

DSC680-T301 Applied Data Science
4/16/2023
Soma Shekar Vayuvegula
Project 2: Credit Card Fraud Prediction

Topic

As the world progresses toward digitization, we rely on digital payments. While digital payments are on the rise, frauds in digital payments especially using a credit card, are increasing, and cybersecurity has become a crucial part of the credit card industry.

Credit Card Fraud Prediction System will bear considerable benefits to the customers and credit card firms so that we can reduce financial risks and losses due to these frauds.

Business Problem

Customers expect to have a secure banking experience, and providing a solid fraud-detecting system has become an integral part of the credit card industry. I would like to build an automated system that detects fraudulent patterns and flags and temporarily blocks such transactions for customers to review, approve, or deny.

The main aim of this project is to build a prediction system using different models to track various patterns, abort abnormal transactions, and notify customers about suspected fraudulent activity on their credit cards.

Datasets

The datasets used in this project will be sourced from Kaggle, a public-domain dataset. I am planning to use data from two different sources and utilize the following data:

trans_date_trans_time – Transaction date and time.

cc_num – Credit Card Number.

merchant – Merchant name where the transaction has been done.

category – Category name of the transaction.

amt – Amount transacted.

first – First name of the credit card user.

last - Last name of the credit card user.

gender – Gender of the credit card user.

street – Street address of the credit card user.

Methods

As part of this project, I will be building the below models:

1. Random Forest: Random Forest is a Supervised Learning algorithm. This is widely used for Classification and Regression problems. It builds multiple decision trees and takes a majority vote to classification.

Random Forest algorithm is used to classify fraud transactions in real-time or batch processes.

2. Decision Tree Approach: Decision Tree is a Supervised Learning used to make predictions based on how the previous set of questions is answered.
Constructing a proper decision tree can be helpful in stock price prediction. We can see if the price of the stock will rise or fall.

3. XG Boost Algorithm: XGBoost is a decision-tree-based ensemble Machine Learning algorithm. It uses a gradient-boosting framework.

This is a popular and open-sourced gradient-boosted trees algorithm. As this algorithm has gained popularity over time, I plan to implement it in this project and see if it performs better than the other two algorithms listed above.

In the above-proposed models, I will be implementing SMOTE technique to increase the number of records in the datasets in a balanced way to overcome issues with imbalanced datasets.

Ethical Considerations

Though this project's credit card transaction data is simulated data, it can match real-time PII data.

Credit card fraud detection might not work as expected with smaller datasets. As we are using a smaller dataset, I am not confident that the model accuracy is as expected.

Challenges/Issues

As we are using the dataset for 1 Jan 2019 - 31 Dec 2020, we might be missing the newer patterns that have come up post-Dec 2020.

As we are using public domain datasets, we cannot expect that this covers all the possible fraudulent transactions.

While working with the real-time datasets for credit card fraud detection, we will be working with massive datasets which might need higher CPU capacities and scaling of the algorithm implemented. We need to keep in mind the scaling of the model to support the substantial real-time datasets.

References

- SHENOY, K. (n.d.). *Credit Card Transactions Fraud Detection Dataset*. Kaggle. <https://www.kaggle.com/datasets/kartik2112/fraud-detection?select=fraudTrain.csv>
- XIANG, N. (n.d.). *Credit Card Fraud Analysis and Modeling*. Kaggle. <https://www.kaggle.com/code/nathanxiang/credit-card-fraud-analysis-and-modeling>