FRAUD SHIELD
SECURING CREDIT CARD TRANSACTIONS

Madhavi Ghanta, March 16, 2024
BELLEVUE UNIVERSITY
DSC 680 Spring 2024

TOPIC

Financial institutions and individuals are at serious risk from fraudulent activities in digital transactions. The goal of this project is to create a sophisticated machine-learning system that can quickly identify credit card fraud and protect financial institutions as well as cardholders.

BUSINESS PROBLEM

The possibility of significant financial losses and security breaches associated with credit card theft makes it a serious problem. Here are some instances where having strong fraud detection systems is important:

- Banks and clients may suffer significant financial losses as a result of fraudulent transactions. Quick detection lessens financial harm by preventing unwanted transactions.
- Incidents of fraud reduce consumer confidence in institutions. Long-term relationships are fostered by implementing excellent fraud detection, which gives clients peace of mind about the security of their financial assets.
- Preventing sensitive consumer data from getting into the wrong hands is made easier by identifying and reducing fraud.

Robust fraud detection systems are necessary due to the possibility for financial losses and security breaches resulting from credit card fraud. The goal of this project is to create an automated system with high accuracy that can discriminate between real and fraudulent transactions. It hopes to accomplish these goals by lowering monetary losses, boosting consumer confidence, and fortifying credit card transaction security.

DATASET

This dataset is sourced from Kaggle(Shenoy, K. (2020, August 5)). From January 2019 to December 2020, this dataset includes both authentic and fraudulent credit card transactions together with stimulated credit card data. It comprises of 1000 client transactions with 800 vendors. Numerous transaction attributes are included in this dataset, including transaction amounts, timestamps, merchant information, cardholder specifications, and geography data.

METHODS

Both supervised and unsupervised machine learning techniques are used in this project. Models like Random Forest, Gradient Boosting, and Neural Networks will be trained to efficiently categorize transactions in the supervised domain. For this project, I want to develop the following models.

- Logistic Regression: The benefits of using a Logistic Regression model for credit card fraud detection are its applicability for smaller datasets, interpretability, and computing efficiency. It delivers clear probability estimates, serves as a baseline, and sheds light on the effects of features. Performance on intricate, non-linear data patterns, however, might be constrained by its linear nature.
- Random Forest: Benefits of using a Random Forest model for credit card fraud detection include handling imbalanced data, identifying complicated patterns, robustness against

noise and outliers, and ensemble learning. It is a good option due to its versatility, ease of tuning, and feature importance analysis.

- 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.

For problems involving the detection of credit card fraud, Random Forest and Gradient Boosting are two ensemble techniques that work well. Whereas Gradient Boosting gradually reduces the errors of earlier trees, Random Forest creates a variety of trees without supervision.

- Neural Networks: Because neural network (NN) models can identify complex non-linear patterns in data, they are advantageous in the detection of credit card fraud. Their proficiency in feature learning from unprocessed data enables them to identify intricate and dynamic fraud patterns.

ETHICAL CONSIDERATIONS

Several ethical considerations are crucial for this project:

- Data Privacy: Even with the simulation, privacy protection is crucial. Prevent unintentional disclosure of Personally Identifiable Information (PII) by making sure that simulated data does not resemble actual customer information.
- Informed Consent: Transparency solves issues and fosters confidence. Even if anonymized, knowing if consent was received for real-based simulations enhances ethical integrity.

- Intent and Use: It is crucial to use simulated data ethically, forbidding any malevolent or destructive aim.

- Data Security: The simulated dataset should be subjected to strong security measures and given the same consideration as actual customer data.

- Stakeholder Implications: To guarantee ethical alignment, the perspectives of stakeholders such as credit card issuers, consumers, and regulators should be taken into account.

CHALLENGES & ISSUES

Challenges:

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

Issues:

Achieving real-world applicability and guaranteeing security for simulated data could be problematic topics. The intricacies of real-world fraud scenarios may not be fully captured by simulated datasets, which could restrict the model's applicability in practical settings.

REFERENCES:

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