Transaction Fraud Detection

Written by: Choudhry Rafay, Max Enderlein, Saad Saleem

Focus

In today's digital world, detecting and preventing online fraud is a necessity. As more people transition to online banking and e-commerce, fraudsters continuously develop new methods to exploit vulnerabilities in payment systems. The ability to detect fraudulent transactions in real time is critical to maintaining consumer trust and financial stability. Unlike traditional security systems, machine learning offers a proactive and dynamic approach to detecting anomalies that could indicate fraud. The goal of this project is to develop an accurate machine learning classifier to predict whether an online transaction is fraudulent based on transaction characteristics.

Data Sources

All data used in this project was sourced from MIT.edu. Please note that the data is synthetically generated and is intended for illustrative purposes only. It should not be interpreted as a factual data set. The dataset contains 1 million transactions, which resulted in a highly imbalanced dataset. Each transaction is described using 16 variables.

Analytical Approach

The initial step involved data preprocessing. The dataset was examined for duplicate entries and missing values, though none were present. All numeric features were standardized using standard scaling to ensure comparability. Given the class imbalance, a stratified sample of 2,000 records was selected to preserve the proportion of fraud and non-fraud cases and allow effective model training and evaluation.

Exploratory data analysis revealed that certain variables had stronger correlations with fraudulent behavior. Notably, the time of transaction, the age of the user's account, and the

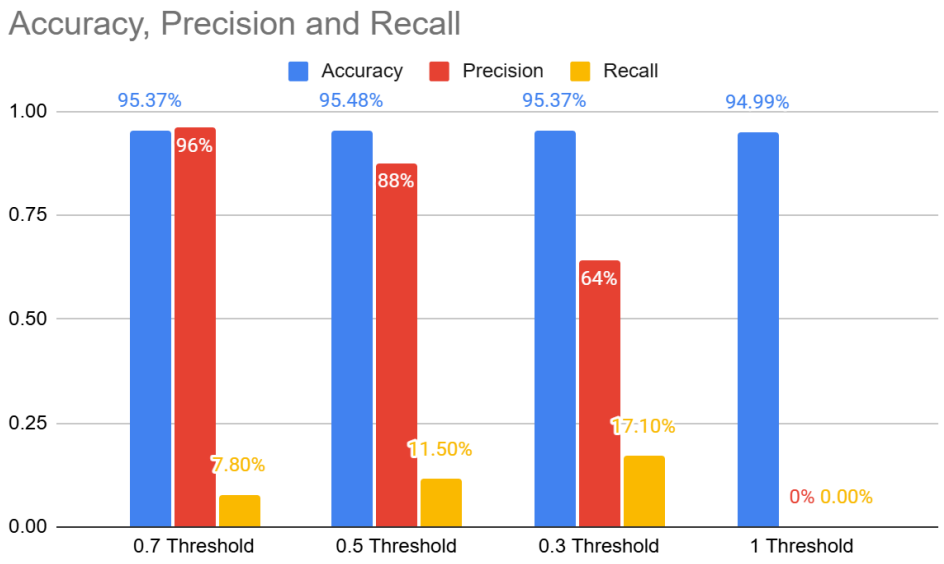
transaction amount were frequently associated with fraud. These insights informed the selection of features emphasized during model development. We used 3 models: logistic regression, clustering, and random forest.

Logistic regression:

Logistic regression was used as a baseline due to its simplicity and interpretability. While it performed reasonably well on clearly fraudulent cases, it failed to detect many subtle fraud instances and demonstrated low recall, especially at higher classification thresholds.

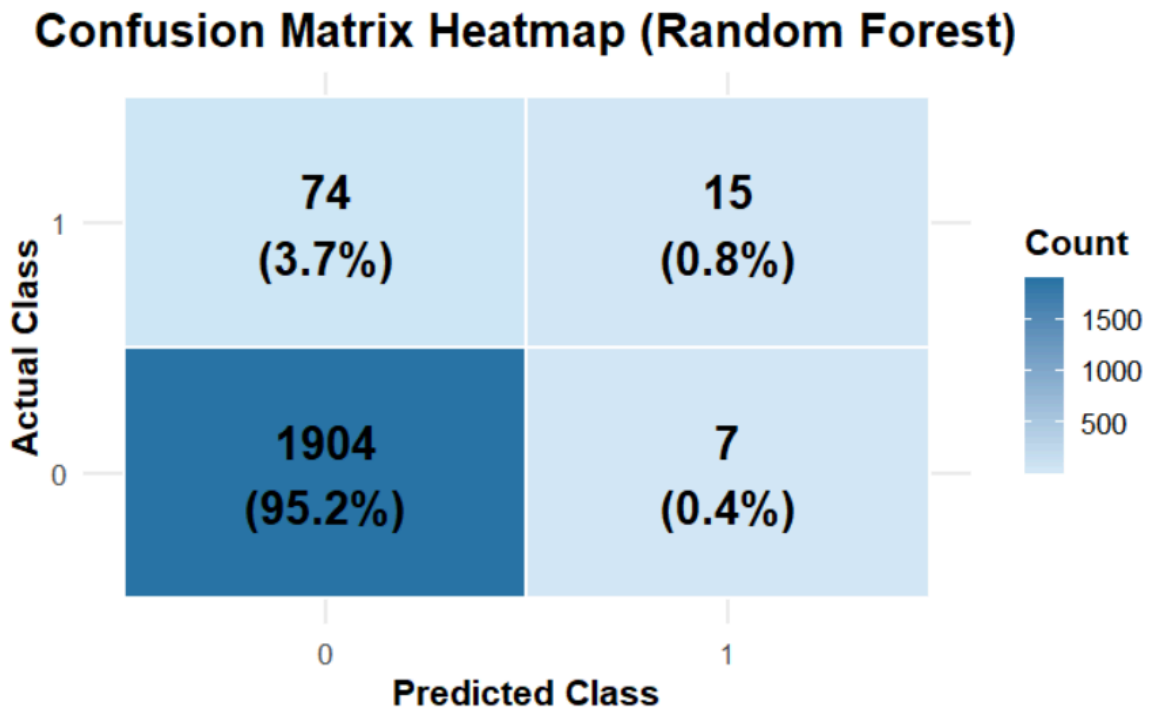
Clustering:

Clustering methods, including K-means and hierarchical clustering, were used in an unsupervised manner. Both methods revealed groups of transactions composed entirely of fraudulent activity, indicating that clustering could be a useful tool. We were aware how hard it can be to detect fraud, so instead of using 2 clusters, we used 3 in hopes we'd have 1 cluster for non-fraud, 1 for fraud, and 1 for suspicious that could be fraud, or like a 50/50 split of fraud and not fraud. This way, some suspicious transactions could've been detected as fraud if they were harder to catch. Unfortunately, it didn't quite end up this way, as it was 2 non-fraud clusters and 1 obvious fraud cluster.



Random Forest:

The random forest model yielded the highest overall classification accuracy. However, its performance remained skewed due to the imbalanced nature of the dataset. While it achieved over 95 percent overall accuracy and high precision for the fraud class, it detected only a modest portion of actual fraud cases, highlighting the difficulty of achieving both high recall and precision in imbalanced settings. Nonetheless, the random forest model outperformed logistic regression in identifying more nuanced fraudulent transactions.



Findings

Initial training of all three classifiers on the full set of predictors yielded varying performance. As expected, the logistic regression model performed the weakest, with an accuracy around 96% but a low recall of 7.6% at a 70% threshold 88% accurate with a 11.5% recall at a 50% threshold and 64% accurate with a 17% recall at a 30% threshold, missing many fraudulent transactions due to the class imbalance. Upon further investigation, we observed that certain

transaction features had a bigger influence on fraud detection, particularly the age of the user's account, the time of the purchase, and the amount of the purchase.

Impacts

The models developed in this project demonstrate high potential for detecting fraudulent transactions. Our results indicate that a carefully tuned logistic model trained on a balanced dataset with a reduced set of critical features can achieve high performance even in highly imbalanced scenarios. This can aid financial institutions in automating fraud detection while minimizing both missed fraud cases and unnecessary false alarms.

The methodology presented is generalizable and could be applied to other financial fraud datasets with minimal adjustments. However, deployment in real-world environments would require integration with real-time data pipelines and continuous retraining to adapt to evolving fraud tactics.

Additionally, while our dataset was anonymized and static, working with live transaction data would necessitate strict privacy protocols and data governance measures. Further work could focus on incorporating time-series patterns, user behavior profiling, and deep learning models.