

Fraudulent Transaction Prediction

SpringBoard Final Capstone Project





1

MOTIVATION

2

DATASET/EDA

3

FEATURE ENGINEERING

4

MODELING

5

CONCLUSION

1

MOTIVATION

- **Fraudulent transactions are increasing because:**
 - More transactions online than ever
 - No human interactions so possible for criminals to impersonate.
- **Tackling this problem is ever more important**
 - To protect the finance of the consumer and the companies.
 - To provide a peace of mind to the consumer



E-COMMERCE



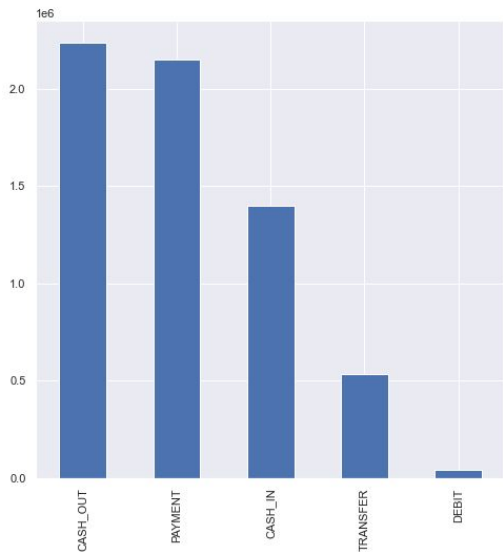
- **Source of data:**
 - Kaggle : <https://www.kaggle.com/datasets/vardhansiramdasu/fraudulent-transactions-prediction>
- **Nature of data:**
 - Credit card transaction
 - 650K rows and 10 columns
 - Contains features of the kind
 - Amount
 - Transfer type
 - Data is labeled

2

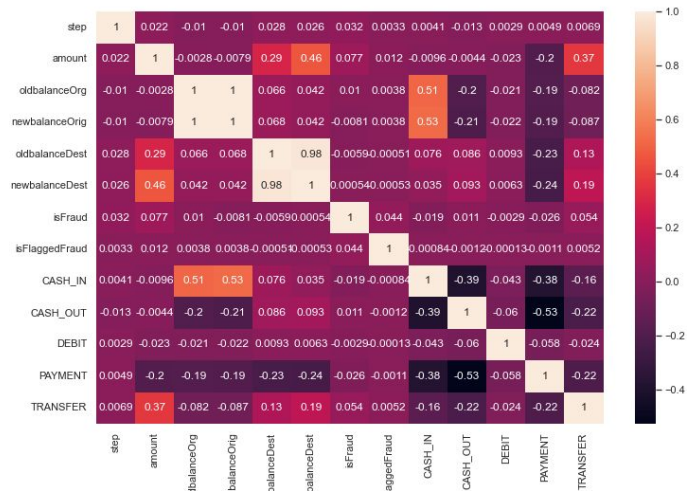
DATASET/EDA

CORRELATION

- Transaction Type



- Correlation between features



- **Checked for multicollinearity using**
 - Using VIF (Variance Inflation factor)
- **Ignored columns that had VIF score of greater than 12**
- **# of Total features**
 - 11 (In this case all features were selected)

3

FEATURE ENGINEERING

BALANCE DATA

- The data had target feature heavily skewed (1000:1)

```
fraud_df['isFraud'].value_counts(normalize = True)
```

```
0    0.998709  
1    0.001291  
Name: isFraud, dtype: float64
```

- Balanced the data using resampling

Considered 5 models

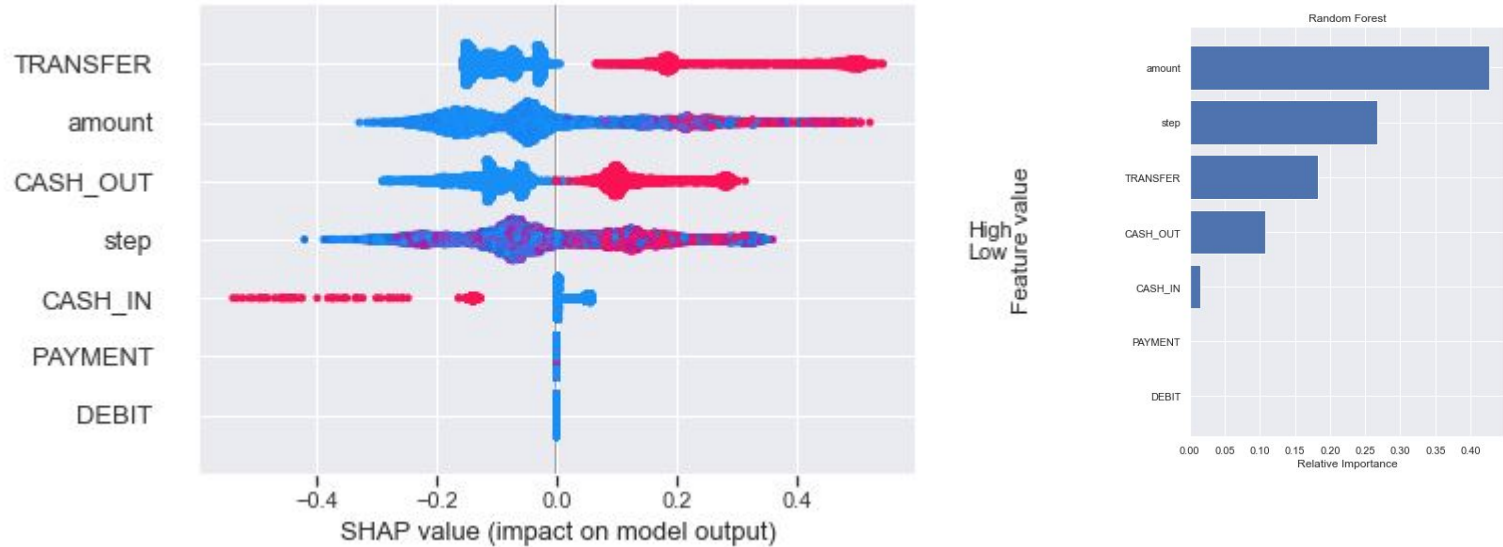
Model	AUC Score
Logistic Regression	0.6162
Random Forest	0.9599
XGBoost	0.9962
Decision Tree	0.8856
Naive Bayes	0.7291

- **Used GridSearch CV**
 - to tune the hyperparameters
- **Picked two top model:**
 - Random Forest
 - XGBoost

4

MODELING

RANDOM FOREST

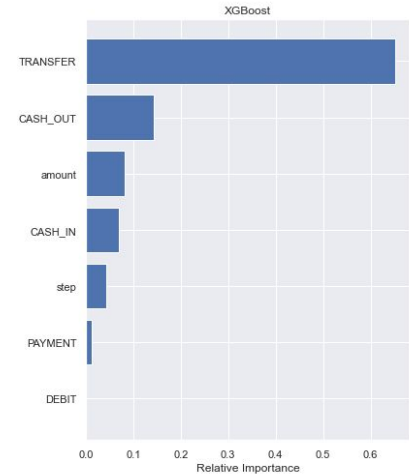
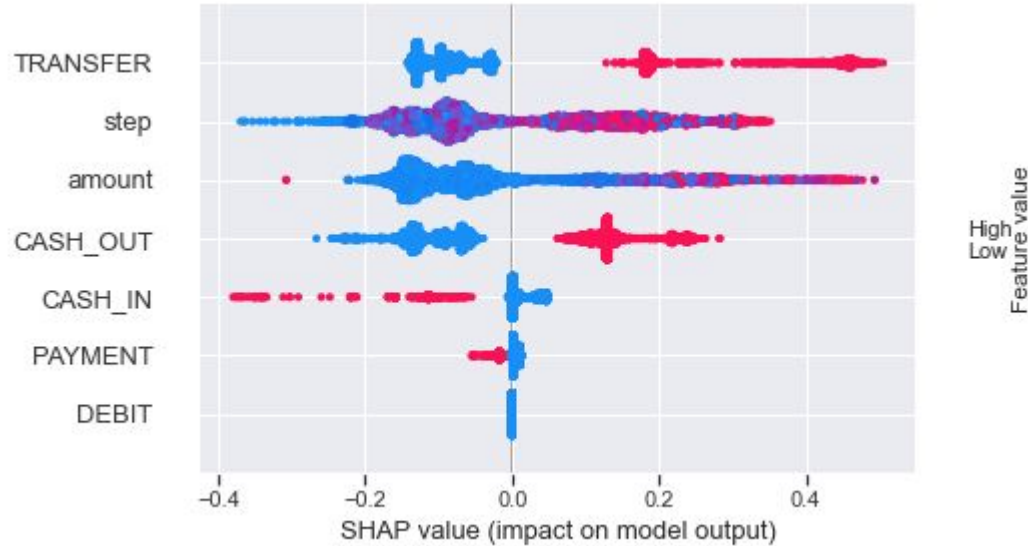


- Inbuilt feature importance tool (right) shows amount to be the most important features but the SHAP analysis (left) shows transfer, and cash out as the most impactful features in this model.

4

MODELING

XGBOOST



- Similarly in XGBoost, inbuilt tool (right) shows transaction type transfer to be most impactful and the SHAP analysis (left) agrees with it.

- In both the models the SHAP analysis shows that transaction type 'TRANSFER' and 'CASH_OUT' and the amount of money during the transaction are the best features to help determine if the transaction is fraudulent.
- Banks and ecommerce platforms should implement second factor authentication for transactions like 'TRANSFER' and 'CASH_OUT' and also flag if the amount in a transaction is sufficiently large.