**Predicting Credit Card Fraud**

Kanagaraj Ramasamy

College of Science & Tech, Bellevue University

Predictive Analytics 630

Professor Andrew Hua

28 February 2024

**Introduction**

Credit card usage across the globe is expanding rapidly. As usage grows, fraudulent transactions increase along with that. The number of credit card holders worldwide grew to 1.25 billion in 2023 (Clearlypayments, 2023). The USA alone accounts for 166 million credit cards. While credit card usage increases, credit card fraud gradually increases yearly. The total value of credit card fraud soared to an eye-popping $36 billion and is expected to reach $43 billion by 2026 (Rej, M, 2023). Most of the time, financial institutions take responsibility for the money lost due to credit card fraud. The financial institution must address the credit card fraud problems to prevent revenue loss and reputation damage. Credit card fraud is happening in many ways, including card-not-present fraud, credit card application fraud, skimming, and lost/stolen cards (Synovus Financial Corp, 2023). This project aims to identify fraudulent credit card transactions using machine learning techniques.

**Data Selection**

The dataset is extracted from the open-source website Kaggle. The dataset simulated credit card transactions from 1st Jan 2019 to 31st Dec 2020. It has 555,719 observations and 23 attributes. It has one attribute for the transaction amount, and one attribute implies whether the transaction is fraudulent. The remaining attributes are information about the transactions, such as card number, merchant locations, customer address, transaction reference number, and time of the transactions. (Kartik Shenoy, 2020).

**Methods:**

The dataset was highly imbalanced, so stratified sampling was used for training and testing data split. Categorical features such as gender, state, and category are encoded using OneHotEncoder. All the numerical features are standardized using StandardScaler. Credit card fraud detection is a supervised classification problem. Multiple classification algorithms, such as decision trees, random forest, XGBClassifier, and logistic regression, were tested for this project. XGBClassifier model was selected to identify fraudulent credit card transactions over other models because of its better performance metrics.

**Results:**

A comparison of the performance metrics of the different models is below:

A screenshot of a graph

Description automatically generated

Identifying fraud is very important, and we can't miss any fraudulent transactions because financial institutions will lose money. Performance metrics recall takes higher priority than others. From the above result, XGBClassifier has better recall, roc\_auc, f1\_score, and accuracy.

A bar graph with text

Description automatically generated

XGBClassifier model identified that features grocery POS, gas transport, transaction amount, transaction hour, travel, grocery net, female gender, age, state VA, and entertainment are the top 10 feature importances.

**Conclusion**

The project's primary goal is to identify credit card fraud transactions to prevent revenue loss and reputation damage to financial institutions. The XGBClassifier model yielded better results with an accuracy greater than 99 %. A feature importance plot has been created to identify the features which have a higher effect on the model. The feature grocery point-of-sales transactions are at the top of the list, followed by gas transport, amount, transaction hour, and travel.

From the EDA, we noticed that all the fraudulent transactions have amounts less than 150 and that most credit card fraudulent transactions happen between 10 PM and 3 AM. Also, we noted that grocery POS, gas transport, and grocery net have a lot of fraudulent transactions. Transactions involving the ages 34, 52, 32, 61, and 43 accounted for more fraudulent transactions. A good starting point is to use the XGBClassifier model to flag the transactions as suspicious before processing. If the transaction falls into the above important feature category, it will likely be fraudulent.

**References**

Clearlypayments - *The growth of the credit card industry in 2023*, *Credit Card Processing and Merchant Account*. Available at: https://www.clearlypayments.com/blog/growth-of-credit-card-industry-in-2023.

Rej, M. (2023) *Credit Card Fraud Statistics (2023)*, *Merchant Cost Consulting*. Available at: https://merchantcostconsulting.com/lower-credit-card-processing-fees/credit-card-fraud-statistics/ (Accessed: 09 December 2023).

Synovus Financial Corp. (2023, November 30). *What are common types of credit card fraud?*. Synovus. https://www.synovus.com/personal/resource-center/financial-safety-and-security/types-of-credit-card-fraud/

Kartik Shenoy. “Credit Card Transactions Fraud Detection Dataset. (2020, August 5). Kaggle. https://www.kaggle.com/datasets/kartik2112/fraud-detection