

# Fraud Transaction Classification

By Asteway Kebede

## Abstract

The goal of this project is to build a classification machine learning model to predict fraud transactions. The global eCommerce fraud losses in 2021 were estimated to be \$20 billion. Therefore, having an application that can detect and warn consumers is a vital step in reducing fraud transactions.

For this project, a pipeline was built to collect the data and store it in an SQL database then performing machine learning to predict if the transaction was fraud or not. The final model was saved and deployed online as an application allowing a user to toggle the threshold rate of the model. The new results of the model are updated showing the precision and recall as well as the confusion matrix of the updated model.

---

## Design

We saw an 18% growth in global eCommerce fraud losses from 2020 to 2021. Due to the recent pandemic, many customers and consumers found eCommerce to be convenient in many ways. However, that also meant more opportunities for the fraudsters and they were able to increase the amount they stole.

Therefore, for this project a classification machine learning model using RandomForest Classifier to predict fraud transactions. A pipeline was built to collect the data and perform machine learning to predict if the transaction was fraud and deployed an app to show the results.

---

## Data

The data is from Vesta who is an eCommerce payment solution corporation.

- Multiple csv files
- Over 590,000 Transactions & 394 Features
  - The features include data card information, amount, fraud/notFraud, time/date and many more features engineered by Vesta.

To access and view a detailed description of the dataset, click here [Dataset](#)

---

## Machine Learning Algorithms

Machine Learning classification algorithms:

1. Logistic Regression
  2. Random Forest
  3. K-neighbors
- 

## TOOLS

The following tools were used in this project:

1. SQL, Python & Pandas to clean, explore and generate the final modeling data
  2. Matplotlib and Seaborn to generate visualizations
  3. SKLearn to build Machine Learning classification models and measuring metrics
  4. Streamlit to develop the app
  5. Heroku to deploy the app
  6. Docker to create a smooth pipeline
- 

## Communication

The findings and slide deck accompanying this project's presentation are accessible in this GitHub repository.