IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 6, May 2025
Impact Factor: 7.67

Fraud Detection on Bank Payments Using Machine Learning

Suresh[1], **Alla Bala Bhanu**[2], **Prathibha S**[3], **Preethi J P**[4], **Rajeshwari Bhandari**[5]
Assistant Professor, Department of Computer Science and Engineering[1]
Students, Department of Computer Science and Engineering[2,3,4,5]
Rao Bahadur Y Mahabaleshwarappa Engineering College, Bellary, Karnataka, India
Corresponding author:sureshkaradi1699@gmail.com

Abstract: *Financial fraud poses a major threat to organizations, with traditional detection methods often proving costly, slow, and inaccurate. The rise in digital transactions highlights the need for intelligent, automated solutions. This study presents a fraud detection model using the Random Forest Classifier to identify fraudulent bank transactions. Experiments conducted on the Banksim dataset show that the model achieves 99% accuracy in both training and testing phases, outperforming conventional approaches. These results underscore the potential of machine learning to enhance the efficiency and effectiveness of fraud detection in the banking sector.*

Keywords: Financial fraud Detection, Machine Learning, Random Forest Classifier, Banksim Dataset, Automated Fraud Detection, Banking Security, Fraudulent transactions, Artificial Intelligence, Classification Algorithms, Data Mining

I. INTRODUCTION

Fraud detection in financial transactions has become essential for banking institutions, especially in the context of immediate payments. To develop effective solutions, it is crucial to review existing literature and assess current methodologies.

Two prominent approaches include: (1) analysis from a research perspective and (2) feasibility studies focused on fraud discovery. Most fraud detection techniques leverage artificial intelligence, such as pattern matching algorithms [1].

However, challenges such as inaccurate and overlapping data persist. Genuine transactions may exhibit patterns similar to fraudulent ones, leading to misclassifications, even with highly accurate models [2].

Effective fraud detection systems must balance the cost of fraud losses with the resources required for fraud prevention. These systems aim to prevent unauthorized payments or misuse of resources by identifying suspicious behavior. Fraud manifests in various forms across multiple domains, and modern detection methods often integrate diverse datasets to differentiate between legitimate and illegitimate transactions. Factors considered in this classification process include IP address, geolocation, device credentials, and transactional metadata [3].

In practical applications, merchants and financial institutions employ data-driven decision systems, using internal data along with sets of business rules or algorithmic logic to flag potential fraud. These systems can be fully automated by linking directly to user accounts, enabling analysts to respond quickly and efficiently.

II. LITERATURE SURVEY

Fraud detection has become increasingly crucial in the financial sector due to the rising volume of digital transactions and the sophisticated tactics employed by fraudsters.

Traditional methods are often inadequate to address these evolving threats, prompting the exploration of advanced techniques such as machine learning.

Various approaches, including supervised and unsupervised learning methods like decision trees, clustering, and anomaly detection, have been studied to enhance fraud detection systems.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 6, May 2025
Impact Factor: 7.67

Feature engineering, which involves extracting meaningful patterns from transaction data such as user behaviour, plays a vital role in building effective models.

Addressing the challenge of imbalanced datasets, techniques like resampling, synthetic data generation, and cost-sensitive learning have been employed.

Evaluation metrics such as precision, recall, F1-score, and ROC-AUC are used to measure the performance of fraud detection models. Implementing real-time fraud detection systems requires balancing accuracy and latency to ensure timely and accurate detection. Emerging trends in this field include the use of deep learning techniques and blockchain technology, which offer promising avenues for developing more secure and efficient fraud detection solutions.

2.1 Purpose/justification of Project

The primary aim of this project is to develop a comprehensive fraud detection system using machine learning techniques, specifically tailored for bank payments. The objectives and justification of this project encompass theoretical, practical, and educational impacts on hardware, software, and users, as detailed below:

i. Theoretical Impact: Enhance the understanding of fraud detection mechanisms by investigating and comparing various machine learning algorithms in the context of bank payments. Contribute to the existing body of knowledge on the interplay between transaction attributes and fraudulent activities, providing insights into the effectiveness of different models and techniques.

ii. Practical Impact: Improve the security of financial transactions by implementing a real time fraud detection system that can identify and prevent fraudulent activities, thus reducing financial losses for banks and their customers. Develop a robust and adaptable model that can be integrated into existing banking systems, ensuring continuous improvement and the ability to adapt to evolving fraud patterns and tactics.

iii. Educational Impact: Offer a valuable case study for students and researchers interested in applying machine learning to cybersecurity and fraud detection, illustrating practical applications and challenges.Serve as an educational resource for implementing machine learning algorithms, feature engineering, and managing imbalanced datasets in real-world scenarios.

iv. Impact on Hardware: Leverage existing banking infrastructure and hardware to implement the fraud detection system, optimizing resource usage and ensuring efficient processing of transaction data.Explore opportunities for deploying specialized hardware, such as GPUs, to accelerate the training and deployment of machine learning models.

v. Impact on Software: Develop a user-friendly software tool capable of processing large volumes of transaction data and detecting fraudulent activities with high accuracy.Ensure seamless integration of the fraud detection system with existing banking platforms, providing real-time alerts and detailed reports to facilitate prompt action against fraud.

vi. Impact on Users: Enhance user trust and confidence in the security of their financial transactions by demonstrating the effectiveness of the fraud detection system.Provide a user-friendly interface for bank employees to monitor and respond to potential fraud cases swiftly and efficiently, thereby improving the overall customer experience.

III. METHODOLOGY

A. SYSTEM DESIGN

To design a fraud detection system for bank payments using machine learning, start by collecting and preprocessing transaction data to ensure its quality.

Train machine learning models using supervised and unsupervised learning techniques, then evaluate their performance with appropriate metrics. Implement a real-time fraud detection engine, combining rules-based systems with machine learning insights. Generate alerts for flagged transactions and provide tools for investigation, while also taking automated actions when necessary.

Continuously monitor and maintain the models and system to ensure scalability and compliance with regulations, and develop user interfaces for fraud management and customer interaction. This comprehensive approach ensures effective fraud detection.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 6, May 2025
Impact Factor: 7.67

B. SYSTEM ARCHITECTURE:

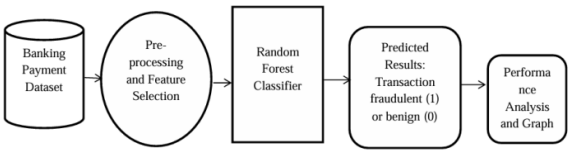
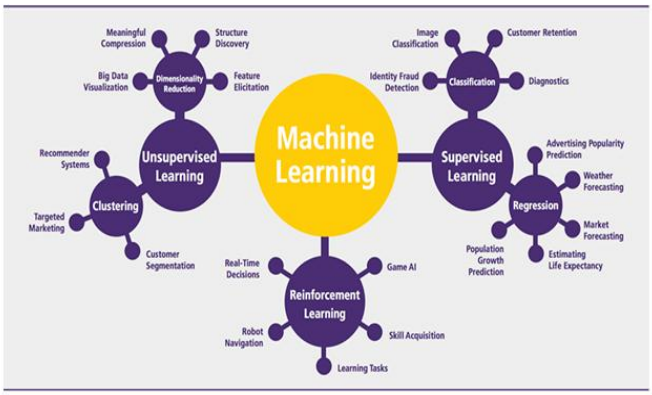


Fig : Architecture of Fraud Detection on Bank Payments

1. Banking Payment Dataset: The process begins with a banking payment dataset, represented by a cylindrical database icon. This dataset contains the payment information, which will be analysed for potential fraudulent activity.
2. Pre-processing and Feature Selection: The next step involves pre-processing the data and selecting relevant features. This is represented by an oval shape in the diagram. Pre-processing might include tasks like data cleaning, normalization, and feature engineering to prepare the data for the machine learning model.
3. Random Forest Classifier: The core of the system is the Random Forest Classifier, shown as a rectangular box. This machine learning model is used to classify transactions as either fraudulent (1) or benign (0). Random Forest is an ensemble learning method that operates by constructing multiple decision trees and combining their outputs.
4. Predicted Results: After the classification, the results are displayed in another rectangular box labelled "Predicted Results". This shows the outcome of the classification, indicating whether each transaction is fraudulent or not.
5. Performance Analysis and Graph: Finally, the performance of the classifier is analysed and visualized, as represented by a smaller rectangular box. This step involves evaluating the accuracy, precision, recall, and other performance metrics of the model, and visualizing these results in graphs.

C. ALGORITHM SPECIFICATION



Machine learning Algorithms

Machine learning can be grouped into two broad learning tasks: Supervised and Unsupervised. There are many other algorithms Supervised learning An algorithm uses training data and feedback from humans to learn the relationship of given inputs to a given output.

For instance, a practitioner can use marketing expense and weather forecast as input data to predict the sales of cans. You can use supervised learning when the output data is known. The algorithm will predict new data.

We used the Random Forest Classifier machine learning algorithm.

We got an accuracy of 99.7% on the training set so we implemented this algorithm.

Hardware requirements:

➢ System : Pentium i3 Processor.

➢ Hard Disk : 500 GB.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 6, May 2025
Impact Factor: 7.67

➢ Monitor : 15'' LED
➢ Input Devices : Keyboard, Mouse
➢ Ram : 4 GB

Software Requirements:
➢ Operating system : Windows 10.
➢ Coding Language : Python.
➢ Web Framework : Flask

D. ALGORITHM PROCEDURE

It works in four steps:
Select random samples from a given dataset.
Construct a decision tree for each sample and get a prediction result from each decision tree. Perform a vote for each predicted result. Select the prediction result with the most votes as the final prediction.Random forests also offer a good feature selection indicator. Scikit-learn provides an extra variable with the model, which shows the relative importance or contribution of each feature in the prediction. It automatically computes the relevance score of each feature in the training phase.Then it scales the relevance down so that the sum of all the scores is 1.
This score will help you choose the most important features and drop the least important ones for model building.

IV. CONCLUSION

Financial fraud can occur in a variety of financial contexts, including the corporate, banking, insurance, and taxes sectors. Financial fraud has recently raised concerns among businesses and industries.
Financial fraud continues to exist despite several attempts to eradicate it, which has a negative impact on society and the economy because daily losses from fraud amount to very huge sums of money.
Machine-learning-based technologies can now be used intelligently to identify fraudulent transactions by examining a significant amount of financial data, thanks to the development of artificial intelligence. In this article, we published a study that thoroughly analyzed and summarized the body of knowledge on ML-based fraud detection. This study uses the Random Forest Classifier methodology, which extracts, synthesis, and reports results using well defined methods.

V. ACKNOWLEDGMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of the people who made it possible, whose constant guidance and encouragement crowned our effort with Success.
We express our sincere gratitude to our Principal Dr. T. Hanumanth Reddy for giving us an opportunity to carry out our academic project.
We wish to place on record our gratitude thanks to Dr. H. Girisha, Head of the Department, Computer Science and Engineering RYMEC, Ballari for providing encouragement and guidance.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 6, May 2025
Impact Factor: 7.67

REFERENCES

[1] S. Delecourt and L. Guo, "Building a robust mobile payment fraud detection system with adversarial examples," in 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), pp. 103–106, IEEE, 2019.

[2] T. Alquthami, A. M. Alsubaie, and M. Anwer, "Importance of smart meters data processing – case of saudi arabia," in 2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA), pp. 1–5, addIEEE, 2019.

[3] O. Adepoju, J. Wosowei, S. lawte, and H. Jaiman, "Comparative evaluation of credit card fraud detection using machine learning techniques," in 2019 Global Conference for Advancement in Technology (GCAT), pp. 1–6, IEEE, 2019.

[4] S. Khatri, A. Arora, and A. P. Agrawal, "Supervised machine learning algorithms for credit card fraud detection: A comparison," in 2020 10th International Conference on Cloud Computing, Data Science Engineering (Confluence), pp. 680–683, IEEE, 2020.

[5] V. Jain, M. Agrawal, and A. Kumar, "Performance analysis of machine learning algorithms in credit cards fraud detection," in 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 86–88, IEEE, 2020.

[6] A. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga, and N. Kuruwitaarachchi, "Real-time credit card fraud detection using machine.