Fraud Detection in Financial Transactions using Unsupervised Consensus Model

Achyut Pillai*
University of California San Diego
Department of Electrical and Computer Engineering†
apillai@ucsd.edu

December 15, 2024

Abstract

This analysis investigates the use of unsupervised machine learning methods to identify suspicious financial transactions in a simulated dataset of bank transactions as fraudulent. Using clustering algorithms, probabilistic models, and statistical measures, the study demonstrates the efficacy of consensus-based anomaly detection paired with an LLM. The results include comparisons of the original and PCA-transformed data. The basis of the model is provided through anomaly scoring, covariance measures, and PCA[2] (Principal Component Analysis).

1 Introduction

Fraud detection in financial systems is imperative to reduce monetary losses and ensure trust between users and banks. Fraudulent transactions often exhibit irregular patterns that differ significantly from legitimate ones. This analysis focuses on detecting these anomalies using unsupervised machine learning methods due to the lack of labeled data.

The primary hypothesis behind this study is that combining multiple unsupervised machine learning models will be effective in detecting anomalous transactions that may be fraudulent. Using techniques such as PCA to simplify the dimensionality of the data and examining how different features correlate, we can optimize model performance, reduce noise in the data, and provide an easily interpretable result to an end user.

The anomaly detection threshold will be set at the 95th percentile based on industry observations. According to a TransUnion report, approximately 4.6% of global digital transactions are suspected to be fraudulent, a number that has increased significantly in recent years[3]. As such, we can approximate 5% of transactions as anomalies, aligning our anomaly thresholding with real-world fraud rates.

The process involves:

- Exploratory Data Analysis and Feature Selection.
- Preprocessing and Dimensionality Reduction.
- Unsupervised Anomaly Detection Using a Consensus Model.
- Evaluation and Interpretation of Results.

*PID: **A16646336**

†ECE 225A - Probability and Statistics for Data Science

2 Data Description and Preprocessing

The dataset, obtained from Kaggle[1], consists of simulated bank transaction data offering a comprehensive insight into transaction behavior. Features include transaction amounts, time, location, customer information and other behavioral markers. As this specific dataset lacks fraud labels, unsupervised methods must be employed.

2.1 Exploratory Data Analysis and Feature Selection

Feature correlations were analyzed using the correlation matrix in Figure 2. Based on this analysis, features such as **CustomerOccupation**, **DeviceID**, and **TransactionDuration** were dropped for the following reasons:

- **Redundancy:** Features exhibiting relatively high correlation with others were considered redundant.
- **Irrelevance:** Features with very low correlation to other variables were deemed insignificant for detecting anomalies.

The Pearson correlation coefficient was used to measure these feature relationships:

$$\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}, \tag{1}$$

where $\text{Cov}(X,Y)$ is the covariance between features X and Y, and σ_X, σ_Y are their standard deviations.

2.2 Preprocessing Steps

Some key preprocessing steps taken include:

- Regenerating random previous transaction dates as data was flawed. *(For most accurate results, better data where this is not required should be found)*
- Converting categorical features into numerical features
- Normalization of numerical features to ensure comparable scales.
- PCA-based dimensionality reduction to identify significant components in a more interpretable manner.

Figure 1 shows the distribution of transaction amounts, highlighting a right-skewed pattern that will inform the anomaly models.

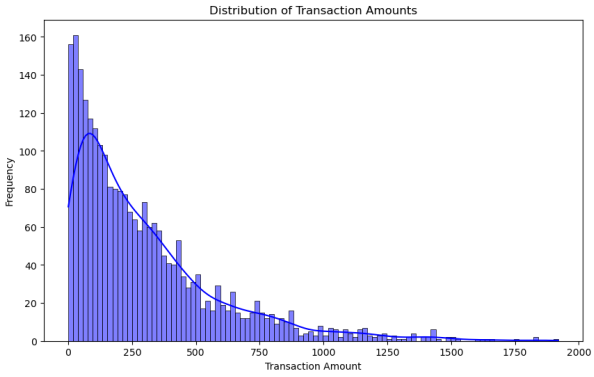

Figure 1: Distribution of transaction amounts.

Figure 2 illustrates feature correlations, which were analyzed to guide PCA and model selection.

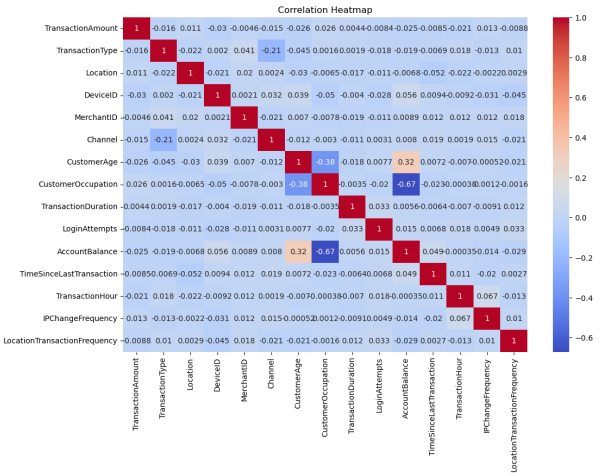

Figure 2: Correlation heatmap of key features.

3 Models and Methodology

When choosing a model, we can only consider unsupervised machine learning models as we have no labels on whether certain transactions are fraudulent. Therefore, the model selection process must prioritize key factors that align with the requirements of fraud detection in financial transactions. These factors include scalability, sensitivity to anomalies, and the ability to handle high-dimensional data, which can be analyzed through statistical tools such as distance measurements and scoring methods.

3.1 Mathematical Foundations

The anomaly detection models rely on statistical relationships within the dataset.

- **Covariance Matrix:** Captures relationships between features.

$$\Sigma = \mathbb{E}[(X - \mu)(X - \mu)^T]. \quad (2)$$

- **Mahalanobis Distance:** A measure of multivariate distance for anomaly detection.

$$d_M(x) = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)}. \quad (3)$$

- **Consensus Anomaly Scoring:** Aggregates scores from multiple models. Let S_{ij} be the anomaly score for sample i given by model j, then the consensus score is:

$$S_i^{\text{consensus}} = \frac{1}{N} \sum_{j=1}^{N} w_j S_{ij}, \quad (4)$$

where w_j are model-specific weights.

3.2 Consensus Model

A consensus framework was employed that combines five unsupervised models, chosen for their complementary strengths:

- **Isolation Forest[4]:** Efficiently isolates anomalies using random partitioning. It is computationally efficient for large datasets and effective in detecting global anomalies.

- **Local Outlier Factor (LOF)[5]:** Identifies local density deviations, capturing anomalies in regions with sparse data points.

- **One-Class SVM[8]:** Learns a hyperplane that encloses the majority of the data, flagging deviations as anomalies. It works well when normal patterns dominate the dataset.

- **DBSCAN[6]:** Detects anomalies as points that do not belong to dense clusters, effective for non-linear and complex data distributions.

- **KMeans[7]:** Clusters data and flags anomalies as points far from cluster centroids, offering a baseline approach for anomaly detection.

The consensus combines the strengths of these models, ensuring robustness and reducing biases inherent in individual techniques. In the weighted model weights were assigned based on each model's strengths. Isolation Forest (3) was given the highest weight for its efficiency in detecting global anomalies. LOF (2) and One-Class SVM (2) were weighted equally for their ability to detect local anomalies and dominant normal patterns, respectively. KMeans (2) serves as a versatile baseline, while DBSCAN (1) was given lower weight due to its sensitivity to parameters.

4 Results and Analysis

Figure 3 and Figure 4 illustrate the anomaly score distributions.

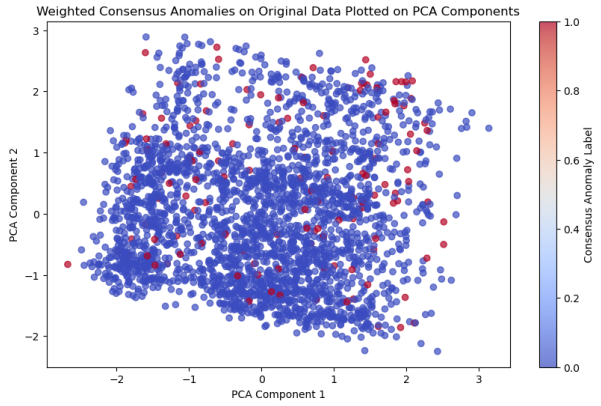

Figure 3: Weighted consensus anomaly scores on original data.

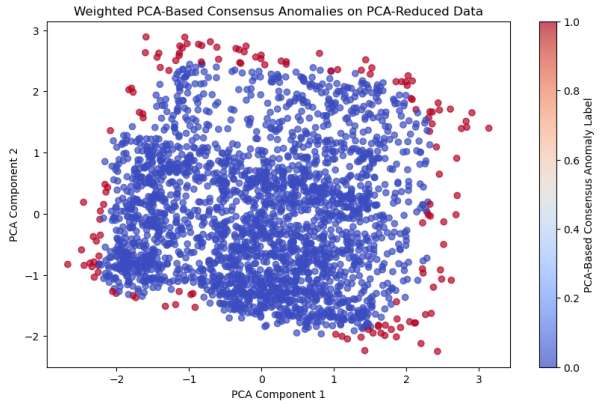

Figure 4: Weighted consensus anomaly scores on PCA-reduced data.

PCA transformation measured slightly fewer anomaly detection results:

- **Original Data:** 166 anomalies detected out of 2512 samples.
- **PCA Data:** 131 anomalies detected out of 2512 samples.
- **Weighted Original Data:** 166 anomalies detected out of 2512 samples.
- **Weighted PCA Data:** 142 anomalies detected out of 2512 samples.

As we don't have ground truth labels on fraud, we are unable to tell which model performs the best, but looking at the distribution of anomalies in Figure 4 we can see that the outliers of the PCA components are chosen. To better understand whether these declared anomalies are truly anomalies we can use an LLM to rationalize the decisions made.

4.1 LLM Integration

The LLM provided an in-depth analysis of one flagged transaction, highlighting key anomalies:

- **High Transaction Amount:** Large transaction amount compared to account balance
- **Peak Hour Transaction:** Transaction occurred during a peak hour (6:00 PM)
- **High IPChangeFrequency:** 9 IP address changes
- **Multiple Login Attempts:** 3 login attempts in a short period
- **Unusual IP Address and Location:** Inconsistent IP address and location.

> "This transaction exhibits several potential anomalies and fraud indicators, which warrant further investigation to determine the likelihood of fraudulent activity. It is essential to consider these anomalies in conjunction with other factors and patterns to ensure a comprehensive fraud detection approach."

5 Conclusion

This study highlights the use of multiple unsupervised models and weighted consensus scoring for fraud detection. Anomalies were effectively detected in both original and PCA-reduced datasets. Thresholding at the 95th percentile aligned with real-world fraud rates, ensuring practical relevance. Future work could integrate dynamic anomaly thresholds, deep-learning-based autoencoders, supervised learning with new datasets, and further LLM optimization.

Code Availabilty: The code and datasets used with all plots and full LLM output can be found: https://github.com/achyutpillai/FraudDetection/tree/main

References

[1] Kaggle, *Bank Transaction Dataset for Fraud Detection*, https://www.kaggle.com/datasets/valakhorasani/bank-transaction-dataset-for-fraud-detection

[2] Jolliffe, I. T., *Principal Component Analysis*, Springer, 2002.

[3] TransUnion, *Digital Fraud Attempts Spike 80% Globally from Pre-Pandemic*, https://newsroom.transunion.com/transunion-report-finds-digital-fraud-attempts-spike-80-globally-from-pre-pandemic/#:~:text=The%20study%20showed%20that%204.6,the%20rates%20found%20in%202019.

[4] Liu, F. T., Ting, K. M., & Zhou, Z. H., *Isolation Forest*, 2008.

[5] Breunig, M. M., et al., *LOF: Identifying Density-Based Local Outliers*, 2000.

[6] Ester, M., et al., *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*, 1996.

[7] MacQueen, J., *Some Methods for Classification and Analysis of Multivariate Observations*, 1967.

[8] Schölkopf, B., et al., *Support Vector Method for Novelty Detection*, 2001.