International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Secure Banking Transactions Through Real-Time Fraud Detection Using Apache Kafka and Machine Learning

THAKUR NIVEDITA SINGH[1], Dr.P.JOHN PAUL[2]
[1]Student, Department of Computer Science and Engineering, Ellenki College of Engineering and Technology, Patelguda, Patancharu, Hyderabad, Telangana.
[2]Principal & Professor, Department of CSE, Ellenki College of Engineering and Technology, Patelguda, Patancharu, Hyderabad, Telangana.

***Abstract*— The rapid growth of digital banking has increased the number of online transactions, making fraud detection an important challenge for financial institutions. This project presents a real-time bank transaction fraud detection system that combines Apache Kafka with Machine Learning to identify suspicious transactions as they occur. Apache Kafka is used to stream transaction data continuously, while a Random Forest classifier analyzes each transaction and predicts whether it is genuine or fraudulent. Before training the model, the transaction dataset is cleaned, encoded, and normalized to improve prediction accuracy. A Flask-based web application provides an easy-to-use interface for generating transaction streams and displaying fraud detection results in real time. The proposed system reduces the delay associated with traditional batch-processing methods and enables faster decision-making. Its modular design also allows the system to handle increasing transaction volumes, making it suitable for modern banking environments that require secure, reliable, and scalable fraud detection solutions.**

***Keywords* — Real-Time Fraud Detection, Bank Transactions, Apache Kafka, Machine Learning, Random Forest, Data Streaming, Financial Fraud Detection, Transaction Monitoring, Flask, Predictive Analytics.**

I. INTRODUCTION

The rapid growth of digital banking, online payment platforms, and mobile financial services has significantly increased the volume of electronic transactions across the world. While these advancements have improved the speed and convenience of financial services, they have also created new opportunities for cybercriminals to perform fraudulent activities such as unauthorized transactions, identity theft, account takeovers, and payment fraud. Financial institutions are therefore under increasing pressure to detect fraudulent transactions quickly and accurately before they result in financial losses. Traditional fraud detection systems mainly rely on predefined rules and historical transaction analysis, which often fail to detect newly emerging fraud patterns and introduce delays in identifying suspicious activities. Consequently, there is a growing demand for intelligent systems capable of monitoring transactions continuously and making accurate decisions in real time. Recent studies have emphasized the importance of combining real-time monitoring with advanced analytical techniques to strengthen the security of financial systems [4, 5, 8].

Real-time monitoring has become a key requirement in modern banking environments where thousands of transactions are processed every second. Streaming technologies allow financial organizations to capture, process, and analyze transaction data immediately after it is generated, thereby reducing the response time for fraud detection. Apache Kafka has emerged as one of the most reliable distributed event-streaming platforms due to its high throughput, scalability, and fault-tolerant architecture. Kafka enables seamless communication between producers and consumers, allowing continuous processing of banking transactions without significant latency. Such real-time processing frameworks support rapid identification of suspicious activities and help organizations respond before fraudulent transactions are completed. Research on transaction monitoring and security has shown that real-time auditing and continuous event processing significantly improve operational transparency and fraud prevention capabilities [5]. Similar advancements in cybersecurity and intrusion detection have also demonstrated the effectiveness of continuous monitoring for identifying abnormal system behavior [2].

Machine Learning has become an essential technology for fraud detection because of its ability

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

to learn complex transaction patterns from historical data and identify anomalies automatically. Unlike traditional rule-based methods, machine learning algorithms continuously improve their prediction capability by learning from diverse transaction records. Various techniques, including Decision Trees, Support Vector Machines, Random Forest, and ensemble learning methods, have been successfully applied to financial fraud detection. Among these algorithms, Random Forest has gained considerable attention because of its robustness, high classification accuracy, and ability to handle noisy and imbalanced datasets effectively. Artificial intelligence further enhances fraud detection by enabling intelligent decision-making and reducing the dependency on manually designed rules. Previous research has highlighted the growing importance of AI and machine learning in solving complex real-world problems, including fraud analysis and financial crime detection [3, 4, 6, 7].

The proposed work presents a Real-Time Financial Transaction Fraud Detection System Using Apache Kafka and Machine Learning, where Apache Kafka is employed for continuous transaction streaming and a Random Forest classifier is used to identify fraudulent transactions in real time. Transaction data are preprocessed through encoding, normalization, and feature transformation before being supplied to the machine learning model for prediction. A Flask-based web application provides an interactive interface for generating transaction streams and visualizing prediction results. The integration of streaming technology with machine learning enables rapid fraud detection while maintaining high scalability and reliability. This architecture provides an efficient solution for modern banking systems by minimizing detection delays, improving prediction accuracy, and supporting secure financial transaction processing in large-scale digital environments [4, 5, 6, 8].

II. RELATED WORK

"Fraud and Financial Crime Detection Model Using Malware Forensics," A. C. Kim, S. Kim, W. H. Park, and D. H. Lee [1]
This study proposed a fraud detection framework that combines malware forensic analysis with machine learning techniques to identify financial crimes in digital banking systems. The authors focused on analyzing malware behavior that targets banking applications and online financial transactions. By extracting behavioral patterns from malware attacks, the proposed model was able to distinguish between legitimate and fraudulent activities with improved accuracy. The framework integrated forensic investigation with intelligent classification methods to detect suspicious transactions before they could cause financial loss. Experimental results demonstrated that combining malware analysis with machine learning significantly enhanced fraud detection performance compared to conventional security methods. The research highlighted the importance of understanding cyberattack behavior in developing effective fraud detection systems and provided valuable insights into protecting modern financial services from increasingly sophisticated cyber threats.

"FraudMiner: A Novel Credit Card Fraud Detection Model Based on Frequent Itemset Mining," K. R. Seeja and M. Zareapoor [2]
This research introduced FraudMiner, a credit card fraud detection model based on frequent itemset mining techniques. The proposed approach analyzed historical transaction records to discover common purchasing patterns among genuine users and identify unusual transaction behavior that could indicate fraud. Instead of relying only on predefined rules, the model used association rule mining to detect rare combinations of transaction attributes associated with fraudulent activities. Experimental evaluation showed that FraudMiner achieved better precision and recall while reducing false positive rates compared to several traditional fraud detection techniques. The study demonstrated that mining frequent transaction patterns can effectively improve fraud detection performance without increasing computational complexity. The proposed model offers a practical solution for financial institutions by enabling early detection of fraudulent transactions while maintaining high accuracy in real-world banking environments.

"Applications of Artificial Intelligence in Machine Learning: Review and Prospect," S. Das, A. Dey, A. Pal, and N. Roy [3]
This review paper presented a comprehensive overview of the applications of artificial intelligence in machine learning across various domains. The authors discussed how AI techniques improve the capability of machine learning algorithms by enabling systems to learn complex patterns from large datasets with minimal human intervention. The paper reviewed different learning approaches, including supervised, unsupervised, and reinforcement learning, along with their applications in healthcare, finance, image processing, natural language processing, and cybersecurity. It also highlighted the growing importance of intelligent decision-making systems in solving real-world problems efficiently. The authors emphasized future research directions, including handling large-scale data, improving prediction accuracy, and developing

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

more adaptive learning models. This work demonstrated that artificial intelligence has become a fundamental technology for building intelligent systems capable of addressing complex computational challenges.

"Organizational Transactions with Real-Time Monitoring and Auditing," Rui Pedro F. M., Dinis Santos H. M., and Santos C. [4]
This paper focused on real-time monitoring and auditing of organizational transactions to improve operational transparency and security. The authors proposed a framework capable of continuously monitoring business transactions while maintaining detailed audit logs for future analysis. The system enabled organizations to detect abnormal activities immediately instead of relying solely on periodic audits. By integrating continuous monitoring mechanisms with transaction processing systems, the proposed approach improved compliance, accountability, and operational efficiency. The study discussed various architectural components required for implementing real-time auditing and emphasized the importance of immediate event analysis for preventing fraudulent activities. Experimental observations indicated that continuous transaction monitoring significantly enhanced the reliability of organizational information systems while reducing the risk of financial fraud and unauthorized operations. The proposed framework provides valuable support for organizations requiring secure and transparent transaction management.

"Analysis on Credit Card Fraud Detection Techniques: Based on Certain Design Criteria," M. Zareapoor, R. S. K., and M. A. Alam [5]
This paper presented a comparative analysis of various credit card fraud detection techniques based on important design criteria such as detection accuracy, computational efficiency, scalability, adaptability, and implementation complexity. The authors reviewed traditional statistical approaches, rule-based systems, anomaly detection methods, and machine learning algorithms used in financial fraud detection. Each technique was evaluated by considering its advantages, limitations, and suitability for real-world banking applications. The study concluded that machine learning methods generally outperform conventional rule-based approaches because they can identify previously unseen fraud patterns and adapt to evolving transaction behaviors. However, the authors also pointed out challenges such as class imbalance, false alarms, and high computational requirements in fraud detection systems. The research provides useful guidance for selecting appropriate fraud detection techniques according to different application requirements and operational constraints.

III. DATASET DETAILS

The dataset used in this project consists of bank transaction records collected for the purpose of developing and evaluating a real-time fraud detection system. It contains both legitimate and fraudulent transactions, allowing the machine learning model to learn the differences between normal banking behavior and suspicious activities. Each transaction record includes multiple attributes such as transaction amount, transaction type, account details, customer information, location, timestamp, merchant details, and other financial features that help characterize transaction behavior. The dataset contains both numerical and categorical features, requiring appropriate preprocessing before model training. Since real-world financial datasets often contain missing values, duplicate records, inconsistent formats, and imbalanced class distributions, preprocessing is necessary to improve data quality. The dataset is labeled into two classes: Legitimate (Non-Fraud) and Fraudulent, making it suitable for supervised machine learning. These labeled records enable the Random Forest classifier to identify hidden patterns associated with fraudulent transactions. The dataset serves as an important resource for developing intelligent fraud detection systems capable of minimizing financial losses while improving the security of digital banking environments.

Before training the machine learning model, the dataset undergoes several preprocessing steps to enhance prediction accuracy and ensure consistency. Missing values and duplicate records are removed, while categorical attributes are converted into numerical form using appropriate encoding techniques. Numerical features are normalized to bring all attributes to a common scale, preventing bias during model training. The processed dataset is then divided into training and testing subsets using an 80:20 ratio, where 80% of the transaction records are used to train the Random Forest classifier and the remaining 20% are reserved for performance evaluation. Apache Kafka is subsequently used to stream transaction data in real time, and the trained model predicts whether each incoming transaction is genuine or fraudulent. This combination of preprocessing, machine learning, and real-time streaming enables the proposed system to provide accurate, scalable, and efficient fraud detection for modern banking applications.

IV. PROPOSED METHODOLOGY

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

The proposed methodology presents a real-time bank transaction fraud detection system that integrates Apache Kafka with a Machine Learning (Random Forest) classifier to identify fraudulent transactions as they occur. Initially, the Zookeeper service is started to coordinate the distributed Kafka environment, followed by the Kafka server, which acts as the messaging platform for real-time transaction streaming. The transaction dataset is then loaded into the Jupyter Notebook environment, where all the required Python libraries and machine learning packages are initialized. Since the dataset contains both numerical and categorical transaction attributes, preprocessing techniques are applied to convert non-numeric values into numerical format. The processed data are further normalized to improve model performance and ensure consistency among all features. Finally, the dataset is divided into training and testing subsets using a 70:30 ratio, where the training data are used to build the fraud detection model while the testing data are reserved for performance evaluation.

After preprocessing, the normalized transaction features are supplied to the Random Forest classifier for model training. Random Forest is selected because of its robustness, high classification accuracy, and ability to handle heterogeneous financial transaction data. During training, the classifier learns the patterns associated with genuine and fraudulent transactions by constructing multiple decision trees and combining their predictions. After model training, predictions are performed on the testing dataset, and performance metrics such as accuracy are calculated to evaluate the effectiveness of the classifier. Although the demonstration dataset contains only five transaction records, the system architecture is designed to support large-scale banking datasets where higher prediction accuracy can be achieved. The trained Random Forest model is then integrated with Apache Kafka so that it can perform fraud prediction on incoming transaction streams in real time.

The complete system architecture illustrates the workflow of the proposed real-time fraud detection framework. It includes transaction dataset loading, preprocessing, feature normalization, Random Forest model training, Kafka producer configuration, Kafka consumer implementation, real-time fraud prediction, and web-based visualization. Apache Kafka Producer continuously publishes transaction records to Kafka topics, while the Kafka Consumer receives these transactions and forwards them to the trained machine learning model for classification. A Flask-based web application provides an interactive interface where users can generate Kafka transaction streams, consume published data, and visualize fraud prediction results instantly. The system also displays transaction details along with the predicted status as either Normal or Fraud, enabling users to monitor banking transactions efficiently.

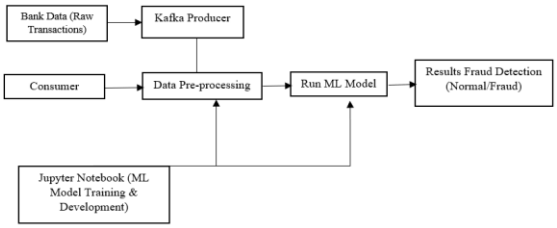


Figure 1. System Architecture for Real-Time Bank Transaction Fraud Detection Using Apache Kafka and Machine Learning

The proposed fraud detection methodology is designed to identify fraudulent banking transactions by combining streaming technology with intelligent machine learning techniques. Initially, the bank transaction dataset is uploaded into the system, and preprocessing operations are performed to clean the data, remove inconsistencies, convert categorical attributes into numerical values, and normalize feature values. After preprocessing, the dataset is divided into training and testing datasets using a 70:30 ratio. The Random Forest classifier is trained using the training data to learn transaction behavior and identify fraud patterns effectively. Once the model is trained, Apache Kafka Producer continuously publishes transaction records as streaming events, while the Kafka Consumer retrieves these events and sends them to the trained Random Forest model for prediction. Finally, the Flask-based web application displays the prediction results in real time by classifying each transaction as Normal or Fraud. The overall framework provides a scalable, efficient, and low-latency solution for real-time financial fraud detection suitable for modern digital banking environments.

V. RESULT AND DISCUSSION

The experimental results demonstrate the effectiveness of the proposed Real-Time Bank Transaction Fraud Detection System using Apache Kafka and the Random Forest machine learning algorithm. Initially, the bank transaction dataset was loaded into the Jupyter Notebook environment and preprocessed by converting categorical attributes into numerical values and normalizing the feature values. The processed dataset was divided into

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

training and testing datasets using a 70:30 ratio. The Random Forest classifier was then trained to distinguish between legitimate and fraudulent banking transactions. After training, the model successfully predicted the transaction class on the testing dataset. Although the demonstration dataset contains only five transaction records, the proposed framework successfully validates the integration of Apache Kafka with machine learning for real-time fraud detection. The trained model was subsequently connected with Kafka Producer and Consumer modules to perform fraud prediction on streaming transaction data.

The Flask-based web application enables users to generate Kafka transaction streams, consume published transactions, and visualize prediction results in real time. The system successfully classified each incoming transaction as Normal or Fraud, demonstrating the capability of the proposed architecture to support continuous transaction monitoring with minimal processing delay. The experimental observations indicate that integrating Apache Kafka with Random Forest provides a scalable and efficient framework for real-time banking fraud detection.

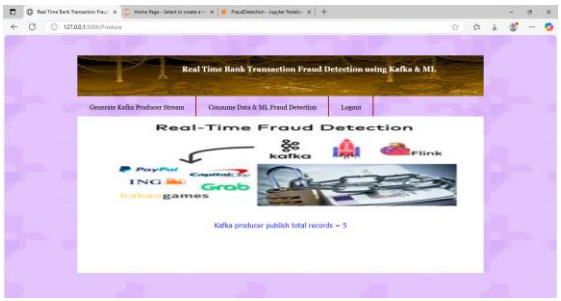
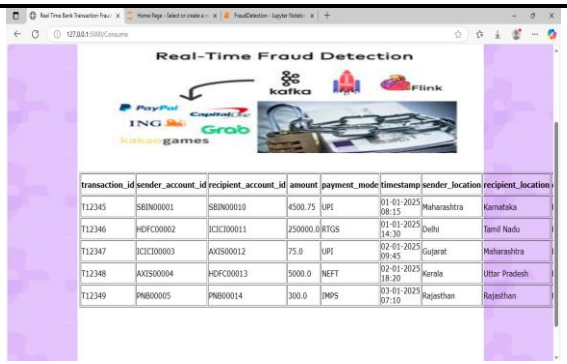


Figure [2] Main Dashboard of the Proposed System

Figure 2 illustrates the home page of the Real-Time Bank Transaction Fraud Detection System. The dashboard provides options for user login, Kafka producer stream generation, Kafka consumer execution, and fraud prediction. It serves as the central interface for accessing all functionalities of the proposed application.



transaction_id	sender_account_id	recipient_account_id	amount	payment_mode	timestamp	sender_location	recipient_location
T12345	SBIN00001	SBIN00010	4500.75	UPI	01-01-2025 08:15	Maharashtra	Karnataka
T12346	HDFC00002	ICICI00011	250000.0	RTGS	01-01-2025 14:30	Delhi	Tamil Nadu
T12347	ICICI00003	AXIS00012	75.0	UPI	02-01-2025 09:45	Gujarat	Maharashtra
T12348	AXIS00004	HDFC00013	5000.0	NEFT	02-01-2025 18:20	Kerala	Uttar Pradesh
T12349	PNB00005	PNB00014	300.0	IMPS	03-01-2025 07:10	Rajasthan	Rajasthan

Figure 3: Kafka Producer and Consumer for Real-Time Transaction Streaming

Figure 3 illustrates the integration of Apache Kafka with the machine learning model. The Kafka Producer continuously publishes banking transaction records, while the Kafka Consumer retrieves the streaming data and forwards each transaction to the trained Random Forest classifier for fraud prediction. This enables real-time processing of banking transactions.

id	recipient_account_id	amount	payment_mode	timestamp	sender_location	recipient_location	device_fingerprint	transaction_frequency	Predicted Status
	SBIN00010	4500.75	UPI	01-01-2025 08:15	Maharashtra	Karnataka	DF001	5	Normal
	ICICI00011	250000.0	RTGS	01-01-2025 14:30	Delhi	Tamil Nadu	DF002	1	Fraud
	AXIS00012	75.0	UPI	02-01-2025 09:45	Gujarat	Maharashtra	DF003	12	Normal
	HDFC00013	5000.0	NEFT	02-01-2025 18:20	Kerala	Uttar Pradesh	DF004	3	Fraud
	PNB00014	300.0	IMPS	03-01-2025 07:10	Rajasthan	Rajasthan	DF005	7	Normal

Figure 4: Real-Time Fraud Prediction Results

Figure 4 shows the prediction results generated by the proposed system. After consuming transaction records from Kafka, the Random Forest classifier predicts each transaction as either Normal or Fraud. The prediction results are displayed instantly through the Flask web interface, enabling users to monitor banking transactions in real time.

DISCUSSION

The obtained results demonstrate the successful integration of Apache Kafka and Machine Learning for real-time bank transaction fraud detection. The transaction dataset was preprocessed to improve data quality before training the Random Forest classifier. The trained model was integrated with Kafka Producer and Consumer modules, enabling real-time analysis of streaming transaction data. The Flask web application successfully displayed fraud

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

prediction results immediately after transaction processing, allowing users to monitor banking transactions continuously. Although the demonstration dataset contained only five records, the proposed architecture verified the feasibility of combining distributed streaming technology with machine learning for fraud detection. The modular design of the system supports large-scale deployment where significantly higher prediction accuracy can be achieved using larger datasets. Overall, the proposed framework provides a practical, scalable, and efficient solution for real-time financial fraud detection in modern digital banking environments.

VI. CONCLUSION

This project successfully presents a real-time bank transaction fraud detection framework by integrating Apache Kafka with Machine Learning techniques. The use of Kafka enables efficient and scalable streaming of transaction data, while the Random Forest algorithm provides reliable and interpretable fraud predictions. The system demonstrates how real-time data ingestion, intelligent classification, and web-based visualization can be combined into a unified solution. Despite being validated on a small dataset, the architecture proves to be modular, extensible, and suitable for deployment in large-scale banking environments. The project highlights the potential of real-time predictive intelligence in strengthening financial security and minimizing fraud-related losses.

REFERENCES

1. Bhadane, A., & Mane, S. B. (2017-Feb). State of research on phishing and recent trends of attacks.I-Manager's Journal on Computer Science, 5(4), 14-35. https://doi.org/10.26634/jcom.5.4.14608

2. D, A. K., & Venugopalan, S. R. (2017/09). INTRUSION DETECTION SYSTEMS: A REVIEW. International Journal of Advanced Research in Computer Science, 8(8), 356-370. https://doi.org/10.26483/ijarcs.v8i8.4703

3. Das, S., Dey, A., Pal, A., & Roy, N. (2015). Applications of artificial intelligence in machine learning: Review and prospect. International Journal of Computer Applications, 115(9) https://doi.org/10.5120/20182-2402

4. Kim, A. C., Kim, S., Park, W. H., & Lee, D. H. (2014/01/). Fraud and financial crime detection model using malware forensics. Multimedia Tools and Applications, 68(2), 479-496. https://doi.org/10.1007/s11042-013-1410-3

5. Rui Pedro, F. M., Dinis Santos, H., M., & Santos, C. (2013). Organizational transactions with real time monitoring and auditing. The Learning Organization, 20(6), 390-405.https://doi.org/10.1108/TLO-09-2013-0048

6. Seeja, K. R., & Zareapoor, M. (2014). FraudMiner: A novel credit card fraud detection model based on frequent itemset mining. The Scientific World Journal, 2014. https://doi.org/10.1155/2014/252797

7. Yang, Q., Hu, X., Cheng, Z., Kang, M. (2014). Machine Learning Based Prediction and Prevention of Malicious Inventory Occupied Orders. International Journal of Mobile Computing and Multimedia Communications, 6(4), 56-72. https://doi.org/10.4018/IJMCMC.2014100104

8. Zareapoor, M., R, S. K., & Alam, M. A. (2012). Analysis on credit card fraud detection techniques: Based on certain design criteria. International Journal of Computer Applications, 52(3) https://doi.org/10.5120/8184-1538

9. Akinapalli, S. (2026). An Ai-Powered Data Trust And Quality Scoring Framework For Enterprise Decision Intelligence Systems. International Journal of Data Science and IoT Management System, 5(1), 946-950.