Journal Publication of International Research for Engineering and Management (JOIREM)

Volume: 04 Issue: 6 | June-2026

ISSN (O) 3107-6696

FRAUD DETECTION IN FINANCIAL TRANSACTIONS USING DATA SCIENCE

Mrs. M. Vasuki[1], M. Gothai Alias Sabari Eswary[2]

[1]Professor, Department of Master Computer Application, Sri Manakula Vinayagar Engineering College (Autonomous), Madagadipet, Puducherry – 605107, India

[2]PG Student, Department of MCA, Sri Manakula Vinayagar Engineering College (Autonomous), Madagadipet, Puducherry – 605107, India

Email: gothaisabarieswary2003@gmail.com

ABSTRACT- The rapid growth of digital banking and online payment systems has increased the volume of financial transactions worldwide. While these technologies provide convenience and efficiency, they have also created opportunities for fraudulent activities. Traditional fraud detection methods mainly depend on manual monitoring and predefined rules, which are often ineffective in identifying complex and emerging fraud patterns. Therefore, there is a need for intelligent systems that can analyze transaction behavior and detect suspicious activities in real time.

his study presents a fraud detection system using Data Science and Machine Learning techniques to identify fraudulent financial transactions. The proposed system analyzes transaction attributes such as transaction amount, transaction type, sender balance, and receiver balance to predict the likelihood of fraud. A Random Forest classifier is employed to improve prediction accuracy, while a rule-based risk assessment mechanism helps categorize transactions into different risk levels. The system also provides real-time monitoring, analytics dashboards, alert generation, and report management features to support effective fraud investigation and decision-making.

Keywords: Financial Fraud Detection, Data Science, Machine Learning, Random Forest, Transaction Monitoring, Fraud Analytics.

I. INTRODUCTION

The rapid expansion of digital banking, online payment platforms, and electronic fund transfer systems has transformed the financial industry. Millions of transactions are processed every day through these digital services, providing customers with faster and more convenient financial operations. However, the increasing use of online financial systems has also led to a significant rise in fraudulent activities, creating major challenges for banks and financial institutions.

Financial fraud is one of the most serious threats faced by the banking sector. Fraudulent transactions can result in substantial financial losses, affect customer trust, and damage the reputation of organizations. As transaction volumes continue to grow, manually monitoring and verifying every transaction becomes increasingly difficult. Therefore, efficient fraud detection mechanisms are essential for maintaining secure financial operations.

Traditional fraud detection approaches mainly depend on predefined rules and manual investigation techniques. Although these methods can identify certain suspicious activities, they often fail to detect newly emerging fraud patterns and complex transaction behaviors. This limitation creates a need for intelligent systems capable of analyzing large volumes of transaction data automatically and accurately.

Recent advancements in Data Science and Machine Learning have provided effective solutions for fraud detection. Machine learning algorithms can analyze historical transaction records, identify hidden patterns, and detect unusual activities that may indicate fraudulent behavior. These techniques enable organizations to improve fraud detection accuracy while reducing manual effort.

In this study, a fraud detection system based on Data Science techniques is proposed to identify suspicious financial transactions. The system utilizes transaction-related attributes and applies a Random Forest classification model to predict fraudulent activities. In addition, the system provides real-time monitoring, risk assessment, alert generation, and analytical reporting features to support effective fraud management and decision-making.

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 6 | June-2026
ISSN (O) 3107-6696

II. RELATED WORK

Several studies have been conducted to improve fraud detection in financial transactions using data mining and machine learning techniques. Traditional fraud detection systems mainly depend on predefined rules and manual verification processes. Although these methods can identify known fraud patterns, they often struggle to detect newly emerging fraudulent activities and complex transaction behaviors.

In recent years, machine learning algorithms have been widely applied in fraud detection systems. Classification techniques such as Logistic Regression, Decision Trees, and Random Forest have been used to analyze transaction data and identify suspicious activities. These approaches help financial institutions detect fraud more accurately by learning patterns from historical transaction records.

Among various machine learning models, Random Forest has gained significant attention due to its ability to handle large datasets and improve prediction accuracy. The algorithm combines multiple decision trees to generate more reliable predictions and reduce the risk of overfitting. Several studies have reported that Random Forest performs better than traditional classification techniques when applied to financial fraud detection problems.

Table 1: Comparison of Existing Fraud Detection Approaches

Study	Method	Accuracy	Limitation
Kumar et al.	Logistic Regression	85%	Low accuracy
Singh et al.	Decision Tree	88%	Overfitting
Chen et al.	Random Forest	92%	No explainability

III. PROPOSED SYSTEM

The proposed system is designed to detect fraudulent financial transactions by analyzing transaction data using Data Science and Machine Learning techniques. The system helps financial institutions identify suspicious activities at an early stage and reduce potential financial losses. By utilizing historical transaction records, the system can recognize patterns associated with fraudulent behavior and provide accurate predictions.

The system accepts transaction-related information such as transaction type, transaction amount, sender balance, and receiver balance as input. These attributes are processed and analyzed before being provided to the machine learning model. Proper preprocessing techniques are applied to ensure that the data is suitable for prediction and analysis.

A Random Forest classification algorithm is used as the primary fraud detection model. The model is trained using transaction records and learns the characteristics of both legitimate and fraudulent transactions. Based on the input data, the model predicts whether a transaction is safe or potentially fraudulent.

The proposed system also includes a risk assessment mechanism that assigns a risk score to each transaction. Depending on the calculated score, transactions are categorized into different risk levels such as Safe, Suspicious, and Fraudulent. This allows investigators to focus on high-risk transactions and take appropriate action when necessary.

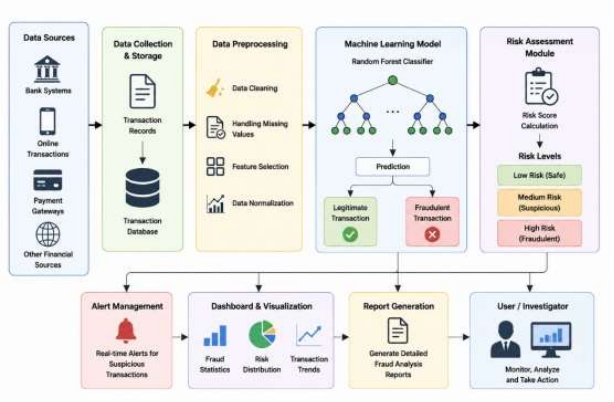

Figure 1. Architecture of the Proposed Fraud Detection System

Table 2. Summary of System Modules

Module	Function

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 6 | June-2026
ISSN (O) 3107-6696

Transaction Analysis	Collect and analyze transaction data
Preprocessing	Clean & normalize data
ML Module	Predict fraudulent transactions
Monitoring	Display real-time data
Risk Assessment	Calculate transaction risk levels
Dashboard Module	Display analytics and monitoring data
Alert Module	Generate fraud alerts
Report Module	Generate fraud reports

IV. DATASET DESCRIPTION

The dataset used in this study consists of financial transaction records collected for fraud detection analysis. It contains information related to various transaction activities performed through digital banking and online payment systems. The dataset includes both legitimate and fraudulent transactions, allowing the machine learning model to learn different transaction patterns and behaviours.

Each transaction record contains several important attributes such as transaction type, transaction amount, sender balance, receiver balance, and transaction status. These features provide valuable information about the nature of a transaction and help identify unusual activities that may indicate fraudulent behaviour. The selected attributes play a significant role in improving the accuracy of fraud prediction.

Before training the machine learning model, the dataset was pre-processed to improve data quality and consistency. Data cleaning techniques were applied to remove unnecessary information and handle missing values. Feature selection and normalization methods were also performed to ensure that the dataset was suitable for effective analysis and model training.

The processed dataset was divided into training and testing sets for model evaluation. The training dataset was used to build the Random Forest classification model, while the testing dataset was used to measure prediction performance. The availability of both normal and fraudulent transaction records helped the model learn transaction patterns effectively and improve fraud detection accuracy.

Table 3. Dataset Features Used for Fraud Detection

Feature	**Description**
Transaction Type	Type of transaction
Amount	Transaction amount
Sender Balance	Sender account balance
Receiver Balance	Receiver account balance
Transaction Status	Status of transaction
Fraud Label	Fraud or legitimate transaction

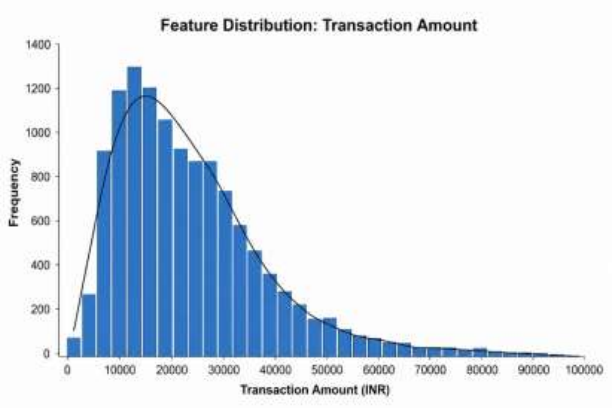

Figure 2. Feature Distribution: Transaction Amount

V. METHODOLOGY

The proposed fraud detection system follows a structured approach to identify suspicious financial transactions using

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 6 | June-2026
ISSN (O) 3107-6696

machine learning techniques. Initially, transaction records are collected from the dataset and subjected to preprocessing operations such as data cleaning, handling missing values, and feature selection. These steps help improve the quality of data and ensure better model performance.

After preprocessing, the selected features are used to train the Random Forest classification model. The model learns patterns from historical transaction records and distinguishes between legitimate and fraudulent transactions. The trained model is then evaluated using testing data to measure its prediction accuracy and effectiveness in fraud detection.Once the model is validated, it is integrated into the fraud monitoring system for real-time analysis. Incoming transactions are analyzed, assigned risk scores, and categorized into different risk levels.

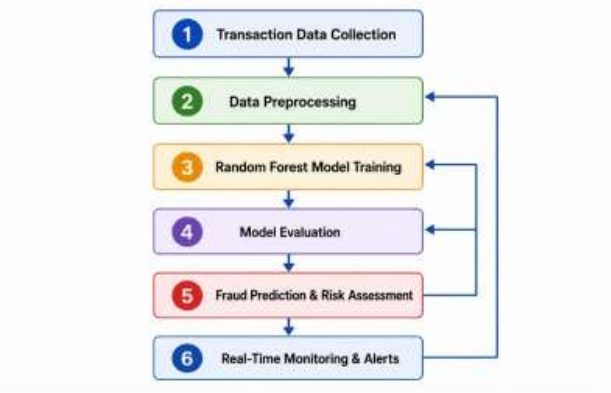

Figure 3. Machine Learning Workflow for Fraud Prediction

VI. MACHINE LEARNING MODELS

Machine Learning plays an important role in identifying fraudulent financial transactions by analyzing historical transaction data and detecting hidden patterns. In this study, the Random Forest algorithm is used as the primary classification model due to its high accuracy and ability to handle large datasets effectively. The model is trained using transaction features such as transaction type, transaction amount, sender balance, and receiver balance to classify transactions as legitimate or fraudulent.

Random Forest is an ensemble learning technique that combines multiple decision trees to improve prediction performance and reduce overfitting. During the prediction process, each decision tree provides a classification result, and the final output is determined through majority voting. The trained model evaluates incoming transactions, calculates fraud probability, and supports the risk assessment process. This approach helps improve fraud detection accuracy while minimizing false predictions.

Table 4. Comparison of Machine Learning Models used in the System

Model	Description	Performance
Logistic Regression	Basic classification model	Moderate accuracy
Decision Tree	Rule-based prediction model	Good Accuracy
Random Forest	Ensemble classification model	Higher accuracy

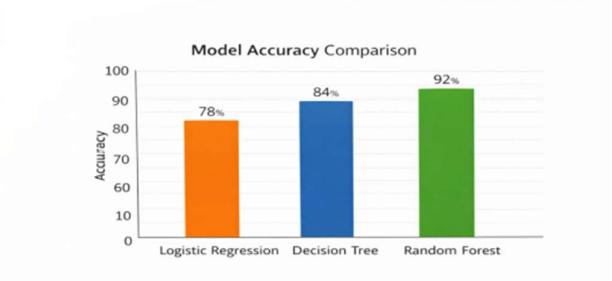

Figure 4. Model Accuracy Comparison

VII. IMPLEMENTATION

The proposed fraud detection system was developed using Python and Flask to provide an interactive and user-friendly web application. The system integrates machine learning techniques with transaction monitoring features to identify suspicious financial activities. A web-based dashboard was designed to allow users to view transaction details, fraud statistics, and risk analysis reports efficiently.

During implementation, transaction data was collected and preprocessed before being used for model training. Data cleaning and feature selection techniques were applied to improve data quality and prediction accuracy. The Random Forest algorithm was trained using historical transaction records and integrated into the application to classify transactions as legitimate or fraudulent.

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 6 | June-2026
ISSN (O) 3107-6696

The developed system provides real-time fraud detection, risk assessment, alert generation, and reporting functionalities. Users can monitor transaction activities through dashboards and receive notifications for high-risk transactions. The implementation successfully combines machine learning and data analytics to support effective fraud detection and decision-making processes.

VIII. EXPLAINABLE AI USING SHAP

xplainable Artificial Intelligence (XAI) helps users understand how machine learning models make predictions. In this study, SHAP (SHapley Additive exPlanations) is used to interpret the predictions generated by the Random Forest model. SHAP provides insights into the contribution of each feature towards the final prediction result.

The SHAP technique evaluates transaction features such as transaction amount, transaction type, sender balance, and receiver balance. By analysing feature contributions, the system identifies which factors have the greatest influence on fraud prediction. This improves transparency and helps users understand why a transaction is classified as legitimate or fraudulent.

The integration of SHAP enhances the reliability and interpretability of the proposed fraud detection system. It allows analysts to validate model predictions, identify important transaction characteristics, and support better decision-making. As a result, the system not only detects fraud effectively but also provides meaningful explanations for its predictions.

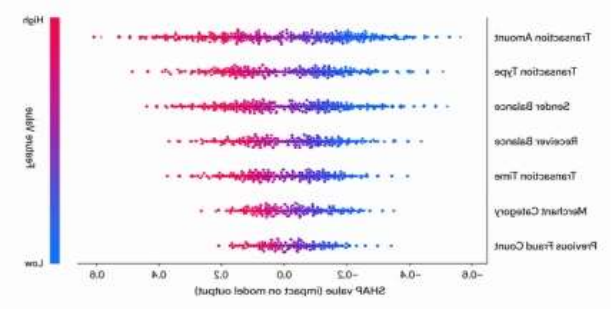

Figure 5. SHAP Summary Plot showing feature importance

IX. RESULTS AND ANALYSIS

The proposed fraud detection system was evaluated using transaction data to measure its effectiveness in identifying fraudulent activities. The Random Forest model demonstrated strong classification performance by accurately distinguishing between legitimate and fraudulent transactions. The evaluation results indicate that the model is capable of detecting suspicious transaction patterns with a high level of reliability.

The analysis of transaction records showed that features such as transaction amount, transaction type, sender balance, and receiver balance had a significant influence on fraud prediction. The SHAP-based explainability analysis further highlighted the contribution of these features in the decision-making process. This improved the transparency of the model and helped users understand the factors responsible for fraud classification.

The experimental results confirm that the proposed system can effectively support real-time fraud detection and risk assessment. The integration of machine learning and explainable AI improves both prediction accuracy and interpretability. Overall, the developed system provides a reliable solution for monitoring financial transactions and reducing the impact of fraudulent activities.

X. CONCLUSION

The proposed fraud detection system successfully utilizes Machine Learning techniques to identify fraudulent financial transactions and support secure transaction monitoring. By analyzing transaction patterns and important financial attributes, the system can effectively distinguish between legitimate and fraudulent activities. The Random Forest model demonstrated reliable performance in detecting suspicious transactions with improved accuracy.

The integration of Explainable AI using SHAP enhances the transparency and interpretability of the prediction process. The developed system provides real-time fraud detection, risk assessment, and analytical insights that can assist financial institutions in preventing financial losses. Overall, the proposed solution offers an efficient and dependable approach for fraud detection in modern digital payment systems.

XI. FUTURE WORK

The proposed fraud detection system can be further enhanced by incorporating advanced deep learning techniques and larger real-world transaction datasets. Future improvements may include the integration of adaptive learning models that continuously update their knowledge based on new transaction patterns and emerging fraud trends. This would improve the system's ability to detect complex and previously unseen fraudulent activities.

In addition, the system can be extended with real-time cloud deployment, automated alert mechanisms, and advanced visualization dashboards for fraud analysis. The inclusion of additional transaction attributes and behavioral analysis techniques may further improve prediction accuracy and decision-making. These enhancements will contribute to the development of a more intelligent, scalable, and efficient fraud detection framework.

REFERENCES

1] A. Ngai, Y. Hu, Y. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decision Support Systems*, vol. 50, no. 3, pp. 559–569, 2011.

[2] C. Phua, V. Lee, K. Smith, and R. Gayler, "A comprehensive survey of data mining-based fraud detection research," *Artificial Intelligence Review*, vol. 34, no. 1, pp. 1–14, 2010.

[3] B. Baesens, V. Van Vlasselaer, and W. Verbeke, *Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques*. Hoboken, NJ, USA: Wiley, 2015.

[4] D. Hand and W. Henley, "Statistical classification methods in consumer credit scoring: A review," *Journal of the Royal Statistical Society*, vol. 160, no. 3, pp. 523–541, 1997.

[5] R. Bolton and D. Hand, "Statistical fraud detection: A review," *Statistical Science*, vol. 17, no. 3, pp. 235–255, 2002.

[6] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY, USA: Springer, 2009.

[7] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[8] F. Provost and T. Fawcett, *Data Science for Business*. Sebastopol, CA, USA: O'Reilly Media, 2013.

[9] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Burlington, MA, USA: Morgan Kaufmann, 2012.

[10] S. Lundberg and S. Lee, "A Unified Approach to Interpreting Model Predictions," in *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 4765–4774.

[11] P. Domingos, "A few useful things to know about machine learning," *Communications of the ACM*, vol. 55, no. 10, pp. 78–87, 2012.

[12] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[13] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

[14] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. Sebastopol, CA, USA: O'Reilly Media, 2019.

[15] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. New York, NY, USA: Springer, 2013.