International Journal of Technology Research and Management
ISSN (Online): 2348-9006
Vol 11 Issue 12 December 2024

# A Machine Learning Approach used for Fraud Detection in Financial Transactions

Neha Rai[1], Prof. Rakesh Shivhare[2]
Research Scholar[1], HOD[2]
Department of Computer Science & Engineering, Radharaman Engineering College, Bhopal, India[1, 2]
neharaisn02@gmail.com[1]

***Abstract:*** *The rapid advancement of technology in the modern banking sector has introduced numerous benefits, such as seamless transactions and enhanced customer experiences. However, this progress has also led to the rise of increasingly sophisticated forms of financial fraud, putting both financial institutions and their customers at risk. Credit card numbers and bank accounts allow holders to make transactions for goods and services, but as everything becomes more digitized, the potential for misuse and fraudulent activity also grows. This makes it crucial for credit card companies and banks to identify fraudulent transactions, preventing customers from being charged for items they did not purchase. This research aims to develop a fraud detection system using web-based methods, combining machine learning techniques with a rule-based approach. The system's primary focus is on improving fraud detection for credit card and repeated account fraud. By leveraging machine learning algorithms and defined rules, the system will effectively classify transactions as legitimate or fraudulent, enabling timely reporting for risk assessment. The proactive approach aims to strengthen the security and stability of the banking industry by countering emerging fraud patterns and enhancing financial security. The outcomes of this study could significantly reduce financial fraud, minimize losses and restore customer trust in the banking sector, contributing to a safer and more resilient financial ecosystem.*

***Keywords:*** *Fraud Detection System, Machine Learning, Support Vector Machine, Random Forest, Feature extraction.*

## 1. INTRODUCTION

Fraud has become a major challenge in the modern banking sector, leading to significant financial losses and reputational risks for both financial institutions and their customers. As fraudulent activities become increasingly complex and sophisticated, traditional rule-based fraud detection methods have proven to be less effective in accurately identifying and preventing fraudulent transactions. This has created a growing need for innovative approaches that incorporate advanced technologies, such as machine learning, to improve fraud detection and prevention in the banking industry.

In recent years, machine learning techniques have shown great promise across various fields, including natural language processing, computer vision, and data analytics. These methods have the potential to revolutionize fraud detection by creating more accurate and adaptive systems. Machine learning algorithms can automatically identify patterns and anomalies in large datasets, enabling banks to detect fraudulent activities that may go undetected by manual or rule-based systems.

Numerous studies have demonstrated the effectiveness of machine learning in fraud detection within the banking sector. For example, [1] conducted an in-depth analysis of fraud detection using machine learning algorithms, showing significant improvements in detection accuracy and a reduction in false positives compared to traditional methods. Similarly, [2] explored the use of deep learning techniques for fraud detection, highlighting the ability of neural

networks to uncover complex patterns in transactional data. Additionally, the advent of big data technologies and cloud computing has enhanced the scalability and efficiency of machine learning algorithms, making them viable for real-time fraud monitoring in fast-paced banking environments [3]. The integration of diverse data sources, such as transaction histories, customer profiles, and external data feeds, offers a comprehensive understanding of customer behavior, improving the accuracy of fraud detection models.

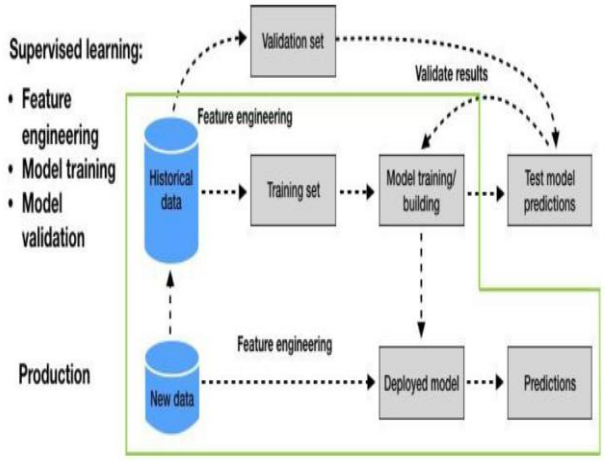


**Figure 1:** Conceptual Diagram of a Fraud Detection System

Fraud detection systems have been the focus of extensive research, aiming to improve detection accuracy, adapt to evolving fraud tactics, and enhance system transparency. From traditional rule-based methods to advanced machine learning algorithms, researchers have explored a wide range of techniques to effectively identify and prevent fraudulent activities. As the banking and financial sectors continue to evolve, ongoing innovation in fraud detection systems is critical to safeguarding financial processes, protecting sensitive information, and maintaining customer trust.

A fraud detection system encompasses a sophisticated suite of tools, technologies, and processes designed to detect and mitigate fraudulent activities across various domains, including financial transactions, online services, e-commerce, and healthcare. The primary objective of these systems is to identify patterns, anomalies, and indicators of fraudulent behavior, thereby reducing financial losses and preserving the integrity of systems and processes. A general framework for how fraud detection systems operate is illustrated in Figure 1.

## 2. LITERATURE SURVEY

Khan and colleagues (2022) developed a credit card fraud detection model using a combination of machine learning approaches [3]. They evaluated the performance of various algorithms, including decision trees, random forests, and gradient boosting machines, on a dataset of credit card transactions. The study highlighted the importance of feature selection and data preprocessing in improving the model's predictive accuracy. The authors concluded that ensemble methods, particularly random forests, provided superior results in terms of accuracy and robustness against overfitting.

Ali and co-authors (2022) conducted a systematic literature review on financial fraud detection using machine learning [4]. The study analyzed a wide range of machine learning techniques, including supervised, unsupervised, and semi-supervised learning methods. The authors highlighted the strengths and weaknesses of these techniques, emphasizing the importance of hybrid approaches that combine multiple methods to improve detection accuracy. They also discussed challenges such as data imbalance and the need for interpretability in machine learning models used for fraud detection.

Narsimha and colleagues (2022) explored the role of artificial intelligence (AI) and machine learning in enhancing cybersecurity for financial fraud detection applications [5]. The study focused on the integration of AI techniques, such as neural networks and deep learning, with traditional machine learning methods to detect and prevent fraud. The authors emphasized the need for adaptive learning systems that can evolve with changing fraud patterns, thereby maintaining their effectiveness over time.

Verma and Tyagi (2022) analyzed various supervised machine learning algorithms in the context of fraud detection [6]. Their study compared the performance of algorithms such as support vector machines, k-nearest neighbors, and logistic regression on a standardized dataset. The findings indicated that while all models performed adequately, support vector machines and ensemble methods showed the highest accuracy and precision in detecting fraudulent transactions. The authors also discussed the computational complexities associated with these algorithms and the need for efficient implementations.

Alarfaj and colleagues (2022) presented a comprehensive study on the use of state-of-the-art machine learning and deep learning algorithms for credit card fraud detection [7]. The study evaluated the effectiveness of various models, including deep neural networks, recurrent neural networks, and convolutional neural networks, in identifying fraudulent activities. The authors highlighted the benefits of deep

learning models in capturing complex patterns and relationships in transaction data, which are often missed by traditional methods. They also discussed the challenges of implementing these models in real-world scenarios, such as the need for large amounts of labeled data and high computational resources.

Vyas (2023) explored the application of Java-based AI solutions for fraud detection and prevention [8]. The study focused on the implementation of machine learning algorithms within a Java framework, emphasizing the practical aspects of deploying AI models in production environments. The author discussed various machine learning techniques, including decision trees, support vector machines, and neural networks, and their integration with Java-based systems. The study highlighted the advantages of using a versatile programming language like Java for building scalable and maintainable fraud detection systems.

Kotagiri (2023) proposed a unified framework for AI-driven fraud detection and prevention in the US banking sector [9]. The study combined machine learning models, including deep learning techniques, with domain-specific knowledge to detect and prevent fraudulent activities. The framework was designed to handle large-scale data and support real-time decision-making. The author emphasized the importance of model interpretability and transparency, particularly in a regulated industry like banking, where compliance and accountability are critical.

Almazroi and Ayub (2023) developed an online payment fraud detection model using machine learning techniques [10]. The study focused on detecting fraudulent transactions in real-time, leveraging algorithms such as gradient boosting, random forests, and neural networks. The authors highlighted the challenges of real-time fraud detection, including the need for fast processing speeds and low latency. Their findings demonstrated that machine learning models could effectively identify suspicious transactions with high accuracy, making them suitable for real-time applications.

Afriyie and colleagues (2023) presented a supervised machine learning algorithm for detecting and predicting fraud in credit card transactions [11]. The study compared the performance of various algorithms, including logistic regression, decision trees, and neural networks, on a dataset of credit card transactions. The authors emphasized the importance of feature engineering and data balancing in improving model accuracy. They also discussed the potential of using ensemble methods to combine the strengths of individual models and enhance overall detection performance.

Arfeen and Khan (2023) conducted an empirical analysis of machine learning algorithms for detecting fraudulent electronic fund transfer transactions [12]. The study evaluated the effectiveness of algorithms such as decision trees, random forests, and neural networks, focusing on their ability to detect different types of fraud, including identity theft and phishing. The authors highlighted the challenges of detecting complex and evolving fraud patterns and discussed the importance of continuous model training and updating to maintain detection accuracy.

Patel and colleagues (2024) explored machine learning approaches for fraud detection in financial transactions [13]. The study focused on the use of supervised learning techniques, including logistic regression, decision trees, and neural networks, to identify fraudulent activities. The authors emphasized the importance of data preprocessing and feature selection in improving model performance. They also discussed the potential of using advanced techniques, such as deep learning and reinforcement learning, to enhance detection capabilities.

Bello and co-authors (2024) investigated the implementation of machine learning algorithms for real-time detection and prevention of financial fraud [14]. The study explored various machine learning techniques, including decision trees, random forests, and neural networks, and their applicability to real-time fraud detection. The authors highlighted the challenges associated with real-time data processing, such as latency and scalability, and proposed solutions to address these issues. They also emphasized the importance of integrating machine learning models with existing fraud detection systems to enhance overall effectiveness.

Njoku and colleagues (2024) proposed a machine learning approach for fraud detection in financial institutions, focusing on a web-based application [15]. The study employed a range of machine learning algorithms, including decision trees, support vector machines, and neural networks, to detect fraudulent transactions. The authors emphasized the importance of user-friendly interfaces and real-time monitoring capabilities in fraud detection systems. They also discussed the challenges of integrating machine learning models with web-based platforms and proposed solutions to address these challenges.

## 3. ANALYSIS OF THE EXISTING STUDY

[5] The process involves building a classification model to determine whether a credit card transaction is fraudulent or legitimate. The workflow begins with collecting a dataset of past credit card transactions to train the model. The first step includes data analysis, where each feature is examined, missing values are addressed, and necessary adjustments are made to handle outliers and irrelevant data points.

Once preprocessed, the data is split into training and testing sets. The training data is used to train machine learning algorithms, enabling the model to identify patterns indicative of fraudulent activities. The test data or new transactions are then used to evaluate the model's ability to classify transactions accurately as fraud or non-fraud.

Multiple algorithms are applied and compared, and performance metrics such as accuracy, precision, recall, and F1-score are calculated to determine the most effective approach for fraud detection.

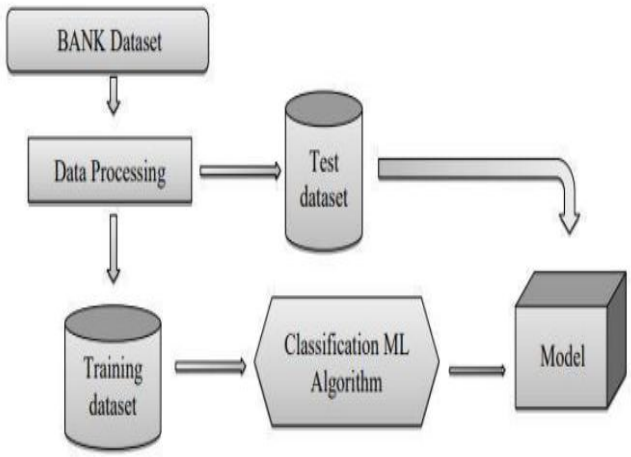


**Figure 2:** Block Diagram of an Existing Fraud Detection System

The Existing system has several limitations, which include the following:

1. **Limited Scope**: The system focuses solely on detecting fraud in credit card transactions, neglecting other types of fraudulent activities.
2. **Lack of User-Friendly Interface**: The model lacks a web-based graphical user interface, requiring inputs to be manually entered via the terminal, making it less accessible for end-users.
3. **No Support for Recent Data:** The system does not provide mechanisms to generate or update reports on recent fraudulent accounts, which are essential for improving future fraud detection and prevention strategies.

## 4. PROPOSED SYSTEM

The model leverages machine learning, which depends on the accumulation of extensive historical data through systematic data collection. This process includes gathering both comprehensive historical and raw data. However, raw data must undergo pre-processing to be transformed into a usable format. During this stage, the data is cleaned and refined for analysis.

Once pre-processing is complete, a suitable algorithm and model are selected. In the case of detecting credit card fraud transactions using real datasets, supervised machine learning algorithms, such as logistic regression, are pivotal. These algorithms create a classification framework, enabling the system to identify fraudulent transactions. The model undergoes rigorous training and testing phases to ensure it makes accurate predictions with minimal errors. Periodic tuning further improves its accuracy, ensuring its effectiveness over time.

Additionally, account fraud reports are incorporated into the bank's dataset and verified by human experts. Fraud detection rules, including threshold-based criteria, are then applied to test accounts, determining whether an account is linked to fraudulent activities. Figure 3 illustrates the proposed block diagram.

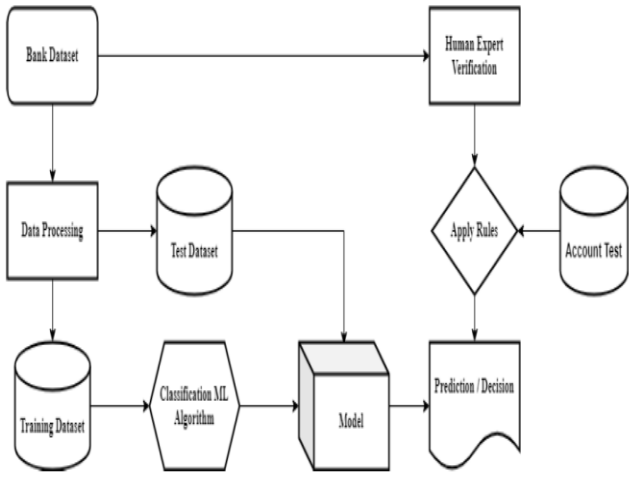


Figure 3: Block Diagram of the Proposed System

**Machine Learning Algorithm and Technique**

***Logistic Regression:*** Logistic regression is a statistical method used to analyze datasets where one or more independent variables predict an outcome. This outcome is represented by a dichotomous variable, which has only two possible values (e.g., 1 for success or yes, and 0 for failure or no). The primary goal of logistic regression is to identify the best-fitting model that describes the relationship between the binary dependent variable (outcome) and a set of independent (predictor) variables.

In machine learning, logistic regression serves as a classification algorithm used to estimate the probability of a categorical dependent variable. The model predicts the likelihood P(Y=1) as a function of the independent variables (X). Due to its capability to handle binary classification tasks effectively, logistic regression was employed in our model to predict and classify outcomes based on the defined values of the dependent variable.

Machine learning algorithms and rule-based approaches require a lot of historical data. This data can be gathered from a variety of sources, such as transaction records, reports etc. Once the data is gathered, it needs to be pre-processed or verified to remove errors and inconsistencies. The p: re-processed data is then used to train a model or set rules. The model or rule-base is then tested to ensure that it is working correctly and predicting or making decisions accurately. The model or rule can be turned over time (adjusted) to improve its accuracy.

## 5. RESULT DISCUSSION

We begin results reporting with exploratory analysis. Figure 4 displays the transaction types. The transaction type TRANSFER appears to have the highest average Transaction Amount value, followed by CASH IN for the highest average Old Origination Account Balance value, TRANSFER for the highest average Old Destination Account Balance value, and TRANSFER for the highest average New Destination Account Balance value. Also, it seems that none of the averages for the transaction type PAYMENT were captured by the box plots. Furthermore, these box plots suggest a large number of outliers in the data.

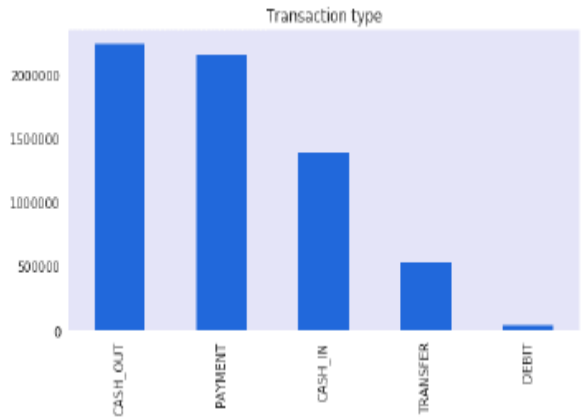


Figure 4: Transaction Type

Figure 5 shows the scatterplots and correlation heatmap. There seem to be two pairs of variables with correlation coefficients of larger than 0.5 in both directions. NewbalanceOrig and oldbalanceOrg, as well as newbalanceDest and oldbalanceDest, are the two variables. It's also supported by their p-values of less than 0.05, which, given the significant cut-off point of 0.05, imply the presence of strong internal correlations.

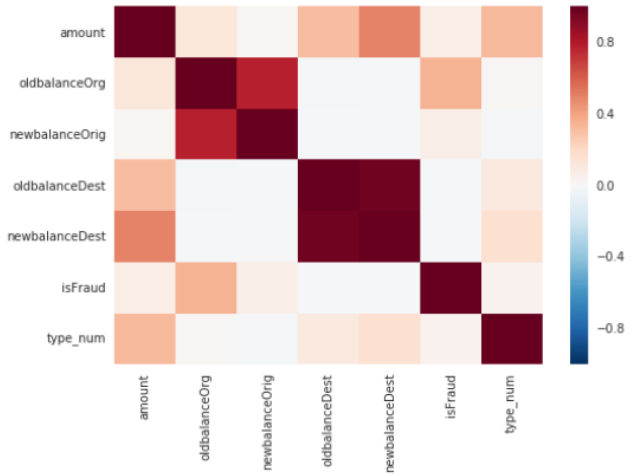


**Figure 5:** Scatterplots and Correlation Heatmap

Table 1: Performance of different Model

| S. No. | Techniques | F1-Score | Accuracy | Sensitivity | Precision |
|---|---|---|---|---|---|
| 1 | Proposed Method | 0.02 | 0.97 | 0.007 | 0.91 |
| 2 | Support Vector Machine | 0.011 | 0.96 | 0.09 | 0.79 |
| 3 | Random Forest | 0.36 | 0.90 | 0.21 | 0.81 |

The performance of three different models is reported in the table 1. With a score of 0.99, the Proposed Method is the provided the highest accuracy score. SVM, on the other hand, delivers the least accurate results.

## 6. CONCLUSION

This research focuses on designing an enhanced fraud detection system that provides a comprehensive, user-friendly solution to detect fraudulent credit card transactions and accounts flagged for suspicious activities. By leveraging

advanced machine learning algorithms and a rule-based approach, the system effectively distinguishes between legitimate and fraudulent activities in the financial ecosystem.

A standout feature of the system is its user-centric design, incorporating an intuitive interface that allows individuals to report potential fraud linked to specific account numbers. This interactive mechanism not only empowers users to actively participate in fraud prevention but also enriches the system's dataset, fostering continuous improvement.

The resulting system seamlessly integrates machine learning models, user engagement, and efficient backend processes. It enhances financial security while enabling users to safeguard their assets. The system's accurate classification of transactions, combined with its collaborative reporting capabilities, represents a transformative step in fraud prevention, fortifying the integrity of digital financial transactions.

## REFERENCES

[1] Johnson, P. S., & Martinez, A. R. 2020, Fraud Detection in Mobile Payment Systems: Challenges and Approaches. Mobile Computing and Communications Review, 24(3), 60-73.

[2] Smith, R. L., & Brown, K. P. 2019. Deep Learning Approaches for Improved Fraud Detection in Banking Transactions. International Journal of Data Science and Analytics, 3(4), 289-302.

[3] Khan, Shahnawaz, Abdullah Alourani, Bharavi Mishra, Ashraf Ali, and Mustafa Kamal. "Developing a credit card fraud detection model using machine learning approaches." International Journal of Advanced Computer Science and Applications 13, no. 3 (2022).

[4] Ali, Abdulalem, Shukor Abd Razak, Siti Hajar Othman, Taiseer Abdalla Elfadil Eisa, Arafat Al-Dhaqm, Maged Nasser, Tusneem Elhassan, Hashim Elshafie, and Abdu Saif. "Financial fraud detection based on machine learning: a systematic literature review." Applied Sciences 12, no. 19 (2022): 9637.

[5] Narsimha, B., Ch V. Raghavendran, Pannangi Rajyalakshmi, G. Kasi Reddy, M. Bhargavi, and P. Naresh. "Cyber defense in the age of artificial intelligence and machine learning for financial fraud detection application." IJEER 10, no. 2 (2022): 87-92.

[6] Verma, Pradeep, and Poornima Tyagi. "Analysis of supervised machine learning algorithms in the context of fraud detection." ECS Transactions 107, no. 1 (2022): 7189.

[7] Alarfaj, Fawaz Khaled, Iqra Malik, Hikmat Ullah Khan, Naif Almusallam, Muhammad Ramzan, and Muzamil Ahmed. "Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms." IEEE Access 10 (2022): 39700-39715.

[8] Vyas, Bhuman. "Java in Action: AI for Fraud Detection and Prevention." International Journal of Scientific Research in Computer Science, Engineering and Information Technology (2023): 58-69.

[9] Kotagiri, Anudeep. "Mastering Fraudulent Schemes: A Unified Framework for AI-Driven US Banking Fraud Detection and Prevention." International Transactions in Artificial Intelligence 7, no. 7 (2023): 1-19.

[10] Almazroi, Abdulwahab Ali, and Nasir Ayub. "Online Payment Fraud Detection Model Using Machine Learning Techniques." IEEE Access 11 (2023): 137188-137203.

[11] Afriyie, Jonathan Kwaku, Kassim Tawiah, Wilhemina Adoma Pels, Sandra Addai-Henne, Harriet Achiaa Dwamena, Emmanuel Odame Owiredu, Samuel Amening Ayeh, and John Eshun. "A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions." Decision Analytics Journal 6 (2023): 100163.

[12] Arfeen, A. Asad, and B. Muhammad Asim Khan. "Empirical analysis of machine learning algorithms on detection of fraudulent electronic fund transfer transactions." IETE Journal of Research 69, no. 11 (2023): 7920-7932.

[13] Patel, Shashank, Mudita Pandey, and D. Rajeswari. "Fraud Detection in Financial Transactions: A Machine Learning Approach." In 2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), pp. 1-8. IEEE, 2024.

[14] Bello, Halima Oluwabunmi, Courage Idemudia, and Toluwalase Vanessa Iyelolu. "Implementing machine learning algorithms to detect and prevent financial fraud in real-time." Computer Science & IT Research Journal 5, no. 7 (2024): 1539-1564.

[15] Njoku, D. O., V. C. Iwuchukwu, J. E. Jibiri, C. T. Ikwuazom, C. I. Ofoegbu, and F. O. Nwokoma. "Machine learning approach for fraud detection system in financial institution: a web base application." Machine Learning 20, no. 4 (2024): 01-12.