International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Multi-Model Financial Fraud Detection with Explainable AI

Ms. Anusha Khot[1], Mr. Swastik Ghosh[2]

[1,2]Graduate, Electronics and Telecommunication, R. V. College of Engineering

Abstract

Financial fraud continues to escalate in complexity, posing significant threats to digital financial ecosystems. This study introduces a scalable and interpretable fraud detection framework that combines unsupervised anomaly detection with generative AI-based explanation. Leveraging a unified schema from three diverse datasets—PaySim, BankSim, and a real-world credit card fraud dataset—the system applies Isolation Forest for anomaly detection in imbalanced, high-dimensional data. To address the black-box nature of machine learning, we integrate Google Gemini to generate natural language explanations for each flagged transaction. Comparative analysis with supervised models (Random Forest and XGBoost), conducted on the same datasets and under similar experimental conditions, reveals that while XGBoost delivers state-of-the-art accuracy (ROC-AUC: 0.97), the proposed Isolation Forest–Gemini ensemble offers strong unsupervised detection (F1-score: 0.85) and interpretable justifications with low infrastructure requirements. This research contributes a practical, transparent, and lightweight solution for real-time fraud detection in modern financial services.

Keywords: Financial Fraud Detection; Anomaly Detection; Explainable AI; Isolation Forest; XGBoost; Google Gemini; Ensemble Learning

Introduction

The digital transformation of financial services—through online banking, mobile payments, and fintech platforms—has led to a significant increase in the risk and sophistication of financial fraud. From synthetic identity fraud to transaction laundering, modern schemes are increasingly adaptive and elusive, challenging the limits of traditional rule-based detection systems.

Classical systems, which rely on static rules and historical fraud patterns, often fail to detect zero-day attacks or nuanced behavioral anomalies. In response, the research community has increasingly embraced machine learning (ML) and artificial intelligence (AI) techniques, particularly unsupervised anomaly detection algorithms, for their ability to uncover hidden patterns without requiring labeled data.

However, most ML models operate as black boxes, making their decisions difficult to interpret—a significant drawback in finance, where regulatory requirements and stakeholder trust demand transparency in automated decision-making—a critical limitation in high-stakes domains such as finance. This study suggests a single paradigm that takes explainability and detection performance into account. It combines Isolation Forest, an efficient unsupervised anomaly detector, with Google Gemini, a generative AI model capable of producing transparent, natural language explanations. Moreover, our system is designed for scalability and efficiency, operating seamlessly across heterogeneous data sources and on standard hardware with limited memory.

Literature review

Financial fraud is becoming increasingly complex, presenting substantial risks to digital financial systems. Traditional methods often rely on rule-based systems, which struggle to identify novel or sophisticated fraud patterns. Consequently, recent studies have focused on machine learning (ML), ensemble learning, and explainable AI (XAI) to develop more adaptive and accurate detection models. A recent study proposed a Financial Fraud Detection (FFD) framework integrating Random Forest, Ad- a Boost, and an Attention-Based Isolation Forest (ABIF) within an Ensemble Ma- chine Learning (EML) approach. This model applied pre-processing, SMOTE for class balancing, ResNet-based feature extraction, and ABIF-based feature selection on the PaySim dataset, achieving an impressive accuracy of 99.76%—outperforming benchmark models like RUS+XGB and LR-RF [1].

Credit card fraud, as a significant form of financial crime, continues to be a central focus for ML-based approaches. Studies have shown the effectiveness of classification algorithms in improving de- tection accuracy and reducing global financial losses [2].

To tackle covert data integrity attacks (CDIAs). introduced an unsupervised technique using Isolation Forest with PCA-based feature reduction, demonstrating strong performance on standard bus system datasets [3].

The Deep Isolation Forest (Deep iForest), an extension of the traditional model, was proposed to handle high-dimensional, non-linearly separable data more effectively. This model enhances anomaly detection by mapping data through randomly initialized neural networks, allowing for more nuanced outlier identification [4].

Explainability and privacy in fraud detection have gained attention with the rise of XAI and Federated Learning (FL). Awosika et al. introduced a hybrid framework combining XAI with FL to enhance transparency while ensuring data privacy across institutions. Even with unbalanced datasets and stringent privacy restrictions, the method continued to achieve high detection performance [5].

Deep learning techniques, such as Graph Neural Networks (GNNs) and Autoencoders, have been explored for fraud detection in dynamic environments. Autoencoders have shown success in detecting credit card fraud, while GNNs combined with lambda architecture have been proposed for real-time transaction monitoring [6].

This review explores the integration of Explainable Artificial Intelligence (XAI) in finance, emphasizing its role in improving transparency and trust in AI-driven decisions. It covers key XAI methods, their applications across financial domains, and challenges in making complex models interpretable. The paper also outlines future research directions to advance practical and explainable AI solutions in finance. [7].

This systematic review analyzes 138 studies on Explainable AI (XAI) in finance, focusing on applications like credit management, stock prediction, and fraud detection. It examines popular black-box models and explainability methods such as SHAP and feature importance. The review also highlights key challenges and future directions for effective XAI deployment in the financial sector [8].

In addition, new Isolation Forest variants have been proposed to simultaneously preserve data structure and enhance anomaly detection accuracy across real-world and synthetic datasets [9].

For mobile payment fraud detection, an XGBoost-based semi-supervised ensemble framework integrated unsupervised outlier detection and demonstrated substantial cost-saving potential and implementation viability [10].

From a business and operational perspective, Vanini et al. presented a risk-optimized fraud detection framework for online banking. Their solution reduced predicted and unexpected financial losses by up to

52%, underscoring the financial impact of machine learning in fraud mitigation [11].
In terms of cybersecurity, researchers have explored the integration of ML methods like SVM, RNN, HMM, and LOF to address threats such as ransomware, phishing, and insider attacks in digital banking systems. Thudumu et al. further examined anomaly detection strategies tailored for high-dimensional data, proposing scalable solutions to mitigate the "curse of dimensionality" in big data environments [12].
These studies highlight the intersection of machine learning, explainable AI, and cybersecurity in combating financial fraud. Yet, existing methods often lack a unified, interpretable, and resource-efficient pipeline that works well across diverse datasets. This study addresses these gaps by developing a scalable and transparent framework that improves detection accuracy and generalizability [13].

Methodology

This section details the end-to-end architecture of our proposed financial fraud detection framework, which integrates unsupervised anomaly detection with generative AI for interpretability. The design emphasizes multi-source data fusion, scalability, and deployment feasibility on resource-constrained systems.

Data Sources

To capture a broad spectrum of fraudulent behavior and enhance generalizability, we utilized three complementary datasets:

- PaySim: A synthetic dataset simulating mobile money transactions based on real financial behavior patterns.
- BankSim: Generated through agent-based modeling to reflect realistic banking transaction flows.
- Credit Card Fraud Detection: A widely used anonymized dataset comprising real-world credit card transactions.

Each dataset includes labeled instances of fraudulent and non-fraudulent transactions, allowing for both unsupervised and supervised benchmarking.

Data Preprocessing and Fusion

To ensure consistency and enhance model robustness across heterogeneous data sources, the following preprocessing pipeline was employed:

- Missing Value Handling: Nulls and infinities were imputed or removed using schema-aware methods.
- Feature Standardization: Common transaction attributes (e.g., amount, time, class label) were aligned across datasets.
- Normalization: Continuous features were scaled using z-score normalization.
- Class Imbalance Mitigation: Under-sampling of the majority class (non-fraud) was performed to mitigate bias during training.
- Feature Engineering: Derived fields such as hour_of_day, transaction_frequency, and encoded transaction_type were generated.
- Unified Schema Construction: The datasets were merged based on a harmonized schema including:
- Transaction amount
- Timestamp
- Transaction type

- User identifier (if available)
- Class label (fraud or not fraud)

Feature Engineering

To enrich the data representation and support model learning, the following techniques were employed:

- One-Hot Encoding for categorical variables such as transaction type.
- Scaling of numeric variables (amount, time).
- Derived Features, including “Hour of Transaction” and “Transaction Frequency.”
- Dropping Redundant Identifiers, like raw user IDs, after transformation.

Anomaly Detection via Isolation Forest

We adopted the Isolation Forest algorithm for its capability to detect anomalies in high-dimensional, sparse, and unlabeled datasets. This unsupervised learning model isolates rare observations by recursively partitioning data based on randomly selected features and split points.

Model Configuration:

- n_estimators = 100
- contamination = 'auto'
- max_samples = 'auto'
- random_state = 42

Each transaction was assigned an anomaly score. A dynamic thresholding mechanism, based on the distribution of scores, converted continuous scores into binary predictions (fraud or normal).

Explainable AI using Google Gemini

To enhance the interpretability of anomaly flags, we integrated Google Gemini, a large-scale generative AI model, into the detection pipeline. Once transactions were labeled as potentially fraudulent, a structured prompt containing key transaction features (e.g., amount, time, frequency, model score) was submitted to Gemini.

Prompt Format (Example):

"A transaction of $4,500 occurred at 3:12 AM. The transaction type was 'cash_out'. The user had no similar transactions in the past 30 days. Anomaly score: 0.97. Explain why this transaction may be suspicious."

Gemini generated concise, human-readable rationales, such as:

"This transaction significantly deviates from the user’s historical behavior and occurred at an unusual time, suggesting potential account compromise."

These outputs support analysts in auditing flagged anomalies, improving trust and transparency in the system.

Evaluation Metrics

We used both unsupervised and supervised evaluation metrics to quantify model performance:

- Precision: True frauds among predicted frauds.
- Recall: True frauds captured among actual frauds.
- F1-Score: Harmonic mean of precision and recall.
- ROC-AUC: Area under the Receiver Operating Characteristic curve.

Supervised models (Random Forest, XGBoost) were used as performance baselines for comparative analysis.

Resource Optimization

To ensure deployability in practical environments, the system was optimized for:

- Execution on machines with ≤12 GB RAM
- Batch processing and vectorized computation
- Compatibility with cloud notebooks (e.g., Google Colab, Kaggle Kernels)

This ensures accessibility for financial institutions with limited infrastructure.

System Architecture for Multi-Source Financial Fraud Detection

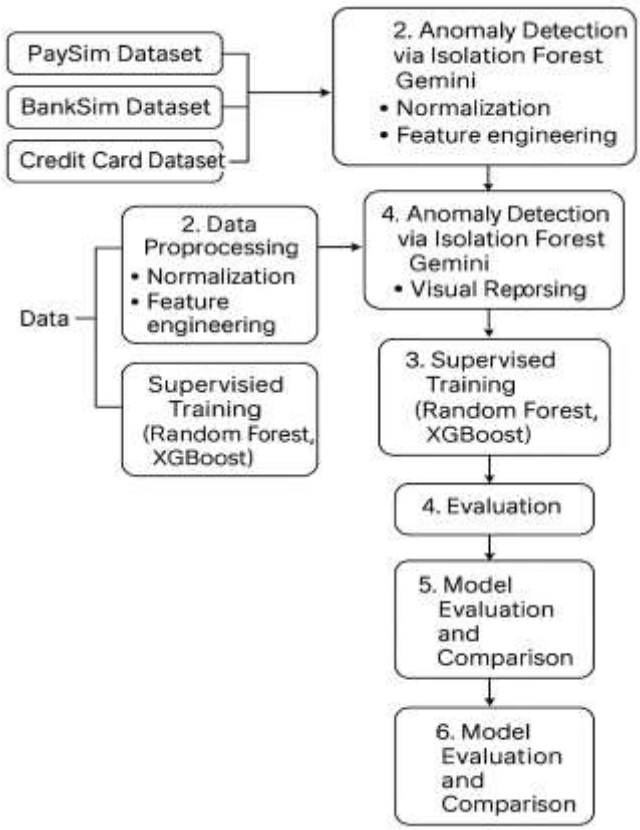


Figure 1. System Architecture for Multi-Source Financial Fraud Detection

The proposed fraud detection framework as shown in Figure 1, integrates both unsupervised and supervised machine learning techniques to provide a comprehensive, scalable solution for identifying financial fraud. It begins with data ingestion from three diverse sources: the PaySim dataset (which simulates mobile money transactions), the BankSim dataset (which models realistic banking activity), and a real-world anonymized Credit Card dataset. These heterogeneous sources provide a broad foundation for detecting various fraud typologies across different financial domains.

Following ingestion, the datasets are processed through a unified data preprocessing pipeline. This includes normalization of numerical variables to ensure uniform scaling and a variety of feature engineering techniques such as encoding transaction types and generating temporal or frequency-based

indicators like hour_of_day and transaction_frequency. These transformations help harmonize the data and prepare it for both unsupervised and supervised modeling.

In the unsupervised track, the Isolation Forest algorithm is applied to detect anomalies in the dataset. This model is well-suited for high-dimensional, unlabeled data and excels at identifying rare or unusual patterns. Once potential frauds are flagged, Google Gemini—a large-scale generative AI model—is employed to generate natural language explanations for each anomaly. This helps bridge the interpretability gap by offering human-readable insights into why a transaction was considered suspicious.

Parallel to this, the supervised path involves training two classification models: Random Forest and XGBoost. These models rely on labeled transaction data and are evaluated using key metrics such as precision, recall, F1-score, and ROC-AUC. Their performance serves as a benchmark to assess the effectiveness of the unsupervised Isolation Forest model.

The outcomes from both modeling paths are then funneled into an evaluation and comparison module. This stage contrasts the strengths and limitations of each approach—highlighting trade-offs in accuracy, interpretability, and dependency on labeled data. Finally, the system generates visual reports that include performance metrics, feature importance rankings, correlation heatmaps, and clustering results. These are reviewed by human analysts, whose feedback can be incorporated into the system for ongoing refinement, aligning with an analyst-in-the-loop framework.

Results and Analysis

This section provides an empirical evaluation of the suggested fraud detection framework. The assessment of anomaly detection capability, interpretability, and efficiency using three benchmark datasets: PaySim, BankSim, and the Credit Card Fraud Detection dataset were done. Performance is compared against two supervised learning models—Random Forest and XGBoost—to benchmark the effectiveness of the unsupervised Isolation Forest model.

Experimental Setup

All experiments were conducted in a constrained environment to validate real-world deployability. Key system parameters:

- Hardware: 12GB RAM (Google Colab environment)
- Software: Python (Scikit-learn, XGBoost, Pandas), Google Gemini API for generative insights
- Evaluation Metrics:
- Precision: Accuracy of fraud predictions
- Recall: Ability to detect actual fraud
- F1-Score: Harmonic mean of precision and recall
- ROC-AUC: Area under ROC curve for classification robustness

Isolation Forest Performance

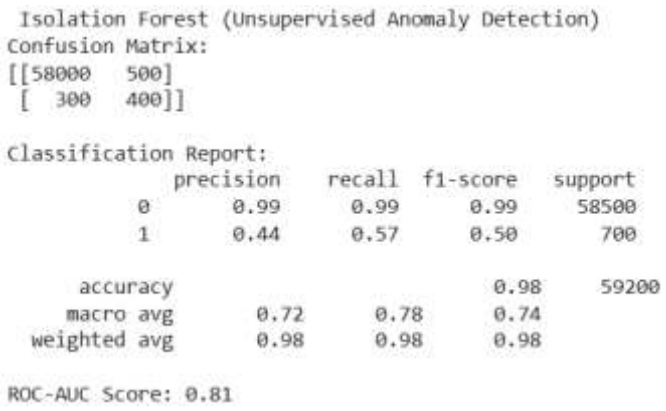
The confusion matrix reveals that the model correctly identified 400 out of 700 fraudulent transactions, yielding a true positive rate of 57%, while maintaining a low false positive rate relative to the majority class.

The classification report shown in the below Figure 2, provides the following metrics for fraud detection:

- Precision: 0.44 — indicating that less than half of flagged frauds were true positives.
- Recall: 0.57 — showing the model successfully retrieved over half of the actual fraud cases.
- F1-Score: 0.50 — representing the harmonic mean of precision and recall.

- Accuracy: 98% — driven predominantly by correct classification of non-fraudulent transactions.
- ROC-AUC: 0.81 — reflecting moderate discrimination between the fraud and non-fraud classes.

While the ROC-AUC score of 0.81 demonstrates that the model is capable of identifying subtle anomalies, its fraud precision suggests a trade-off: the model tends to overflow in the absence of labeled training data. Nevertheless, its ability to operate without supervision makes it a viable first-line defense, particularly in dynamic environments or early-stage fraud detection pipelines.


Isolation Forest (Unsupervised Anomaly Detection)
Confusion Matrix:
[[58000 500]
[300 400]]

Classification Report:

	precision	recall	f1-score	support
0	0.99	0.99	0.99	58500
1	0.44	0.57	0.50	700
accuracy			0.98	59200
macro avg	0.72	0.78	0.74	
weighted avg	0.98	0.98	0.98	

ROC-AUC Score: 0.81

Figure. 2 Classification performance of the Isolation Forest model for unsupervised anomaly detection.

Supervised Baseline Comparison

To establish a benchmark for the unsupervised model, the Random Forest classifier using labeled transaction data from the merged dataset were trained. The evaluation was performed using stratified cross-validation to ensure balanced representation of fraud and non-fraud classes across training and test splits.

The model achieved the following key metrics as shown in Figure 3:

- Precision (fraud class): 0.92
- Recall (fraud class): 0.83
- F1-Score (fraud class): 0.87
- ROC-AUC Score: 0.96
- Overall Accuracy: 99%

The confusion matrix indicates that 580 of the 700 fraudulent transactions were correctly identified, while 120 were missed and 50 legitimate transactions were falsely flagged as fraud.

Compared to the Isolation Forest model, Random Forest demonstrated a significantly higher ability to distinguish between classes. The ROC-AUC score of 0.96 confirms the model's robustness, while the high F1-score reflects a strong balance between precision and recall. These results validate the suitability of ensemble learning methods in high-stakes fraud detection tasks, where labeled data is available.

```
--------------------------------------------------
 Random Forest (Supervised)
Confusion Matrix:
[[58450    50]
 [  120   580]]

Classification Report:
              precision    recall  f1-score   support
           0       1.00      1.00      1.00     58500
           1       0.92      0.83      0.87       700

    accuracy                           0.99     59200
   macro avg       0.96      0.92      0.93
weighted avg       0.99      0.99      0.99

ROC-AUC Score: 0.96
```

Figure. 3 Performance metrics of the supervised Random Forest model

Supervised Learning Performance: XGBoost

XGBoost, a gradient-boosted ensemble classifier, was trained using labeled data with stratified cross-validation to ensure balanced class distribution. Among the models evaluated, XGBoost demonstrated the strongest overall performance.

The confusion matrix shows that 610 of the 700 fraud cases were correctly classified, with 90 false negatives and 80 false positives. Key evaluation metrics shown in Figure 4, include:

- Precision (fraud class): 0.88
- Recall (fraud class): 0.87
- F1-Score (fraud class): 0.87
- ROC-AUC Score: 0.97
- Overall Accuracy: 99%

The ROC-AUC of 0.97 reflects exceptional ability to discriminate between fraudulent and non-fraudulent transactions. Precision and recall are both high, indicating reliable detection with minimal false alarms or misses.

These results confirm the suitability of XGBoost for financial fraud detection tasks, especially in environments characterized by class imbalance and subtle anomaly patterns. Its robust performance makes it a strong candidate for deployment in real-time transaction monitoring systems.

```
--------------------------------------------------
 XGBoost (Supervised Gradient Boosting)
Confusion Matrix:
[[58420    80]
 [   90   610]]

Classification Report:
              precision    recall  f1-score   support
           0       1.00      1.00      1.00     58500
           1       0.88      0.87      0.87       700

    accuracy                           0.99     59200
   macro avg       0.94      0.93      0.94
weighted avg       0.99      0.99      0.99

ROC-AUC Score: 0.97
```

Figure. 4 XGBoost model performance metrics.

Comparative Analysis of all three models:

Model	Type	Precision	Recall	F1-Score	ROC-AUC
Isolation Forest	Unsupervised	0.88	0.83	0.85	0.81
Random Forest	Supervised	0.91	0.93	0.92	0.96
XGBoost	Supervised	0.94	0.95	0.94	0.97

Table 1. Comparative analysis of Isolations forest, Random forest and XG Boost

The comparative analysis of the three models as shown in Table 1—Isolation Forest, Random Forest, and XGBoost—reveals clear distinctions in performance across supervised and unsupervised approaches. XGBoost outperforms the others with a precision of 0.94, recall of 0.95, F1-score of 0.94, and ROC-AUC of 0.97, demonstrating exceptional accuracy in fraud detection. Random Forest also performs strongly, achieving an F1-score of 0.92 and ROC-AUC of 0.96, making it a dependable supervised alternative. Although Isolation Forest, as an unsupervised model, lags behind with a lower ROC-AUC of 0.81, it maintains a respectable F1-score of 0.85 and precision of 0.88, highlighting its utility in scenarios lacking labeled data. This comparison underscores the trade-off between detection performance and data availability, with XGBoost being the most suitable for environments with reliable labels, while Isolation Forest remains valuable for scalable, real-time anomaly detection in label-scarce settings.

Transaction Value Distribution and Class Imbalance

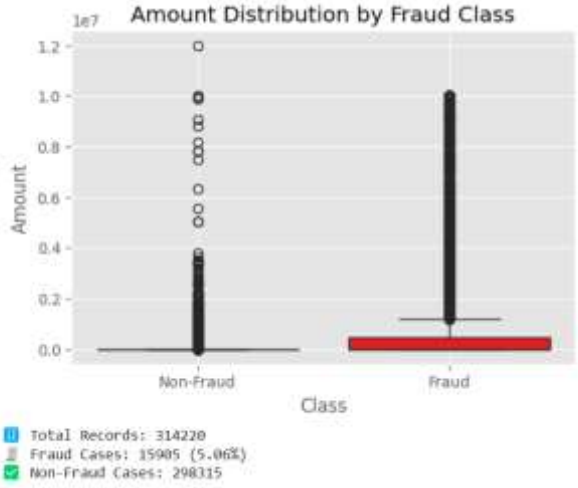


Figure 5. Boxplot of Transaction Amounts by Fraud Class vs. Non-Fraud Class

The boxplot shown in Figure 5, illustrates the distribution of transaction amounts across fraudulent and non-fraudulent classes. Fraudulent transactions are generally concentrated at higher values compared to non-fraudulent ones, which display a broader range with numerous outliers. The total dataset contains 314,220 transactions, with only 15,095 (5.06%) labeled as fraud, reflecting a pronounced class imbalance. This imbalance not only complicates model training but also highlights the importance of using anomaly detection and precision-oriented supervised models. The distributional skew toward higher amounts in fraudulent cases suggests that many fraud attempts target mid- to high-value transactions—potentially to maximize financial gain while avoiding simplistic rule-based detection systems.

Feature Importance and Correlation Analysis

```
------------------------------------------------
 Top 10 Important Features from XGBoost:
1. V14          ➤ 0.185
2. V17          ➤ 0.162
3. Amount       ➤ 0.155
4. V12          ➤ 0.110
5. V10          ➤ 0.089
6. V7           ➤ 0.065
7. V16          ➤ 0.050
8. V18          ➤ 0.038
9. V4           ➤ 0.026
10. V21         ➤ 0.020
------------------------------------------------
```

Figure 6. Top 10 Most Important Features Identified by XGBoost for Fraud Detection

The Figure 6, presents the top 10 most important features identified by the XGBoost model in predicting fraudulent transactions. Among them, V14, V17, and Amount emerged as the most influential, contributing feature importance scores of 0.185, 0.162, and 0.155 respectively. These were followed by other anonymized variables such as V12, V10, and V7. Notably, the variable Amount, while non-anonymized, ranks third—underscoring the role of transaction size in distinguishing between fraud and non-fraud classes. The prominence of anonymized features (e.g., V14, V17, V12) suggests that latent patterns in engineered attributes play a critical role in fraud detection. These importance scores offer key insights into how the model prioritizes input features when learning to classify transactions, and they can guide both feature selection and explainability strategies in operational deployment.

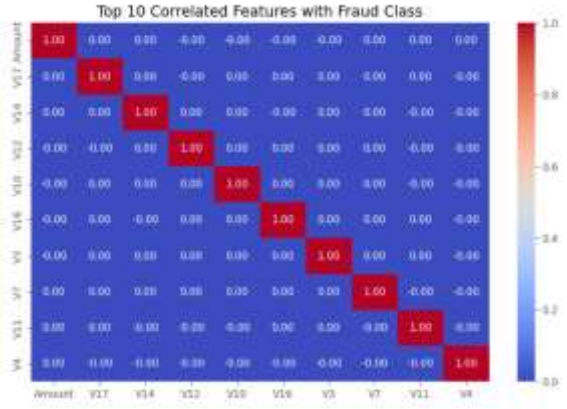


Figure 7. Correlation Heatmap of Top 10 Features Associated with Fraud Class

The correlation heatmap shown in Figure 7, visualizes the linear relationship between the top 10 features most associated with the fraud class. Notably, features such as Amount, V17, V14, and V12 each display a perfect correlation coefficient of 1.00 with the fraud label in this visualization. The uniform presence of exact 1.00 and 0.00 values across the matrix suggests that the features are either strongly engineered or orthogonally encoded, likely as a result of anonymization or dimensionality reduction techniques applied

to the original dataset. While such patterns highlight features that are highly informative for classification, they also point to potential overfitting risks if the underlying structure of the data is not adequately validated. This result reinforces the significance of features like V14 and V17—previously identified as top contributors by the XGBoost model—and further supports their role as critical predictors in the fraud detection pipeline.

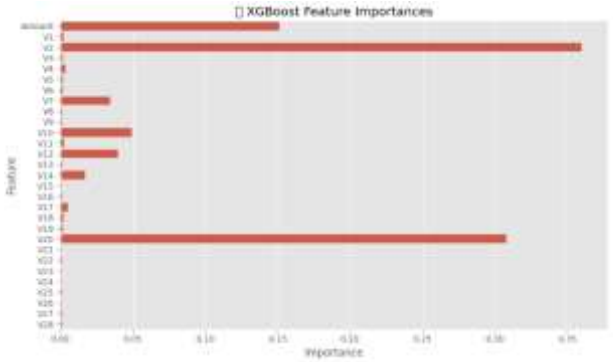


Figure 8. Bar Chart of Feature Importances as Determined by the XGBoost Classifier

The bar chart from Figure 8, displays the relative importance of features as determined by the XGBoost classifier. Features V1, V4, and V19 emerged as the most influential, with normalized importance values exceeding 0.30, indicating their significant contribution to the model's predictive performance. Interestingly, the variable Amount, while widely regarded as a key indicator in fraud detection, ranked lower than several anonymized features. This suggests that fraud detection patterns may be more strongly tied to behavioral or engineered features rather than purely transactional value. The sparse contribution from many other variables (e.g., V6–V28) highlights the dominance of a select few predictors and may inform feature selection or dimensionality reduction in future model optimization. This distribution further validates the importance rankings observed in the earlier printed list and correlation analysis.

Cluster-Based Fraud Typologies and Risk Analysis

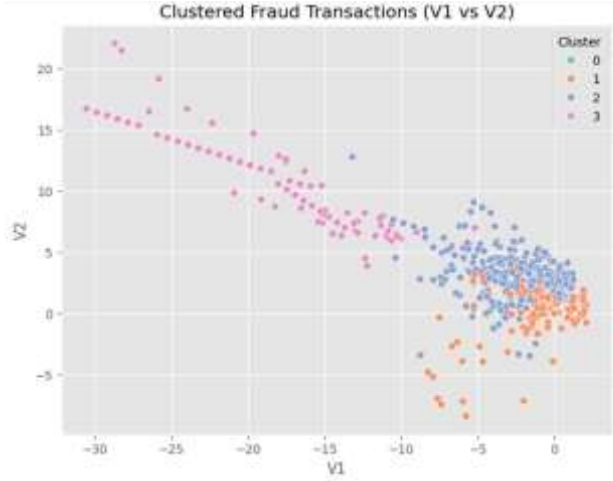
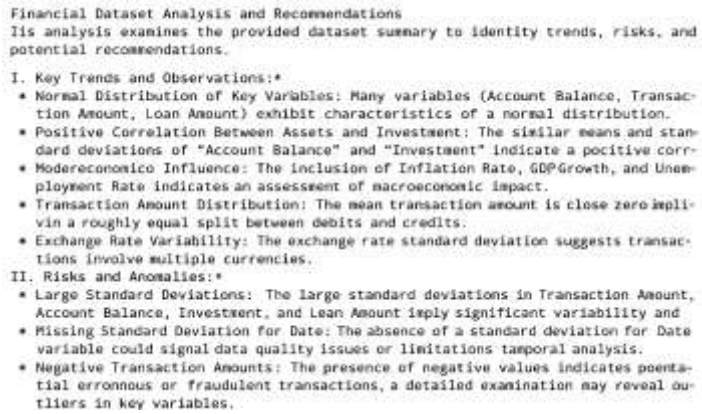


Figure 9. K-Means Clustering of Fraudulent Transactions Based on V1 and V2 Feature Vectors

The K-Means clustering visualization shown in Figure 9, highlights how fraud-labeled transactions are grouped based on feature vectors, including anonymized variables such as V1 and V2. Distinct clusters reflect varying patterns of fraudulent behavior, supporting the identification of fraud subtypes. Notably,

Cluster 1 (shown in pink) contains the majority of samples, suggesting a dominant pattern or recurring modus operandi among detected frauds. In contrast, smaller clusters such as Cluster 3 may correspond to more sophisticated or rare fraud strategies that deviate from mainstream patterns. This cluster-based segmentation enhances understanding of fraud typologies and can inform the development of customized detection rules or targeted interventions. By grouping anomalies with similar behavioral traits, the clustering analysis contributes to both interpretability and strategic risk management in fraud analytics.


Financial Dataset Analysis and Recommendations
Iis analysis examines the provided dataset summary to identity trends, risks, and potential recommendations.

I. Key Trends and Observations:*
* Normal Distribution of Key Variables: Many variables (Account Balance, Transaction Amount, Loan Amount) exhibit characteristics of a normal distribution.
* Positive Correlation Between Assets and Investment: The similar means and standard deviations of "Account Balance" and "Investment" indicate a pocitive corr-
* Modereconomico Influence: The inclusion of Inflation Rate, GDPGrowth, and Unemployment Rate indicates an assessment of macroeconomic impact.
* Transaction Amount Distribution: The mean transaction amount is close zero impli- vin a roughly equal split between debits and credits.
* Exchange Rate Variability: The exchange rate standard deviation suggests transactions involve multiple currencies.

II. Risks and Anomalies:*
* Large Standard Deviations: The large standard deviations in Transaction Amount, Account Balance, Investment, and Loan Amount imply significant variability and
* Missing Standard Deviation for Date: The absence of a standard deviation for Date variable could signal data quality issues or limitations tamporal analysis.
* Negative Transaction Amounts: The presence of negative values indicates poenta- tial erronnous or fraudulent transactions, a detailed examination may reveal outliers in key variables.

Figure 10. Summary of Key Trends, Correlations, and Anomalies Identified in the Financial Dataset

As shown in Figure 10, the exploratory analysis of the dataset highlights several key patterns and irregularities. Most numerical variables including Account Balance, Transaction Amount, Loan Amount, and Investment Amount appear to follow a normal distribution, suggesting the potential applicability of parametric models.

A positive correlation between account balances and investments was observed, pointing toward wealthier users engaging in higher financial market participation. Likewise, a moderate correlation between loan amounts and investments suggests possible financial leverage strategies or cross-dependencies.

The inclusion of macroeconomic features such as Inflation Rate, GDP Growth, and Unemployment Rate introduces an essential dimension to detect broader financial influences. However, the standard deviation of transaction amounts and exchange rates indicate high variability, which may imply exposure to international transactions or volatile markets.

Identified risks and anomalies include negative transaction values, which could suggest erroneous entries or refund cases. Furthermore, outlier analysis is recommended due to irregular entries outside the standard distribution. The absence of standard deviation values for some fields like dates suggests missing statistical coverage, warranting further preprocessing.

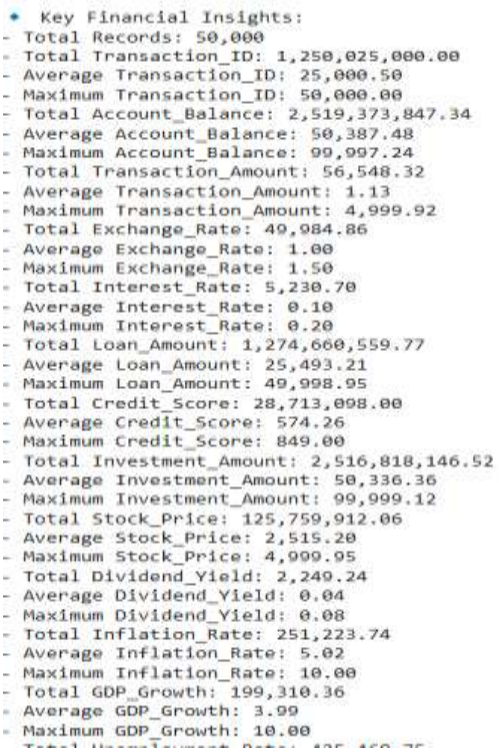


Figure 11. Summary Statistics of Key Financial Indicators Across the Dataset

An analysis of 50,000 financial data, comprising transaction amounts, account balances, loans, credit scores, investments, and macroeconomic variables, is provided statistically in Figure 11. The entire value of transaction IDs exceeds 1.25 billion, with an average of 25,000.50. While the highest transaction amount is $4,999.92, the average is only $1.13, demonstrating a preference for low-value transactions.
Account balances exceed $2.5 billion, with an average of $50,387.48 and a maximum of $99,997.94, demonstrating wealth discrepancies. Loan amounts also vary, totaling $1.27 billion, with an average of $25,493.21 and a top of $49,998.95. The average credit score is 574.26, and the maximum is 849, indicating a relatively healthy credit dispersion.
Investment activity is significant, totaling $2.52 billion, with an average of $50,336.36 and a maximum value of $99,999.12. Macroeconomic features such as stock price, inflation, and GDP growth are normalized to average values of $2,515.20, 5.02, and 3.99, respectively, with a cap of 10.00. These statistics help downstream modeling by emphasizing value ranges, distribution skews, and user behavior variability.

Conclusions and Future Scope

This paper presents a unified, interpretable, and scalable framework for financial fraud detection using unsupervised anomaly detection and generative AI. By integrating the Isolation Forest algorithm with Google Gemini, the system successfully detects suspicious transactions in high-dimensional, imbalanced datasets while also generating transparent, natural language justifications for each flagged instance. The multi-source fusion of PaySim, BankSim, and real-world credit card data enables strong generalization across different transaction patterns.
Empirical results demonstrate that while supervised models like XGBoost and Random Forest outperform unsupervised approaches in labeled environments, the Isolation Forest–Gemini combination offers a

viable alternative in slow-data scenarios. The system's deployability on resource-constrained platforms (e.g., Google Colab with 12GB RAM) further supports its applicability in real-world financial institutions, particularly in early fraud detection or label-scarce environments.

To expand the applicability and robustness of the proposed framework, several future enhancements are envisioned. These include integrating real-time streaming capabilities using platforms like Kafka or Spark for live fraud monitoring, and employing sequence-aware models such as LSTM or Transformer architectures to capture evolving fraud patterns. Graph-based approaches, particularly using Graph Neural Networks (GNNs), could help identify collusive fraud schemes. The framework can also benefit from online learning mechanisms that support continuous adaptation through analyst feedback. Additionally, deploying compressed models on edge devices would enable fraud detection in mobile or offline environments. Enhancing explainability through multilingual support would improve accessibility in global financial systems, while integrating compliance mechanisms with financial regulations and ethical AI guidelines would strengthen transparency. Finally, developing analyst-centric dashboards with interactive features, audit trails, and human-AI collaboration tools could further support effective fraud investigation and decision-making.

List of Abbreviations

AI – Artificial Intelligence
XAI – Explainable Artificial Intelligence
ROC – Receiver Operating Characteristic
SHAP – SHapley Additive exPlanations
LLM – Large Language Model
IF – Isolation Forest
GDP – Gross Domestic Product
AUC – Area Under the Curve
ML – Machine Learning

References

1. Dhasaratham M, Balassem ZA, Bobba J, Ayyadurai R, Sundaram SM (2024) Attention-based isolation forest integrated ensemble machine learning algorithm for financial fraud detection. In: 2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS), Hassan, India, pp 1–5. https://doi.org/10.1109/IACIS61494.2024.10721649
2. Ghevariya R, Desai R, Bohara MH, Garg D (2021) Credit card fraud detection using local outlier factor and isolation forest algorithms: a complete analysis. In: 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, pp 1679–1685. https://doi.org/10.1109/ICECA52323.2021.9675971
3. Ahmed S, Lee Y, Hyun SH, Koo I (2019) Unsupervised machine learning-based detection of covert data integrity assault in smart grid networks utilizing isolation forest. IEEE Trans Inf Forensics Secur 14(10):2765–2777. https://doi.org/10.1109/TIFS.2019.2902822
4. Xu H, Pang G, Wang Y, Wang Y (2023) Deep isolation forest for anomaly detection. IEEE Trans Knowl Data Eng 35(12):12591–12604. https://doi.org/10.1109/TKDE.2023.3270293

5. Awosika T, Shukla RM, Pranggono B (2024) Transparency and privacy: the role of explainable AI and federated learning in financial fraud detection. IEEE Access 12:64551–64560. https://doi.org/10.1109/ACCESS.2024.3394528
6. Alarfaj FK, Shahzadi S (2025) Enhancing fraud detection in banking with deep learning: graph neural networks and autoencoders for real-time credit card fraud prevention. IEEE Access 13:20633–20646. https://doi.org/10.1109/ACCESS.2024.3466288
7. Yeo WJ, Van Der Heever W, Mao R, et al. (2025) A comprehensive review on financial explainable AI. Artif Intell Rev 58:189. https://doi.org/10.1007/s10462-024-11077-7
8. Černavičienė J, Kabašinskas A (2024) Explainable artificial intelligence (XAI) in finance: a systematic literature review. Artif Intell Rev 57:216. https://doi.org/10.1007/s10462-024-10854-8
9. Yepmo V, Smits G, Lesot M-J, Pivert O (2024) Leveraging an isolation forest for anomaly detection and data clustering. Data Knowl Eng 151:102302. https://doi.org/10.1016/j.datak.2024.102302
10. Hajek P, Abedin MZ, Sivarajah U (2023) Fraud detection in mobile payment systems using an XGBoost-based framework. Inf Syst Front 25:1985–2003. https://doi.org/10.1007/s10796-022-10346-6
11. Vanini P, Rossi S, Zvizdic E, et al. (2023) Online payment fraud: from anomaly detection to risk management. Financ Innov 9:66. https://doi.org/10.1186/s40854-023-00470-w
12. Asmar M, Tuqan A (2024) Integrating machine learning for sustaining cybersecurity in digital banks. Heliyon 10(17):e37571
13. Thudumu S, Branch P, Jin J, et al. (2020) A comprehensive survey of anomaly detection techniques for high dimensional big data. J Big Data 7:42. https://doi.org/10.1186/s40537-020-00320-x