Blossom Bank

Online Fraud Detection Using Machine Learning

A report by: Bolanle Oyelowo

Executive summary

An international financial services business with its headquarters in London, UK, Blossom Bank, commonly known as BB PLC, provides retail and investment banking, pension management, asset management, and payments services.

Problem Statement

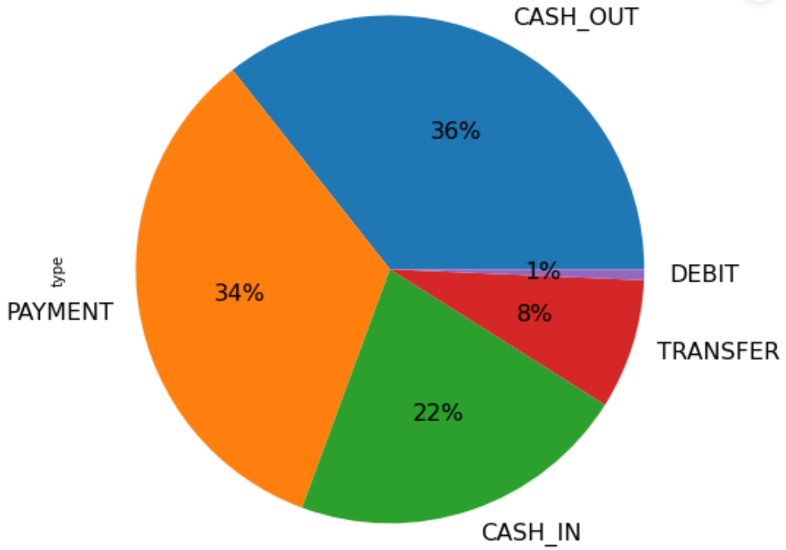
Blossom Bank's goal is to create a machine learning model that can foresee online payment fraud. The bank would be protected against financial losses by using supervised machine learning to identify current fraudulent transactions and comprehend their behaviours, as well as unsupervised machine learning to predict and flag future fraudulent transactions. These techniques would also help me to recommend strategies and policies that will ensure Blossom Bank's continued security.

Insights

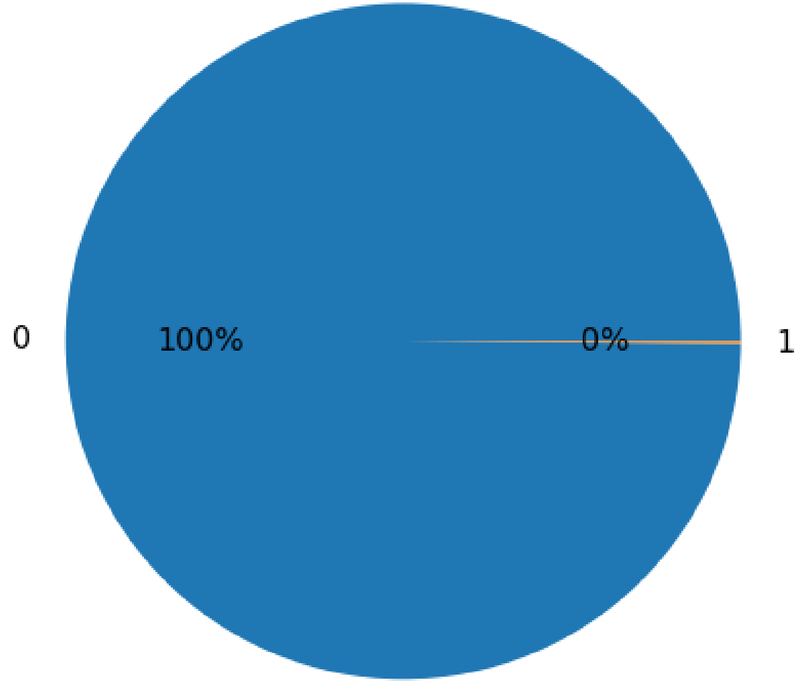
- The data provided by Blossom Bank has Fraudulent and Non-fraudulent transactions.
- The data provided has 1048575 records and 10 fields/columns.
- There are fives types of transaction which are: Cash_in, Cash_out, Transfer, Payment and Debit.
- Out of the five types of transaction, Cash_out and transfer have fraudulent transactions.
- From the description of data and pie chart, it was observed that the most common type of transaction is "CASH_OUT", with a frequency of 373641 with 36% as shown on the pie chart, followed by Payment and Debit is the least common type of transaction used.
- It was observed that there are more non-fraudulent transactions than fraudulent transactions. Non-fraudulent transaction is almost 100%.

Visualization

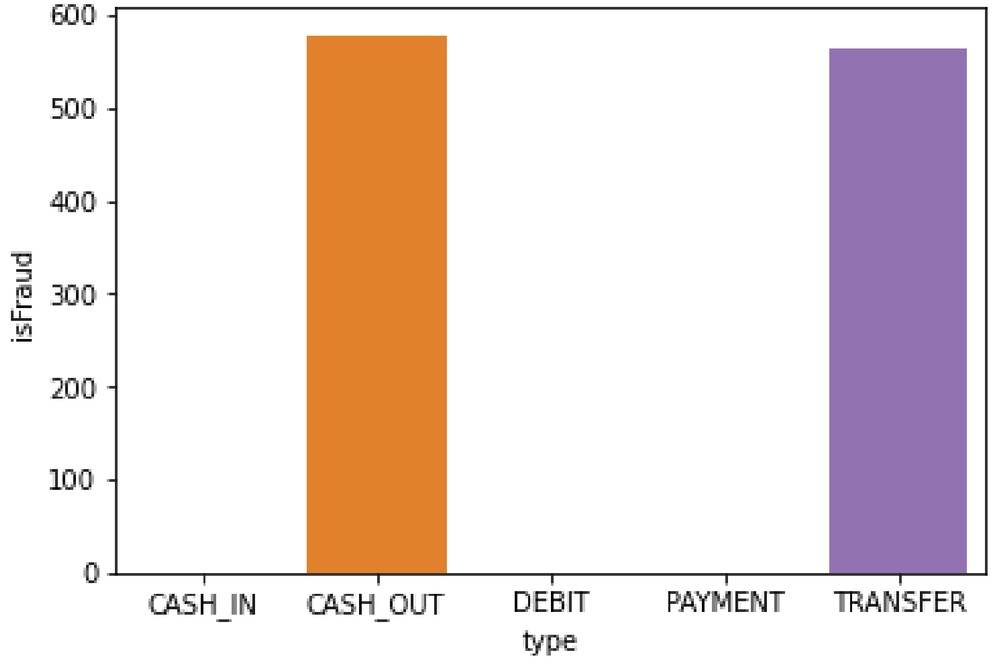
(1)



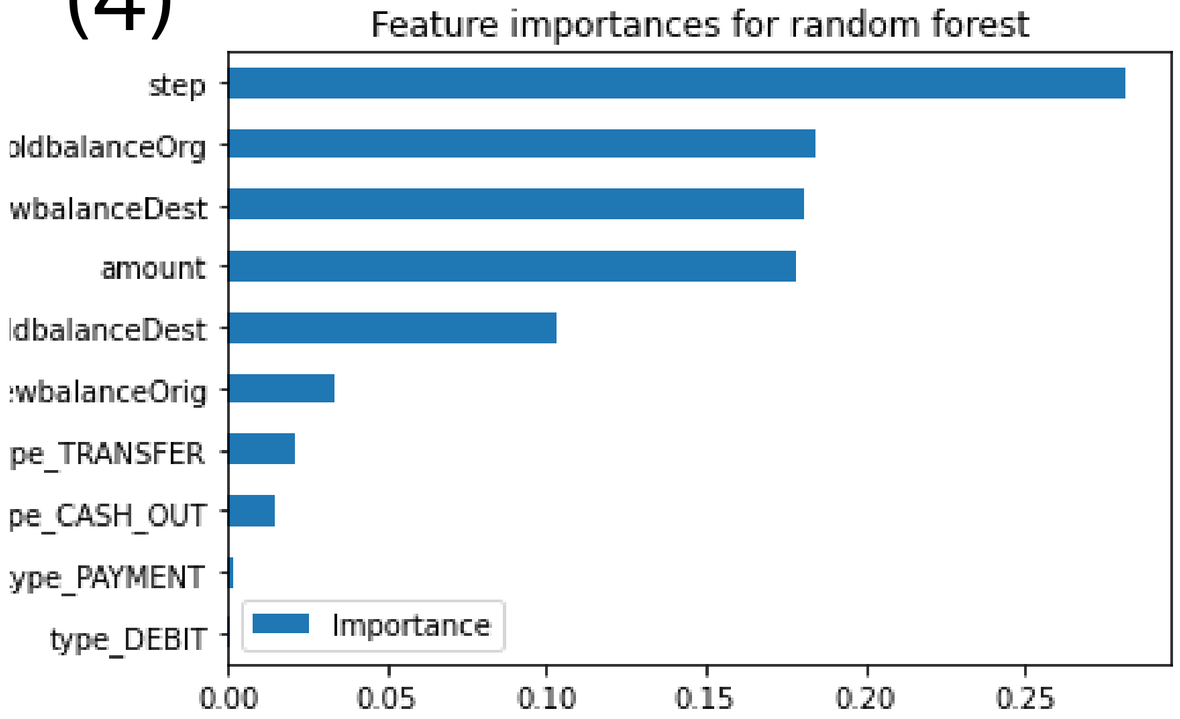
(2)



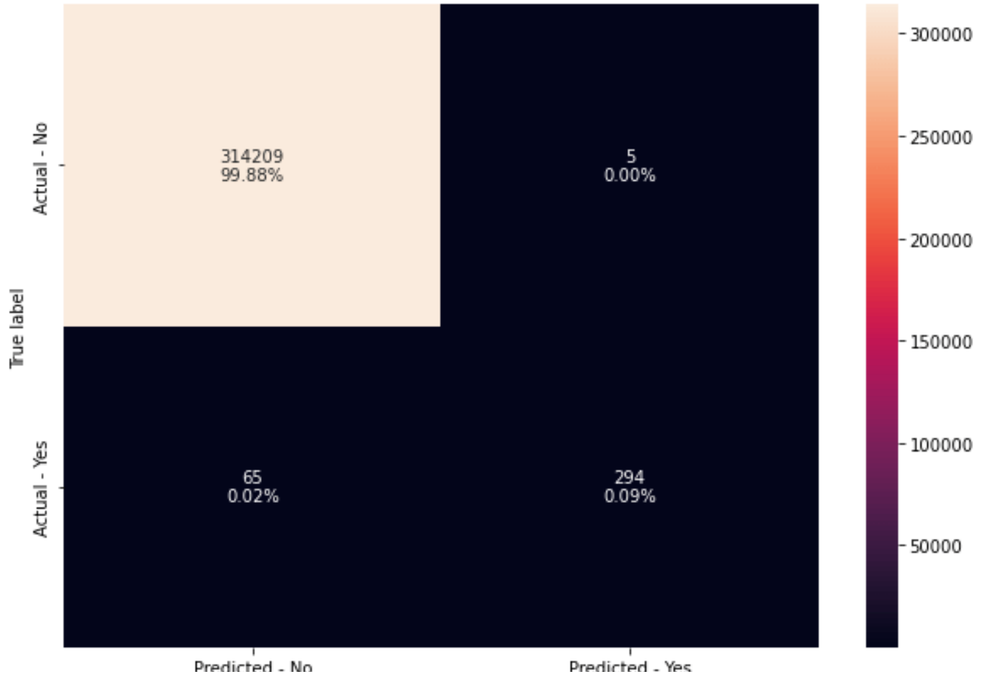
(3)



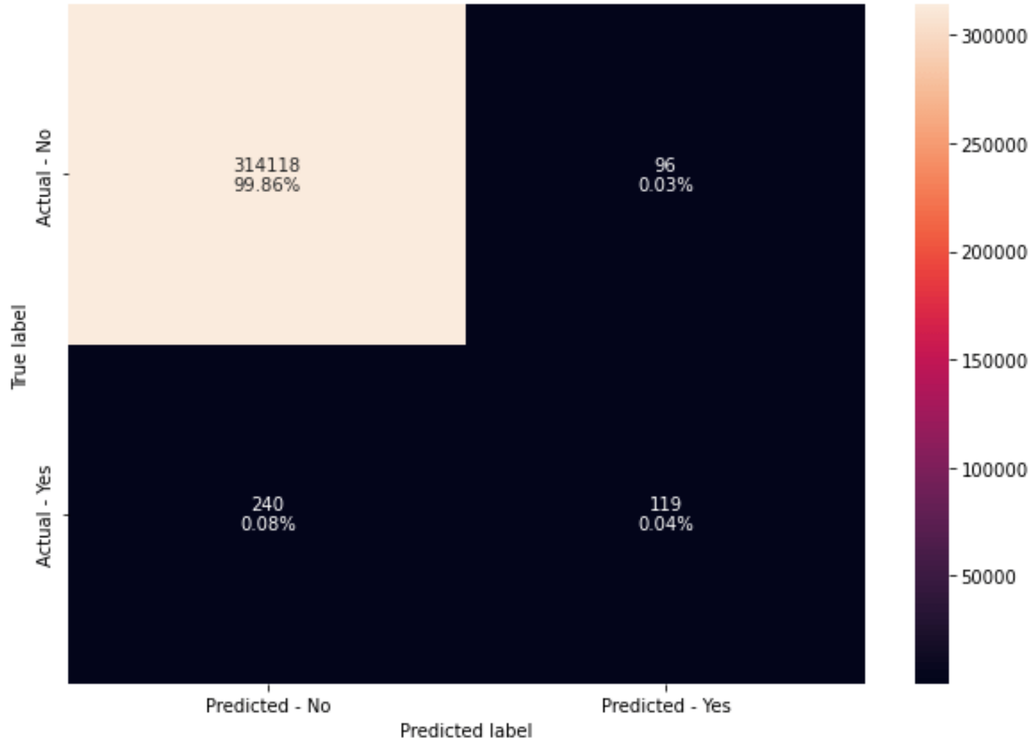
(4)



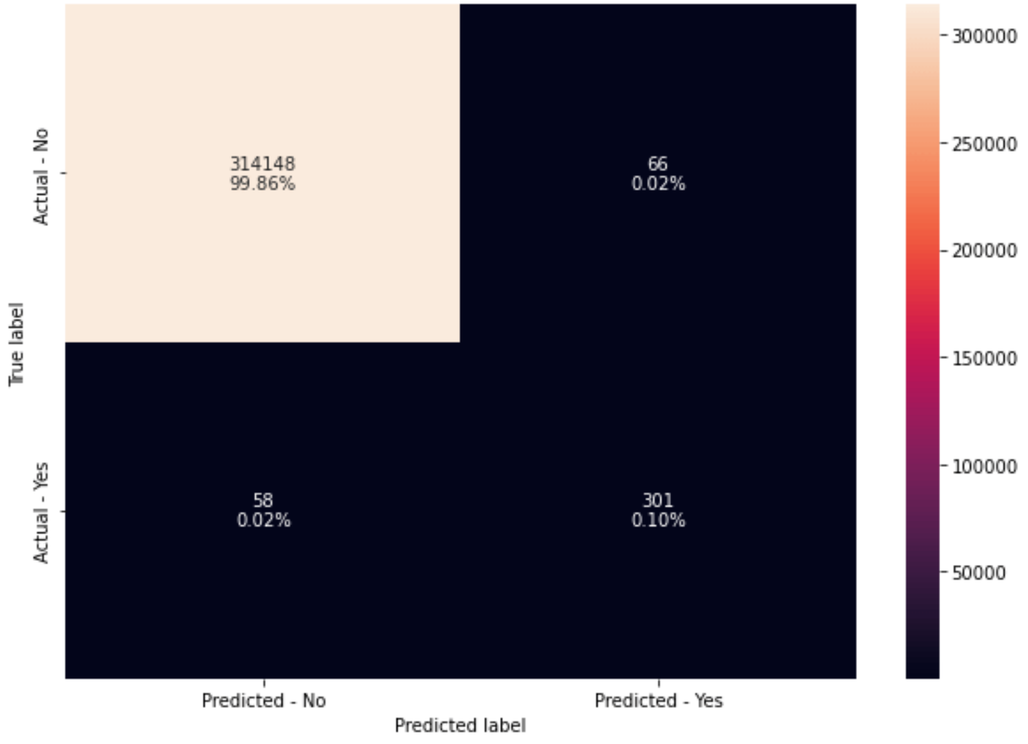
(5)


314209
99.88%
5
0.00%
65
0.02%
294
0.09%
Actual - No
Actual - Yes
True label
Predicted - No
Predicted - Yes
300000
250000
200000
150000
100000
50000

(6)


314118
99.86%
96
0.03%
240
0.08%
119
0.04%
Actual - No
Actual - Yes
True label
Predicted - No
Predicted - Yes
Predicted label
300000
250000
200000
150000
100000
50000

(7)


314148
99.86%
66
0.02%
58
0.02%
301
0.10%
Actual - No
Actual - Yes
True label
Predicted - No
Predicted - Yes
Predicted label
300000
250000
200000
150000
100000
50000

(8)


step
amount
oldbalanceOrg
newbalanceOrig
oldbalanceDest
newbalanceDest
isFraud
1 -0.026 -0.0068 -0.0072 -0.0023 -0.02 0.045
-0.026 1 0.0049 -0.0011 0.22 0.31 0.13
-0.0068 0.0049 1 1 0.093 0.064 0.0038
-0.0072 -0.0011 1 1 0.095 0.064 -0.0094
-0.0023 0.22 0.093 0.095 1 0.98 -0.0076
-0.02 0.31 0.064 0.064 0.98 1 -0.0005
0.045 0.13 0.0038 -0.0094 -0.0076 -0.0005 1
step
amount
oldbalanceOrg
newbalanceOrig
oldbalanceDest
newbalanceDest
isFraud

Visualization Summary

(1) A piechart representing the percentage of each type of transaction.

(2) A piechart showing the distribution of fraudulent transactions against non-fraudulent transactions.

(3) The bar chart shows the avenue at which fraudulent actions take place.

(4) This graph shows the level of importance of eacf metric (feature)

(5) The heatmap shows results of predicting fraudulent transactions with the Random Forest Model.While 294 transactions were actually fraudulent as predicted, 65 transactions were actually fraudulent even with the fact that it was predicted to be non-fraudulent. 5 transactions were actually non-fraudulent even when it was predicted to be fraudulent. 314148 transactions were actually non-fraudulent as predicted.

(6) The heatmap shows results of predicting fraudulent transactions with the Logistic Regression Model Model.While 119 transactions were actually fraudulent as predicted, 240 transactions were actually fraudulent even with the fact that it was predicted to be non-fraudulent. 96 transactions were actually non-fraudulent even when it was predicted to be fraudulent. 314118 transactions were actually non-fraudulent as predicted.

(7) The heatmap shows results of predicting fraudulent transactions with the Decision Tree Model.While 301 transactions were actually fraudulent as predicted, 58 transactions were actually fraudulent even with the fact that it was predicted to be non-fraudulent. 66 transactions were actually non-fraudulent even when it was predicted to be fraudulent. 314148 transactions were actually non-fraudulent as predicted.

(8) The heatmap shows the relationship/correlation between all metrics. It shows that oldbalanceOrg and newbalanceOrd are highly correlated and newbalanceDest and oldbalanceDest are highly correlated.

Blossom Bank

Recommendations

There should be a transfer limit on a daily basis so as to curb the rate at which fraudulent transactions take place through transfers.

Proper investigation should be carried out by Blossom bank in the process of Cash_out over the counter.

Blossom bank should provide additional ways of securing accounts. For instance, an OTP (One time password) should be directed to the account holder's email address in addition to a secret question being answered especially for transfers.Google Authenticator should also be implemented.

Blossom bank should encourage active cyber-security and Improve controls by implementing continuous auditing and monitoring.

Blossom Bank

THANK YOU