

Credit Card Fraud Detection

BY : NOUF , YARA



Table of contents

- | | | | |
|-----------|---------------------|-----------|----------------------|
| 01 | Introduction | 05 | preprocessing |
| 02 | Dataset | 06 | MODELS |
| 03 | EDA | 07 | Conclusion |

Dataset



Features

- Step
- Type
- Amount
- nameOrig
- oldbalanceOrg
- newbalanceOrig
- nameDest
- oldbalanceDest
- newbalanceDest
- isFraud
- isFlaggedFraud



Rows

6M >

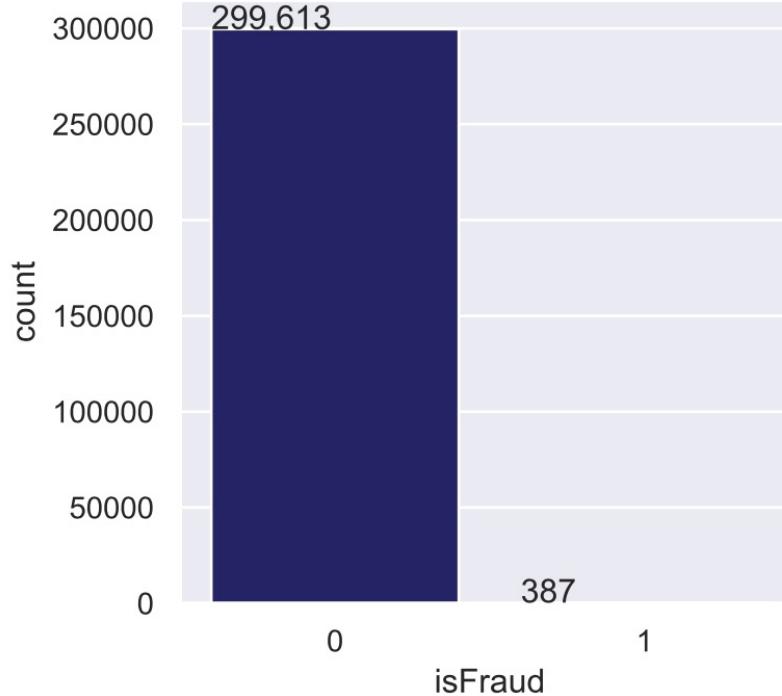
Synthetic Financial Datasets : An Overview

Ever since the advent of internet the digital revolution has risen and has crepted into all aspects to our lives. One of the most important digital revolution happened in financial system and especially transacting money to someone from any part of the world digitally. Digital transactions have become a part of daily life like purchasing a product online, sending money to friends, depositing cash in bank account, investment purposes etc., They had a lot of benefits so does paved way for fraudulent activities. People started using digital money transactions medium to launder money and make the money look like it comes from a legal source.



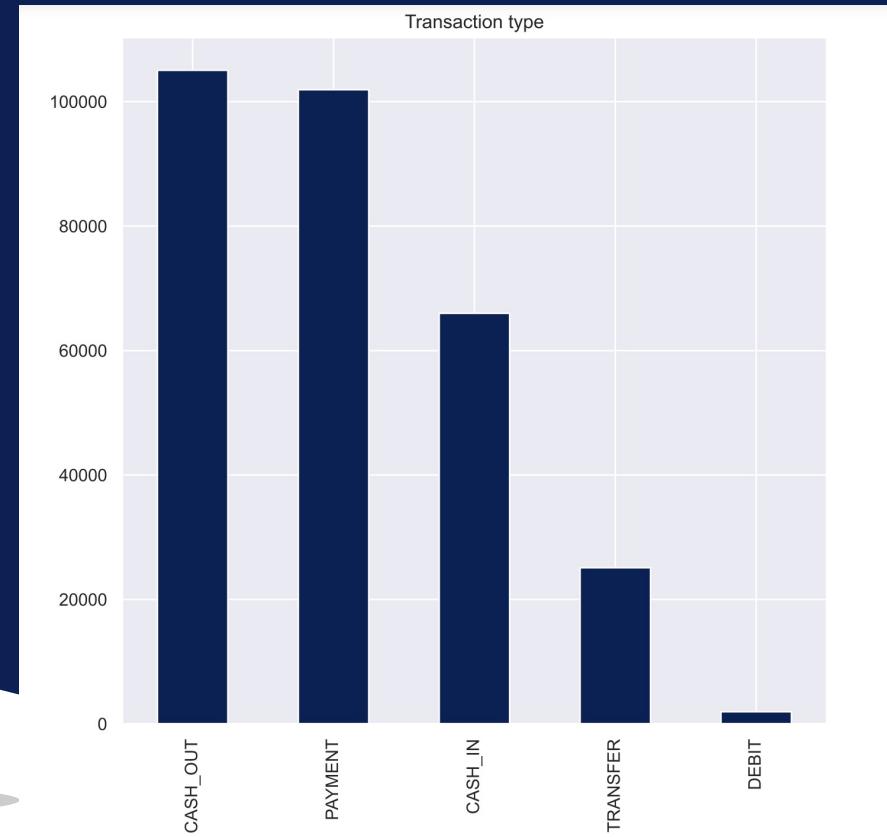


EDA!



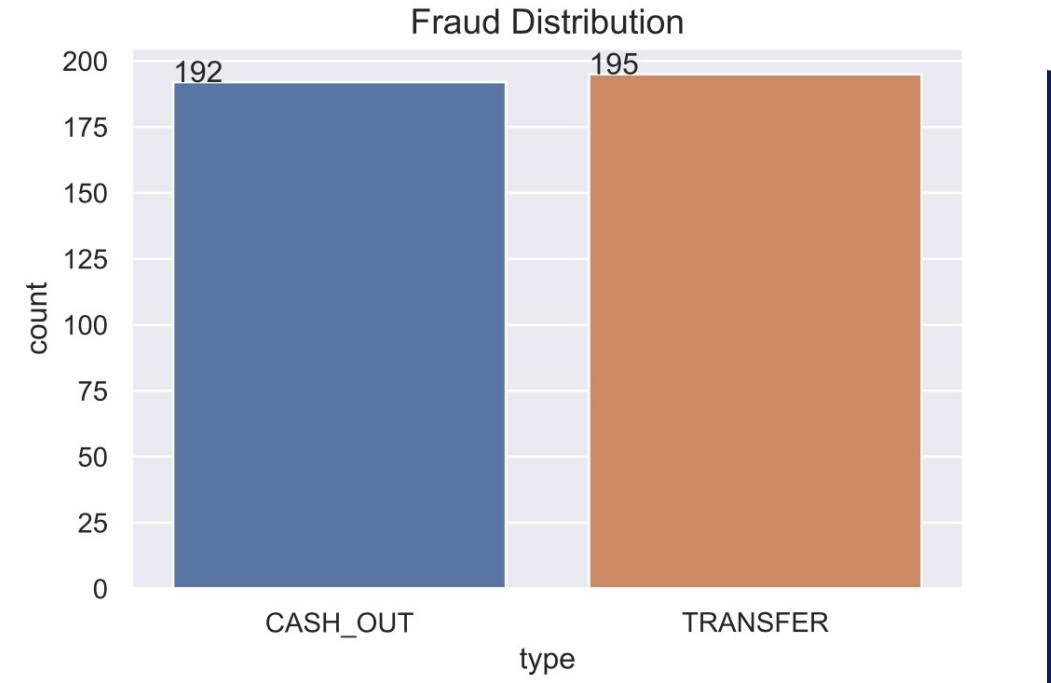
0 represents # not fraud
1 represents # fraud

Types of Transactions and the rate of each of them per day

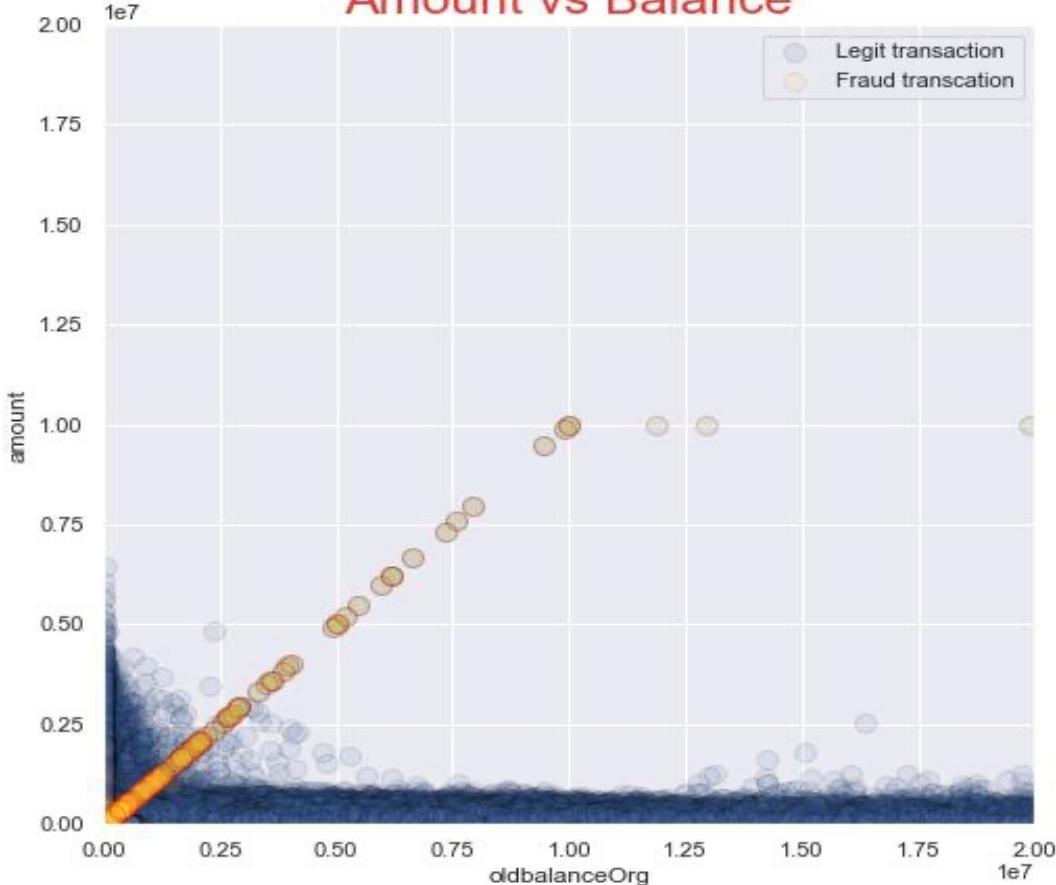




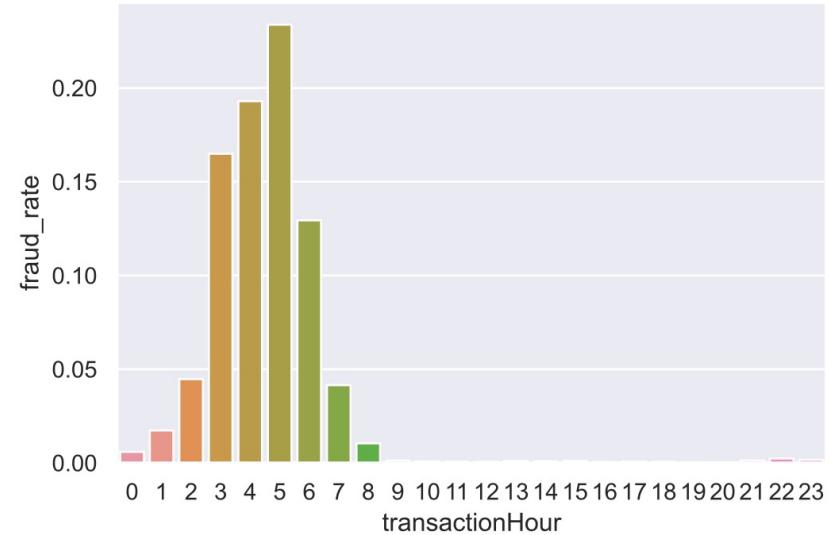
Fraud Distribution



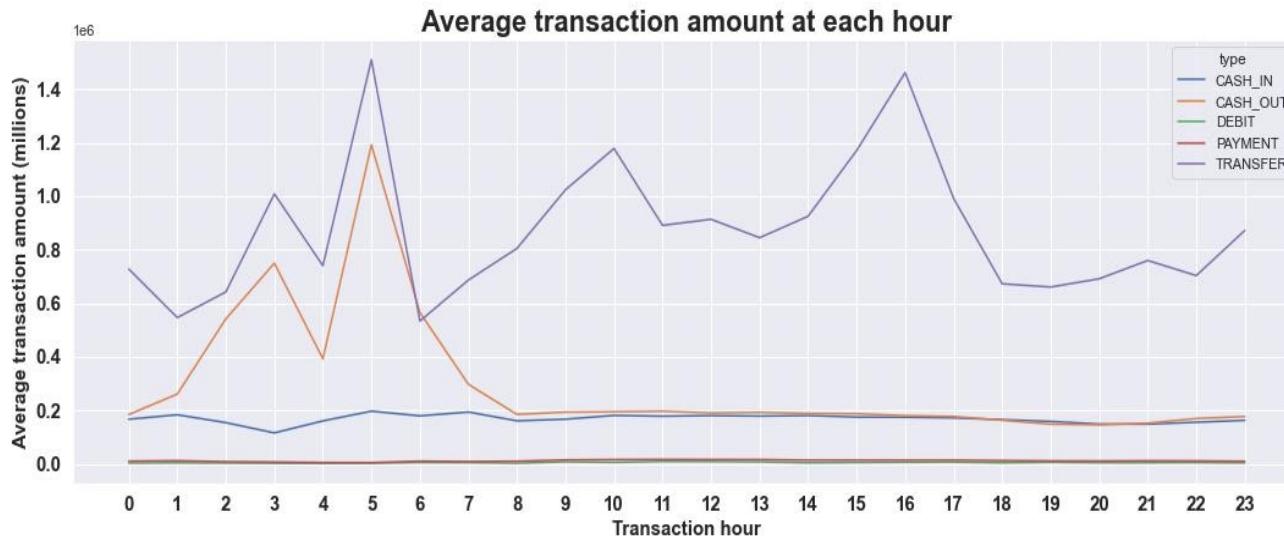
Amount vs Balance



Fraud and time



Average Transaction amount at each hour



MODEL



Random Forest Classifier

	precision	recall	f1-score	support
0	1.00	1.00	1.00	86739
1	1.00	1.00	1.00	250
accuracy			1.00	86989
macro avg	1.00	1.00	1.00	86989
weighted avg	1.00	1.00	1.00	86989


```
[[86739      0]
 [      0    250]]
```

Credit Card Fraud Detection



Dataset



Features

- V1, V2, ... V28
- **Amount**
- **Class**
- **Time**

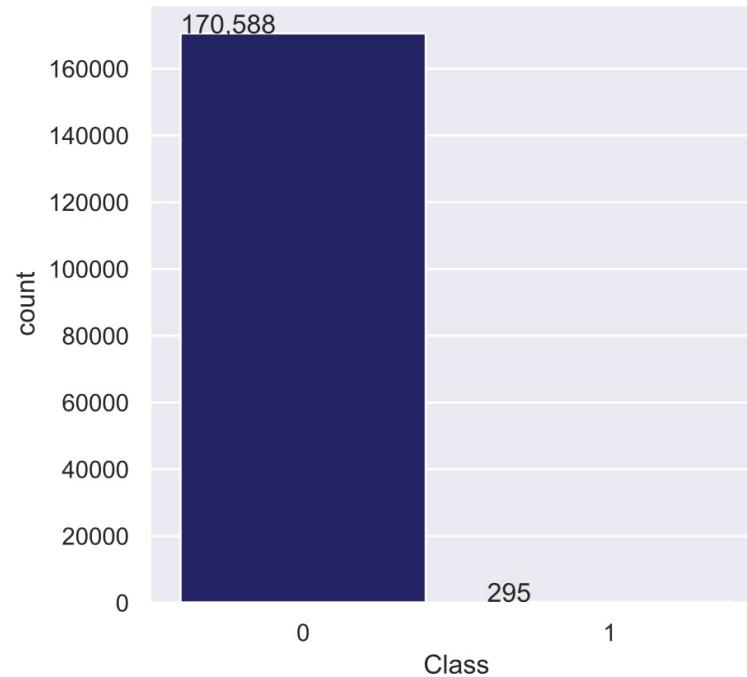


Rows

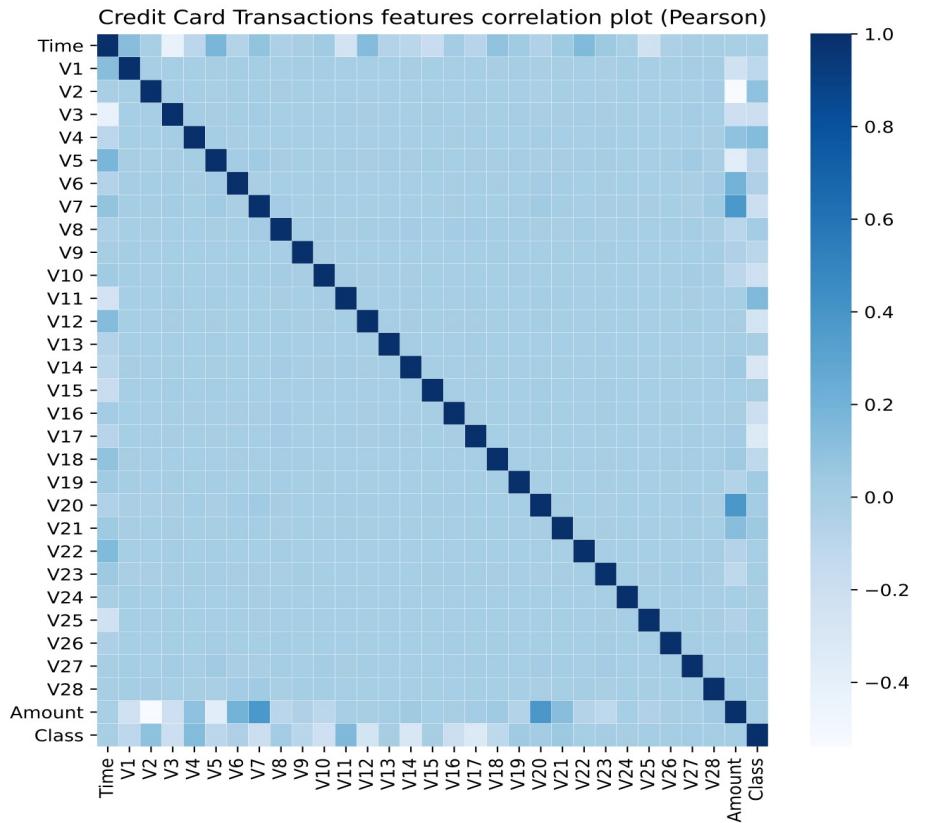
250K >



EDA!



0 represents # not fraud
1 represents # fraud



Resampling





SMOTE

1 → 170588
0 → 170588



Models

Logistic Regression

KNN

Train

	precision	recall	f1-score	support
0	1.00	1.00	1.00	170588
1	0.90	0.64	0.74	295
accuracy			1.00	170883
macro avg	0.95	0.82	0.87	170883
weighted avg	1.00	1.00	1.00	170883
[[170566 22] [107 188]]				

	precision	recall	f1-score	support
0	1.00	1.00	1.00	170588
1	0.95	0.81	0.87	295
accuracy			1.00	170883
macro avg	0.98	0.90	0.94	170883
weighted avg	1.00	1.00	1.00	170883
[[170576 12] [57 238]]				

val

	precision	recall	f1-score	support
0	1.00	1.00	1.00	56863
1	0.86	0.61	0.71	99
accuracy			1.00	56962
macro avg	0.93	0.80	0.85	56962
weighted avg	1.00	1.00	1.00	56962
[[56853 10] [39 60]]				

	precision	recall	f1-score	support
0	1.00	1.00	1.00	56863
1	0.90	0.78	0.83	99
accuracy			1.00	56962
macro avg	0.95	0.89	0.92	56962
weighted avg	1.00	1.00	1.00	56962
[[56854 9] [22 77]]				

SVM

Train

	precision	recall	f1-score	support
0	1.00	1.00	1.00	170588
1	0.85	0.82	0.84	295
accuracy			1.00	170883
macro avg	0.93	0.91	0.92	170883
weighted avg	1.00	1.00	1.00	170883


```
[[170546    42]
 [    53   242]]
```

val

	precision	recall	f1-score	support
0	1.00	1.00	1.00	56863
1	0.83	0.82	0.82	99
accuracy			1.00	56962
macro avg	0.91	0.91	0.91	56962
weighted avg	1.00	1.00	1.00	56962


```
[[56846    17]
 [    18   81]]
```

Voting Classifier(XGB , DT , RF)

Voting Classifier(XGB , LOG, KNN)

Train

	precision	recall	f1-score	support
0	1.00	1.00	1.00	170588
1	0.90	0.76	0.82	295
accuracy			1.00	170883
macro avg	0.95	0.88	0.91	170883
weighted avg	1.00	1.00	1.00	170883


```
[[170563    25]
 [  71   224]]
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	170588
1	0.96	0.81	0.88	295
accuracy			1.00	170883
macro avg	0.98	0.91	0.94	170883
weighted avg	1.00	1.00	1.00	170883


```
[[170579    9]
 [  56   239]]
```

val

	precision	recall	f1-score	support
0	1.00	1.00	1.00	56863
1	0.84	0.73	0.78	99
accuracy			1.00	56962
macro avg	0.92	0.86	0.89	56962
weighted avg	1.00	1.00	1.00	56962


```
[[56849    14]
 [  27   72]]
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	56863
1	0.89	0.77	0.83	99
accuracy			1.00	56962
macro avg	0.95	0.88	0.91	56962
weighted avg	1.00	1.00	1.00	56962


```
[[56854    9]
 [  23   76]]
```

Random Forest Classifier

Train

	precision	recall	f1-score	support
0	1.00	1.00	1.00	170588
1	0.88	0.59	0.71	295
accuracy			1.00	170883
macro avg	0.94	0.80	0.85	170883
weighted avg	1.00	1.00	1.00	170883

[[170565	23]
[120	175]]

Bagging Classifier

val

	precision	recall	f1-score	support
0	1.00	1.00	1.00	170588
1	0.85	0.75	0.80	295
accuracy			1.00	170883
macro avg	0.92	0.88	0.90	170883
weighted avg	1.00	1.00	1.00	170883

[[170548	40]
[73	222]]

	precision	recall	f1-score	support
0	1.00	1.00	1.00	56863
1	0.83	0.54	0.65	99
accuracy			1.00	56962
macro avg	0.91	0.77	0.82	56962
weighted avg	1.00	1.00	1.00	56962

[[56843	20]
[27	72]]

Train

Stacking Classifier(XGB , log , RF)

	0	1.00	1.00	1.00	170588
	1	0.95	0.77	0.85	295
accuracy				1.00	170883
macro avg		0.98	0.88	0.93	170883
weighted avg		1.00	1.00	1.00	170883
[[170577 11] [68 227]]					

Stacking Classifier(XGB , log , KNN)

		precision	recall	f1-score	support
	0	1.00	1.00	1.00	170588
	1	0.97	0.81	0.88	295
accuracy				1.00	170883
macro avg		0.99	0.90	0.94	170883
weighted avg		1.00	1.00	1.00	170883
[[170581 7] [57 238]]					

val

	precision	recall	f1-score	support
	0	1.00	1.00	56863
	1	0.91	0.71	99
accuracy			1.00	56962
macro avg	0.95	0.85	0.90	56962
weighted avg	1.00	1.00	1.00	56962
[[56856 7] [29 70]]				

	precision	recall	f1-score	support
	0	1.00	1.00	56863
	1	0.93	0.76	99
accuracy			1.00	56962
macro avg	0.96	0.88	0.92	56962
weighted avg	1.00	1.00	1.00	56962
[[56857 6] [24 75]]				

Train

XGB

	precision	recall	f1-score	support
0	1.00	1.00	1.00	170588
1	0.94	0.82	0.87	295
accuracy			1.00	170883
macro avg	0.97	0.91	0.94	170883
weighted avg	1.00	1.00	1.00	170883

[[170573	15]	
[54	241]]

DT

	precision	recall	f1-score	support
0	1.00	1.00	1.00	170588
1	0.85	0.76	0.80	295
accuracy			1.00	170883
macro avg	0.92	0.88	0.90	170883
weighted avg	1.00	1.00	1.00	170883

[[170547	41]	
[70	225]]

val

	precision	recall	f1-score	support
0	1.00	1.00	1.00	56863
1	0.89	0.78	0.83	99
accuracy			1.00	56962
macro avg	0.94	0.89	0.91	56962
weighted avg	1.00	1.00	1.00	56962

[[56853	10]	
[22	77]]

	precision	recall	f1-score	support
0	1.00	1.00	1.00	56863
1	0.78	0.74	0.76	99
accuracy			1.00	56962
macro avg	0.89	0.87	0.88	56962
weighted avg	1.00	1.00	1.00	56962

[[56843	20]	
[26	73]]

Testing



SVM

Train

	precision	recall	f1-score	support
0	1.00	1.00	1.00	170588
1	0.85	0.82	0.84	295
accuracy			1.00	170883
macro avg	0.93	0.91	0.92	170883
weighted avg	1.00	1.00	1.00	170883

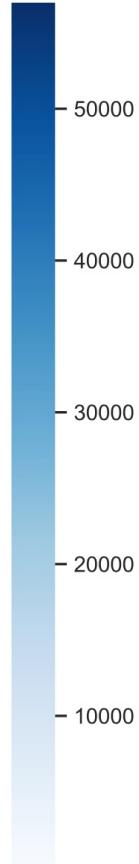
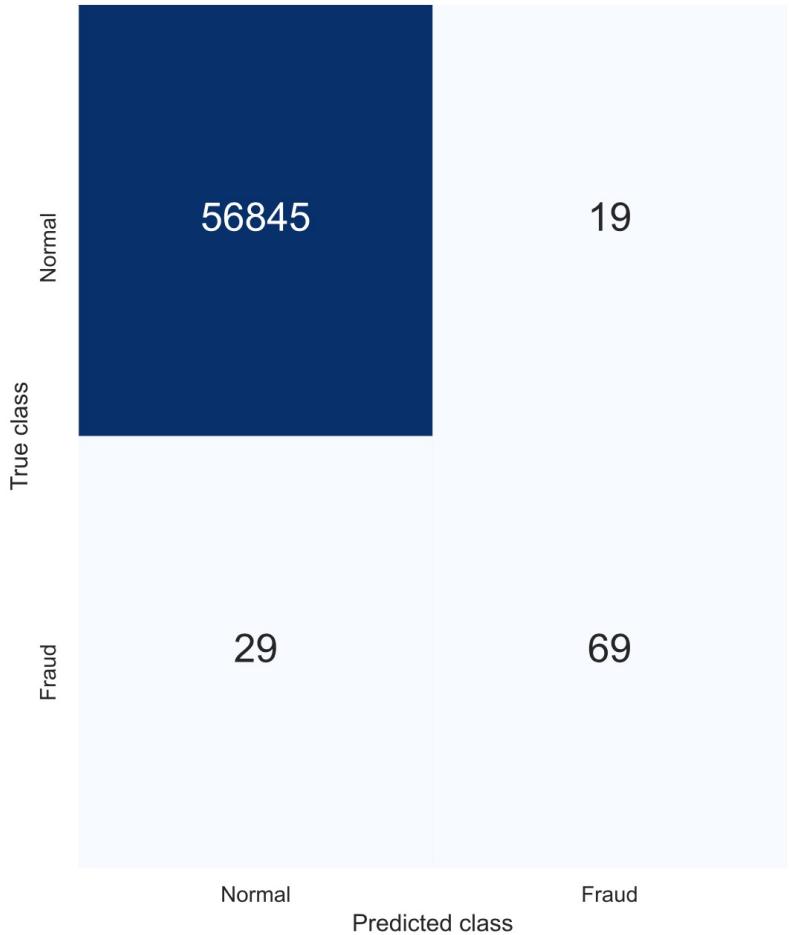

```
[[170546    42]
 [    53   242]]
```

val

	precision	recall	f1-score	support
0	1.00	1.00	1.00	56863
1	0.83	0.82	0.82	99
accuracy			1.00	56962
macro avg	0.91	0.91	0.91	56962
weighted avg	1.00	1.00	1.00	56962


```
[[56846    17]
 [    18   81]]
```

Confusion matrix



SVM

	precision	recall	f1-score	support
0	1.00	1.00	1.00	56864
1	0.78	0.70	0.74	98

Conclusion



- First test results were promising .
- Giving future solutions in the form of a model that predicts these frauds
- Detection and investigation of fraud .
- Bypass the identification of potential fraudulent activity and provide a practical framework for dealing with the presence and effects of fraud.



Thanks!

Do you have any questions?

