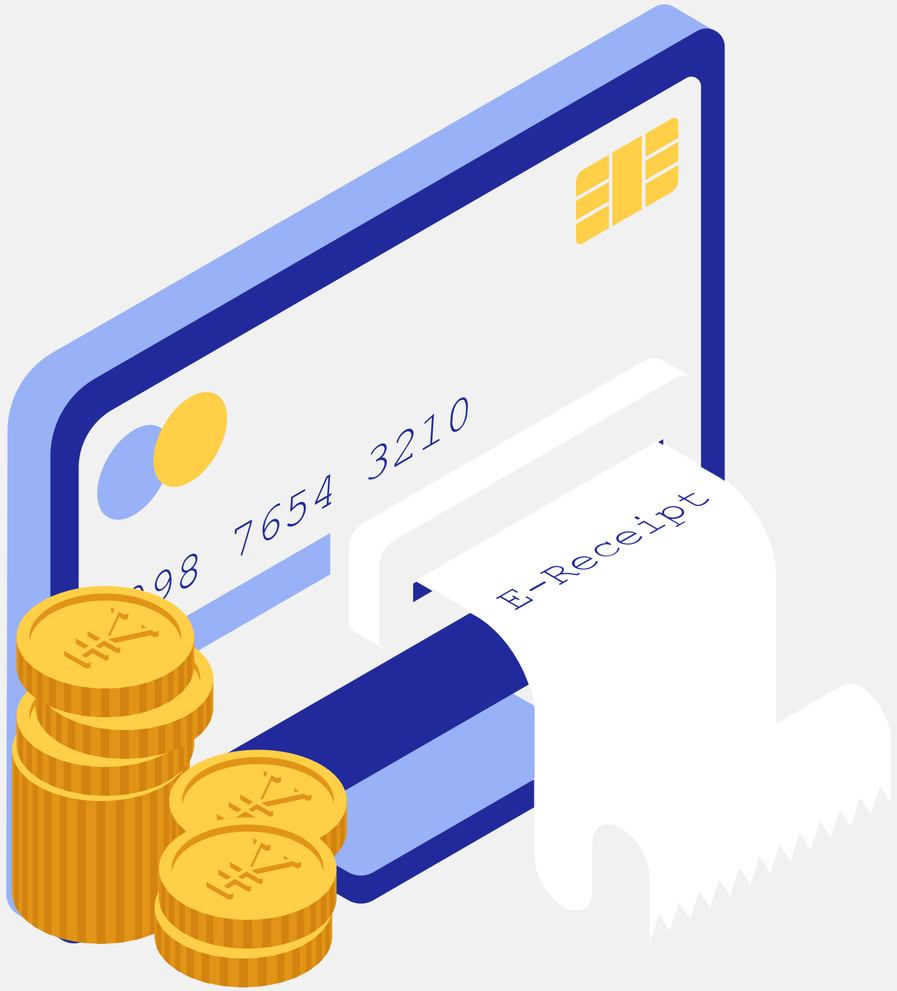


ONLINE FRAUD DETECTION

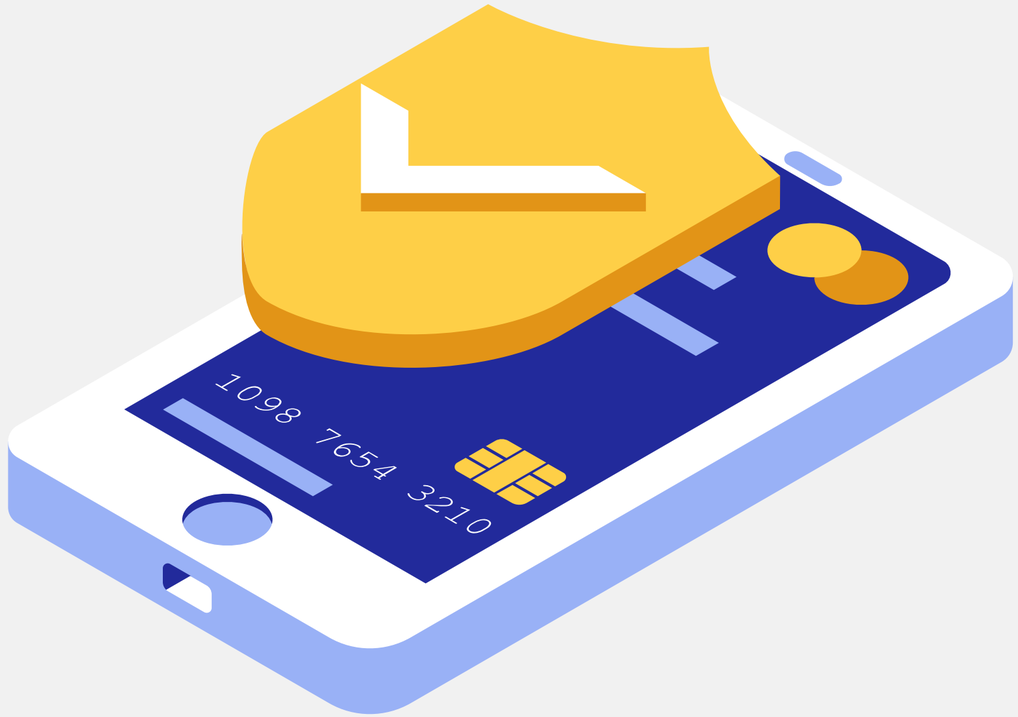
Using Machine Learning Models

by Joshua.S.Ewuji

The Experience

Fraud is an act of wrongful or criminal deception intended to result in financial or personal gain.

Over this act over a billion, as the Federal Trade Trade Commission published fraud complaints in 2021 over 48% of the yearly transaction. While 27% from Other consumer complaints and identity theft complaints 25%.



Blossom Bank, has had records of such occurrences that had been common in their day-to-day activity.

Payment Transaction; is one of the ways of entry into frauds of major transactions. And this is the major service financial institute provides.

Machine Learning; is an automated algorithm to help train the system to detect such occurrences and help prevent further of the same transaction with appropriate models with efficient accuracy.

Problem definition: clearly articulate the problem that is to be solved with your data
mining. How will the business benefit from your solution?

To detect the range of sums with high fraud and train the model towards it.
To detect the validity of the true amount Orig and amount Dest.
To identify the type of transaction with which fraud acts the most.

The steps implemented for the analysis and models

Step 1

- Import libraries needed for analysis and visualization.
- import CSV file as df data from the drive.
- visualize head, tail, and shape and describe
- Identify column uniqueness and label.

Step 2

- verify isnull and identify column types (object, float, int64).
- Identify categorical and numerical data.
- Create two new columns to verify if all amountOrig was in amounts.
- Visualize the most type of transactions made. The highest sum and range.

Step 3

- Import machine models sklearn to convert the categorical data to numerical and split the models into 60% 40%.
- Train the data as X-test, X-train and Y-test and Y-train.
- import other models to test the accuracy of the model introduced.

Step 4

- Machine learning models was imported, and classifiers were also introduced to perform an accurate test of the models.
- classifiers were able to derive score_list, cross-validation and ROC-AUC.

Results

- The new column created detected that not all times amount debit from orig is created to the recipient, also apart from the identified 1 (isfraud) i see more fraud from transactions orig and not credited on newamountdest column.

- The result derived from adaboost model making it 100% model for the machine learning.

- Also went further using other models and classifiers introduced to check the model validity were of score_list, cross validation and ROC-AUC.
- Having Adaboost with 100% accuracy, follow by Random forest, K-nearest Neighbors, SGD, Decision Tree.

Challenges:

- Sourcing for appropiate models for ML.
- Processing time of models and trainings.
- Visualization of the large data columns in a multi-variate way.