

Blossom
Bank

Online Fraud Detection Using Machine Learning



A Report By: Odidi Naomi Tracy

Executive Summary



Blossom Bank

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. The bank's value propositions included ensure the safety and credibility of it's transactions.

Problem Statement

Blossom Bank wants to build a Machine Learning model to detect and predict online payment fraud within it's system. By using supervised machine learning to identify existing fraudulent transactions and understanding their behaviours, and unsupervised machine learning to predict and flag future fraudulent transactions, I will be saving the bank against financial losses and also recommending strategies and policies that ensures they remain secure.



Insights



The distribution charts for the variables **amount**, **newbalanceDest**, and **oldbalanceOrg** were all positively skewed, indicating that the outliers were farther to the right.



Random Forest provided the best accuracy of the three supervised machine learning models used, with a score of 99.97%.



Of the five (5) transaction types the bank operates, PAYMENT, TRANSFER, CASH OUT, DEBIT, CASH IN, CASH OUT was the most conducted transaction type and generated the most revenue at 41.3%.



While there was a strong correlation between the balances, old and new, other variables showed relatively no correlation, suggesting that the results of one did not affect the results of the other.



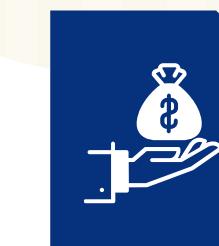
The largest transaction a consumer made was for £1,000,000, and the most transactions a recipient got was 98, indicating that either our customers have several accounts or those transactions were fraudulent.



Of the transaction types, TRANSFER had a higher number of fraudulent transactions followed by CASH OUT.

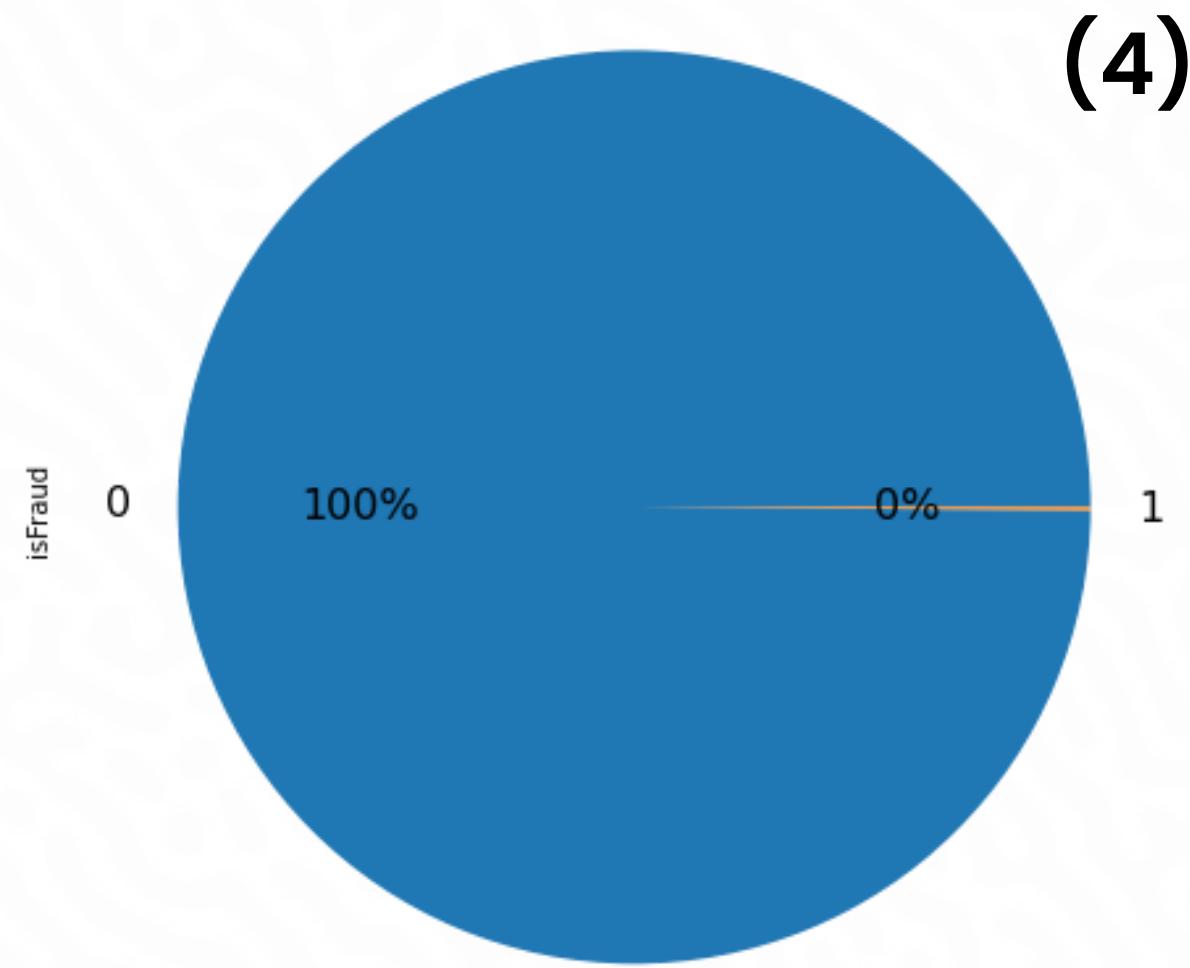
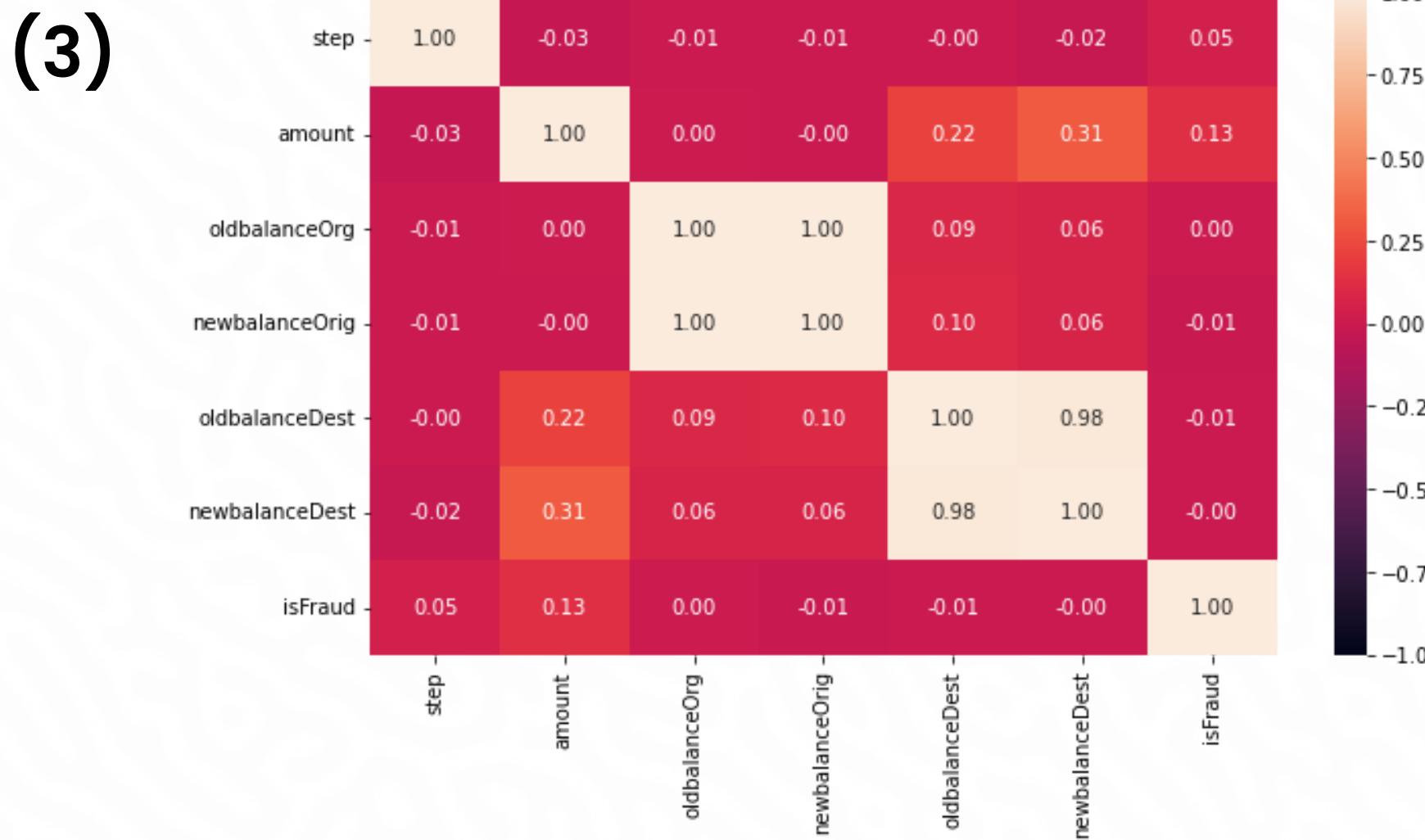
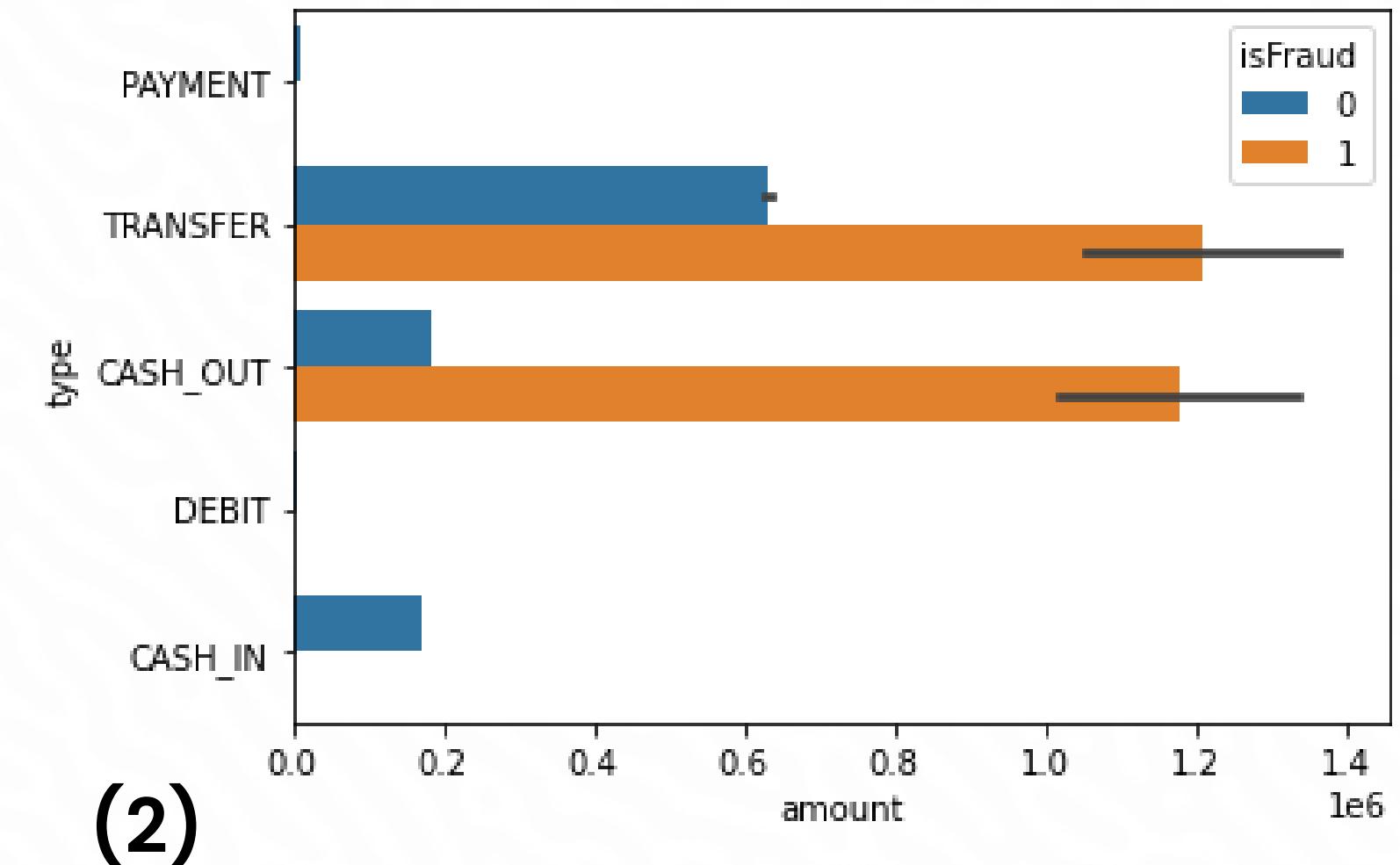
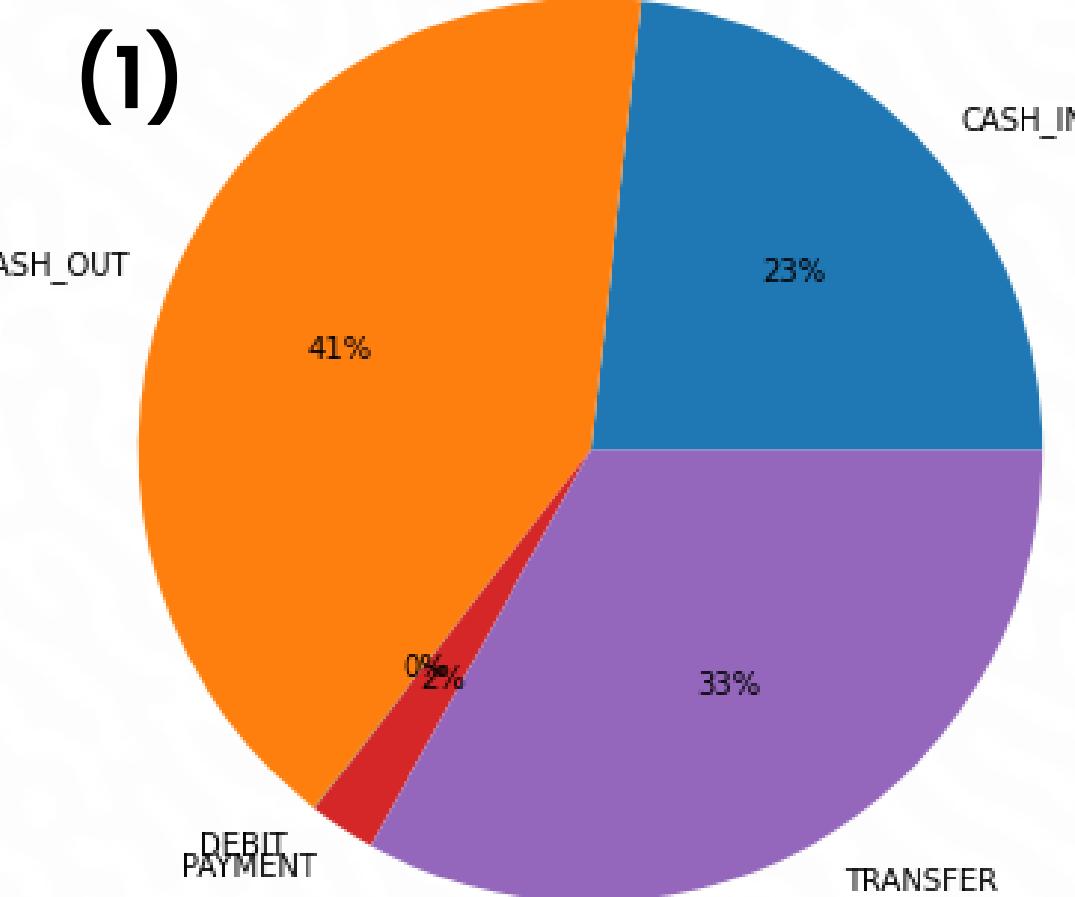


Although 1,048,575 transactions were made, 1142 were fraudulent. This however suggests flaws in our data set.

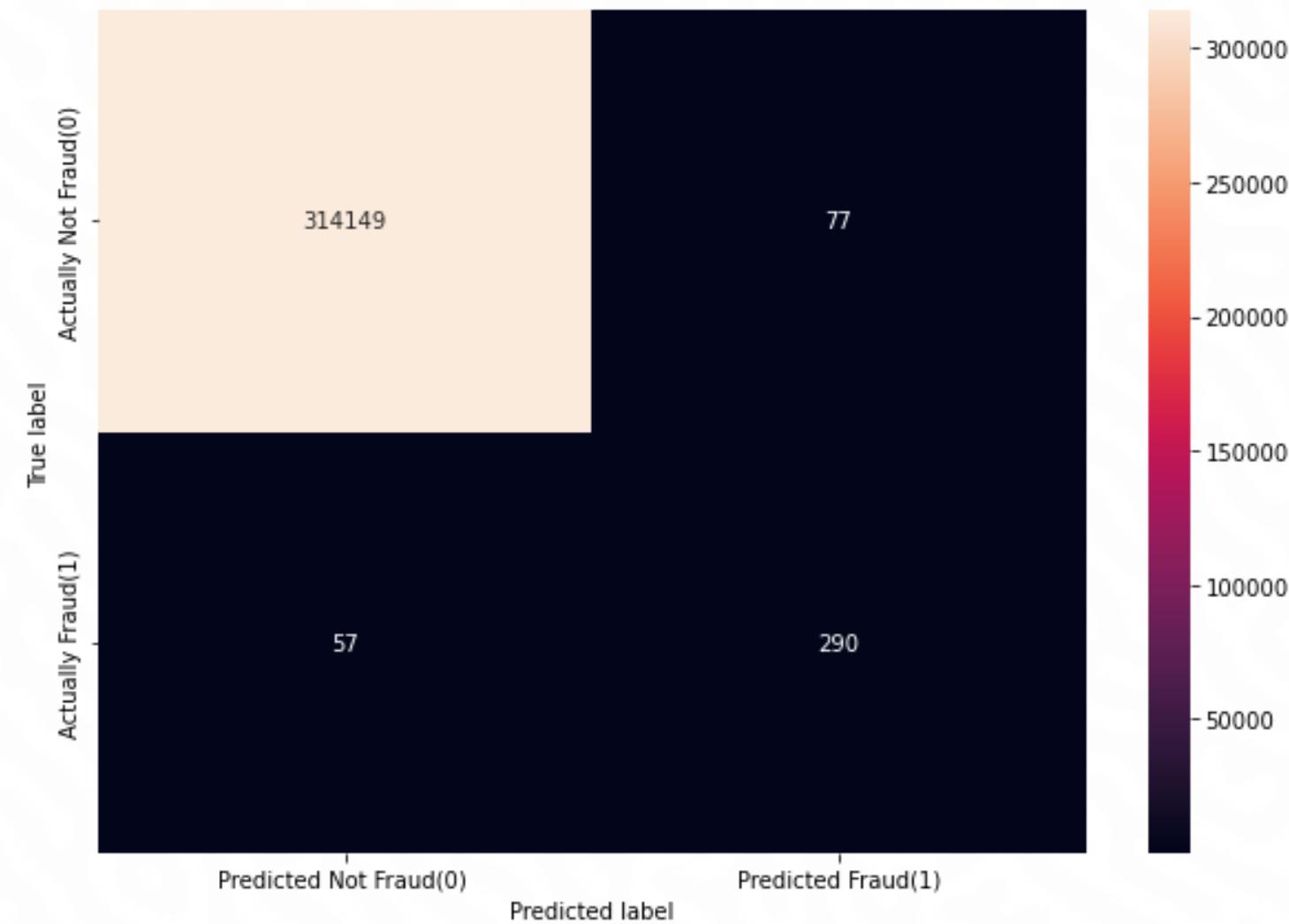


Customer C1334405552 is the top loser with £1277212.77 in losses and customer C615227407 benefitted the most from fraudulent transactions at £10026725.16

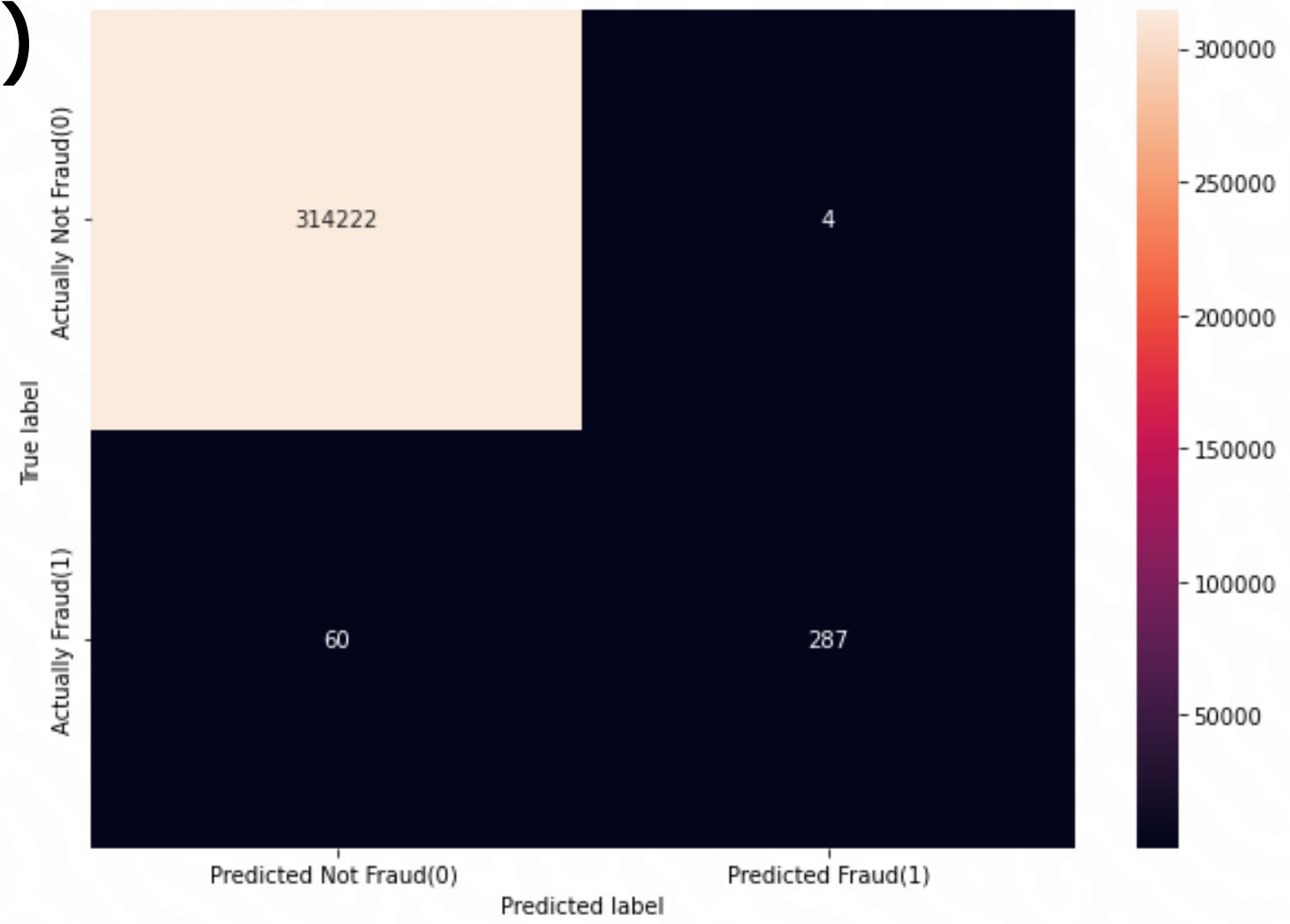
Visualization



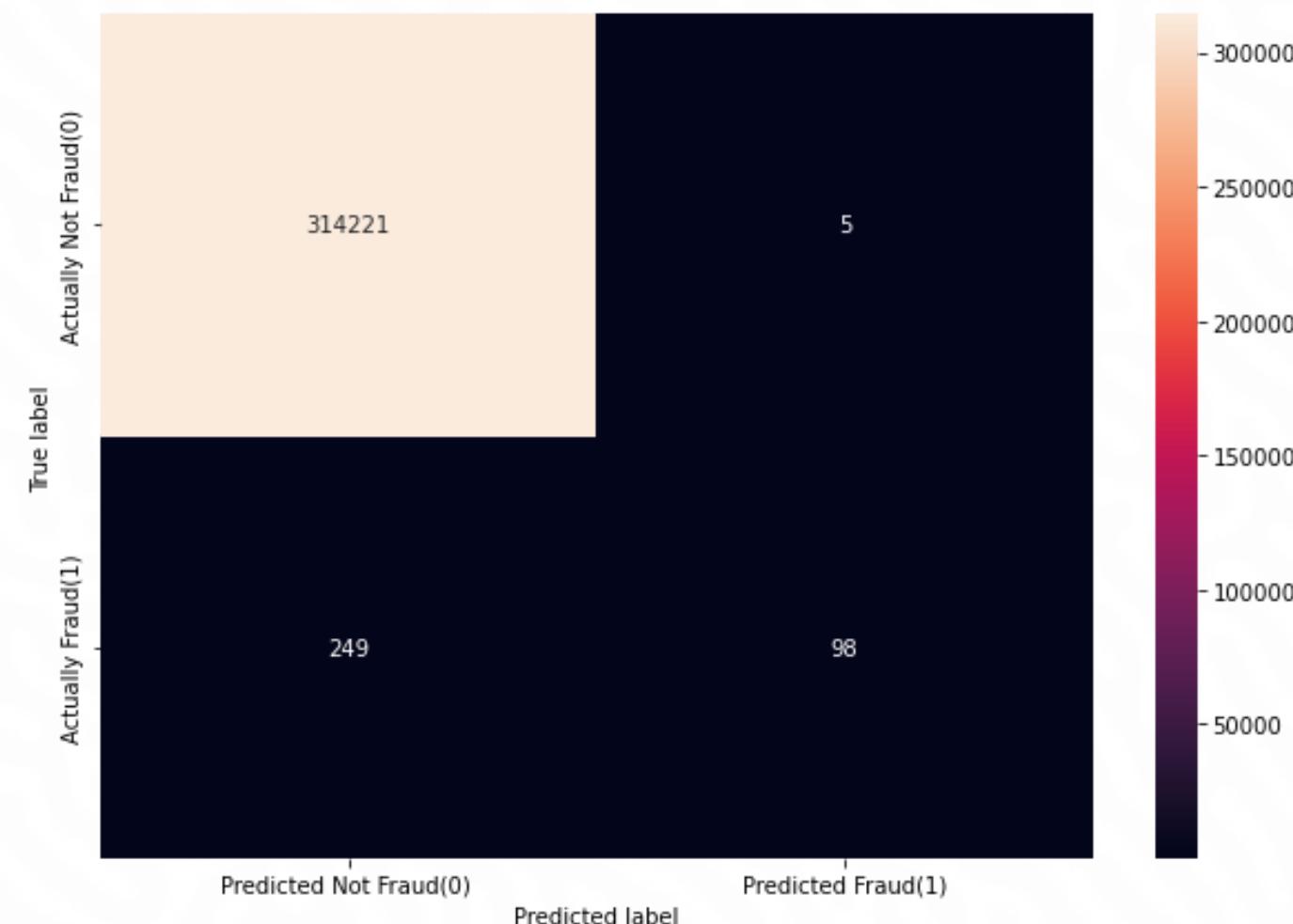
(5)



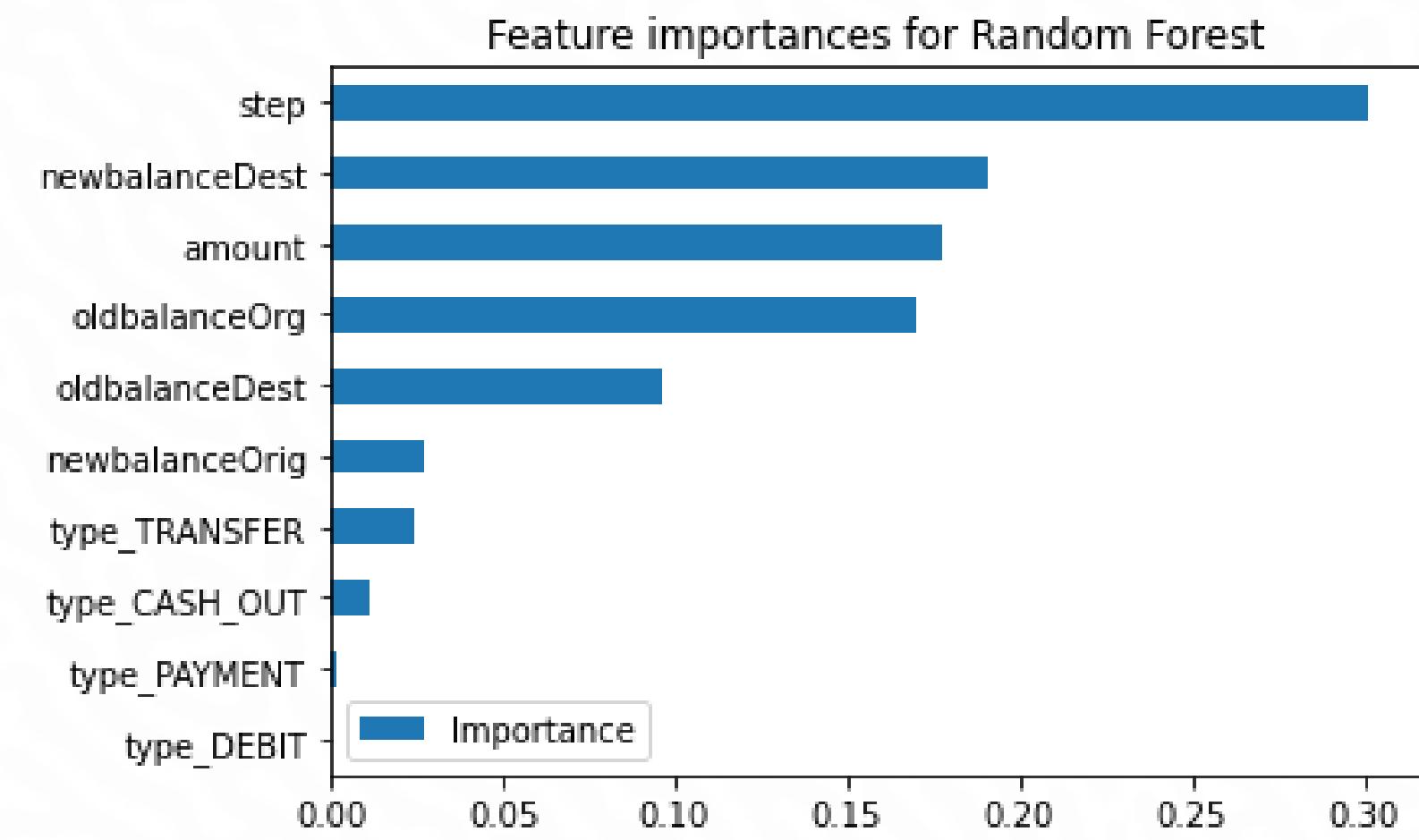
(6)



(7)



(8)



Visualization Summary

(1)

A piechart representing the revenue generated by each transaction type.

(2)

A bar chart representing the transaction types with the most fraudulent activities.

(3)

A heatmap showing the relationship/correlation between all the variables.

(4)

A piechart showing the distribution of fraudulent transactions against normal transactions.

(5)

A heatmap showing the results of predicting fraudulent transactions using a decision tree model. While 57 transactions turned out to be fraudulent despite the prediction stating they were normal, 314149 transactions were in fact normal as expected. The same is true of the 290 predicted fraudulent transactions coming true and the 77 transactions that were actually normal despite the prediction that they were fraudulent.

(6)

A heatmap showing the results of predicting fraudulent transactions with the Random Forest Model. While 60 transactions turned out to be fraudulent despite the prediction stating they were normal, 314222 transactions were in fact normal as expected. The same holds true for 4 transactions that turned out to be normal despite the prediction indicating they were fraudulent and 387 fraudulent transactions that actually occurred.

(7)

A heatmap showing the results of predicting fraudulent transactions with a logistic regression model. While 249 transactions turned out to be fraudulent despite the prediction stating they were normal, 314221 transactions were in fact normal as expected. The same is true for 5 transactions that turned out to be legitimate despite the prediction saying they were fraudulent, and 98 fraudulent transactions were real..

(8)

A feature importance chart displaying the variables' relative weights or levels of relevance for the machine learning model.

Recommendations

01.

I suggest the bank engage data engineers to make sure the dataset is accurate and of high quality. This will enable the business to produce machine learning models that anticipate outcomes more accurately and produce better results overall.

02.

The number of steps necessary for a transaction to be approved by the bank should also be increased. The higher the steps, the better.

03.

As the CASH OUT transaction type reported high fraud activity despite having a minor revenue impact, stronger cash withdrawal regulations should be put in place. I also suggested that cash withdrawal officers undergo regular evaluations.

04.

I recommend that the bank invests in ongoing training and education for their workers so they can stay current on new technologies and combat online fraud within their system.





**Odidi Naomi
Tracy**

About Me

Hi! My name is Odidi Naomi Tracy.

I am a mathematician with a flair for analysis. My areas of expertise are in data analysis and visualization, data modeling, machine learning, business development, and reporting. I excel at picking up new skills and tools quickly. I currently have experience with GitHub, Excel, PowerBI, and Python libraries like Pandas, Numpy, Matplotlib, Seaborn, and ScikitLearn. My portfolio includes projects like Sentiment Analysis, Customer Segmentation Using RMF Model, Online Fraud Detection, and Loan Prediction using Python and Machine Learning techniques like Clustering, Regression & Classification.

I have a love for using data to develop tech-enabled, strategic business solutions that encourage revenue growth. I am a recipient of the Women Techster Fellowship class of 2023, which Tech4Dev offers in partnership with Microsoft and the Islamic Development Bank. I'm currently undergoing a master's program in Actuarial Science.

Connect with me via [LinkedIn](#) or send me a [mail](#). Opportunities in data science, finance, and insurtech are always welcome to me. You can browse my [portfolio](#) and essays on [medium](#).