CREDIT CARD FRAUD DETECTION

Using Classification Models to Identify Fraudulent Transactions

•Student name: William Muthama
•Student pace: part time hybrid
•Scheduled project review date/time: 27th August to 30th August
•Instructor name: Anthonny Muiko
•Blog post URL: https://github.com/WILLY-GUSH/Project_phase3

OVERVIEW

Objective

•The goal of this project is to develop a system that accurately identifies fraudulent credit card transactions in real-time.

•Fraud detection is crucial because it helps prevent significant financial losses for banks and protects customers from unauthorized charges.

BUSINESS AND DATA UNDERSTANDING

Business Problem

•Credit card fraud is a major issue, costing the global economy billions each year. Detecting fraud as it happens can help mitigate these losses.

•The challenge is to distinguish fraudulent transactions from legitimate ones as quickly as possible, without flagging too many false positives (legitimate transactions marked as fraud).

Data Overview

•The dataset consists of over 280,000 transactions, with 31 features including the transaction amount, time, and whether or not the transaction was fraudulent.

•Only 0.17% of the transactions in the dataset are fraudulent, which highlights the challenge: finding a small number of fraudulent transactions in a large pool of legitimate ones.

BUSINESS AND DATA UNDERSTANDING

Classification Model

- We used a classification model, specifically a Random Forest, to predict whether a transaction is fraudulent or not.
- This type of model is trained on historical transaction data, learning patterns that differentiate fraud from legitimate transactions.

Why Classification

- Classification is ideal for this task because it can quickly categorize new transactions as either "fraudulent" or "legitimate" based on learned patterns. This speed is essential for preventing fraud in real-time.

EVALUATION METRICS

Model Performance

We used several metrics to evaluate how well our model performs:

- **Accuracy:** The model is correct about 99.94% of the time. This means it usually identifies transactions correctly, but given the imbalance (very few frauds), this number alone isn't enough.
- **Precision:** When the model says a transaction is fraudulent, it's right about 91% of the time. This is important because we don't want to falsely accuse customers of fraud.
- **Recall:** The model catches about 82% of all fraudulent transactions. This means it's good at identifying fraud, but we aim to catch even more.
- **F1-score:** This balances precision and recall, giving us an overall sense of the model's effectiveness. Our model's F1-score is around 86%, indicating a good balance between identifying fraud and avoiding false alarms.

Visual Representation

A confusion matrix shows how often the model correctly identified fraud versus how often it made mistakes. Most mistakes were false negatives, where the model missed fraud, and some were false positives, where it incorrectly flagged legitimate transactions.

KEY FINDINGS

Model Insights

- The model identified several key features that are most important in detecting fraud. For instance, large transaction amounts at unusual times were often flagged as fraudulent.
- The model also showed that fraudulent transactions often have patterns different from legitimate ones, such as being made in quick succession or in different locations.

Model Performance

- Overall, the model performs well, but it’s not perfect. It misses some fraudulent transactions and occasionally flags legitimate ones as fraud. However, with further refinement, it can become even more reliable.

RECOMMENDATIONS

Business Recommendations

•Implement the model into the bank's transaction monitoring system. This will allow for real-time detection of fraudulent transactions.

•Regularly retrain the model with new data to improve its accuracy over time, as fraud tactics evolve.

NEXT STEPS

Future Improvements

- To enhance accuracy, we recommend collecting more diverse data, especially from newer fraud cases, to keep the model updated with the latest trends in fraud tactics.
- Exploring more advanced models or combining multiple models might also improve performance, especially in catching rare cases of fraud.

Scalability

- This model can be scaled to handle larger datasets as the volume of transactions grows. It can also be adapted for use in different regions or types of financial institutions with some customization.

THANK YOU