Anonymized credit card transactions labeled as fraudulent or genuine
Transformed using PCA
Contains transactions made by European credit card holders in Sep 2013
2 Features were not transformed using PCA which are Time & Amount



Data source is kaggle.

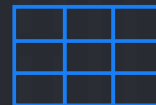


Due to nature classes are highly unbalanced.



Data Shape is:

- 28 Components, 2 Features & 1 Label
- 284807 Rows (Observations)



Data Types:

All features are
Integers of type int64



ML Problem Type is:
Classification



Classification Target:

Feature 'Class' is the response variable, and it takes value 1 in case of fraud and 0 otherwise.



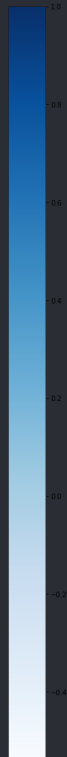
Data Cleaning

- Due to PCA preprocessing Data Cleaning needed was very limited.
- No Nulls found
- No Data Type Inconcistencies
- No Categorical Inconsistencies would exist
- No Illogical Values
- A Number of Duplicates were there (~1000 rows) which were dropped

Visuals



1	0.12	-0.011	-0.42	-0.11	0.17	-0.063	0.085	-0.038	-0.0079	0.031	-0.25	0.13	-0.066	-0.1	-0.18	0.011	-0.074	0.09	0.03	-0.051	0.046	0.14	0.051	-0.016	-0.23	-0.042	-0.0052	-0.0093	-0.011	-0.012	
-0.012	1	0.0069	-0.0081	0.0023	-0.007	0.00041	-0.0092	-0.0012	0.0018	0.00082	0.001	-0.0015	-0.00057	-0.0027	-0.0006	-0.0033	-0.0035	-0.0035	0.00092	-0.0014	0.0028	-0.0014	-0.0013	-0.00072	-0.00022	-0.00068	-0.016	-0.0049	-0.23	-0.094	
-0.011	0.0069	1	0.0053	-0.0015	0.0052	-0.00059	0.0074	0.0029	-0.00027	0.00062	-0.00063	0.0023	0.00068	0.0027	0.0015	0.004	0.0032	0.0025	-0.00036	-0.0013	-0.0049	0.0012	-0.0039	0.0007	-0.0016	0.00025	0.00076	0.0016	-0.53	0.085	
-0.42	-0.0081	0.0053	1	0.0028	-0.0069	-0.0015	0.012	-0.0018	-0.0036	-0.0096	0.0023	-0.0059	0.00011	-0.003	-0.0012	-0.0044	-0.0082	-0.0035	-1.6e-05	-0.0023	0.0035	-0.00027	0.00045	-7.2e-05	0.00043	-9.4e-05	0.00071	0.00013	-0.21	-0.18	
-0.11	0.0023	-0.0015	0.0028	1	0.0017	-0.00088	0.0047	0.00089	0.0022	0.0028	-0.0012	0.0034	0.00018	0.0028	0.00057	0.0033	0.0037	0.0023	-0.00056	0.00032	-0.001	0.00012	0.00073	-0.00012	0.00016	0.00078	0.0013	0.00023	0.1	0.13	
0.17	-0.007	0.0052	-0.0069	0.0017	1	-0.00094	-0.0087	0.0014	-0.0012	-0.006	0.00041	-0.0023	1.9e-05	-0.001	-0.0012	-0.0024	-0.0045	-0.0027	0.00044	-0.0012	0.0016	-0.00056	0.0012	0.0002	6.9e-05	0.00039	-0.00058	-0.00082	-0.39	-0.088	
-0.063	0.00041	-0.00059	-0.0015	-0.00088	-0.00094	1	0.00044	0.003	-0.00073	-0.0022	-0.00021	-0.0012	0.0004	0.00018	-0.00047	0.00012	-0.0017	0.00054	0.00011	-0.00018	-0.0021	0.0011	-0.00076	0.0012	0.0007	-2.8e-05	0.00029	0.00092	0.22	-0.044	
0.085	-0.0092	0.0074	-0.012	0.0047	-0.0087	0.00044	1	-0.0064	-0.0049	-0.014	0.0025	-0.0062	-0.00017	-0.0038	-0.0014	-0.0059	-0.0088	-0.0043	0.00085	-0.0012	0.009	-0.0023	0.0033	-0.00038	-7.2e-05	0.00062	-0.0045	0.0017	0.4	-0.17	
-0.038	-0.0012	0.0029	-0.0018	0.00089	0.0014	0.003	-0.0064	1	0.001	0.00048	0.0047	-0.0044	-0.0014	-0.0084	0.001	-0.0044	-0.0056	-0.0013	-0.00063	0.00027	0.019	-0.0062	0.005	0.00011	1.1e-05	-0.0014	0.00061	-9.9e-05	-0.1	0.033	
-0.0079	0.0018	-0.00027	-0.0036	0.0022	-0.0012	-0.00073	-0.0049	0.001	1	-0.013	-0.00022	-0.0024	0.00075	0.002	-0.00028	-8.6e-05	-0.0023	-0.00037	0.00025	-0.0018	0.00068	0.00078	0.00068	-0.0001	-0.00028	0.0013	0.0082	0.0056	-0.044	-0.094	
0.031	0.00082	0.00062	-0.0096	0.0028	-0.006	-0.0022	0.014	0.00048	-0.013	1	-0.00084	-0.0069	0.0014	0.00017	-0.0023	-0.0037	-0.0079	-0.0025	0.0011	-0.0044	0.0038	-0.00048	0.0019	0.00015	-0.00056	0.0011	0.011	0.0092	-0.1	-0.21	
-0.25	0.001	-0.00063	0.0023	-0.0012	0.00041	-0.00021	0.0025	0.0047	-0.0022	0.00084	1	-0.0056	0.00046	0.0077	-0.00087	0.0048	0.0074	0.0021	-0.00049	-0.00099	-0.0028	-0.00015	-3.7e-05	8e-05	4.7e-05	-0.0002	0.002	0.0026	1.5e-05	0.15	
0.13	-0.0015	0.0023	-0.0059	0.0034	-0.0023	-0.0012	-0.0062	-0.0044	-0.0024	-0.0069	0.0056	1	-0.00055	-0.01	0.00088	-0.0074	-0.013	-0.0035	0.00059	0.0012	0.0033	0.00015	0.00049	0.00059	-0.00018	-0.00014	0.00093	-0.00091	-0.0093	-0.25	
-0.066	-0.00057	0.00068	0.00011	0.00018	1.9e-05	0.0004	-0.00017	-0.0014	0.00075	0.0014	0.00046	-0.00055	1	-0.0011	0.00023	-0.00081	-0.00017	-0.00016	8.6e-05	0.00038	0.00052	1.6e-05	0.00025	4.9e-05	0.00025	-0.0001	-0.0016	-0.0006	0.0052	-0.039	
-0.1	-0.0027	0.0027	-0.003	0.0028	-0.001	0.00018	-0.0038	-0.0084	0.002	0.00017	0.0077	-0.01	-0.0011	1	0.00087	-0.0091	-0.014	-0.0045	0.0016	0.0027	0.0056	-0.0019	0.00067	-2.6e-05	0.00016	-0.0007	-0.0046	-0.0047	0.034	-0.29	
-0.18	-0.0006	0.0015	-0.0012	0.00057	-0.0012	-0.00047	-0.0014	0.001	-0.00028	-0.0023	-0.00087	0.00088	0.00023	0.00087	1	-0.00028	-0.00017	-0.00043	0.00055	-0.00075	-0.0027	-0.0012	0.00097	0.00011	0.00044	-0.002	-0.00064	0.00086	-0.0033	-0.0033	
0.011	-0.0033	0.0004	-0.0044	0.0033	-0.0024	0.00012	-0.00059	-0.0044	-8.6e-05	-0.0037	0.0048	-0.0074	-0.00081	-0.0091	-0.00028	1	-0.0091	-0.0094	0.0024	0.0011	0.0043	-0.00082	0.0012	-0.00048	0.00021	-0.0012	-0.004	0.00016	-0.0045	-0.19	
-0.074	-0.0035	0.0032	-0.0082	0.0037	0.0045	-0.0017	-0.0088	-0.0056	-0.0023	-0.0079	0.0074	-0.013	-0.00017	-0.014	-0.00017	-0.0091	1	-0.0053	0.00099	0.0015	0.0036	-0.00016	0.00067	0.001	-0.00068	0.00016	-0.0034	0.0027	0.0077	-0.31	
0.09	-0.0035	0.0025	-0.0035	0.0023	-0.0027	0.00054	-0.0043	-0.0013	-0.00037	-0.0025	0.0021	-0.0035	-0.00016	-0.0045	-0.00043	0.0054	-0.0053	1	-0.0001	-0.00024	0.0016	-0.00053	0.00024	-0.00071	-0.00056	-0.0006	-0.0042	-0.0013	0.036	-0.11	
0.03	0.00092	-0.00036	-1.6e-05	-0.00056	0.00044	0.00011	0.00085	-0.00063	-0.00025	0.0011	-0.00049	0.00059	8.6e-05	0.0016	0.00055	0.0024	0.00099	-0.0001	1	-0.00027	0.00024	0.0013	0.00038	-0.00011	-8.4e-05	0.00086	-0.00054	0.00035	-0.056	0.034	
-0.051	-0.0014	-0.0013	-0.0023	0.00032	-0.0012	-0.00018	-0.0012	0.00027	-0.0018	-0.0044	-0.00099	0.0012	0.00038	0.0027	-0.00075	0.0011	0.0015	-0.00024	-0.00027	1	-0.0054	-0.0016	-0.0011	-0.0003	-0.00064	-0.00031	4.9e-05	0.0027	0.34	0.021	
0.046	0.0028	-0.0049	0.0035	-0.001	0.0016	-0.0021	0.009	0.0019	0.00068	0.0038	-0.0028	0.0033	0.00052	0.0056	-0.00027	0.0043	0.0036	0.0016	0.00024	0.0054	1	0.0096	-0.0064	0.0012	-0.00087	-0.00087	-0.0052	-0.0044	0.11	0.026	
0.14	-0.0014	0.0012	-0.00027	0.00012	-0.00056	0.0011	-0.0023	-0.0062	0.00078	-0.00048	-0.00015	0.00015	1.6e-05	-0.0019	-0.0012	-0.00082	-0.00016	-0.00053	0.0013	-0.0016	0.0096	1	0.0019	-3.1e-05	0.0002	-0.0015	0.003	0.0014	-0.065	0.0049	
0.051	-0.0013	-0.0039	0.00045	0.00073	0.0012	-0.00076	0.0033	0.005	0.00068	0.0019	-3.7e-05	0.00049	0.00025	0.00067	0.00097	0.0012	0.00067	0.00024	0.00038	-0.0011	-0.0064	0.0019	1	0.00027	-0.00053	-0.00019	-0.002	-0.0032	-0.11	-0.0063	
-0.016	-0.00072	0.0007	-7.2e-05	-0.00012	0.0002	0.0012	-0.00038	0.00011	-0.0001	0.00015	8e-05	0.00059	4.9e-05	-2.6e-05	0.00011	-0.00048	0.001	-0.00071	-0.00011	0.0003	0.0012	-3.1e-05	0.00027	1	-0.00019	0.00057	-0.00088	0.00032	0.0051	0.0072	
-0.23	-0.00022	-0.0016	0.00043	0.00016	6.9e-05	0.0007	-7.2e-05	1.1e-05	0.00028	-0.00056	4.7e-05	-0.00018	0.00025	0.00016	0.00044	-0.00021	-0.00068	-0.00056	-0.00064	-0.00087	0.0002	-0.00053	-0.00019	1	4.8e-05	0.00013	-0.00057	-0.048	0.0032		
-0.042	-0.00068	0.00025	-9.4e-05	0.00078	0.00039	-2.8e-05	0.00062	-0.0014	0.0013	0.0011	-0.0002	-0.00014	-0.0001	-0.0007	-0.002	-0.0012	0.00016	-0.0006	0.00086	-0.00031	-0.00087	-0.0015	-0.00019	0.00057	4.9e-05	1	-0.0033	-0.001	-0.0034	0.0043	
-0.0052	-0.016	0.0076	-0.0071	0.0013	-0.0058	0.00029	-0.0045	0.00061	0.0082	0.011	0.002	-0.00093	-0.0016	-0.0046	-0.00064	-0.00064	-0.004	-0.0034	-0.0042	-0.00054	-4.9e-05	-0.0052	0.003	-0.002	-0.00088	-0.0013	-0.0033	1	-0.014	0.028	0.022
-0.093	-0.0049	0.0016	-0.00013	0.00023	-0.00082	0.00092	-0.0017	-9.9e-05	0.0056	0.0092	0.0026	-0.00061	-0.0006	-0.0047	0.00086	-0.0016	-0.0027	-0.0013	0.00035	0.0027	-0.0044	0.0014	-0.0032	0.00032	-0.00057	-0.001	-0.014	1	0.01	0.0097	
-0.011	-0.23	-0.53	-0.21	0.1	-0.39	0.22	0.4	-0.1	-0.044	-0.1	-1.5e-05	-0.0093	0.0052	0.004	0.0033	0.0045	0.0077	0.036	-0.056	0.34	0.11	-0.065	-0.11	0.0051	-0.048	-0.0034	0.028	0.01	1	0.0058	
-0.012	-0.094	0.085	-0.18	0.13	-0.088	-0.044	-0.17	0.033	-0.094	-0.21	0.15	-0.25	-0.039	-0.29	-0.0033	-0.19	-0.31	-0.11	0.034	0.021	0.026	0.0049	-0.0063	-0.0072	0.0032	0.0043	0.022	0.0097	0.0058	1	
Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24	V25	V26	V27	V28	V29	Amount	Class





Model Tuning – Parameters Used

Random Forest

```
max_depth: None  
max_features: 'auto'
```

XGBoost

```
learning_rate: 0.35  
max_depth: 6  
tree_method: 'auto'  
use_rmm: 'true'
```

LightGBM

```
eval_metric: 'auc'  
learning_rate: 0.08  
max_depth: 13  
tree_method: 'gpu_hist'
```

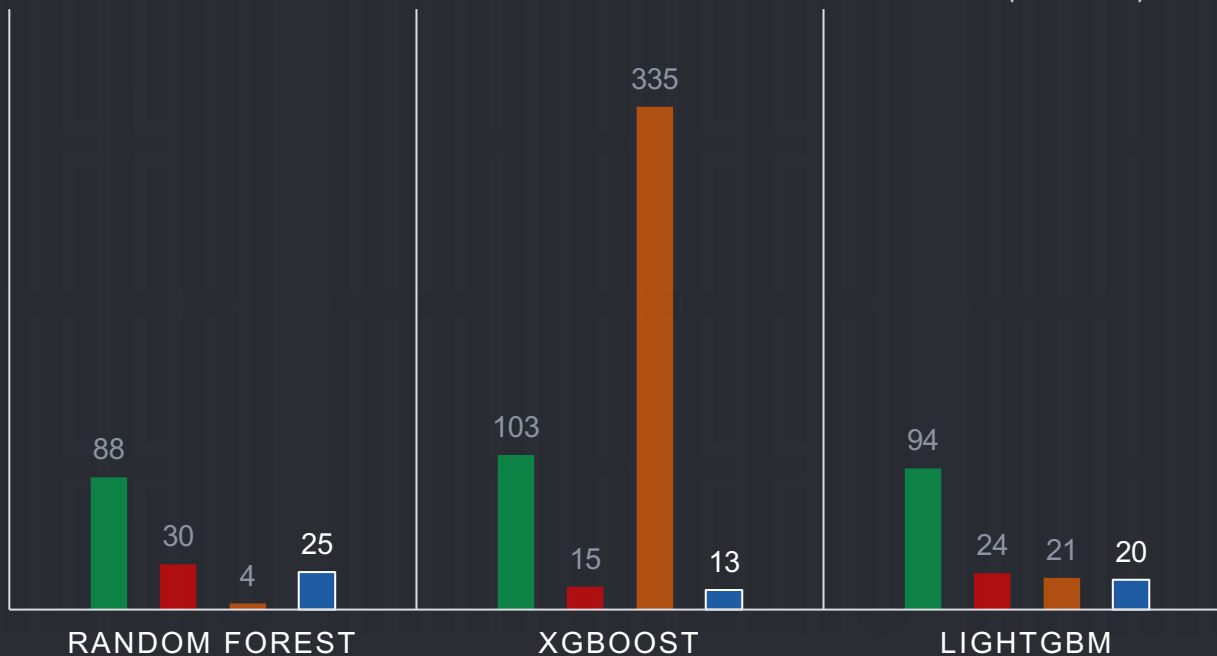


Outcome

MODEL COMPARISON

■ True Positives ■ False Negatives ■ False Positives ■ % True Missed
(1 - Recall) * 100

~71k
#Transactions
Tested





Technical Outcome Summary



Outcome Summary

	Model	precision	recall	f1-score
0 Genuine	Random Forest	1	1	1
	XGBoost	1	0.99	1
	LightGBM	1	1	1
1 Fraud	Random Forest	0.96	<u>0.74</u>	0.83
	XGBoost	0.21	<u>0.86</u>	0.33
	LightGBM	0.83	<u>0.80</u>	0.81



Outcome Summary (without Time)

	Model	precision	recall	f1-score
0 Genuine	Random Forest	1	1	1
	XGBoost	1	1	1
	LightGBM	1	1	1
1 Fraud	Random Forest	0.96	<u>0.75</u>	0.84
	XGBoost	0.24	<u>0.87</u>	0.37
	LightGBM	0.82	<u>0.80</u>	0.81



Confusion Matrix

All features

74%

Recall

96%

Precision

True Label	0 Genuine		
	1 Fraud		
	0 Genuine	70809	5
	1 Fraud	31	87
		Predicted Label	

Time Excluded

75%

Recall

96%

Precision

True Label	0 Genuine		
	1 Fraud		
	0 Genuine	70810	4
	1 Fraud	30	88
		Predicted Label	

Random Forest



Confusion Matrix

All features

86%

Recall

21%

Precision

True Label	0 Genuine	1 Fraud	Predicted Label
	70424	390	
0 Genuine	16	102	
1 Fraud	0	1	

Time Excluded

87%

Recall

24%

Precision

True Label	0 Genuine	1 Fraud	Predicted Label
	70479	335	
0 Genuine	15	103	
1 Fraud	0	1	

XGBoost



Confusion Matrix

All features

80%

Recall

83%

Precision

True Label	0 Genuine	70795		19
	1 Fraud	24	94	
		0 Genuine	1 Fraud	
		Predicted Label		

Time Excluded

80%

Recall

82%

Precision

True Label	0 Genuine	70793		21
	1 Fraud	24	94	
		0 Genuine	1 Fraud	
		Predicted Label		

LightGBM