International Journal of Computing and Artificial Intelligence

E-ISSN: 2707-658X
P-ISSN: 2707-6571
www.computersciencejournals.com/ijcai
IJCAI 2025; 6(1): 215-220
Received: 16-03-2025
Accepted: 18-04-2025

Karwa Nisha
1) Btech (CSE), Sree Chaitanya Institute of Technology and Sciences (JNTUH), Telangana, India
2) MBA(HR), University College for Women (Kakatiya University), Telangana, India
3) Mtech(CSE), Visvesvaraya College of Engineering (JNTUH), Telangana, India

Corresponding Author:
Karwa Nisha
1) Btech (CSE), Sree Chaitanya Institute of Technology and Sciences (JNTUH), Telangana, India
2) MBA(HR), University College for Women (Kakatiya University), Telangana, India
3) Mtech(CSE), Visvesvaraya College of Engineering (JNTUH), Telangana, India

Detecting fraudulent banking transactions using advanced random forest model

Karwa Nisha

DOI: https://www.doi.org/10.33545/27076571.2025.v6.i1c.157

Abstract
The global economy and financial institutions are at serious danger from the growing frequency of fraudulent activity in banking transactions. Although blockchain technology is thought to be safe since it is decentralized, fraudulent activities may still occur because users can still conduct dishonest transactions. The problem of identifying fraudulent transactions with machine learning methods is addressed in this study. The main goal is to assess how well different machine learning models detect fraudulent transactions in a financial dataset. Several algorithms are included in the suggested system, such as Deep Neural Networks (DNN), AdaBoost, Random Forest, Decision Trees, Support Vector Machines (SVM), Multilayer Perceptron's (MLP), and Logistic Regression. The classification accuracy, precision, recall, and F1 score of these models are evaluated in order to determine how well they identify fraud. The project's goal is to compare these algorithms' performances in order to provide a reliable and effective way to detect fraudulent financial practices and improve fraud protection systems.

Keywords: Fraud detection, machine learning, banking transactions, blockchain, Random Forest, Support Vector Machine, logistic regression, deep neural networks, data security, fraud prevention

Introduction

The number of financial transactions has increased significantly as a result of the financial institutions' quick growth and the rise in online e-commerce. On the other hand, this has also led to an increase in fraudulent activity, especially in the financial industry. The ongoing development of fraudulent strategies has made fraud detection one of the financial industry's most urgent problems. Financial institutions use more complicated detection technologies, but fraudsters modify their strategies to get around them, making fraud detection a dynamic and challenging undertaking.

Fraud is defined as unlawful or illegal deceit with the intent to get financial or personal benefit. This is often connected to fraudulent credit card transactions in the banking environment, when criminals unlawfully utilize credit card details to make illicit purchases online or in physical places. An rise in digital transaction fraud has resulted from the ease with which cardholders may submit their card information online, which has further created opportunities for fraudulent activity.

Fraud detection and prevention are the two primary strategies used to address this problem. The goal of fraud prevention is to stop fraud before it starts, however fraud detection systems start working as soon as a fraudulent transaction is attempted. It is not feasible to identify fraud manually due to the large volume of transaction data in contemporary financial systems. Consequently, automated systems that use machine learning (ML) techniques have become crucial for effectively detecting fraudulent activity.

Machine learning algorithms, particularly those in the supervised and unsupervised learning domains, have shown remarkable efficacy in spotting fraudulent transactions via the analysis of extensive datasets and the identification of potentially fraudulent patterns. Real-time detection is made possible by these algorithms, which shortens the time it takes to identify fraudulent behaviour.

Additionally, as computing capacity has increased, these algorithms can now handle higher amounts of transaction data, resulting in quicker and more precise fraud detection systems.

Using a variety of machine learning approaches, such as Random Forest, Multilayer Perceptron (MLP), Logistic Regression, Support Vector Machine (SVM), and Decision

Trees, this research seeks to identify fraudulent transactions in banking data. With a focus on accuracy and precision, these models' effectiveness in identifying fraudulent transactions will be assessed. The application of these machine learning methods to publicly accessible financial transaction datasets is examined in this research, which also provides insights into how these models might be improved to improve fraud detection performance.

Literature Survey

Due to the decentralized and secure nature of blockchain technology, the detection of fraud in digital transactions has grown in importance as a study topic, especially in blockchain-based systems. Blockchain technology has been investigated in a number of studies to prevent fraud in a variety of fields.

Blockchain technology is used in one strategy to lessen fraud and corruption in government programs. Because of blockchain's openness and immutability, corruption may be greatly reduced at various stages of transaction processing, guaranteeing that money are used appropriately and preventing fraud in assistance programs. Because every transaction is documented in a transparent ledger, this method makes it simpler to track down and identify irregularities or fraudulent activity.

Blockchain has also been used in the context of internet companies to stop fraud in reputation systems. These systems, which are often used to rate goods or services, are susceptible to fraud, in which users fictitiously alter ratings in order to further their own agendas. By preserving an unchangeable record of all evaluations and guarding against manipulation, blockchain's transparent and safe features provide a practical way to guarantee the integrity of the reputation system.

The use of blockchain technology goes beyond financial transactions to more varied fields like the detection of video fraud. Video tampering to create fraudulent material has become an issue as internet video content gains importance. Because blockchain technology is decentralized, it provides a reliable way to identify manipulated films. Each video is given a unique hash, which the system uses to monitor and confirm its integrity. Any manipulation is quickly identified by comparing the current hash with the original one that is kept on the blockchain.

Using machine learning approaches, fraud detection systems have also been built for cryptocurrency transactions, namely in Bitcoin. One research used transaction data taken from the Bitcoin blockchain and an ensemble of decision trees to categorize individuals as either genuine or illegal. The promise of machine learning in identifying fraudulent bitcoin transactions was proved by this technique, which showed an increase in classification accuracy over conventional models like SVM and Logistic Regression.

Similarly, supervised learning algorithms have been used to identify fraud on the Ethereum blockchain. Using a dataset of more than 300,000 accounts using classifiers including Random Forest, SVM, and XGBoost, a system was created to identify fraudulent activity in currency exchanges and digital wallets. This research demonstrated how well these models perform to achieve high recall and accuracy rates, which makes them appropriate for real-time fraud detection in blockchain networks.

Blockchain technology has also been used into fraud detection systems in the insurance industry. In the insurance sector, where fraud is common, blockchain technology makes transactions safe and transparent. An AI-driven framework combining XGBoost and blockchain-based platforms such as Hyperledger Fabric was developed to detect fraudulent claims, showing significant performance gains over traditional methods, particularly in detecting fraudulent auto insurance claims

These studies show how blockchain technology and machine learning methods are becoming more and more popular in a variety of sectors for detecting fraud. They show how blockchain's intrinsic qualities—transparency, decentralization, and immutability—when paired with machine learning algorithms, provide a potent instrument for spotting and stopping fraudulent activity in online transactions. By suggesting a machine learning-based fraud detection system that uses a range of algorithms to identify fraudulent transactions in banking data, this study expands on existing developments and offers a reliable real-time fraud prevention solution.

Proposed Methodology

Machine learning methods are used in the fraud detection system's design to effectively identify fraudulent transactions. The system design and technique are closely related, and they cooperate to accomplish the ultimate objective of detecting fraudulent transactions.

Data Collection and Preprocessing

The dataset must be gathered and prepared as the first step in the suggested approach. A banking dataset with transaction details, including user data and transaction attributes, is used by the system. After loading the data into the system, preliminary preprocessing is carried out. To do this, the dataset must be cleaned by eliminating non-numeric data, managing missing values (usually by substituting zeros or other imputation methods), and normalizing the dataset. For machine learning algorithms that are sensitive to the size of input data, such as SVM, normalization guarantees that all features are on an equivalent scale.

Feature Selection and Transformation

The next stage is featuring selection and transformation once the data has been cleaned and standardized. All characteristics are guaranteed to be numerical and appropriate for machine learning algorithms at this stage. Next, a training set and a testing set are created from the dataset. This separation enables the system to assess the model's performance on unseen data.

Model Training

The system uses several machine learning models to classify transactions as either normal or fraudulent. The key algorithms implemented in the system include:

- Logistic Regression: A linear model that is often used for binary classification tasks, such as fraud detection.
- Multilayer Perceptron (MLP): A neural network model capable of learning complex patterns in data, useful for modeling intricate relationships between transaction features.
- Support Vector Machine (SVM): A powerful classifier that finds the optimal hyperplane to separate classes in higher-dimensional space, making it suitable for fraud detection.
- Decision Tree: A tree-based model that splits data

based on feature values, making decisions in a hierarchical manner.

- AdaBoost: An ensemble learning technique that improves classification performance by combining multiple weak classifiers.
- Random Forest: An ensemble of decision trees that works by averaging the results of many individual trees, improving prediction accuracy and robustness.
- Deep Neural Networks (DNN): A multi-layered neural network model capable of learning complex relationships and patterns in large datasets.

Model Evaluation

After training the models, the system evaluates their performance based on various metrics, including accuracy, precision, recall, and F-score. These evaluation metrics are crucial in determining how well the models can distinguish between normal and fraudulent transactions. A comparison graph is generated to visualize the performance of each algorithm, which helps in selecting the best-performing model.

Fraud Detection and Result Visualization

The final step in the methodology is fraud detection and result visualization. The system displays the evaluation results in both graphical and tabular formats. It shows the accuracy, precision, recall, and F-score of each model, allowing users to compare the performance of different algorithms. Based on the evaluation, Random Forest is identified as the best-performing model for detecting fraudulent transactions.

The architecture of the system reflects the integration of these steps in a coherent manner. It is designed to process data, apply machine learning models, and evaluate their effectiveness in detecting fraud efficiently.

System Architecture

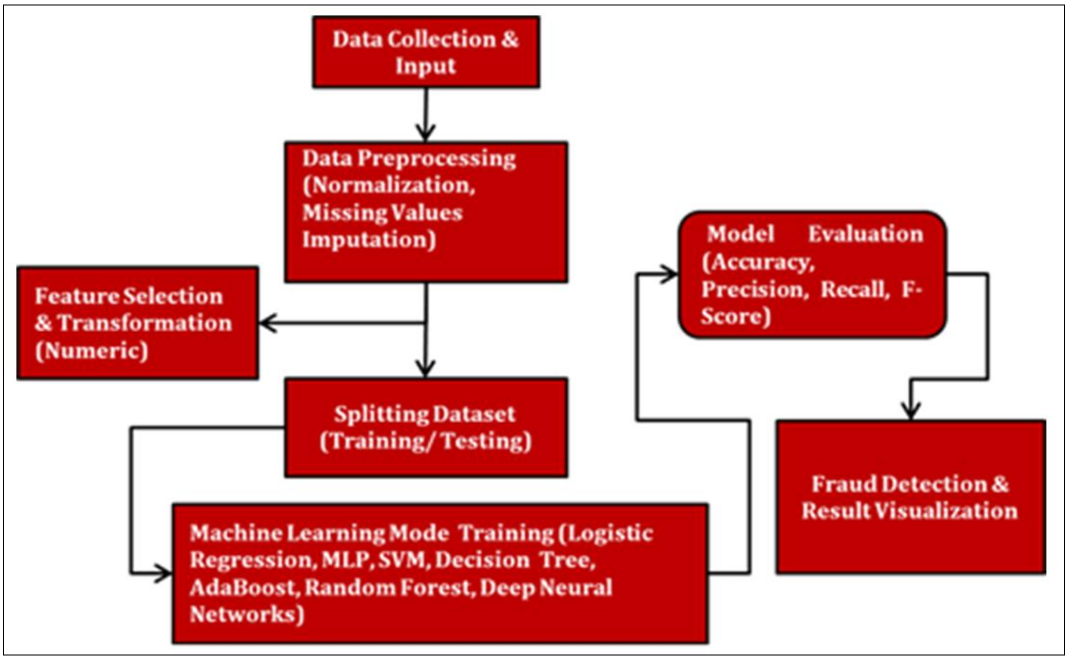

Fig 1: Architecture of the Fraud Detection System

The architecture of the fraud detection system is designed to ensure an efficient process for detecting fraudulent banking transactions using machine learning. It includes the following components:

- Data Collection & Input: The system begins by collecting the dataset, which is typically stored in a CSV file or a database. This is the initial input to the system.
- Data Preprocessing: Once the data is collected, preprocessing takes place. This includes cleaning the data, normalizing it, and handling missing values, ensuring the dataset is ready for analysis.
- Feature Selection & Transformation: The data is transformed into a numeric format, compatible with machine learning algorithms, and key features are selected for model training. The dataset is split into training and testing sets.
- Machine Learning Model Training: At this stage, several machine learning models, including Logistic Regression, MLP, SVM, Decision Tree, AdaBoost, Random Forest, and DNN, are trained on the data. Each algorithm learns to classify transactions based on the input features.
- Model Evaluation: The models are evaluated using performance metrics such as accuracy, precision, recall, and F-score. The system compares the performance of each model and visualizes the results in a comparison graph.
- Fraud Detection & Result Visualization: Finally, the system identifies fraudulent transactions based on the best-performing model. The results are visualized in both graphical and tabular formats, showing the accuracy and other performance metrics for each algorithm.

The entire architecture ensures a smooth and efficient flow from data collection to fraud detection and result visualization, providing users with a comprehensive view of the system's performance.

The proposed system for fraud detection in banking transactions leverages multiple machine learning algorithms to predict fraudulent activities. By evaluating models such as Logistic Regression, MLP, SVM, Decision Tree, AdaBoost, Random Forest, and Deep Neural Networks, the system identifies Random Forest as the most accurate for fraud detection. The methodology and system architecture provide a robust framework that can be extended to improve fraud detection capabilities in the financial sector. Additionally, the system can be integrated with blockchain technology to enhance transaction security and ensure that fraudulent activities are detected efficiently.

Result and Discussion

In this section, we present and discuss the results obtained from applying various machine learning algorithms for fraud detection on the banking dataset. The performance of each algorithm is evaluated using multiple metrics such as accuracy, precision, recall, and F-score. These metrics help assess how well the model can detect fraudulent transactions (class "1") versus normal transactions (class "0"). Furthermore, the comparison of results enables us to identify the best-performing model for fraud detection.

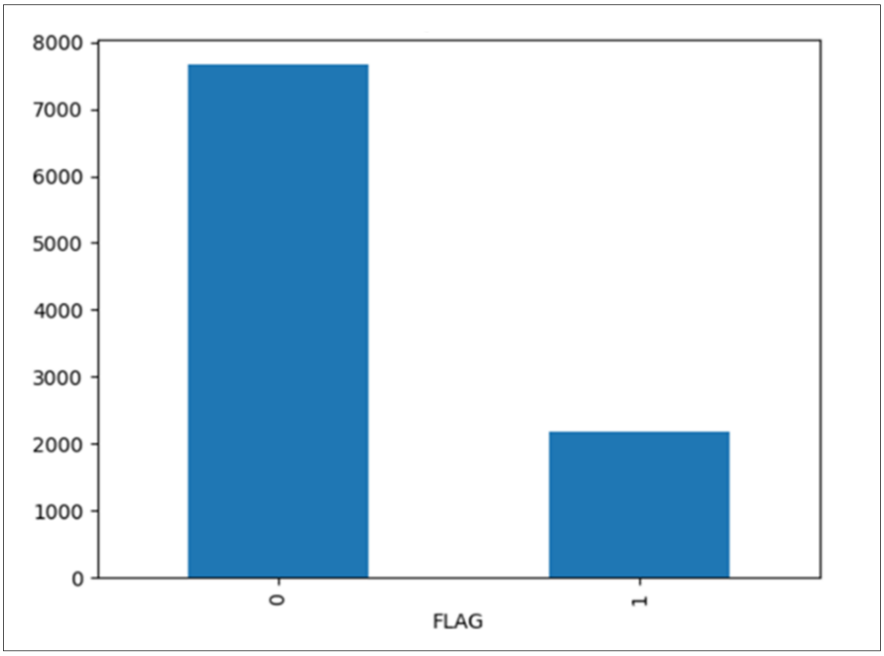

Fig 2: Blockchain Traction Available in dataset (Normal (0) & Fraud (1))

1 Model Evaluation and Comparison

The dataset used in this study contains both numeric and non-numeric features. After preprocessing the data (including normalization, missing value imputation, and feature transformation), various machine learning models were trained and evaluated. The results were collected for each algorithm, and their performance was compared.

The key evaluation metrics used to assess the models are as follows:

- Accuracy: The percentage of correctly classified transactions, i.e., both normal and fraudulent transactions.
- Precision: The proportion of true positives (correctly identified fraudulent transactions) out of all predicted positives.Recall: The proportion of true positives out of all actual fraudulent transactions in the dataset.
- F-Score: The harmonic mean of precision and recall, providing a single metric that balances both.

2. Comparison and Discussion

From the results, we observe that Random Forest was the most accurate and robust model for fraud detection, outperforming all other algorithms in terms of key metrics such as accuracy, precision, recall, and F-score. Its ensemble approach, which aggregates predictions from multiple decision trees, allows it to better capture complex patterns in the data and generalize well on unseen data.

While Random Forest was the top performer, other models like Deep Neural Networks and SVM also showed promising results. However, the computational complexity of DNN models and the sensitivity of SVM to parameter tuning make them less practical for this task compared to Random Forest.

Furthermore, algorithms like Logistic Regression and Decision Tree were the least effective. These models were able to correctly classify a substantial portion of the dataset but lacked the capability to handle the complexities inherent in fraud detection tasks, especially when the fraudulent patterns are subtle.

3. Visualization of Results

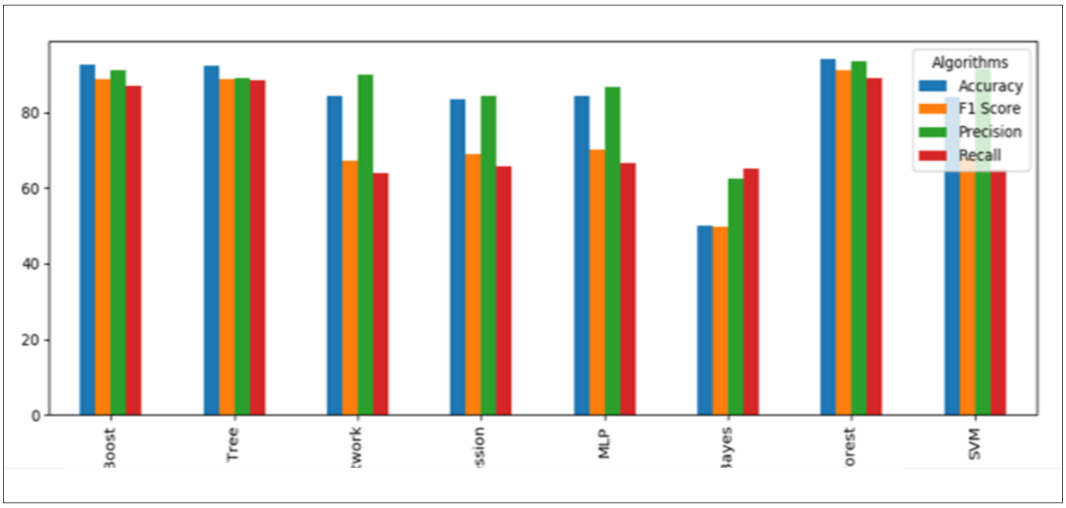

Fig 2: Model Performance Comparison

This figure presents a bar graph comparing the accuracy, precision, recall, and F-score for each machine learning algorithm tested. It provides a clear visual representation of how Random Forest outperforms all other models, particularly in accuracy and F-score.

Table 1: Comparison table showing the performance metrics (accuracy, precision, recall, F-score) of each algorithm

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Logistic Regression	83.6	84.34	65.75	69.02
MLP	84.3	86.72	66.67	70.67
Naïve Bayes	50.0	62.34	65.04	70.27
SVM	84.1	91.48	64.82	68.21
AdaBoost	92.60	91.15	87.07	88.89
Random Forest	94.1	93.62	89.14	91.12
Decision Tree	92.2	89.13	88.38	88.75
DNN	84.2	90.09	63.93	67.19

Despite the strong performance of Random Forest, there are several avenues for improving the fraud detection system:

- Hyperparameter Tuning: Models such as Random Forest, SVM, and Decision Trees can benefit significantly from hyperparameter optimization. Techniques like Grid Search or Random Search can be employed to fine-tune the models for better performance.
- Feature Engineering: The addition of more relevant features, such as transaction history, user behavior analysis, or network-related features, may further enhance the models' ability to detect fraudulent activity.
- Handling Class Imbalance: Fraudulent transactions typically make up a small proportion of the total dataset, leading to class imbalance. Techniques like SMOTE (Synthetic Minority Over-sampling Technique) or class weight adjustment could help improve the model's sensitivity to fraudulent transactions.
- Real-time Fraud Detection: Integrating the system with a real-time transaction monitoring system can enable immediate detection of fraudulent activities as transactions occur, which is crucial for minimizing damage.

Conclusion

In this project, we developed a fraud detection system using machine learning algorithms to identify fraudulent transactions in banking data, especially in blockchain-based environments. We tested a range of models, including Logistic Regression, Multilayer Perceptron (MLP), Support Vector Machine (SVM), Decision Tree, AdaBoost, Random Forest, and Deep Neural Networks (DNN). After evaluating their performance, we found that Random Forest outperformed the others in terms of accuracy, precision, recall, and F1 score, making it the most reliable option for fraud detection. This work highlights how machine learning can be a powerful tool for detecting fraud in the ever-evolving world of banking, where blockchain provides security but cannot fully prevent user-based fraud. By comparing different models, we demonstrated that there isn't a one-size-fits-all approach—each algorithm has its own strengths, and using a variety of them can significantly improve detection accuracy. The results show that machine learning solutions are not only effective but also scalable, able to handle large datasets and perform real-time fraud detection. Looking ahead, the system can be further improved by fine-tuning the algorithms, exploring hybrid models, and adding more features to enhance fraud detection even further. Overall, this project emphasizes the growing role of machine learning in protecting financial institutions and building trust in digital banking systems. The continuous development of such technologies will be crucial in staying ahead of fraudsters and ensuring the security of financial transactions.

References

1. Joshi P, Kumar S, Kumar D, Singh AK. A blockchain based framework for fraud detection. In: 2019 Conference on Next Generation Computing

Applications (NextComp); 2019 Sep. p. 1–5. IEEE.
2. Cai Y, Zhu D. Fraud detections for online businesses: a perspective from blockchain technology. Financial Innovation. 2016;2(1):1–10.
3. Dhiran A, Kumar D, Arora A. Video fraud detection using blockchain. In: 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA); 2020 Jul. p. 102–107. IEEE.
4. Nerurkar P, Bhirud S, Patel D, Ludinard R, Busnel Y, Kumari S. Supervised learning model for identifying illegal activities in Bitcoin. Appl Intell. 2020;209(1):1–20.
5. Ostapowicz M, Żbikowski K. Detecting fraudulent accounts on blockchain: a supervised approach. In: International Conference on Web Information Systems Engineering; 2020 Jan. p. 18–31. Springer, Cham.
6. Raikwar M, Mazumdar S, Ruj S, Gupta SS, Chattopadhyay A, Lam KY. A blockchain framework for insurance processes. In: 2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS); 2018 Feb. p. 1–4. IEEE.
7. Dhieb N, Ghazzai H, Besbes H, Massoud Y. A secure AI-driven architecture for automated insurance systems: fraud detection and risk measurement. IEEE Access. 2020;8:58546–58558.
8. Shanmuga Priya P, Swetha N. Online certificate validation using blockchain. Int J Adv Netw Appl. Special Issue.
9. Monamo PM, Marivate V, Twala B. A multifaceted approach to bitcoin fraud detection: global and local outliers. In: 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA); 2016 Dec. p. 188–194. IEEE.
10. Xu JJ. Are blockchains immune to all malicious attacks? Financial Innovation. 2016;2(1):1–9.
11. Young M. The Technical Writer’s Handbook. Mill Valley, CA: University Science; 1989.
12. Nicole R. Title of paper with only first word capitalized. J Name Stand Abbrev. In press.
13. Elissa K. Title of paper if known. Unpublished.