**MAKEATHON 2023**

By American Express

Team Name: Girl’s Hive

1. Aastha Kanth
2. Pooja Gusain
3. Ruby Chawla
4. Jiya Sharma
5. Ankita Das

**Theme**

**ML Model- Fraud Detection**

Problem Statement:

To build an ML model for detecting credit card fraud by using an imbalanced dataset.

Fraud can be detected in various forms and one of the major forms of it is credit card fraud. Earlier these frauds were detected by human-proposed patterns and rules to rely upon. This was time-consuming as well as the accuracy level of detecting fraud was unmonitored and was incapable of detecting a new fraud pattern.

Solution:

The ML model will be trained on the dataset so as to detect real-time fraud patterns. This will reduce the time consumption as well as the accuracy level of the model will be defined and could be refined over time.

The dataset was taken from Kaggle for building the model:

<https://www.kaggle.com/code/bannourchaker/frauddetection-part1-eda/input>

The dataset consisted of 11 columns out of which 1 column was for the label ‘isFraud’ which consisted of ‘0’ and ‘1’ values indicating,

‘0’ for non-fraudulent transaction

'1’ for fraudulent transactions

Out of the remaining 10 columns, only 6 columns were taken as the features of our model.

Since the label consists of only 2 predictable values 0 and 1, it comes under binary classification and hence, we developed the model based on **Logistic Regression.**

Methodology:

* Architecture Diagram

Data Collection:

The dataset was downloaded from Kaggle and was accessed from a local repository.

Exploratory Data Analysis:

We first noted the shape of the dataset and then regarding the type of each column used in the dataset. We the analyses the value for each

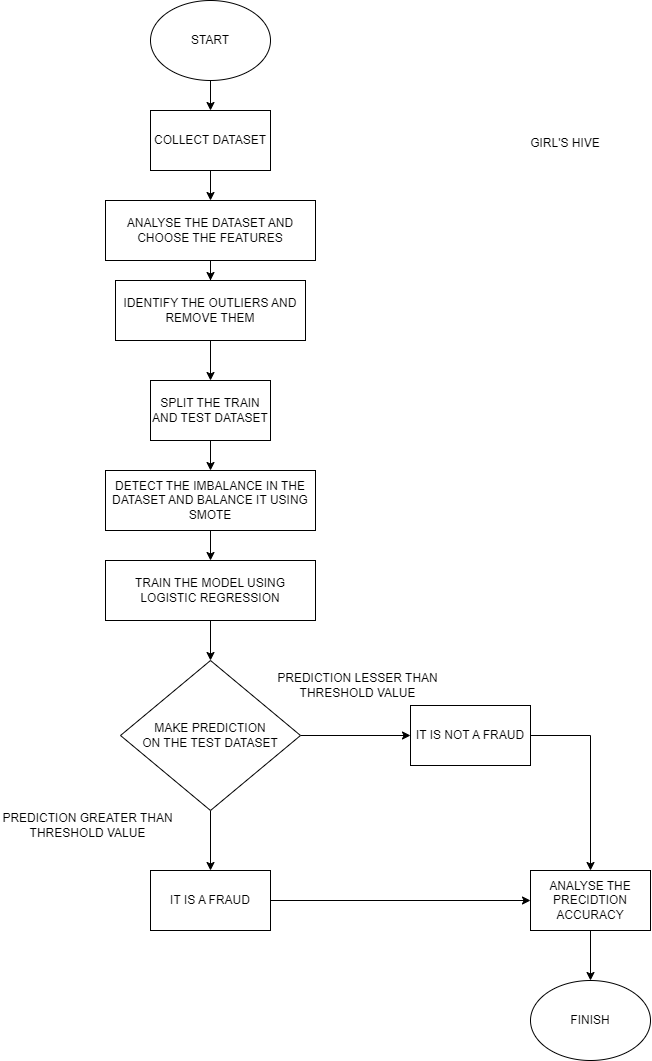
Pre-Processing:

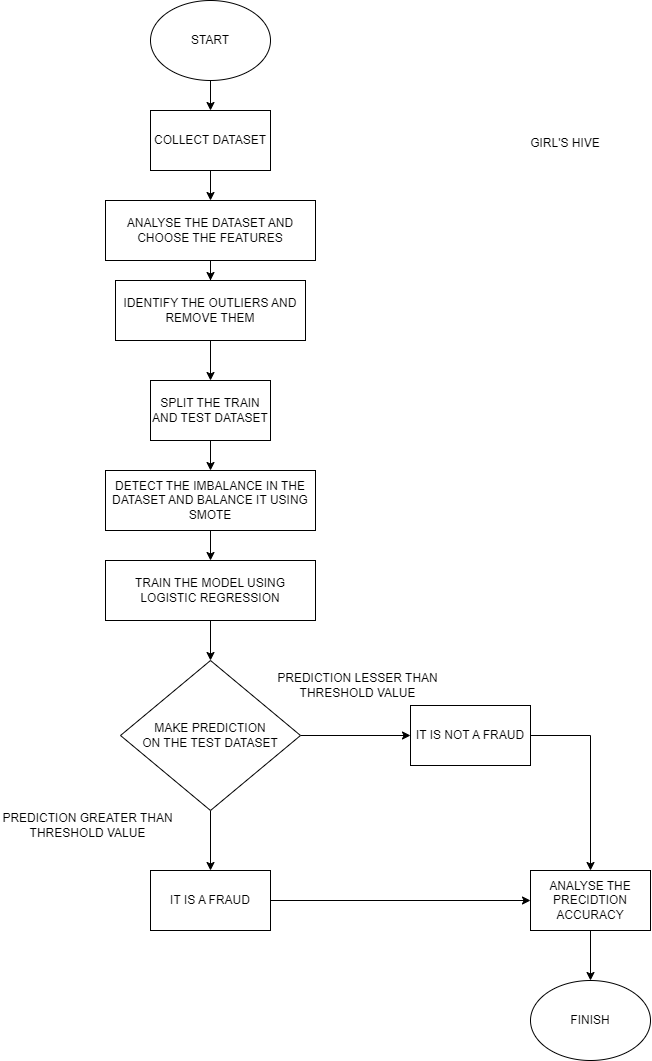
Modeling:

Logistic Regression:

Optimal Result of the model:

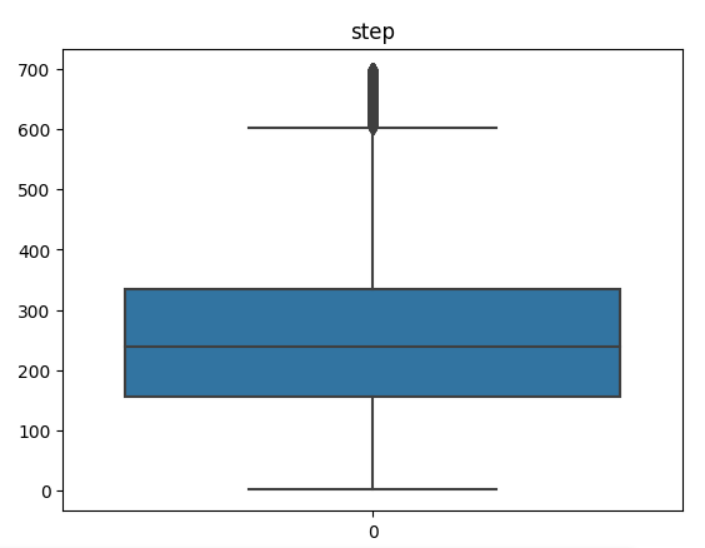
* Flow Chart

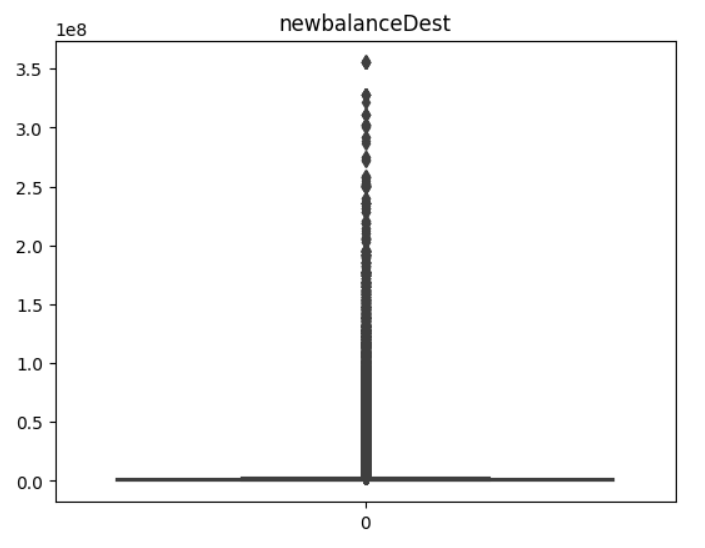
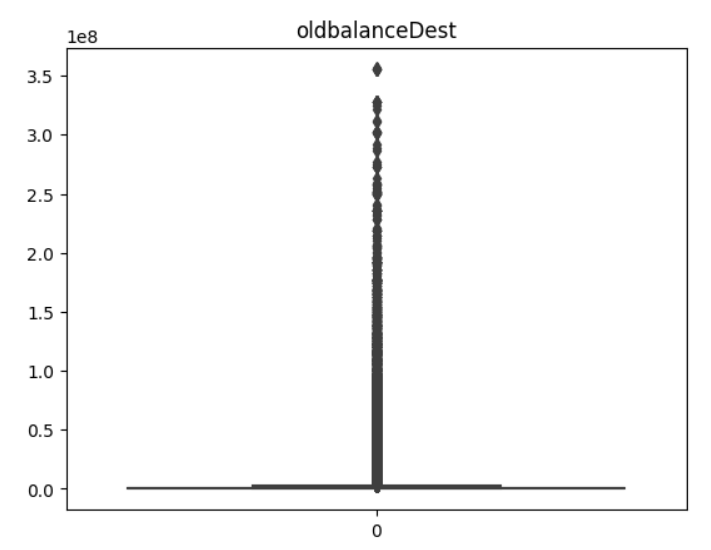
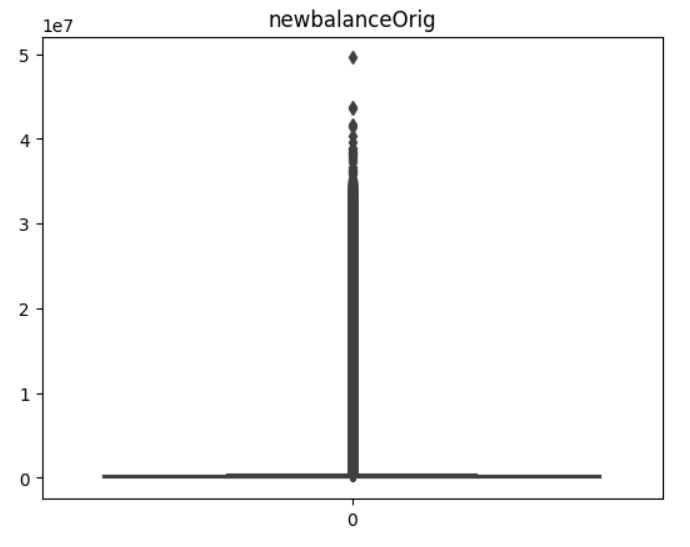
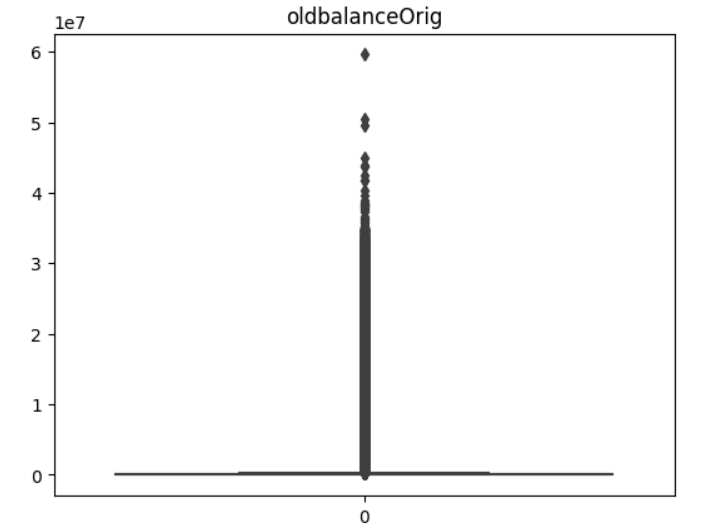
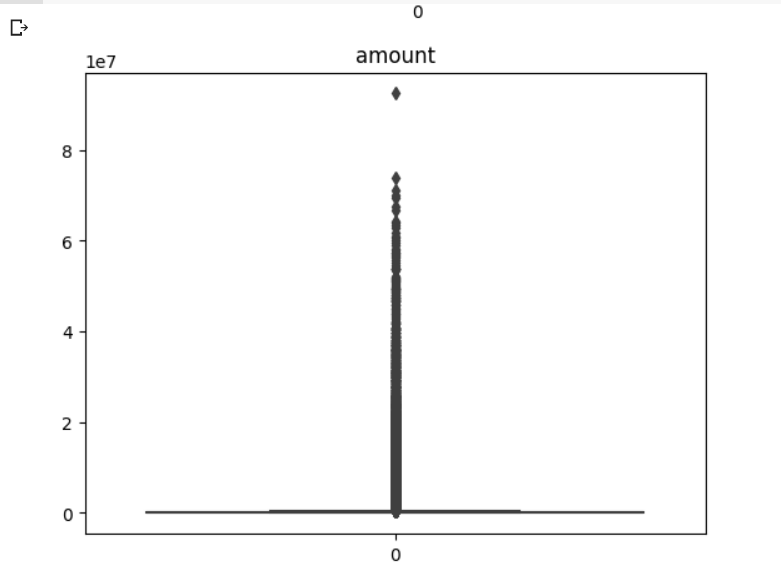




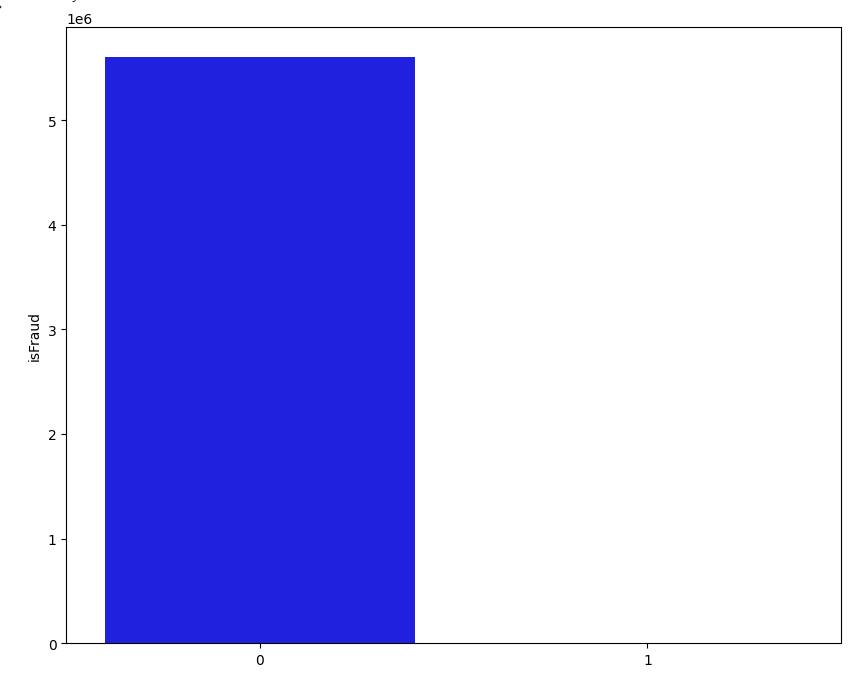
* Wireframes
* Graphical representation

The box plot to view the outliers present in the dataset were obtained as follows:

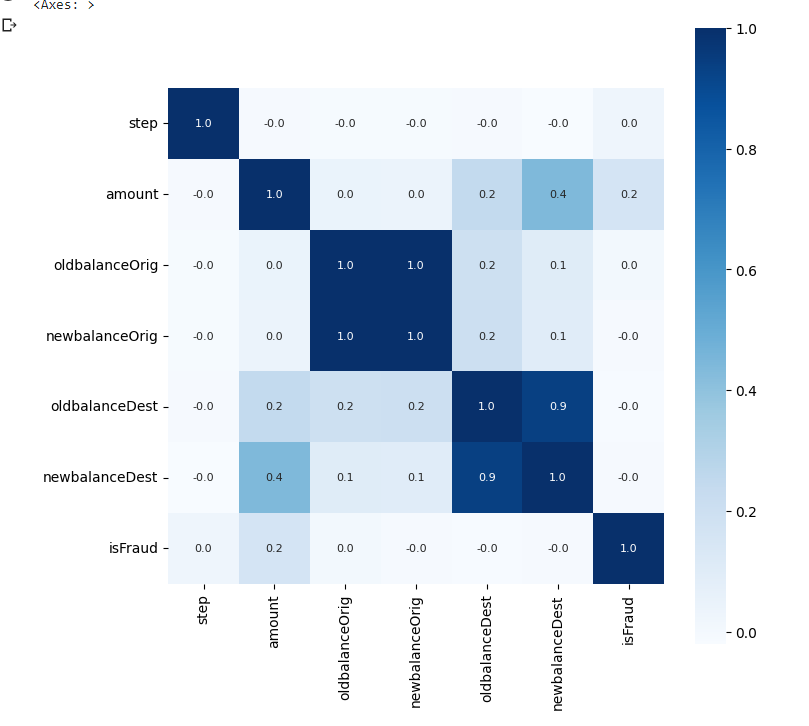




Further the imbalance in the dataset was noted by a bar graph:



Heatmap for the correlation of features:



* Prototype ( Screenshot, Screencast, Image, Video etc. )

Societal Impact/ Novelty:

This model will have a major impact on creating a safe and strong interface for the transaction of money from source to destination account. Thus, the fraud rate via credit cards will get reduced in no time.

The model will learn new patterns and possibilities along with the dataset it encounters in real-time. Making the prediction of the model more accurate.

This will be a great asset for the business and finance sector as the money will undergo a safer transaction route. This will also ensure to build a higher security in the cyber world.

Future Scope:

This model can future be integrated with other algorithms like SVM, random forest, etc to create a hybrid machine learning model in order to achieve a higher rate of accuracy.

This model can further be deployed for a wide range of use and easy accessibility for society. The columns which were not used as a feature in the model like source account number and destination account number will be used to detect the origin of the fraud.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*