Fraud Detection

Delpita Putri

Dataset Fraud.CSV

https://www.kaggle.com/datasets/chitwanmanchanda/fraudulent-transactions-data

PROBLEM?

Dataset : Simulates mobile money transactions based on a sample of real transactions extracted from one month

6.362.620 rows, 11 columns

Save

Fraud

Business Value

Detect Fraud

01

02

Safe Banking

03

Happy Customers

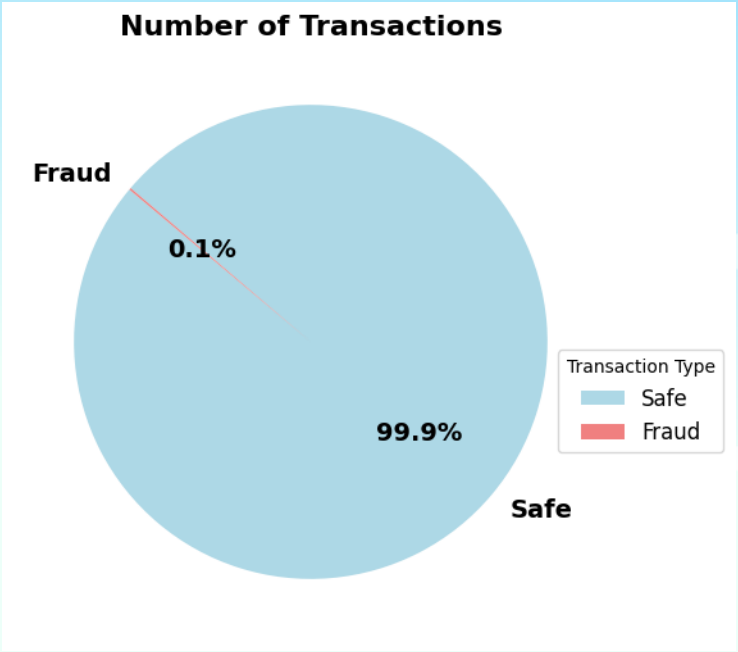
More Customers

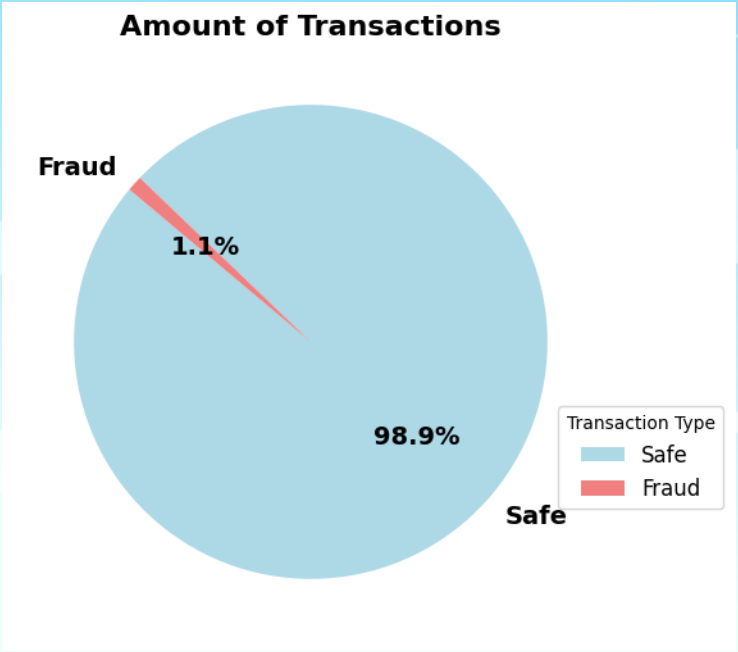
04

05

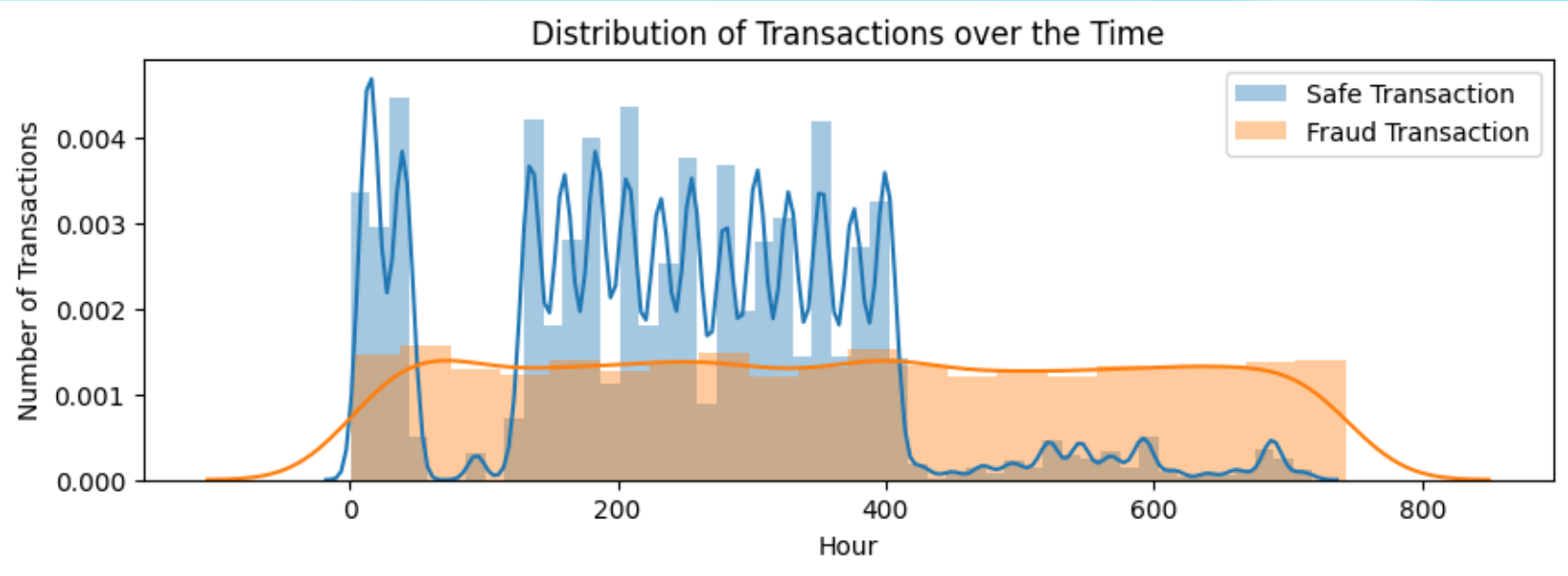
More Profit

Save VS Fraud



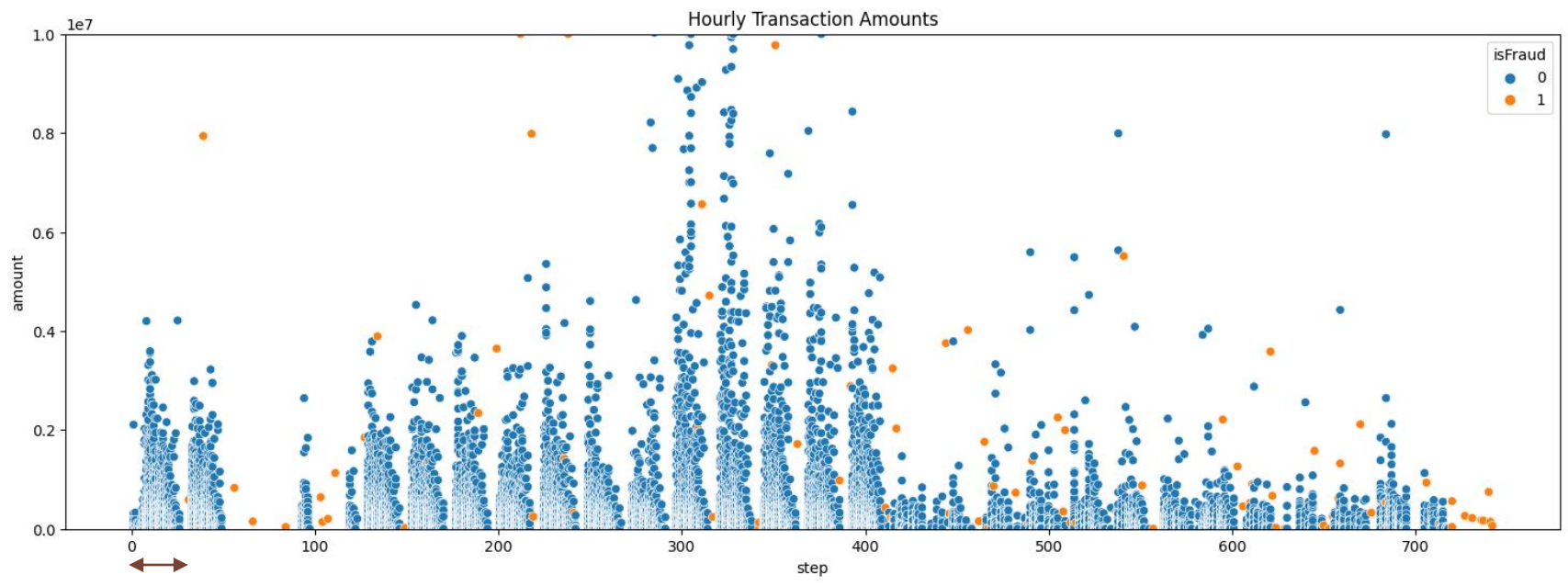


Frequency



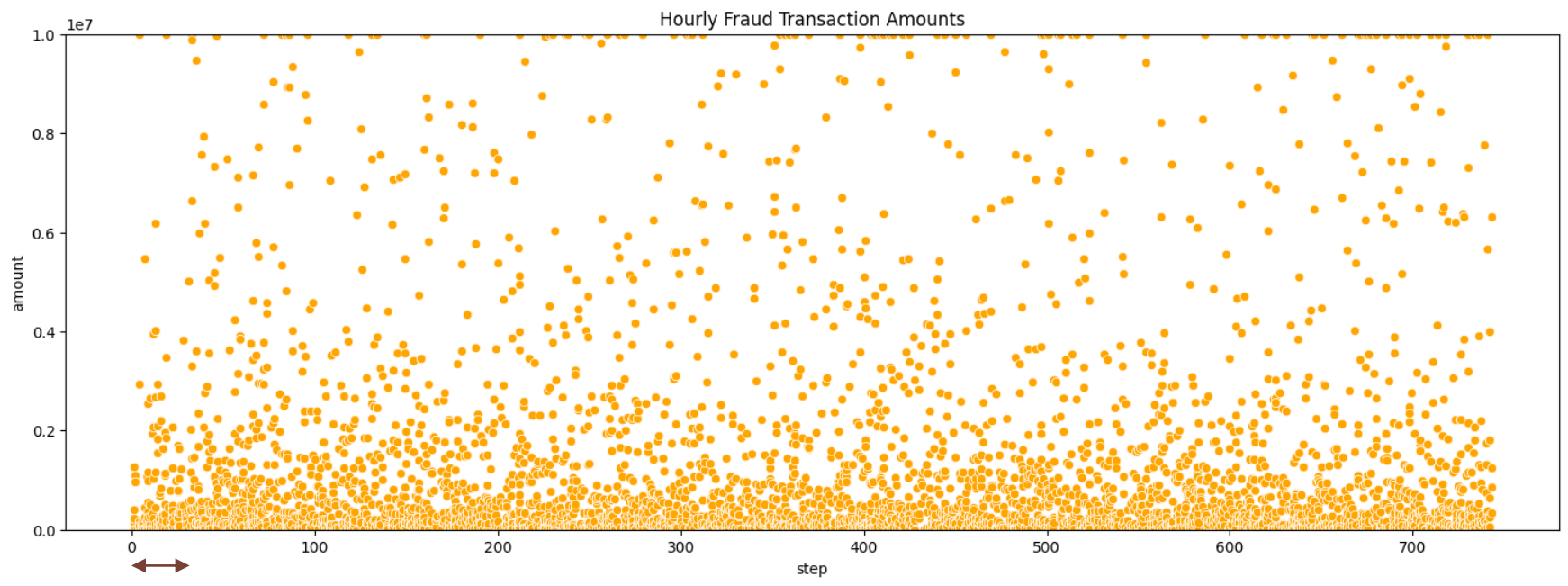
Distrubition plot shows drastic changes in the number of transactions occured each hour

Daily Pattern



Significant Daily Pattern

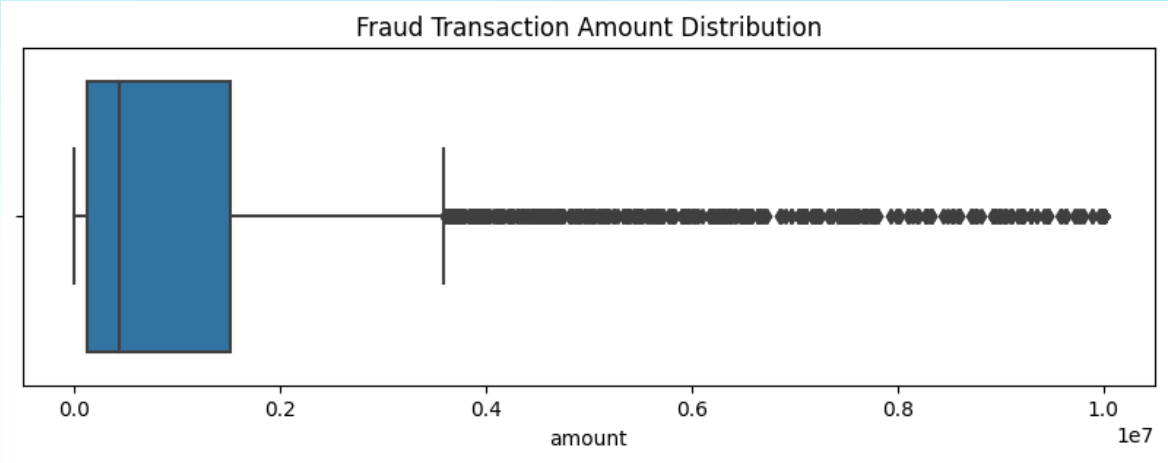
Daily Pattern



24 Hours

No Significant Daily Pattern

Fraud Amount Distribution

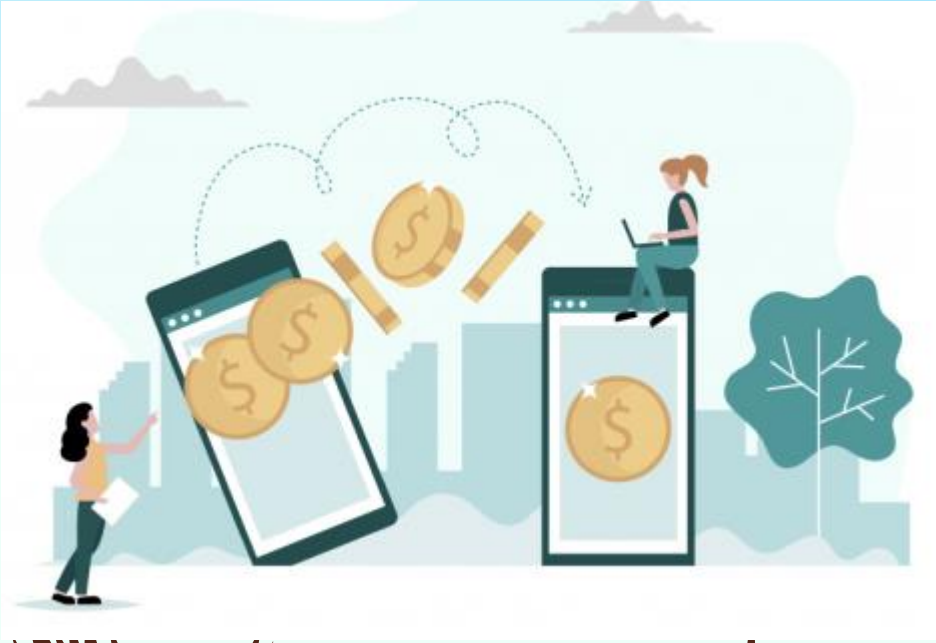



```
1 fraud.amount.describe()
```

```
count    8.213000e+03
mean     1.467967e+06
std      2.404253e+06
min      0.000000e+00
25%      1.270913e+05
50%      4.414234e+05
75%      1.517771e+06
max      1.000000e+07
Name: amount, dtype: float64
```

"Most of the frauds happen below $400,000"

Type of Transactions



- Transfer
- Cash-Out

Descriptive statistics

8.213 Count

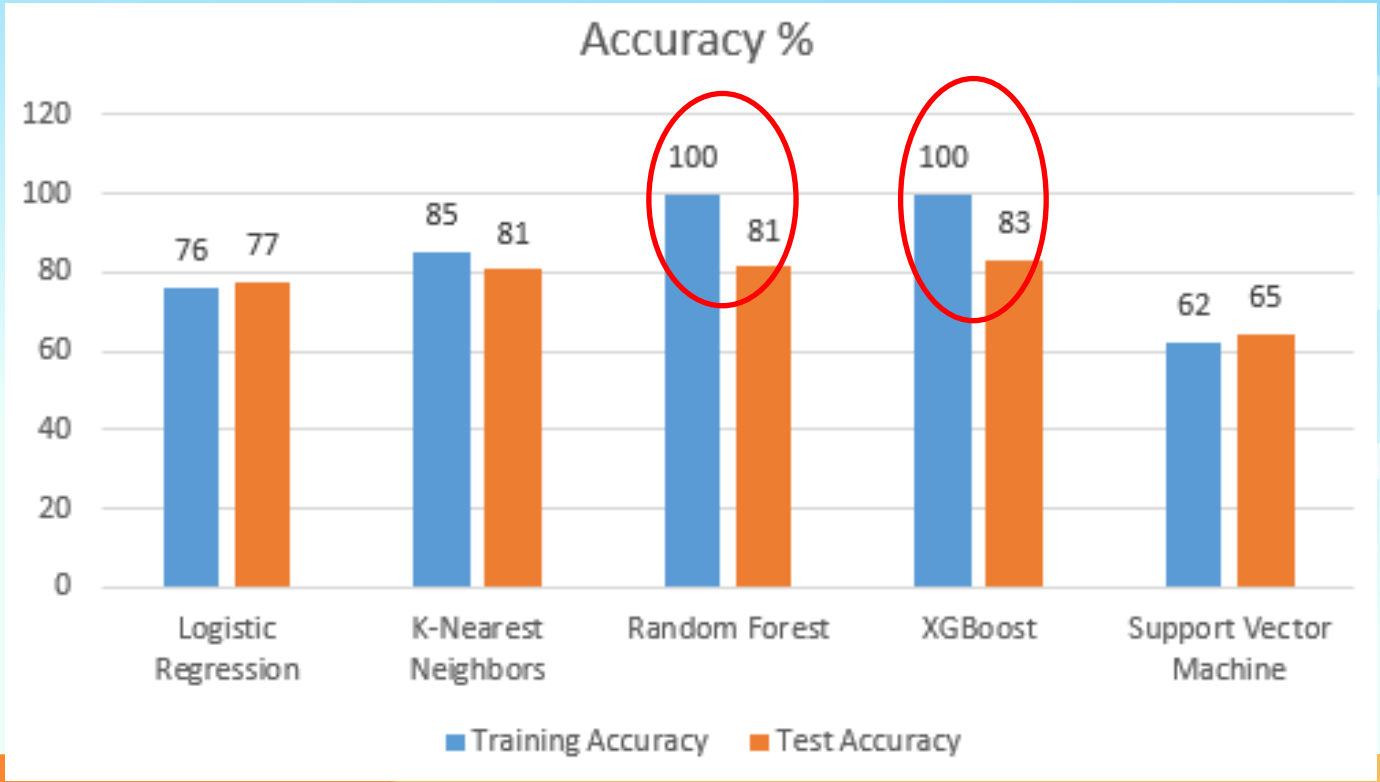
$146 k Mean

$0 Min

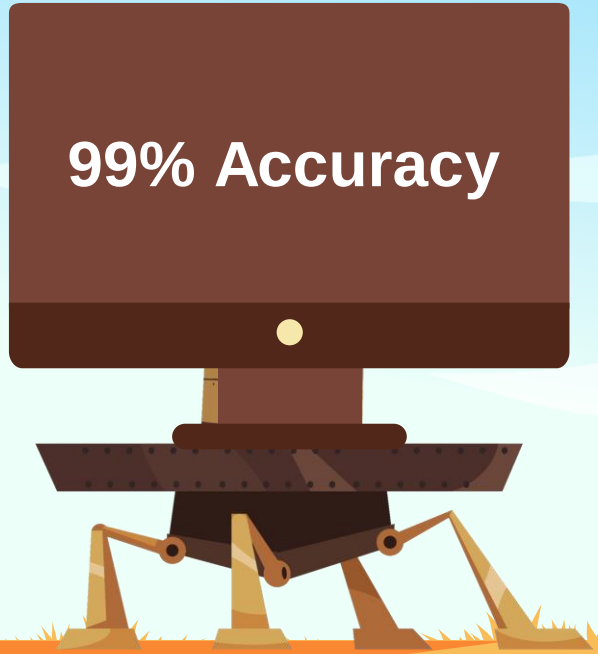
$10M Max

$64 – 10M Range

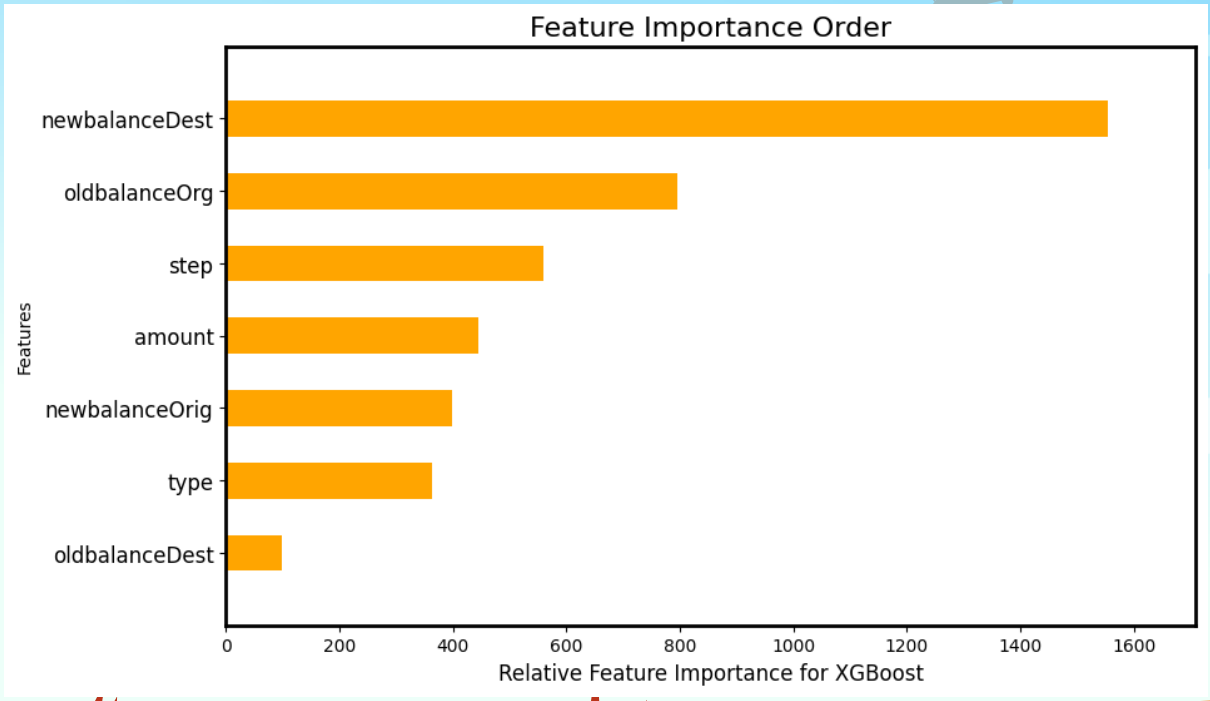
Base Models


Accuracy %
120
100
80
60
40
20
0
76
77
85
81
100
81
100
83
62
65
Logistic Regression
K-Nearest Neighbors
Random Forest
XGBoost
Support Vector Machine
Training Accuracy
Test Accuracy

Best Models



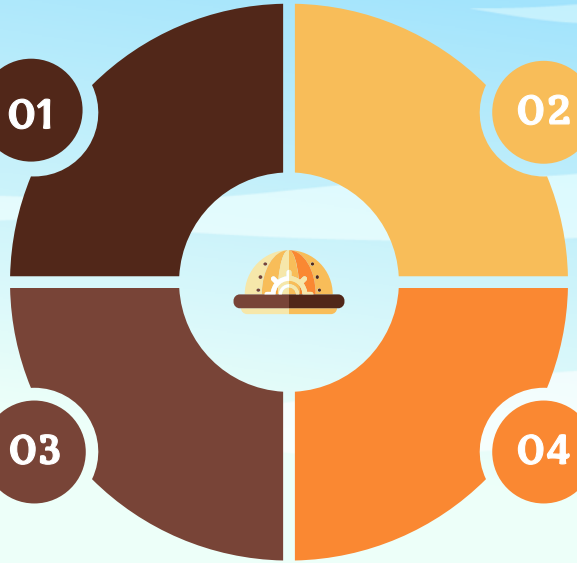
Features


Feature Importance Order
newbalanceDest
oldbalanceOrg
step
amount
newbalanceOrig
type
oldbalanceDest
Features
0
200
400
600
800
1000
1200
1400
1600
Relative Feature Importance for XGBoost

Conclusions

99% acc with XGBoost and Balanced Data

01



02

oldBalanceOrig and newBalanceDest are most influential Features

More fraud transactions in low amounts

03

04

Fraud happens with Transfer and Cash-Out. Debit is very safe

Thank You