Bellevue University

Fraud Shield | Securing Credit Card Transactions

MILESTONE 1 – PROJECT 1

Date: 08/31/2023
Aarti Ramani
SC680-T302 Applied Data Science (2241-1)

Table of Contents

TOPIC 2
BUSINESS PROBLEM 2
DATASET 3
METHODS 3
ETHICAL CONSIDERATIONS 4
CHALLENGES & ISSUES 5
Challenges: 5
Issues: 5
REFERENCES: 5

TOPIC

Fraudulent activities in digital transactions pose a significant threat to both consumers and financial institutions. This project aims to develop an advanced machine-learning solution for rapid credit card fraud detection, ensuring security for both cardholders and financial organizations.

BUSINESS PROBLEM

Credit card fraud poses a severe challenge due to its potential for substantial financial losses and security breaches. The significance of robust fraud detection systems lies in the following areas:

- Fraudulent transactions can lead to substantial financial losses for both banks and customers. Swift detection helps prevent unauthorized transactions, minimizing financial damage.
- Fraud incidents erode customer trust in banks. Implementing effective fraud detection reassures customers that their financial assets are secure, fostering long-term relationships.
- Detecting and mitigating fraud helps protect customers' sensitive data from falling into the wrong hands.

Credit card fraud's potential for financial losses and security breaches necessitates robust fraud detection systems. This project seeks to develop an automated system that can distinguish between legitimate and fraudulent transactions with high accuracy. By doing so, it aims to reduce financial losses, enhance customer trust, and strengthen the security of credit card transactions.

DATASET

This dataset is sourced from Kaggle(Shenoy, K. (2020, August 5)). This dataset contains stimulated credit card data with legitimate and fraudulent credit card transactions from Jan 2019 to Dec 2020. It includes transactions made by 1000 customers with 800 merchants. This dataset includes a range of transaction attributes such as transaction amounts, timestamps, merchant details, geographic information, and cardholder specifics.

METHODS

The project employs a combination of supervised and unsupervised machine learning techniques. In the supervised domain, models such as Random Forest, Gradient Boosting, and Neural Networks will be trained to effectively classify transactions. I plan on building the following models for this project.

- Logistic Regression: Using a Logistic Regression model for credit card fraud detection has benefits including interpretability, computational efficiency, and suitability for smaller datasets. It provides insights into feature impacts, acts as a baseline, and offers transparent probability estimations. However, its linear nature might limit performance on complex, non-linear data patterns.
- Random Forest: Using a Random Forest model for credit card fraud detection offers benefits like ensemble learning, robustness against noise and outliers, handling imbalanced data, and capturing complex patterns. Its ability to handle non-linearity, easy tunability, and feature importance analysis make it a strong choice.
- Gradient Boosting: Gradient Boosting is a powerful method for credit card fraud detection due to its high predictive accuracy, ensemble nature that reduces overfitting,

and the ability to handle complex patterns in the data. It also offers insights into feature importance, aiding in identifying relevant variables for fraud detection.

Both Random Forest and Gradient Boosting are ensemble methods that excel in handling credit card fraud detection tasks. Random Forest builds diverse trees independently, while Gradient Boosting sequentially improves the errors of previous trees.

- Neural Networks: Neural Network (NN) models offer advantages in credit card fraud detection due to their ability to capture intricate non-linear patterns in data. They excel in feature learning from raw data, making them adept at detecting complex and evolving fraud behaviors.

ETHICAL CONSIDERATIONS

Several ethical considerations are crucial for this project:

i. Data Privacy: Despite the simulation, safeguarding privacy is essential. Ensuring simulated data does not resemble real customer information avoids accidental exposure of PII (Personally Identifiable Information).
ii. Informed Consent: Transparency builds trust and addresses concerns. Understanding if consent was obtained for real-based simulations, even if anonymized, adds ethical integrity.
iii. Intent and Use: Ethical utilization of simulated data is paramount, prohibiting any malicious or harmful intent.
iv. Data Security: Robust security measures should be applied to the simulated dataset, treating it with the same importance as real customer data.

v. Stakeholder Implications: The viewpoints of stakeholders like credit card issuers, customers, and regulators to ensure ethical alignment should be considered.

CHALLENGES & ISSUES

Challenges:

The project anticipates challenges arising from imbalanced data, potential data quality issues, and the complexity of feature engineering.

Issues:

Ensuring security for simulated data and achieving real-world applicability could be areas of concern. Simulated datasets might not fully capture the complexities of actual fraud scenarios, potentially limiting the model's effectiveness in real-world situations.

REFERENCES:

Shenoy, K. (2020, August 5). *Credit Card Transactions Fraud Detection Dataset*. Kaggle. https://www.kaggle.com/datasets/kartik2112/fraud-detection?select=fraudTrain.csv