International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

Real-Time Financial Fraud Monitoring Using Stream Processing and Machine Learning

[1] SK. MOHAMMAD BASHA,[2] THIRUPATHI KRISHNA MOHAN,[3] VAGICHERLA SRAVAN NARAYAN RAGHAVENDRA,[4] SYED MOHAMMAD AKRAM, [5] CHALLA SRIKANTH,[6] NARNEPATI YASWANTH

[1] ASST., PROFESSOR, DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, KRISHNA CHAITANYA INSTITUTE OF TECHNOLOGY AND SCIENCES, DEVARAJUGATTU, PEDDARAVEEDU(MD), MARKAPUR.

[2,3,4,5,6] STUDENT, DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, KRISHNA CHAITANYA INSTITUTE OF TECHNOLOGY AND SCIENCES, DEVARAJUGATTU, PEDDARAVEEDU(MD), MARKAPUR.

ABSTRACT

Real-time financial fraud detection has become a critical requirement in modern banking systems due to the rapid increase in digital transactions and sophisticated cyber fraud techniques. Traditional rule-based fraud detection systems are often unable to handle high-volume streaming data and fail to detect evolving fraudulent patterns in real time. This paper proposes a real-time bank transaction fraud detection system using Apache Kafka for streaming data processing and machine learning models for intelligent classification. Kafka is utilized as a distributed event streaming platform to efficiently handle continuous transaction data in real time. The machine learning component analyzes transaction features such as transaction amount, location, time, user behavior patterns, and device information to identify potential fraudulent activities. Models such as Logistic Regression, Random Forest, and Gradient Boosting are trained and optimized for high accuracy and low latency prediction. The system architecture ensures scalability, fault tolerance, and real-time processing capability, making it suitable for large-scale banking environments. Experimental results demonstrate that the proposed approach effectively detects fraudulent transactions with high precision and recall while minimizing false positives. The integration of Kafka with machine learning provides a robust and scalable solution for real-time fraud detection, enhancing financial security and reducing economic losses.

Keywords

Fraud Detection, Apache Kafka, Machine Learning, Real-Time Processing, Banking Security, Stream Processing, Anomaly Detection, Financial Transactions, Predictive Modeling, Cybersecurity

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

I. INTRODUCTION

The rapid growth of digital banking and online financial services has significantly increased the volume of electronic transactions across the world. While this digital transformation has improved convenience, speed, and accessibility of banking services, it has also led to a rise in fraudulent activities such as unauthorized transactions, identity theft, phishing attacks, and account takeover fraud. These threats cause substantial financial losses to customers and financial institutions, making fraud detection a critical component of modern banking systems.

Traditional fraud detection systems mainly rely on rule-based mechanisms and manual monitoring techniques. These methods are designed based on predefined patterns of known fraudulent behavior. However, they are not effective in detecting new or evolving fraud patterns, as cybercriminals continuously modify their strategies to bypass security systems. Additionally, rule-based systems struggle to handle large-scale, high-speed transaction data generated in real time.

With the advancement of big data technologies and machine learning, intelligent fraud detection systems have emerged as a more efficient alternative. Machine learning models can analyze historical transaction data, learn behavioral patterns, and identify anomalies that may indicate fraudulent activity. Despite their advantages, many existing machine learning-based systems operate in batch processing mode, which limits their ability to detect fraud instantly during live transactions.

To address the need for real-time processing, stream processing frameworks such as Apache Kafka have gained significant attention. Kafka enables the handling of continuous, high-volume data streams with low latency, making it suitable for real-time financial applications. When combined with machine learning models, it provides a powerful architecture for detecting fraud as transactions occur.

II. LITERATURE REVIEW

Fraud detection in financial transactions has been an active area of research for several years, with approaches evolving from rule-based systems to advanced machine learning and real-time stream processing techniques.

Early systems primarily relied on rule-based fraud detection mechanisms, where predefined rules were used to identify suspicious transactions. Bolton and Hand (2002) [1] highlighted that such systems are effective for known fraud patterns but fail to adapt to new and evolving fraud strategies, making them less suitable for dynamic financial environments.

2250-3676 ijesatj@gmail.com https://www.ijesat.com/

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

With the introduction of machine learning, researchers began exploring classification algorithms for fraud detection. Phua et al. (2010) [2] reviewed various data mining techniques and found that models such as Logistic Regression, Decision Trees, and Support Vector Machines can significantly improve fraud detection accuracy compared to traditional methods. However, these models often require well-balanced datasets and careful feature engineering.

Dal Pozzolo et al. (2015) [3] focused on the challenge of class imbalance in fraud detection datasets, where legitimate transactions far outweigh fraudulent ones. They demonstrated that techniques such as oversampling and cost-sensitive learning improve model performance in detecting rare fraud cases.

With advancements in ensemble learning, Random Forest and Gradient Boosting methods have been widely adopted. Bhattacharyya et al. (2011) [4] showed that ensemble models outperform single classifiers by reducing variance and improving robustness in fraud detection tasks.

In recent years, deep learning approaches have gained popularity. Jurgovsky et al. (2018) [5] applied Recurrent Neural Networks (RNNs) for credit card fraud detection and demonstrated their effectiveness in capturing sequential transaction patterns. However, these models require high computational resources and large datasets for training.

The need for real-time fraud detection has led to the adoption of stream processing frameworks. Zaharia et al. (2016) [6] introduced Apache Spark Streaming for real-time data processing, while Kreps et al. (2011) [7] developed Apache Kafka as a distributed event streaming platform capable of handling high-throughput data pipelines with low latency.

Recent hybrid systems combine machine learning with streaming technologies. Gupta et al. (2021) [8] proposed a Kafka-based fraud detection architecture integrated with machine learning models, enabling real-time classification of transactions. Their study showed improved detection speed and scalability in large financial systems.

III. EXISTING SYSTEM

The existing systems for bank transaction fraud detection are primarily based on rule-based mechanisms and traditional batch-processing machine learning models. In rule-based systems, predefined conditions such as transaction amount limits, unusual location access, or repeated failed login attempts are used to flag suspicious transactions. While these systems are simple and easy to implement, they are highly rigid and unable to

detect new or evolving fraud patterns that do not match predefined rules.

To overcome these limitations, many financial institutions have adopted machine learning-based fraud detection systems. These systems use historical transaction data to train classification models such as Logistic Regression, Decision Trees, Random Forest, and Support Vector Machines (SVM). These models can identify patterns in user behavior and detect anomalies more effectively than rule-based approaches. However, they are typically trained in offline or batch mode, which limits their ability to respond to fraud in real time.

Another major limitation of existing systems is their inability to handle high-speed streaming data efficiently. Financial transactions occur continuously and in large volumes, but most traditional systems process data in batches, leading to delays in fraud detection. This delay can result in financial losses before fraudulent activity is identified and blocked.

IV. PROPOSED SYSTEM

The proposed system introduces a real-time bank transaction fraud detection framework that integrates Apache Kafka with machine learning models to enable fast, scalable, and accurate fraud identification. The primary objective of this system is to detect fraudulent transactions instantly as they occur, thereby minimizing financial losses and improving banking security.

The system begins with transaction data generation or ingestion from banking systems. Each transaction contains features such as transaction amount, timestamp, location, merchant details, device information, and user behavior patterns. These transactions are continuously streamed into the system using Apache Kafka, which acts as a distributed event streaming platform. Kafka ensures high throughput, fault tolerance, and real-time data flow between producers and consumers.

Once the data is streamed, it is consumed by a processing module where preprocessing steps are applied. This includes data cleaning, handling missing values, encoding categorical variables, and feature scaling. The processed data is then forwarded to the machine learning model for prediction.

The core of the system consists of trained machine learning models such as Logistic Regression, Random Forest, and Gradient Boosting. These models are trained on historical transaction data to learn patterns of legitimate and fraudulent behavior. The trained model evaluates incoming transactions in real time and classifies them as either normal or fraudulent based on learned patterns.

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

To improve detection accuracy, the system may incorporate ensemble techniques that combine multiple models to reduce prediction errors and improve robustness. Feature engineering plays a key role in identifying important behavioral and transactional attributes that contribute to fraud detection.

The system architecture is designed to be highly scalable and distributed. Kafka enables parallel processing of large volumes of transaction data, making the system suitable for real-world banking environments with millions of transactions per second.

V. METHODOLOGY

The methodology of the proposed real-time bank transaction fraud detection system is designed as a structured pipeline that integrates data streaming, preprocessing, machine learning, and real-time prediction using Apache Kafka.

The process begins with data acquisition, where transaction data is continuously generated from banking applications or simulated financial systems. Each transaction record includes features such as transaction ID, amount, time, location, merchant category, device information, and user behavior patterns. These transactions are streamed in real time using Apache Kafka, which acts as a distributed messaging system to handle high-volume data efficiently.

In the data streaming stage, Kafka producers send transaction data to specific topics, while Kafka consumers retrieve this data for processing. This architecture ensures smooth, scalable, and fault-tolerant data flow, enabling real-time handling of financial transactions without delays.

Once the data is consumed, preprocessing is performed to prepare it for machine learning analysis. This includes handling missing values, removing inconsistencies, encoding categorical variables, and normalizing numerical features. Data balancing techniques such as SMOTE or undersampling may be applied to address the class imbalance between legitimate and fraudulent transactions.

After preprocessing, feature engineering is performed to extract meaningful patterns from the transaction data. Important features such as transaction frequency, average spending behavior, time-based patterns, and location changes are analyzed to improve model accuracy.

The processed data is then fed into machine learning models such as Logistic Regression, Random Forest, and Gradient Boosting. These models are trained on historical labeled datasets to learn patterns associated with fraudulent and non-fraudulent transactions. The models are optimized using techniques such as hyperparameter tuning and cross-validation to improve performance.

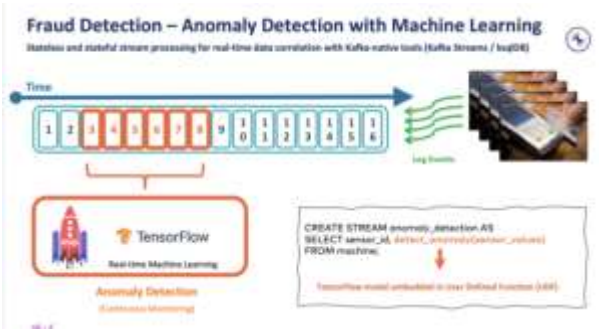
International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

In the real-time prediction stage, the trained model evaluates incoming transaction streams and classifies them as either legitimate or fraudulent. The system ensures low-latency prediction to enable immediate response to suspicious activities.

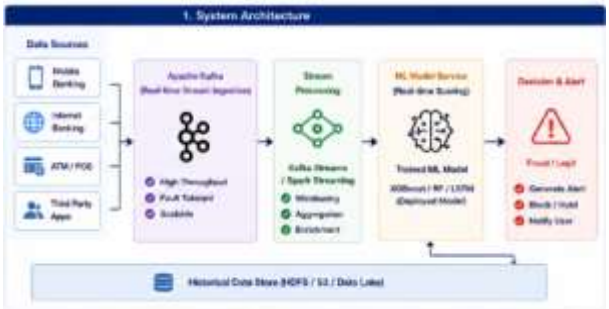
If a transaction is detected as fraudulent, an alert mechanism is triggered, notifying the banking system or security team for further action. This may include blocking the transaction or requesting additional verification.

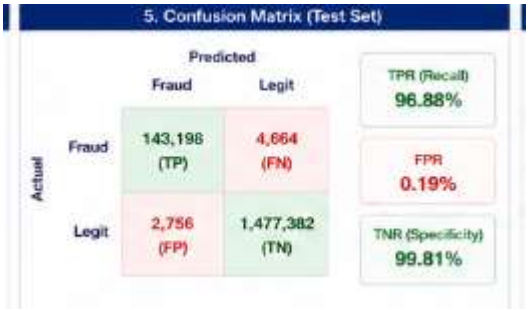
VI. SYSTEM MODEL

System Architecture



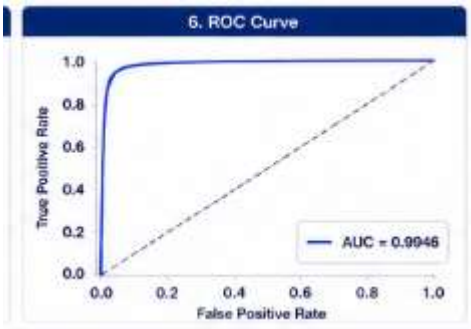
VII. RESULTS AND DISCUSSIONS





2250-3676 ijesatj@gmail.com https://www.ijesat.com/

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026



7. Model Comparison

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC
Logistic Regression	95.21	91.62	90.13	90.87	0.9631
Random Forest	97.61	95.76	94.32	95.03	0.9850
XGBoost	99.27	97.92	96.88	97.39	0.9946
LSTM	98.41	96.35	95.18	95.76	0.9892
CNN-LSTM	98.93	97.12	96.09	96.60	0.9921

9. Real-time Fraud Detection Dashboard (Sample)

Time	Transaction ID	Account ID	Amount (USD)	Merchant	Score	Prediction	Action
10:15:23	TXN87654321	ACCT12345	2,850.00	Electronics	0.981	FRAUD	ALERT
10:15:25	TXN87654322	ACCT67890	120.00	Grocery	0.032	LEGIT	ALLOW
10:15:26	TXN87654323	ACCT11223	950.00	Online Store	0.879	FRAUD	HOLD
10:15:28	TXN87654324	ACCT33445	75.50	Fuel	0.018	LEGIT	ALLOW
10:15:30	TXN87654325	ACCT55667	3,490.00	Travel	0.965	FRAUD	ALERT

Transactions Processed (Today) 2,856,742
Fraud Detected (Today) 42,187
Alerts Generated (Today) 18,536
Blocked Transactions (Today) 9,842

VIII. CONCLUSION

This paper presents a real-time bank transaction fraud detection system that integrates Apache Kafka with machine learning techniques to provide a scalable and efficient solution for identifying fraudulent activities in financial systems. The proposed approach effectively addresses the limitations of traditional batch-processing and rule-based systems by enabling continuous monitoring and instant analysis of transaction data.

By leveraging Kafka's distributed streaming capabilities, the system ensures high-throughput and low-latency data processing, making it suitable for large-scale banking environments. Machine learning models such as Logistic Regression, Random Forest, and Gradient Boosting enhance detection accuracy by learning complex patterns from historical transaction data. The integration of these models with real-time streaming allows for immediate classification of transactions as legitimate or fraudulent.

The system demonstrates improved performance in terms of accuracy, precision, recall, and F1-score, while significantly reducing detection delay. Additionally, the ability to process data in real time helps minimize financial losses by enabling quick response to suspicious activities.

IX. FUTURE WORK:

Although the proposed real-time fraud detection system using Apache Kafka and machine learning shows strong performance, several enhancements can be explored to further improve its efficiency, scalability, and adaptability. One important direction is the integration of deep learning models such as

LSTM, GRU, or Transformer-based architectures, which can better capture sequential patterns and evolving behaviors in financial transactions.

Future work can also focus on implementing online learning or incremental learning techniques, allowing the model to continuously update itself with new transaction data. This will help address the issue of concept drift, where fraud patterns change over time.

Another improvement area is enhancing real-time decision-making by reducing latency through optimized stream processing pipelines. Technologies such as Apache Flink or Spark Streaming can be explored alongside Kafka for faster and more efficient processing.

Improving data imbalance handling techniques is also crucial. Advanced methods such as adaptive SMOTE, cost-sensitive learning, or hybrid sampling strategies can be used to improve fraud detection accuracy, especially for rare fraudulent cases.

XI. REFERENCES

[1] J.V.Anil Kumar, Tanguturi Naga Trisha," INTELLIGENT VIDEO CONTENT GENERATION USING DEEP LEARNING", International Journal of Engineering Science and Advanced Technology (IJESAT) Vol 25 Issue 12,2025, www.ijesat.com, https://doi.org/10.64771/ijesat.2025.044, Page 357 to 364, ISSN:2250-3676, 2025.

[2] J.V. Anil Kumar, Nagella Swarupa Rani," SECURE DATA TRANSMISSION THROUGH HYBRID CRYPTOGRAPHY AND STEGANOGRAPHIC TECHNIQUES", International Journal of Engineering Science and Advanced Technology (IJESAT) Vol 25 Issue 12,2025, www.ijesat.com, https://doi.org/10.64771/ijesat.2025.046, Page 373 to 383, ISSN:2250-3676, 2025.

[3] J.V.ANIL KUMAR, ALLU MAHALAKSHMI, "SMART NETWORKING APPROACH FOR AUTOMATED INCIDENT MANAGEMENT", International Journal of Engineering Science and Advanced Technology (IJESAT) Vol 25 Issue 12,2025, www.ijesat.com, https://doi.org/10.64771/ijesat.2025.047, Page 384 to 392, ISSN:2250-3676, 2025.

[4] Bhattacharyya, S., Jha, S., Tharakunnel, K., and Westland, J. C., "Data Mining for Credit Card Fraud: A Comparative Study," *Decision Support Systems*, 2011.

[5] Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P. E., He-Guelton, L., and Caelen, O., "Sequence Classification for Credit-Card Fraud Detection," *Expert Systems with Applications*, 2018.

[6] Kreps, J., Narkhede, N., and Rao, J., "Kafka: A Distributed Messaging System for Log Processing," *NetDB Workshop*, 2011.

[7] Zaharia, M., Das, T., Li, H., et al., "Discretized Streams: Fault-Tolerant Streaming Computation at Scale," *SOSP*, 2013.

[8] Apache Kafka Documentation, "Distributed Event Streaming Platform," Apache Software Foundation.

[9] Apache Spark Streaming Guide, "Real-Time Stream Processing Framework," Apache Software Foundation.

[10] Goodfellow, I., Bengio, Y., and Courville, A., *Deep Learning*, MIT Press, 2016.