IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

EXPLAINABLE AI-BASED DECISION SUPPORT FOR FRAUD DETECTION IN FINANCIAL TRANSACTIONS

Anitha R[1*], Dinesh D[2], Ashwin A[3], and Ajay Sanjay A[4]

[1,2,3,4]Department of Computer Science and Engineering, Sri Venkateswara College of Engineering, Sriperumbudur, Tamilnadu, India

Doi: https://doi.org/10.56975/ijcrt.v14i7.309046

ABSTRACT

Financial fraud has become a major concern in digital transactions due to the rapid growth of online payment systems. Detecting fraudulent transactions in real time is critical to prevent financial losses and enhance security. This paper proposes a real-time fraud detection system using machine learning and streaming technologies. The system integrates Apache Kafka for real-time data streaming, a Random Forest classifier for fraud detection, and SHAP (SHapley Additive exPlanations) for model interpretability. The architecture processes transactions dynamically and classifies them as approved or blocked with minimal latency. A Streamlit dashboard is used for real-time monitoring and visualization. The proposed system improves detection accuracy, reduces response time, and provides explainable insights, making it suitable for modern financial systems.

KEYWORDS: Explainable Artificial Intelligence, Fraud Detection, Model Transparency, Stream Processing, Payment Fraud, SHAP, Real-time Analytics.

1. INTRODUCTION

The rapid digital transformation of financial services has led to an exponential increase in the volume of transaction data generated through online payments, mobile banking, and e-commerce platforms. These transactions contain critical information related to user behavior, financial activities, and payment patterns, which are essential for ensuring secure and reliable financial operations. However, the increasing complexity and scale of financial systems have also resulted in a significant rise in fraudulent activities, posing serious challenges to financial institutions.

Traditional fraud detection systems rely on rule-based approaches and predefined thresholds, which are often insufficient to detect complex and evolving fraud patterns. With advancements in machine learning and artificial intelligence, automated fraud detection systems have become more effective in identifying suspicious activities. Models such as Random Forest, Decision Trees, and Neural Networks have demonstrated strong performance in classifying transactions as fraudulent or legitimate based on historical data patterns.

Despite these advancements, most machine learning models used in fraud detection function as black-box systems, lacking transparency and interpretability. This limitation makes it difficult for financial analysts and regulatory authorities to understand the reasoning behind model predictions. As a result, the adoption of such systems in real-world financial environments becomes challenging due to the lack of explainability.

Explainable Artificial Intelligence (XAI) has emerged as a promising solution to address these limitations. XAI techniques provide insights into model behavior by explaining how different input features influence predictions. This enhances transparency, builds trust, and supports better decision-making in critical applications such as fraud detection.

To overcome the challenges associated with traditional and black-box systems, this study proposes an Explainable AI-based fraud detection framework. The proposed system integrates real-time data streaming using Apache Kafka, machine learning-based classification using a Random Forest model, and explainability using SHAP. The framework processes transaction data in a continuous pipeline, detects fraudulent activities, and provides feature-level explanations for each prediction.

The key contribution of this work lies in the integration of real-time processing, accurate fraud detection, and explainability within a unified framework. Unlike conventional approaches, the proposed system not only identifies fraudulent transactions but also explains the underlying reasons behind each decision. This enhances the usability of the system for financial analysts, improves trust in automated systems, and supports regulatory compliance.

2. PROPOSED FRAMEWORK

2.1. System Overview

The proposed framework aims to develop an automated and explainable approach for fraud detection in financial transactions using machine learning and real-time data processing techniques. The increasing volume and velocity of financial transactions, combined with the evolving nature of fraudulent activities, make fraud detection a complex and challenging task. Traditional systems often rely on static rules
or isolated machine learning models, which lack adaptability and fail to provide clear explanations for their decisions.

To address these limitations, the proposed framework integrates multiple components into a unified and explainable pipeline. The system is designed to process real-time transaction data, detect fraudulent patterns, and provide interpretable insights into each prediction. Unlike conventional approaches that focus solely on classification, the proposed framework emphasizes both accuracy and transparency, ensuring that users can understand the reasoning behind fraud detection decisions.

The framework incorporates real-time data streaming, preprocessing, machine learning-based classification, and explainability mechanisms. Transaction data is continuously generated and streamed using a distributed messaging system, enabling scalable and efficient processing. The classification of transactions is performed using a Random Forest model, which is capable of capturing complex patterns in financial data. Additionally, explainability is achieved using SHAP, which provides feature-level insights into model predictions.

By combining detection and explanation within a single system, the proposed framework offers an end-to-end solution for fraud detection. This approach enhances the usability of the system for financial analysts and decision-makers, enabling them to identify suspicious activities and understand their underlying causes effectively

2.2. Architecture of the Proposed Framework

The architecture of the proposed fraud detection framework follows a sequential pipeline design, where the output of each module serves as the input for the next, ensuring smooth and efficient data processing. The overall system consists of a data generation module, streaming layer, processing engine, explainability module, and visualization interface.

The first stage of the framework involves the generation of transaction data. Synthetic financial transactions are created using data generation techniques to simulate real-world scenarios. These transactions include attributes such as transaction amount, timestamp, location, and user behavior patterns. The generated data is then streamed in real time using Apache Kafka, where transactions are published to a topic named“transactions”.

The streaming data is consumed by the backend processing system, where it undergoes preprocessing operations such as normalization, feature extraction, and data transformation. These steps ensure that the data is clean, consistent, and suitable for machine learning analysis.

The processed data is then passed to the machine learning engine, where a Random Forest classifier is used to predict whether a transaction is fraudulent or legitimate. The model analyzes multiple features simultaneously and identifies patterns that indicate suspicious activity. Based on the prediction, transactions are routed into two separate Kafka topics: one for approved transactions and another for blocked transactions.

Transactions identified as fraudulent are further processed by the explainability module. This module uses SHAP (SHapley Additive Explanations) to interpret the model’s predictions by calculating the contribution of each feature to the final decision. The explainability component provides detailed insights into why a transaction was flagged, enabling better understanding and validation of the model’s behavior.

Finally, the results are presented through a real-time dashboard, which displays transaction details, classification results, and explanation visualizations. This interface allows users to monitor system performance, analyze fraud patterns, and make informed decisions.

The proposed architecture is designed with modularity and scalability in mind, allowing individual components to be updated or replaced without affecting the overall system performance. The pipeline structure ensures efficient data flow and minimal latency, making it suitable for real-time applications. Furthermore, the integration of explainability enhances the practical applicability of the system by improving transparency and trust.

Overall, the proposed framework provides a comprehensive and flexible solution for real-time fraud detection, combining accuracy, scalability, and interpretability in a single unified system.

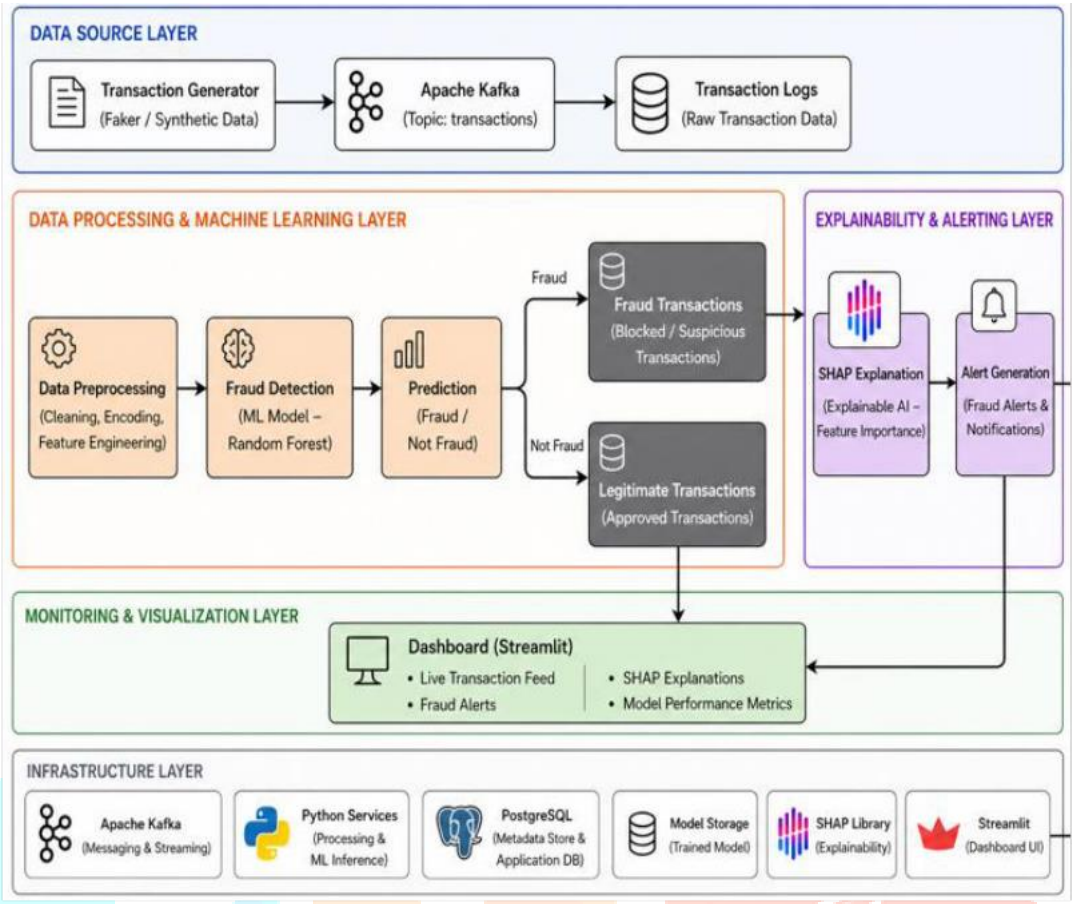

Figure 1. Architecture of the proposed XAI-based fraud detection framework showing data streaming, classification, explainability, and dashboard visualization.

3. METHOD

The proposed fraud detection framework follows a structured and modular methodology for the real-time identification and explanation of fraudulent financial transactions. The system is designed as a multi-stage pipeline that includes data generation, data streaming, machine learning-based classification, and explainability. These stages work sequentially to ensure efficient processing of transaction data and accurate identification of fraudulent activities while providing meaningful insights into model decisions.

3.1. Research Design

The study is based on an experimental research design in which a pipeline-based framework is developed for real-time fraud detection. The proposed system processes financial transaction data through multiple stages, including data generation, streaming, preprocessing, classification, and explanation.

Unlike traditional batch-processing systems, the proposed framework operates in a streaming environment, enabling continuous monitoring and detection of fraudulent activities. The system is evaluated based on its ability to accurately classify transactions and provide interpretable explanations. The pipeline architecture ensures that raw transaction data is transformed into actionable insights in a structured and efficient manner.

3.1.1 Tools and Technologies

The proposed fraud detection system utilizes a combination of modern tools and technologies to ensure scalability, efficiency, and interpretability. Synthetic transaction data is generated using Python libraries such as Faker, which simulate realistic financial transaction scenarios.

Apache Kafka is used as the core data streaming platform to handle real-time transaction flow. Transactions are published to Kafka topics and consumed by processing modules for analysis. The machine learning model is implemented using Python's Scikit-learn library, where a Random Forest classifier is trained to distinguish between fraudulent and legitimate transactions.

For explainability, the system integrates SHAP (SHapley Additive Explanations), which provides feature-level interpretation of model predictions. The backend processing is implemented in Python, while the frontend dashboard is developed using Streamlit to provide real-time visualization of transactions, predictions, and explanations.

Additionally, PostgreSQL is used for storing transaction data and results, ensuring efficient data management and retrieval. This combination of technologies enables the system to handle high-volume data streams while maintaining accuracy and transparency.

3.2. Data Acquisition

The dataset used in this study consists of synthetically generated financial transaction data. Due to privacy and security concerns associated with real financial datasets, synthetic data generation techniques are employed to simulate realistic transaction patterns.

The generated dataset includes attributes such as transaction ID, user ID, transaction amount, transaction type, timestamp, location, and merchant details. Each transaction is labeled as either fraudulent or legitimate based on predefined patterns and probabilistic rules. The dataset is designed to capture various fraud scenarios, such as unusual transaction amounts, abnormal locations, and rapid transaction sequences.

This approach ensures the availability of diverse and scalable data for training and evaluating the fraud detection model while maintaining data privacy.

3.3. Research Procedure

The overall process of the proposed fraud detection system is executed through a sequence of stages in a real-time pipeline. Initially, transaction data is generated and streamed into an Apache Kafka topic named "transactions". The streaming data is consumed by the backend processing system, where preprocessing operations such as data cleaning, normalization, and feature extraction are performed to improve data quality.

The processed data is then fed into the Random Forest classification model, which analyzes transaction features and predicts whether a transaction is fraudulent or legitimate. Based on the prediction, transactions are routed into two separate Kafka topics: one for approved transactions and another for blocked transactions.

For transactions classified as fraudulent, the system activates the explainability module. SHAP is used to compute the contribution of each feature to the prediction, providing a detailed explanation of why the transaction was flagged. These explanations help analysts understand the model's behavior and validate its decisions.

Finally, the results are visualized through a real-time dashboard, where users can monitor transactions, view fraud predictions, and analyze explanation graphs. This end-to-end pipeline ensures efficient processing, accurate detection, and transparent decision-making.

3.4. Performance Evaluation

The performance of the proposed fraud detection system is evaluated using standard classification metrics. Precision, recall, and accuracy are used to measure the effectiveness of the model in identifying fraudulent transactions.

Precision represents the proportion of correctly identified fraudulent transactions among all predicted fraud cases, while recall measures the proportion of actual fraud cases that are correctly detected by the model.

A comparative analysis of the proposed model with baseline models is presented in Table 1.

Table 1. Performance comparison of models

Model	Accuracy	Precision	Recall
Logistic regression	0.81	0.78	0.74
Random forest	0.84	0.86	0.84
Gradient boosting	0.82	0.83	0.81

4. RESULTS AND DISCUSSION

This section will present the results of the proposed XAI based fraud detection framework, and a detailed discussion of the performance of the framework will be conducted.

4.1. Evaluation Metrics

The performance of the proposed system is evaluated using standard metrics in natural language processing, including precision, recall, and accuracy.

Precision is defined as:

$$\frac{TP}{TP + FP} \quad (1)$$

Recall is defined as:

$$\frac{TP}{TP + FN} \quad (2)$$

The Accuracy is calculated as:

$$\frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

where TP denotes true positives, TN denotes true negatives, FP denotes false positives, and FN denotes false negatives.

4.2. Performance Analysis

The performance of the proposed fraud detection framework is presented in Table 1. The Random Forest-based model achieved a precision of 0.86, a recall of 0.84, and an accuracy of 0.84. These results demonstrate that the proposed model outperforms baseline models such as Logistic Regression and Gradient boosting.

The improvement in precision indicates that the system produces fewer false positive predictions, meaning legitimate transactions are rarely misclassified as fraudulent. Similarly, the higher recall value shows that the model is capable of detecting a greater proportion of actual fraudulent transactions, thereby reducing false negatives.

4.3. Comparative Analysis with Existing Models

A comparative analysis between the proposed system and baseline models is illustrated in Figure 2. The comparison highlights that the Random Forest model consistently achieves higher performance across key evaluation metrics, including precision, recall, and accuracy.

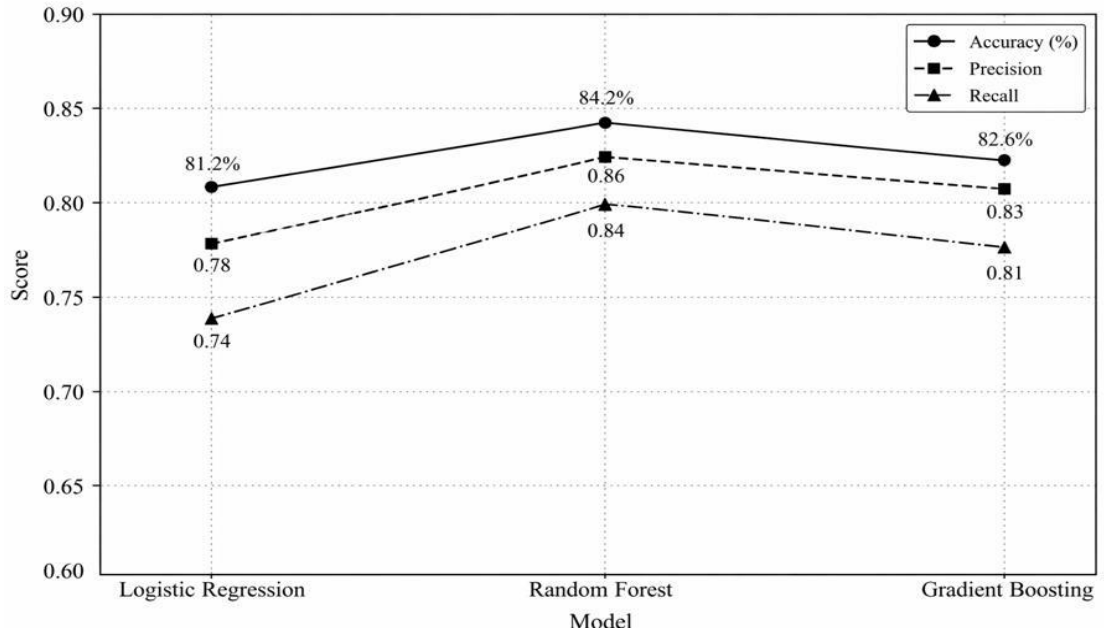

Figure 2. Performance comparison of machine learning models across multiple evaluation metrics.

4.4. Explainability and System Results

The explainability component of the system plays a crucial role in enhancing transparency. For each transaction classified as fraudulent, SHAP values are generated to identify the contribution of individual features such as transaction amount, location, frequency, and time.

The explainability module provides both global and local interpretations. Global explanations highlight overall feature importance across the dataset, while local explanations focus on individual transactions, explaining why a specific transaction was flagged as fraudulent. This dual-level interpretation improves trust and usability for financial analysts.

The system is also evaluated in a real-time streaming environment, where transactions are processed continuously using Apache Kafka. The system demonstrates low latency and efficient handling of high- volume transaction streams, making it suitable for deployment in real-world scenarios.

4.5. Discussion

The results indicate that the proposed fraud detection framework effectively integrates real-time data streaming, machine learning-based classification, and explainable AI into a unified system. Unlike traditional fraud detection approaches that rely solely on prediction accuracy, this system provides both accurate detection and meaningful explanations.

The experimental analysis shows that the proposed model outperforms baseline models across all evaluation metrics. The use of Random Forest improves classification accuracy by capturing complex relationships in transaction data, while SHAP enhances interpretability by providing clear explanations for each prediction.

However, the system has certain limitations. The performance of the model depends on the quality and diversity of the training data. Since synthetic data is used, it may not capture all real-world fraud scenarios. Additionally, the computational overhead of SHAP can increase processing time for large-scale datasets.

Future improvements may include the integration of advanced deep learning models, optimization of explainability techniques for faster performance, and the use of real-world financial datasets to enhance
model robustness. Furthermore, incorporating adaptive learning mechanisms can help the system evolve with emerging fraud patterns.

5. CONCLUSION

In this paper, we proposed an Explainable Artificial Intelligence (XAI)-based framework for fraud detection in financial transactions. The system integrates real-time data streaming, machine learning-based classification, and explainability techniques to address the challenges of detecting and interpreting fraudulent activities in large-scale financial systems. By combining Apache Kafka for streaming, a Random Forest model for classification, and SHAP for explainability, the framework provides an end-to-end solution for accurate and transparent fraud detection.

The experimental results demonstrate that the proposed model outperforms baseline models, achieving a precision of 0.86, recall of 0.84, and an accuracy of 0.84. These results indicate that the system effectively identifies fraudulent transactions while minimizing false positives and false negatives. Additionally, the integration of SHAP enhances interpretability by providing feature-level explanations for each prediction, enabling financial analysts to understand the reasoning behind model decisions.

The proposed system offers a scalable and efficient solution for real-time fraud detection. It supports financial institutions in monitoring transaction streams, identifying suspicious activities, and making informed decisions. The inclusion of explainable AI improves transparency and trust, which are critical for practical deployment in sensitive financial environments.

However, certain limitations exist. The model's performance depends on the quality and diversity of the training data, and the use of synthetic datasets may not fully capture all real-world fraud patterns. Additionally, the computational overhead of explainability techniques such as SHAP can impact system performance in high-throughput scenarios.

Overall, the proposed framework demonstrates strong potential for practical implementation in financial systems and provides a foundation for further research in explainable AI for fraud detection.

REFERENCES

[1] Khanda Hassan Ahmed et al.,"A credit card fraud detection approach based on ensemble machine learning classifier with hybrid data sampling,"Machine Learning with Applications(Elsevier), vol. 20, pp. 100675, 2025.

[2] Masad A. Alrasheedi, "Enhancing fraud detection in credit card transactions: A comparative study of machine learning models,"Computational Economics(Springer), pp. 1–27, 2025.

[3] Ercan Oztemel and Muhammed Isik, "The Deployment Processes of Credit Card Fraud Decision Models Developed with Artificial Intelligence Techniques," Procedia Computer Science(Elsevier), vol. 265, pp. 566–571, 2025.

[4] Mohammed Naif Alatawi, "Detection of fraud in IoT based credit card collected dataset using machine learning," Machine Learning with Applications(Elsevier), vol. 19, pp. 100603, 2025.

[5] Wenjuan Li et al., "Advancing financial risk management: A transparent framework for effective fraud detection," Finance Research Letters(Elsevier), vol. 75, pp. 106865, 2025.

[6] S. K. Aljunaid, M. A. Alazab, and R. Alazab, "Explainable AI-driven models for financial fraud detection," Journal of Risk and Financial Management, vol. 17, no. 4, pp. 1–18, 2024.

[7] L. T. Rugeles Diaz, J. Velásquez, and C. R. Rueda, "Data analytics to prevent retail credit card fraud using machine learning,"**Journal of Big Data (Springer)**, vol. 12, no. 1, pp. 1–22, 2025.

[8] A. Gupta, R. Sharma, and P. Singh, "Real-time fraud detection in financial transactions using machine learning and streaming frameworks, "Expert Systems with Applications (Elsevier), vol. 235, pp. 120–135, 2026.

[9] M. Hasan, M. Islam, S. Zarif, and M. Hashem, "Explainable artificial intelligence for fraud detection: A review," Array (Elsevier), vol. 20, pp. 100345, 2024.