Fast Fraud Screening: Using Lightweight Models to Flag Risk Before Deep Analysis

A. Ahmed, N.Bakare Ayoub, C.M. Cordor, B. Owens
The Erdös Institute Data Science Boot Camp Fall 2025

Introduction:

Financial fraud is a growing challenge. This project proposes a two-stage fraud detection framework: a lightweight surrogate model that screens transactions, prioritizing high recall, then flags transactions that goes through more intensive models/human analysts for verification. This reduces computational load, speeds detection, and focuses resources.

We find that traditional metrics are misleading for our analysis and we propose an alternative metric. We use the Lift metric for our fraud detection assessment, prioritizing recall and predicted positive rates for efficiency. Using J.P. Morgan Chase & Co's synthetic data, the XGBoost models outperformed the Logistic Regression model. XGB achieved 7x Lift and 73% recall, flagging ~3 in 4 fraud cases while incorrectly flagging less than 16% of legitimate transactions.

Stakeholders:

Financial institutions, including banks, lenders, investment firms, and stock traders, are key stakeholders who would benefit from a fraud detection model to minimize financial losses. Consumers also benefit from early detection because it minimizes harm, stress, and long-term damage – even when financial losses are reimbursed.

Dataset:

We utilize the JPMorganChase Payment Data for Fraud Protection, a synthetic dataset designed to safeguard customer privacy. This subject-centric data contains over 1.49 million transactions (electronic transfers, bill payments, deposits, withdrawals) spanning approximately 50 years. Each entry details the transaction amount in U.S. dollars, involved accounts (sender, beneficiary, or both), and other identifying features.

Methods:

We engineer both transaction and graph-derived features to train our lightweight, high-recall fraud screening model.

To analyze the complex network of over 300,000 customer and merchant accounts, we used Python's NetworkX package to construct a multi-directed graph. A multi-directed graph is a network structure composed of nodes, representing sender or beneficiary accounts, and the edges representing transactions. Edge direction shows fund flow – who is sending versus who is receiving the funds – and multiple edges indicate repeated transfers. Fraud often occurs in clusters of accounts cycling funds among themselves, making the graph view effective for uncovering hidden connections and identifying potential fraud rings.

Models: Logistic Regression, XGBoost, Linear Discrimination Analysis, PCA

Key Performance Indicators (KPIs):

Model Performance KPI - Our KPIs include the percentage of fraud transactions correctly flagged (Recall), and a measure of how much better the model is at identifying fraud compared to random guessing (Lift). We examined the trade-offs between recall and lift for our XGBoost model across different thresholds. **The business optimal threshold is 0.36, where the model achieves 84% recall and 4.84 Lift, balancing detection coverage.**

Business KPI - We use synthetic data to create realistic fraud risk KPIs, evaluating the model on false positives, analyst workload savings, simulated loss-avoidance scores, total loss avoided (USD), and total review cost. Key takeaways:

- Our lightweight model offers high speed and low computational overhead, suitable for real-time deployment.
- The model showed a low false positive rate (~16%).
- And a relatively low fraud miss rate (~17%), requiring threshold tuning for recall-first performance.
- Compared to the baseline randomized filter, our filter nets a **$2.16M** financial benefit in deployment.

Results:
Initial results show strong computational performance and low false positive rate, demonstrating feasibility for real time screening. At the current decision threshold (0.36) fraud detection precision remains limited, resulting in a high false positive rate. Future iterations will adjust model tuning and increase feature engineering to increase precision as well. This stage validates the model architecture and provides a foundation for recall-lift optimization in subsequent experiments.

Challenges:
Developing a fraud detection model using the J.P. Morgan dataset presents three main challenges:

1. **Feature Engineering:** New features must be generated from raw transaction data to uncover fraudulent patterns.
2. **Imbalanced Data:** Fraudulent transactions (2.06% of 1.49 million+) are significantly outnumbered by non-fraudulent ones, making it difficult for a model to learn.
3. **Synthetic Data:** Transaction timestamps in this synthetic dataset may follow a pattern, potentially lacking predictive information about fraud, as fraudulent labels were assigned using predefined probabilities.
4. **Normal metrics like accuracy and PR-AUC do not reflect the performance** of our filter well. Instead we use 'Lift,' a standard data science metric that measures efficiency and is defined as follows:

$$\textbf{Lift} = [TP / (TP+FN)] \; / \; [(FP+TP) / (FP+TP+TN+FN)]$$

This metric is similar to recall but replaces Predicted Positive Rate (PPR) in the denominator. Lift prioritizes high fraud detection (recall) and a low PPR, aiming to minimize missed fraud.

Any randomized model has constant Lift=1, for any corresponding recall value. Our model has the flexibility to trade Recall for Lift. For example, with recall reduced to 70%, our XGBoost model boasts 7x the Lift of the baseline. This significantly boosts fraud detection efficiency and reduces operational costs.