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 5, December 2025
Impact Factor: 7.67

UPI Fraud Detection Using Machine Learning

Shrinidhi S Koundinya[1], Shubha K[2], Thushar S[3]
B.E, CSE, Kalpataru Institute of Technology, Tiptur, India [1,2,3]

Abstract: *The rapid adoption of digital payments in India has made the Unified Payments Interface (UPI) a dominant platform for fast and seamless transactions. However, this growth has also led to a significant rise in fraudulent activities, including phishing, social engineering, and unauthorized account access. To address these challenges, this paper proposes a machine learning–based UPI fraud detection system capable of identifying fraudulent transactions in real time. The system employs supervised learning algorithms trained on transactional data to detect suspicious patterns, anomalies, and abnormal user behavior. In addition, behavioral analysis and anomaly detection techniques are incorporated to capture evolving fraud strategies. The proposed architecture integrates Flask for backend services and React.js for frontend visualization, enabling real-time monitoring and intuitive dashboards. The model analyzes key features such as transaction frequency, geolocation inconsistencies, device usage, and merchant risk profiles to proactively flag fraudulent activities. Designed for scalability, the system can be deployed on cloud infrastructure to handle high transaction volumes with low latency. Overall, the proposed solution enhances digital payment security, improves transparency, and supports proactive fraud prevention, contributing to a secure and trustworthy UPI ecosystem.*

Keywords: UPI fraud detection, digital payments, machine learning, online transaction security, financial fraud prevention, real-time monitoring

I. INTRODUCTION

1.1 Problem Statement

The rapid adoption of the Unified Payments Interface (UPI) has transformed digital financial transactions in India by enabling instant, secure, and cashless payments. However, the increasing reliance on UPI has also led to a significant rise in fraudulent activities. Cybercriminals exploit system vulnerabilities and user unawareness to carry out frauds such as phishing attacks, fake payment requests, social engineering, SIM swap fraud, malware-based attacks, and unauthorized account access. Existing fraud detection systems used in digital payment platforms are largely rule-based and reactive in nature. These systems depend on predefined thresholds and static rules, which are effective only for known fraud patterns. As fraudsters continuously modify their techniques to bypass such rules, traditional approaches often fail to detect new and evolving fraud behaviors. Additionally, the massive volume and real-time nature of UPI transactions make manual monitoring impractical. Another major challenge in UPI fraud detection is the imbalance in transaction data, where fraudulent transactions constitute a very small portion compared to legitimate ones. This imbalance increases the risk of false negatives, leading to undetected fraud, and false positives, which inconvenience genuine users. Therefore, there is a critical need for an intelligent, adaptive, and real-time fraud detection system that can accurately identify suspicious UPI transactions and prevent financial losses before they occur.

1.2 Objective

The primary objective of this project is to develop a efficient UPI fraud detection system using machine learning techniques that enhances the security and integrity of digital payment transactions. The system aims to analyze large volumes of UPI transaction data to identify hidden patterns, anomalies, and behavioral deviations associated with fraudulent activities. By leveraging supervised learning algorithms, the model seeks to accurately classify transactions as legitimate or fraudulent while minimizing false positives and false negatives.In addition, the project focuses on enabling real-time fraud detection and alert mechanisms to prevent financial losses before transactions are completed. The system is designed to adapt to evolving fraud strategies through continuous learning and behavioral analysis.

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 5, December 2025 Impact Factor: 7.67

Another key objective is to provide a scalable and user-friendly architecture, integrating backend services and intuitive visualization dashboards, to support effective monitoring by financial institutions. Overall, the project aims to strengthen user trust, improve operational efficiency, and contribute to a secure and reliable UPI-based digital payment ecosystem.

1.3 Scope

This project focuses on the design, development, and evaluation of a machine learning–based UPI fraud detection system.

In-Scope

- Data Collection and Preprocessing: Using historical UPI transaction datasets and performing data cleaning, normalization, and handling of missing values.
- Feature Engineering: Identifying and extracting important features such as transaction amount, transaction frequency, timestamp, geolocation data, device information, payer and payee details, and user behavior patterns.
- Model Development: Implementing supervised learning algorithms such as Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM) for fraud detection.
- System Architecture: Developing backend services using Flask and a frontend interface for transaction input, prediction, and visualization.
- Performance Evaluation: Evaluating models using metrics like accuracy, precision, recall ,and confusion matrix analysis.

Out-of-Scope

- Deployment in live UPI or banking production environments
- Integration with external financial systems or telecom databases.
- Legal investigation or enforcement actions based on detected fraud.

1.4 Project Motivation and Significance

The rapid growth of UPI has made digital payments faster and more convenient, but it has also increased the risk of fraudulent activities such as phishing, fake payment requests, and unauthorized access. Traditional rule-based fraud detection methods are often insufficient to handle the large volume and evolving nature of UPI fraud. This project is motivated by the need for an intelligent and proactive fraud detection system that can identify suspicious transactions in real time. By applying machine learning techniques, the system can learn from transaction data, adapt to new fraud patterns, and reduce financial losses. The significance of this work lies in improving UPI transaction security, strengthening user trust, and supporting financial institutions with a scalable and efficient fraud prevention solution, contributing to a safer digital payment ecosystem.

II. METHODOLOGY

[1] Data Description and Sources

The analysis is based on a UPI transaction dataset consisting of records representing both legitimate and fraudulent digital payment transactions. The dataset includes multiple transaction-level attributes, with the primary objective being the classification of a binary target variable, where Fraud = 1 indicates a fraudulent transaction and Fraud = 0 represents a genuine transaction

Transaction Characteristics: Transaction amount, transaction date and time, transaction frequency, and transaction type.

User and Account Behavior: Historical transaction behavior, transaction count within a given time window, and abnormal spending patterns

Location and Device Information: Geographical location of transactions, geolocation mismatches, device identifiers, and IP-based access patterns.

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 5, December 2025 Impact Factor: 7.67

Merchant and Payment Attributes: Merchant category, payment channel, and risk indicators associated with merchants or receivers.

[2] Data Preprocessing and Cleaning

Effective fraud detection requires high-quality data; therefore, extensive preprocessing was performed to address inconsistencies and improve model performance.

Categorical Standardization: Inconsistent categorical values related to transaction type, device information, and payment channels were standardized to ensure uniform representation.

Missing Value Treatment: Missing values in numerical attributes such as transaction amount and frequency were handled using appropriate statistical methods

Encoding: Categorical variables were converted into numerical form using encoding techniques suitable for machine learning models.

Normalization and Scaling: Numerical features were normalized to ensure uniformity and prevent dominance of high-range values during model training.

Handling Class Imbalance: Since fraudulent transactions constitute a small portion of total transactions, techniques were applied to handle data imbalance and improve fraud detection accuracy.

[3] Feature Engineering

Feature engineering is a critical step in improving the effectiveness of the UPI fraud detection model. After preprocessing, relevant features were extracted and transformed to better represent transaction behavior and fraud indicators. Domain knowledge and exploratory data analysis were used to identify features that significantly contribute to distinguishing fraudulent transactions from genuine ones. Key engineered features include transaction frequency within a specific time window, time-based patterns such as unusual transaction hours, geolocation inconsistencies, and device usage variations. Behavioral features capturing deviations from a user's normal spending habits were also derived. These engineered features enhance the model's ability to capture subtle and evolving fraud patterns that may not be evident from raw transaction data alone. Feature selection techniques were applied to remove redundant and less significant attributes, thereby reducing model complexity and improving computational efficiency. The final feature set ensures a balance between model accuracy and scalability

[4] Model Development

The fraud detection task is formulated as a binary classification problem, where transactions are classified as fraudulent or legitimate. Supervised machine learning algorithms were employed due to the availability of labeled transaction data. Multiple models were trained and evaluated, including Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM). These algorithms were selected for their ability to handle structured transaction data and capture both linear and non-linear relationships. Each model was trained on the pre-processed dataset and optimized using suitable hyperparameters to improve prediction performance. To ensure generalization and robustness, the dataset was divided into training and testing sets. Model performance was evaluated using standard metrics such as accuracy, precision, recall, and confusion matrix analysis, with particular emphasis on minimizing false negatives to prevent undetected fraud. The best-performing model was integrated into the system for real-time fraud detection, enabling proactive identification of suspicious UPI transactions.

III. RESULTS AND DISCUSSION

This section discusses the results obtained from the implementation of the proposed UPI Fraud Detection System using Machine Learning. The performance of different machine learning models is evaluated and compared to determine their effectiveness in detecting fraudulent UPI transactions. Since fraud detection datasets are highly imbalanced, special emphasis is placed on performance metrics beyond accuracy.

The proposed UPI fraud detection system was evaluated using multiple supervised machine learning models, including Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM). Model performance was

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 5, December 2025 Impact Factor: 7.67

assessed using accuracy, precision, recall, F1-score, and confusion matrix analysis to account for the imbalanced nature of fraud data

The results indicate that Logistic Regression achieved the most reliable and consistent performance, particularly in terms of precision and overall classification stability. Its ability to model the probability of fraud and handle binary classification effectively made it suitable for detecting fraudulent UPI transactions with fewer false alarms. Decision Tree models showed interpretability but were more prone to misclassification, while SVM performed well in capturing non-linear patterns but required higher computational complexity.

Confusion matrix analysis showed that Logistic Regression maintained a good balance between false positives and false negatives, which is critical in fraud detection to prevent financial losses while minimizing inconvenience to genuine users. Overall, the results demonstrate that machine learning techniques are effective for UPI fraud detection, with Logistic Regression emerging as a practical and efficient solution for real-time digital payment security.

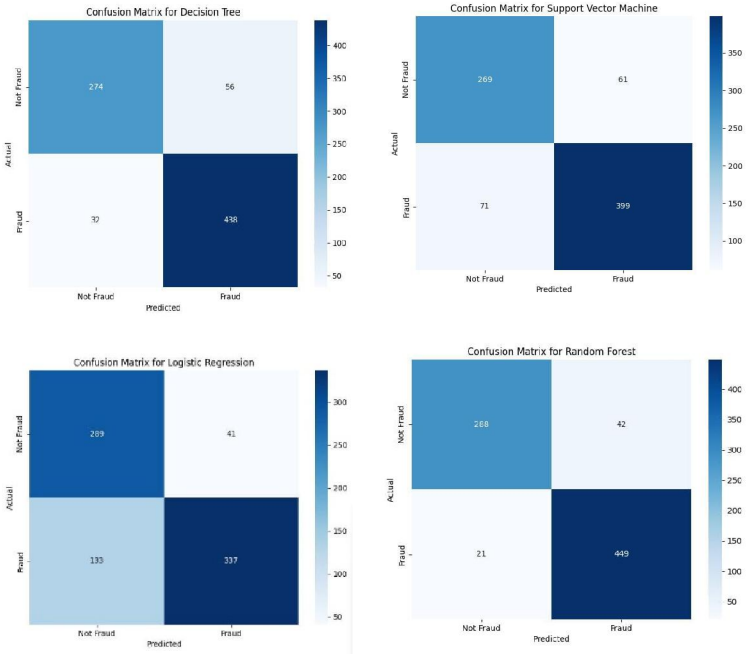

Figure: comparative confusion matrices of different machine learning models

REFERENCES

[1]. Mohammed, E., and Far, B., "Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study," IEEE Transactions on Big Data, vol. 6, no. 3, pp. 558–570, 2020, doi:10.1109/TBDATA.2019.2946178.

[2]. Randhawa, K., et al., "Credit Card Fraud Detection Using AdaBoost and Majority Voting," IEEE Access, vol. 7, pp. 15285–15294, 2019, doi:10.1109/ACCESS.2019.2896971.

[3]. Sorournejad, S., et al., "A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective," Journal of Information Security and Applications, vol. 48, pp. 102–118, 2019, doi:10.1016/j.jisa.2019.04.005.

[4]. Singh, T., Di Troia, F., Vissagio, C. A., and Stamp, M., "Support Vector Machines and Malware Detection," Computers & Security, vol. 86, pp. 208–223, 2019, doi:10.1016/j.cose.2019.06.005.

[5]. Wedge, C., Canter, R., Rubio, M., et al., "Solving the False Positives Problem in Fraud Prediction Using Automated Feature Engineering," Journal of Machine Learning Research, vol. 20, no. 1, pp. 1–22, 2019.

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 5, December 2025
Impact Factor: 7.67

[6]. Chaudhary, G., et al., "Fraud Detection in Credit Card Transactions Using Machine Learning Techniques," Journal of Big Data, vol. 7, Article no. 90, 2020, doi:10.1186/s40537-020-00329-1.

[7]. Brown, L., and Wang, P., "A Comparative Study of Support Vector Regression for Predicting Financial Markets," Finance and Technology, vol. 15, pp. 77–85, 2020, doi:10.1016/j.financetech.2020.05.003.

[8]. Gupta, K., et al., "A Hybrid Machine Learning Model for Fraud Detection," Expert Systems with Applications, vol. 143, Article no. 113005, 2020, doi:10.1016/j.eswa.2019.113005.

[9]. Kavitha, J., Indira, G., Kumar, A., Shrinita, A., and Bappan, D., "Fraud Detection in UPI Transactions Using Machine Learning," International Journal of Advanced Research, 2021.

[10]. Lakshmi, K. K., Gupta, H., and Ranjan, J., "UPI Based Mobile Banking Applications – Security Analysis and Enhancements," International Journal of Computer Applications, 2020.

[11]. Dhanwani, D. C., Tonpewar, A., Ikhar, D., Ladole, K., and Mahant, S., "Online Fraud Detection System," International Journal of Engineering Research, 2019.

[12]. Kumar, A., and Singh, R., "Performance Analysis of Supervised Learning Algorithms for Fraud Detection in UPI Transactions," International Journal of Computer Science, 2020.

[13]. Sharma, P., et al., "Anomaly Detection in Digital Transactions Using Deep Learning Techniques," IEEE Access, vol. 9, pp. 54832–54859, 2021.

[14]. Verma, S., and Patel, R., "Ensemble Learning Techniques for Fraud Detection in Digital Payments," Expert Systems with Applications, 2019.

[15]. Gupta, A., and Singh, S., "Deep Learning Architectures for Financial Fraud Detection," Applied Sciences, MDPI, vol. 15, no. 3, 2021.

[16]. Deshmukh, P., et al., "Predictors of Fraud Risk in UPI Transactions Using Penalized Logistic Regression," Procedia Computer Science, 2019.

[17]. Kaggle, "UPI Fraud Detection Dataset," [Online]. Available: https://www.kaggle.com

[18]. Pedregosa, F., et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[19]. McKinney, W., "Data Structures for Statistical Computing in Python," Proceedings of the 9th Python in Science Conference, pp. 51–56, 2010.

[20]. Waskom, M. L., "Seaborn: Statistical Data Visualization," Journal of Open Source Software, vol. 6, no. 60, p. 3021, 2021