Vijay Bhaskar, 2025, 13:6
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal
www.ijset.in

UPI Fraud Detection Using Machine Learning

Vijay Bhaskar Reddy, Abhishek Kumar, Hrithik, Manjunath Shetty, guide - Mrs. Sukanya
Department of CSE Navodaya Institute Of Technology Raichur, India

Abstract - The rapid increase in digital transactions, particularly through the Unified Payments Interface (UPI), has been accompanied by an alarming rise in fraudulent activities targeting unsuspecting users. This project aims to address this gap by developing an effective fraud detection system using machine learning. The proposed system utilizes four popular machine learning algorithms: Random Forest, Logistic Regression, Decision Tree, and Support Vector Machine (SVM) to classify transactions as either legitimate or fraudulent. By analyzing historical transaction data, the system detects anomalies or suspicious behavior indicative of fraud and flags such transactions in real-time. Experimental results indicate that machine learning offers the advantage of continuously improving detection accuracy as the system learns from new transaction data.

Keywords - UPI, Fraud Detection, Machine Learning, Random Forest, SVM, Logistic Regression, Decision Tree.

I. INTRODUCTION

The introduction of the Unified Payments Interface (UPI) has revolutionized the way digital payments are made in India, providing a quick, secure, and seamless means of transferring funds between accounts. UPI has become the backbone of the digital payments ecosystem, facilitating transactions through smartphones and digital banking systems. However, the rise in UPI usage has led to a parallel increase in fraudulent activities, which have become a significant concern for both users and financial institutions.

UPI fraud manifests in various techniques, including phishing, unauthorized transactions, and social engineering attacks. Traditional methods of fraud detection often rely on predefined rules or heuristics, which may not be sufficient to identify emerging fraud patterns. Given the dynamic nature of fraudulent activities and the large volume of transactions processed daily, there is a pressing need for more sophisticated fraud detection systems that can adapt to new trends.

Machine learning (ML) has emerged as a powerful tool for addressing this challenge, as algorithms are capable of analyzing vast amounts of transaction data to identify patterns and detect anomalies indicating fraudulent behavior. Unlike traditional rule-based approaches, machine learning models can learn from historical data and adapt to changing fraud tactics, making them more effective at identifying fraud in real-time.

This paper proposes the use of machine learning techniques for UPI fraud detection, specifically exploring Random Forest, Logistic Regression, Decision Tree, and Support Vector Machine (SVM). The primary objective is to identify the most effective model for detecting UPI fraud with minimal false positives and high accuracy.

II. L ITERATURE REVIEW

Machine learning has proven to be a powerful tool for detecting fraud in various domains, including banking and e-commerce. Supervised learning algorithms are widely used because they learn from labeled data where transactions are tagged as fraudulent or legitimate.

Random Forest Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes predicted by individual trees. It is highly effective for classification tasks due to its ability to handle complex datasets with multiple features and is less prone to overfitting compared to individual decision trees. Studies have demonstrated its

effectiveness in achieving high accuracy and low false positives in online payment systems.

Decision Tree Decision Trees create a tree-like structure where each node represents a decision based on a feature. They are popular for their simplicity and interpretability, allowing for the identification of patterns such as sudden changes in transaction behavior. However, they are prone to overfitting, which can be mitigated by ensemble methods.

Logistic Regression Logistic Regression is a statistical model used for binary classification that predicts the probability of a transaction being fraudulent. It is computationally efficient and suitable for real-time fraud detection, particularly when the relationship between features and the target is linear.

Support Vector Machine (SVM) SVM finds the optimal hyperplane to separate different classes in the feature space. It is particularly effective in high-dimensional spaces and works well with non-linear decision boundaries, making it suitable for detecting sophisticated fraud patterns.

III. PROPOSED METHODOLOGY

The proposed system is designed to be scalable, real-time, and adaptive29. It consists of data collection, preprocessing, model development, and deployment phases.

Data Collection and Preprocessing

The dataset includes features such as Transaction Amount, Transaction Time, Merchant Information, User Location, Device Information, and Transaction History 30.

Data preprocessing involves handling missing values, encoding categorical variables, and normalizing numerical features.

- Handling Missing Data: Imputation or deletion strategies are used to handle incomplete records32.
- Feature Encoding: Categorical features like location and device ID are converted into numerical form using Label Encoding or One-Hot Encoding 33.
- Feature Scaling: Numerical values are normalized or standardized to ensure consistency, which is crucial for algorithms like SVM and Logistic Regression.

Machine Learning Algorithms

Four algorithms were implemented and trained on the preprocessed dataset:

- Random Forest: Utilizes bootstrapping and feature randomization to build multiple trees and aggregates results via majority voting 35.
- Logistic Regression: Models the probability of fraud using the sigmoid function.
- Decision Tree: Recursively splits data to minimize impurity (Gini or entropy).
- SVM: Uses a kernel trick (e.g., RBF) to handle non- linear separations and maximize the margin between classes38.

Deployment

The selected model is deployed using the Flask web framework39. The system provides a User Interface (UI) for transaction input and a backend that processes data and returns real-time fraud predictions40

Results and Discussion

The system was tested using a dataset split into 70% training and 30% testing sets. The models were evaluated based on Accuracy, Precision, Recall, and F1-Score.

Performance Analysis Table I summarizes the performance of the four algorithms. Random Forest achieved the highest performance across all metrics.

IV. CONCLUSION

In this project, we explored the use of machine learning techniques for detecting fraud in Unified Payments Interface (UPI) transactions. With the rapid growth of digital payments, UPI systems have become increasingly vulnerable to fraudulent activities, making effective and timely fraud detection essential.

By analyzing transaction patterns and applying machine learning algorithms, the proposed system is

able to distinguish between legitimate and fraudulent transactions with improved accuracy. Features such as transaction amount, frequency, time, and user behavior play a crucial role in identifying suspicious activities. The results demonstrate that machine learning models can significantly reduce false positives while maintaining high detection rates.

Overall, the study highlights that machine learning-based fraud detection systems are efficient, scalable, and adaptable to evolving fraud patterns. Implementing such systems can enhance the security of UPI platforms, protect users from financial losses, and build greater trust in digital payment ecosystems. Future improvements may include using real-time data, advanced deep learning models, and continuous model updates to further strengthen fraud detection performance.

Acknowledgments

I thank all my colleagues who provided a deep insight and expertise that greatly assisted in the document development.

I personally thank my Professor for the assistance with the initial drafts and the core structure of the proposal for the comments that greatly im- proved the manuscript.

I would also like to show my gratitude to the University for sharing its pearls of wisdom with me during the course of the development of this pro- posal. I am immensely grateful to everyone.

REFERENCES

1. Bhattacharyya, S., Jha, S., & Kumar, S. (2011). Credit card fraud detection using machine learning techniques. International Journal of Computer Applications, 34(3), 26-30.
2. Chawla, N. V., & Bowyer, K. W. (2002). SMOTE: Synthetic minority over- sampling technique. Journal of Artificial Intelligence Research, 16, 321- 357.
3. Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit risk models via machine-learning algorithms. Journal of Banking & Finance, 34(11), 2767-2787.
4. Liao, Y. L., & Lin, C. Y. (2016). A review of fraud detection techniques for electronic transactions. Journal of Computational Science, 19(3), 288-299.
5. Zaslavsky, A., & Hodge, V. J. (2018). Data mining for fraud detection: Tech- niques, challenges, and applications. Journal of Data Science, 8(1), 23-40.
6. Verma, P., & Bansal, J. (2018). Machine learning for fraud detection in financial transactions. In Proceedings of the International Conference on Computational Intelligence (pp. 165-174).
7. Zhang, H., & Zhou, X. (2019). Financial fraud detection using machine learning techniques. Springer Handbook of Computational Intelligence, 1397-1414.
8. Dhanalakshmi, R., & Srinivasan, S. (2014). Credit card fraud detection us- ing decision tree and SVM algorithms. International Journal of Computer Applications, 101(14), 23-28.
9. Jha, S., & Verma, S. (2017). Application of machine learning in the detec- tion of fraud in mobile transactions. International Journal of Computer Applications, 160(6), 36-41.
10. Su, Z., & Sun, H. (2017). Machine learning techniques for fraud detection in payment systems. Journal of Computer Science and Technology, 32(3), 472-484.
11. Dutta, S., & Das, S. (2015). Using machine learning for fraud detection in e-commerce transactions. Proceedings of the International Conference on Information and Communication Technology for Intelligent Systems, 61-65.
12. Chandwani, R., & Nisar, K. (2020). A comprehensive review on fraud de- tection techniques using machine learning. International Journal of Ad- vanced Research in Computer Science, 11(5), 30-35.
13. Soni, P., & Jindal, V. (2018). Machine learning approach for real-time fraud detection in digital banking. Journal of Computer Applications, 49(4), 15-22.
14. Li, Z., & Ma, Y. (2016). A novel fraud detection approach using SVM for credit card transaction. Proceedings of the International Conference on Network and Information Systems, 153-157.
15. Rani, S., & Kaushik, V. (2019). Data mining techniques for fraud detection in online

payments. International Journal of Computer Science & Network Security, 19(8), 65-72.
16. Rao, A., & Kulkarni, P. (2019). Fraud detection in mobile payments using machine learning. International Journal of Data Science, 18(2), 211-218.
17. Gupta, M., & Sharma, R. (2021). An efficient machine learning model for detecting fraudulent UPI transactions. Advances in Artificial Intelligence and Data Science, 11(3), 118-126.
18. Malik, M., & Sharma, R. (2017). Fraud detection using Random Forest algorithm in online financial systems. International Journal of Computer Science and Network Security, 17(10), 47-51.
19. Yadav, V., & Sharma, K. (2019). Application of machine learning algo- rithms for fraud detection in digital payments. International Journal of Emerging Technologies, 11(1), 34-42.
20. Kumar, R., & Saini, H. (2018). A comparative analysis of machine learning algorithms for fraud detection. International Journal of Artificial Intelli- gence and Applications, 9(3), 55-63.