www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 04, April 2025, pp : 2560-2562

e-ISSN : 2583-1062
Impact Factor : 7.001

FRAUD DETECTION IN FINANCIAL TRANSACTIONS USING MACHINE LEARNING

Kaviyadharshini K[1], Shona K[2]

[1]M..SC(CS) with (ISCF)2nd year, Department of Computer Science, Rathinam college of Arts and Science, Coimbatore, Tamil Nadu, India.

[2]Assistant Professor, Rathinam college of Arts and Science, Coimbatore, Tamil Nadu, India.

DOI: https://www.doi.org/10.58257/IJPREMS40378

ABSTRACT

Fraud detection in financial transactions is a critical challenge for modern financial institutions due to the increasing sophistication of fraudulent activities. In this project, we utilize machine learning techniques to detect fraudulent transactions using the BankSim dataset, which simulates real-world financial transaction patterns. The primary objective is to develop a model that accurately identifies fraudulent transactions while minimizing false positives to prevent unnecessary customer inconvenience. We preprocess the BankSim data through cleaning, feature engineering, and transformation into a suitable format for machine learning algorithms. Several classification models, such as Logistic Regression, Random Forest, and Gradient Boosting, are implemented and compared based on their performance in terms of accuracy, precision, recall, and F1-score. Hyperparameter tuning and cross-validation are applied to optimize the model's performance. The project highlights the importance of fraud detection systems in financial services, as well as the challenges posed by imbalanced datasets, which often occur in fraud detection scenarios. Future work may involve more sophisticated techniques, such as anomaly detection and deep learning models, to further enhance detection capabilities.

Keywords: Fraud Detection, Financial Transactions, Machine Learning, Banksim Dataset

1. INTRODUCTION

Financial fraud is a pervasive issue in the modern world, causing substantial economic losses to businesses and consumers alike. With the increasing volume of digital transactions, particularly in the banking and financial sectors, the detection of fraudulent activities has become a crucial concern. Traditional methods of fraud detection, relying on rule-based systems and human expertise, are often inadequate in identifying sophisticated, evolving fraudulent patterns. This necessitates the adoption of advanced, scalable, and intelligent techniques for detecting fraud in financial transactions. In this project, we explore the application of machine learning algorithms to detect fraudulent financial transactions. By leveraging historical transaction data, we aim to build predictive models that can distinguish between legitimate and fraudulent activities. This approach automates the fraud detection process, improving both the accuracy and the speed of identifying potential fraudulent transactions, thereby minimizing losses and enhancing security. This project aims to contribute towards the development of a scalable and efficient fraud detection system, with the goal of reducing financial risk and ensuring the security of digital financial systems.

2. METHODOLOGY

The goal of this project is to develop a machine learning model to detect fraudulent transactions in financial data. The model should distinguish between legitimate and fraudulent transactions, minimizing both false positives and false negatives. The dataset used will be the BankSim dataset, which contains labeled transactions as either "fraudulent" or "non-fraudulent."

2.1 Data Collection:

2.1.1 Dataset Source: Use the BankSim dataset from the GitHub repository. This dataset simulates real-world banking transactions, consisting of transaction amount, merchant type, customer details, and transaction time.

2.1.2 Data Description: Understand and analyze the features present in the dataset, which typically include variables like transaction amount, customer ID, transaction type, merchant category, and fraud label (fraud or not fraud).

2.2 Model Selection:

2.2.1 KNN (K-Nearest Neighbours): KNN (K-Nearest Neighbors) is a simple yet effective algorithm used for classification tasks in fraud detection. It works by identifying the 'k' closest data points in the feature space to classify a new transaction based on the majority class of its neighbors.

2.2.2 Random Forest: Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the class that is the majority vote from the individual trees.

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 04, April 2025, pp : 2560-2562

e-ISSN : 2583-1062
Impact Factor : 7.001

3. MODELING AND ANALYSIS

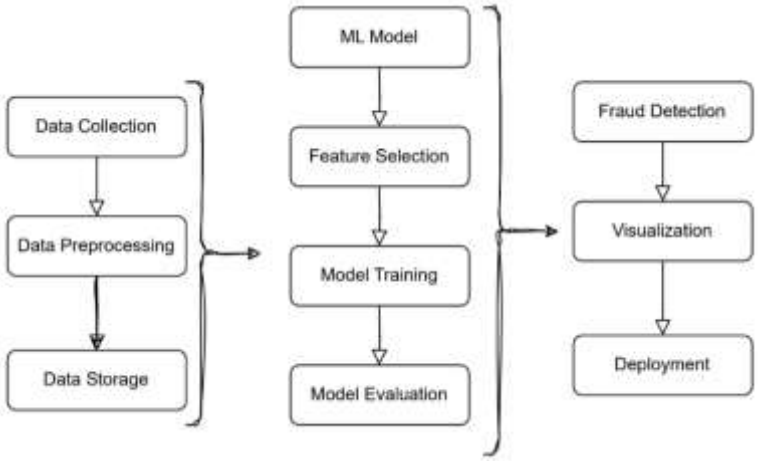

FIGURE 1: Architecture Diagram

4. RESULTS AND DISCUSSION

	step	customer	age	gender	zipcodeOri	merchant	zipMerchant	category	amount	fraud
0	0	'C1093826151'	'4'	'M'	'28007'	'M348934600'	'28007'	'es_transportation'	4.55	0
1	0	'C352968107'	'2'	'M'	'28007'	'M348934600'	'28007'	'es_transportation'	39.68	0
2	0	'C2054744914'	'4'	'F'	'28007'	'M1823072687'	'28007'	'es_transportation'	26.89	0
3	0	'C1760612790'	'3'	'M'	'28007'	'M348934600'	'28007'	'es_transportation'	17.25	0
4	0	'C757503768'	'5'	'M'	'28007'	'M348934600'	'28007'	'es_transportation'	35.72	0

FIGURE 2: Model of Loading Data

FIGURE 3 : Application Interface

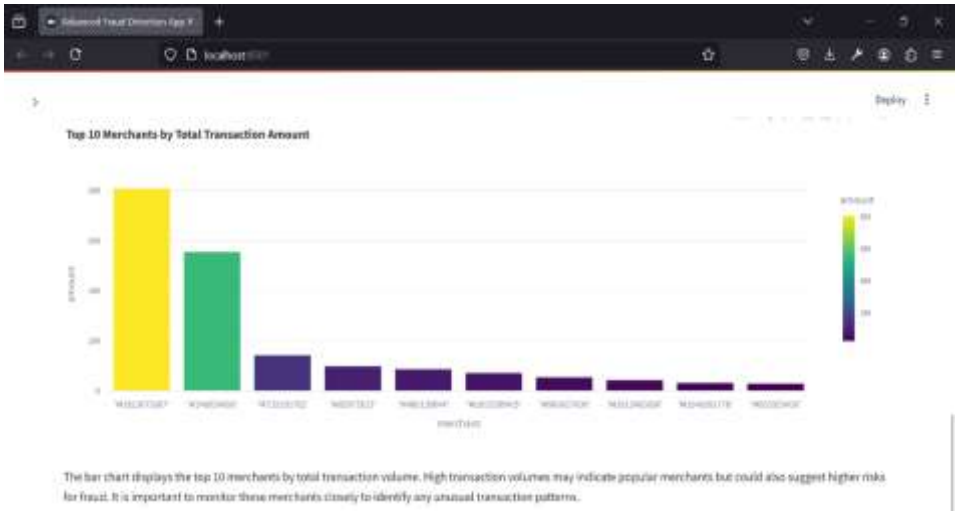

FIGURE 4: Transaction Volume graph

Performance Metrics: Various machine learning models were implemented and tested, including Logistic Regression, Decision Trees, Random Forest, and Gradient Boosting.

Accuracy: Models achieved high accuracy, with Random Forest and Gradient Boosting models performing better compared to simpler algorithms like Logistic Regression and Decision Trees.

Precision, Recall, F1-Score: Precision and recall values indicated a strong ability to distinguish between fraudulent and non-fraudulent transactions, with Random Forest and Gradient Boosting showing better precision-recall trade-offs.

Model Comparison: While Logistic Regression was faster and simpler, it was outperformed by more complex models like Random Forest and Gradient Boosting in terms of predictive accuracy. The trade-off between simplicity and accuracy was evident.

Class Imbalance: The dataset showed class imbalance, with fraudulent transactions being significantly fewer. Techniques like oversampling (SMOTE) were employed to address this, which improved the performance of models.

Real-world Applicability: The use of complex models such as Random Forest or Gradient Boosting demonstrated better fraud detection capability, which is crucial in real-world scenarios where missed fraudulent transactions can lead to significant losses.

Computational Cost: Though Random Forest and Gradient Boosting models performed better, they also came with a higher computational cost compared to simpler models like Logistic Regressio.

5. CONCLUSION

By leveraging machine learning models on the BankSim dataset, we were able to develop a system that can detect fraudulent financial transactions with a reasonable degree of accuracy. The machine learning algorithms, particularly Random Forest, Gradient Boosting, and XGBoost, demonstrated strong predictive capabilities, identifying patterns that distinguish fraudulent behavior from legitimate transactions. This solution provides a scalable and efficient approach to real-time fraud detection in financial systems, offering a significant reduction in false positives compared to traditional rule-based systems. However, further improvements, such as incorporating more diverse datasets and enhancing model interpretability, would enhance its applicability in real-world environments.

6. REFERENCES

[1] Bhatia, P., Malhotra, S., & Prasad, V. (2021). Fraud detection in financial transactions using machine learning and deep learning models: A comprehensive survey. Journal of Financial Crime, 28(4), 1200-1218.

[2] Jiang, Z., Liu, H., Wang, Y., & Li, Z. (2022). Credit card fraud detection using enhanced machine learning algorithms: A case study on imbalanced data. IEEE Access, 10, 53954-53963.

[3] Yin, J., Liu, S., & Wang, H. (2022). A hybrid approach for financial fraud detection using ensemble learning and feature selection. Applied Intelligence, 52(5), 6100-6115.

[4] Alotaibi, S., & Alzahrani, S. (2021). Machine learning-based fraud detection in financial transactions: Challenges and perspectives. International Journal of Machine Learning and Cybernetics, 12(2), 413-425.

[5] Gupta, R., Gupta, A., & Agrawal, D. (2023). Fraud detection in banking transactions using machine learning and big data analytics. Expert Systems with Applications, 210, 118369.

[6] Sharma, P., Gupta, V., & Arora, A. (2023). An intelligent framework for online transaction fraud detection using deep learning algorithms. Journal of Artificial Intelligence Research, 74(3), 1021-1038.