BLOSSOM BANK

ONLINE FRAUD DETECTION USING MACHINE LEARNING

Chike Chukwumah

TABLE OF CONTENT

INTRODUCTION

- Fraud detection is the process of detecting scams and preventing fraudsters from obtaining money or property through false means. Fraud is a serious business risk that needs to be identified and mitigated in time.

- BLOSSOM BANK also known as BB PLC is a multinational financial services group, that offers retail and investment banking, pension management, asset management, and payment services with its headquarter in London, UK.

EXECUTIVE SUMMARY

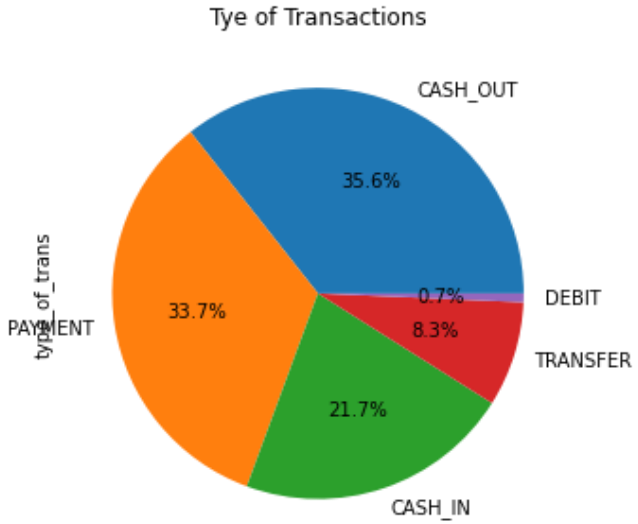
Blossom Bank also known as BB PLC is a multinational financial services group, that offers retail and investment banking, pension management, asset management and payments services, headquartered in London, UK

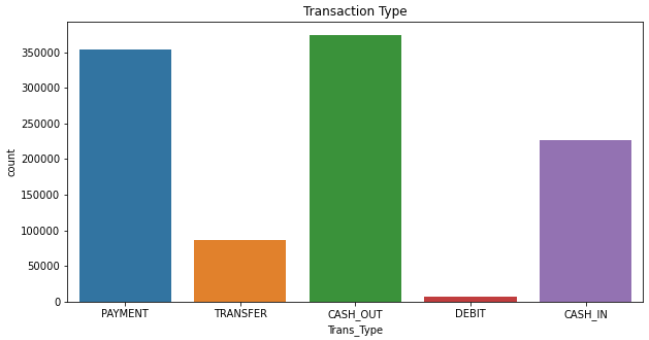
Blossom Bank wants to build a Machine Learning model to predict online payment fraud.

Insights and visualization was done using the concept of Univariate Analysis, Bivariate Analysis and Multivariate Analysis after wrangling the data

Recommendation is based on the ML models used to train the data set and feature engineering and classification for accurate prediction.

EXPLORATORY DATA ANALYSIS



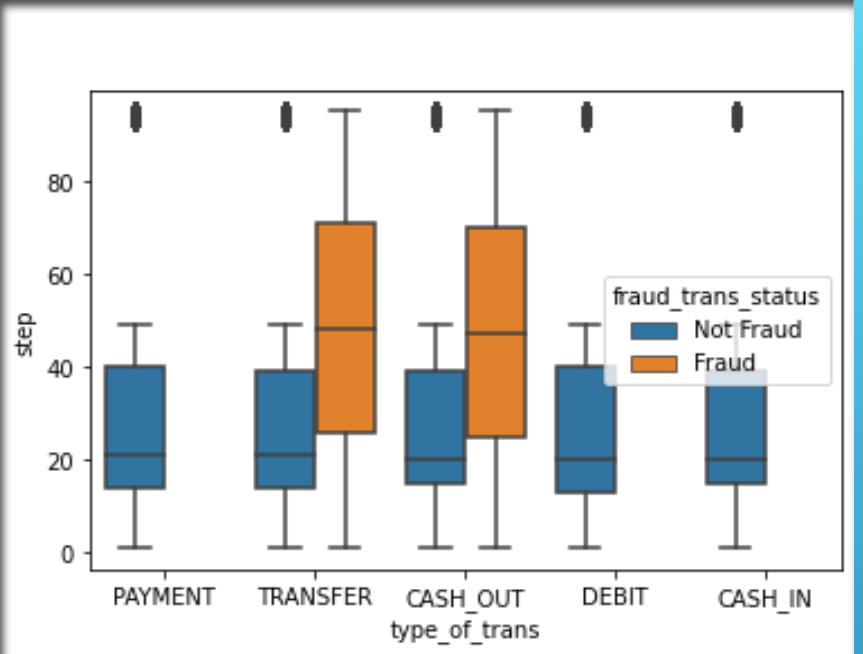


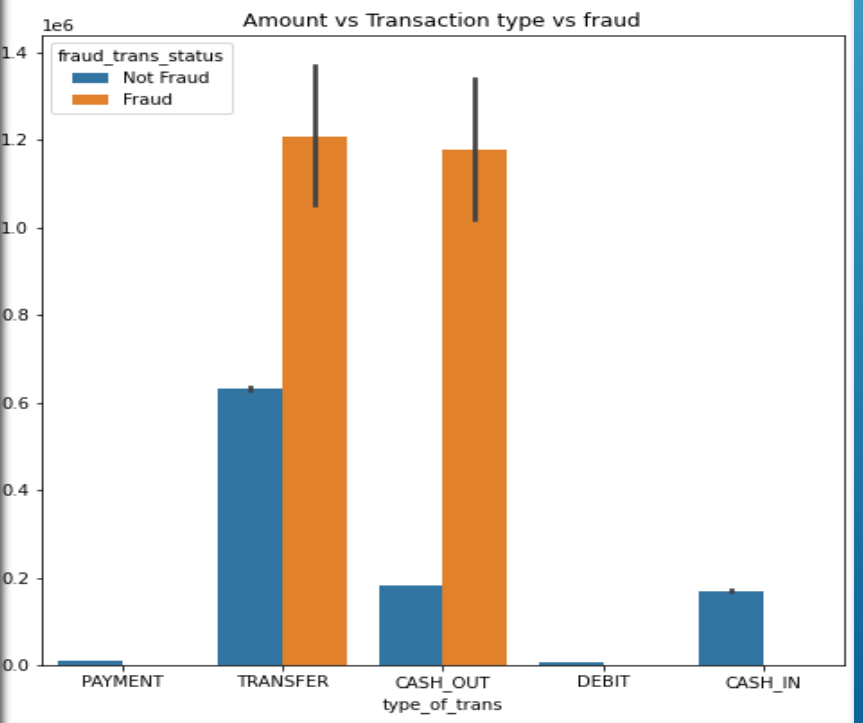
▸They were five main transaction types the bank use as payment type. They include: PAYMENT, TRANSFER, CASH OUT, DEBIT, CASH IN, CASH OUT.

▸ CASH OUT transaction type generated the largest number of transaction and also accounted for 35.6% of the total transaction type.

▸ PAYMENT Type transactions accounted for 33.7% and

▸ DEBIT and TRANSFER transaction were the least payment type accounting for 0.7% and 8.3% respectively.

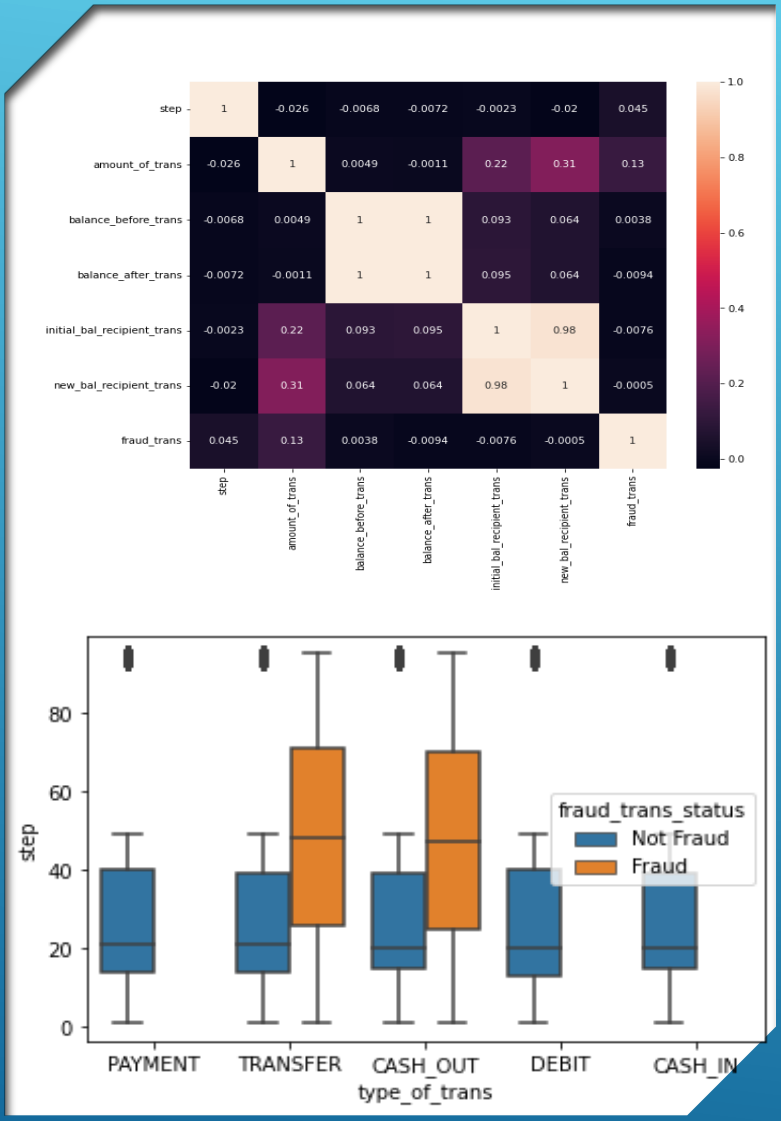




▸TRANSFER type of transactions were noticed to have more fraudulent transactions than other transaction type.

▸This was followed by CASH OUT transaction type.

▸PAYMENT, DEBIT and CASH IN transactions were not fraudulent transactions.



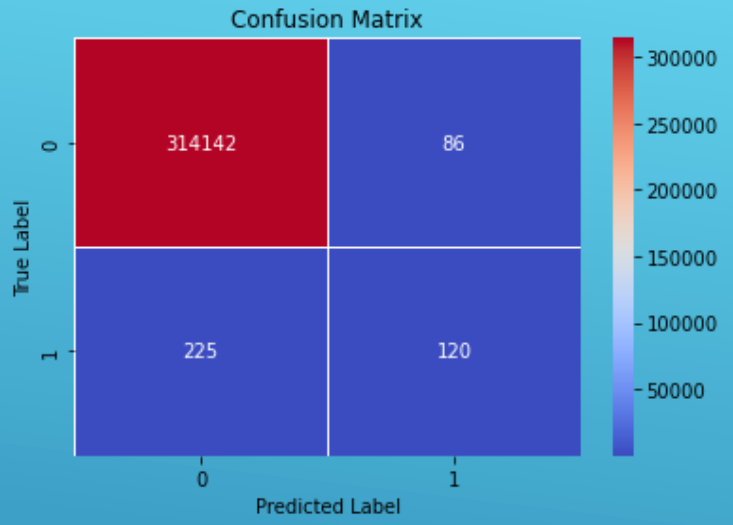
▸Correlation between variables were explored. We noticed a high correlation between balance before and after transaction are highly correlated compared to others.

▸TRANSFER type of transactions were noticed to have more fraudulent transactions than other transaction type.

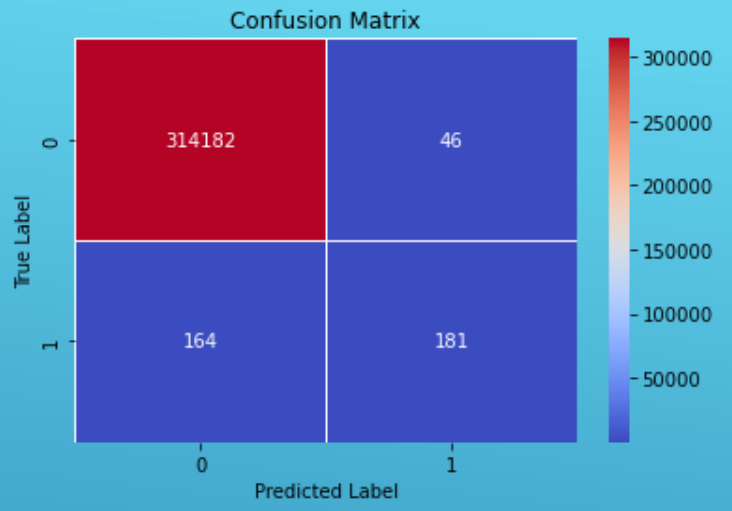
▸This was followed by CASH OUT transaction type.

▸PAYMENT, DEBIT and CASH IN transactions were not fraudulent transactions.

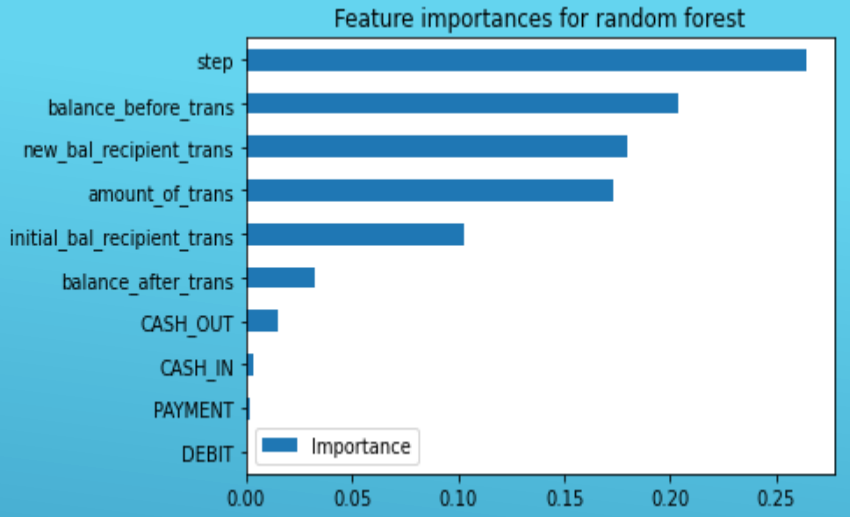
MODEL EVALUATION



Logistic Regression – Precision of 38%, Recall of 58% and accuracy of 0.9990

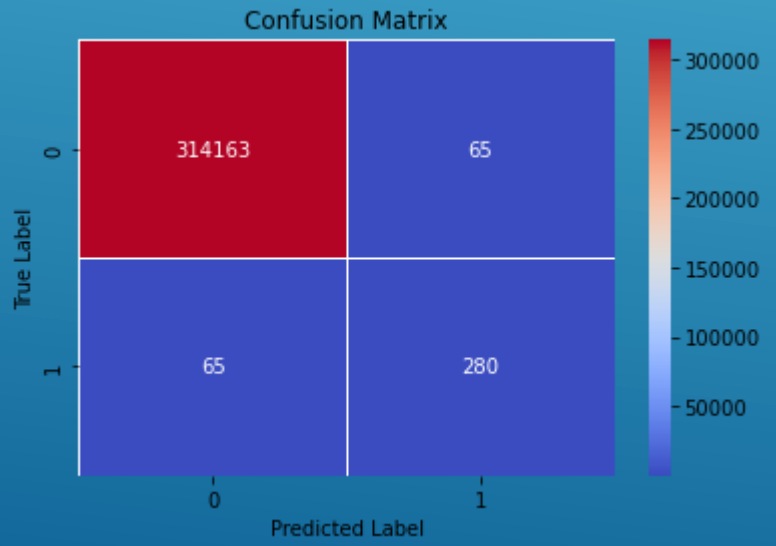


KNN – Precision of 52%, Recall of 80% and accuracy of 0.9993

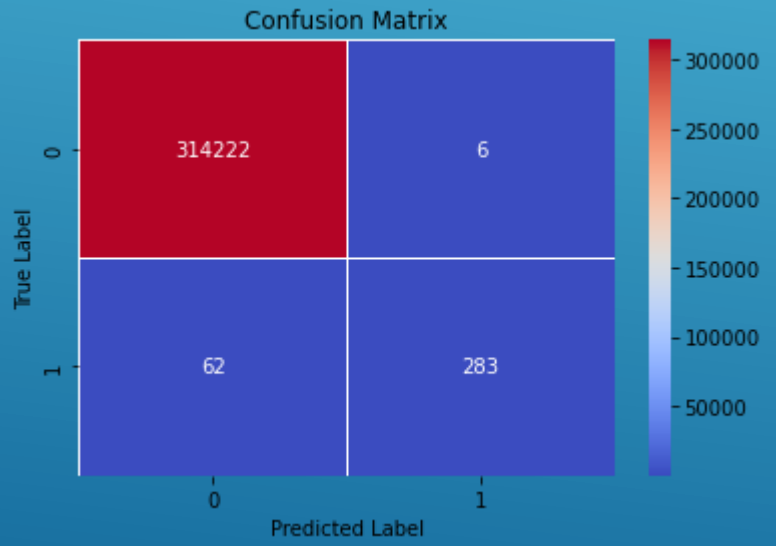


Feature importance shows how much each features or variables will contribute to the model prediction, and we can see that step has a larger relevance followed by balance before transaction in that order.

Confusion matrix shows the performance of a classification model on a given set for which true values are known



DecisionTreeClassifier – Precision of 81%, Recall of 81% and accuracy of 0.9995



RandomForestClassifier – Precision of 82%, Recall of 98%, and accuracy of 0.9997

0

CONCLUSION AND RECOMMENDATION

Random Forest performs best with recall and precision of 98% and 82% respectively, which is important for our problem statement where false negative is our priority.

Random Forest Classifier model should be deployed by Blossom bank because for this business problem.

Since CASH OUT type of Transaction is the highest, I would recommend a cash limit policy to mitigate fraud in this type of transaction.

Repeated wrong PIN or Password – Blossom bank should halt the transaction and alert the customer immediately.

Increased cybersecurity for banking websites and mobile applications.

Blossom bank should enable Two factor authentication for Transfer type transaction to confirm genuity of the transaction with the customer.

THANK YOU!!!