IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 12, December 2025

DOI: 10.17148/IJIREEICE.2025.131213

UPI FRUAD DETECTION USING MACHINE LEARNING

Divyarani S Y [1], Sinchana K S [2], Minchu K S [3], Monisha B M [4]

Department of Artificial Intelligence and Machine Learning Engineering,

G. Madegowda Institute of Technology,

Bharthinagara, Maddur TQ, Mandya dist-571422, India

Abstract: With the rapid growth of digital payment systems in India, the Unified Payments Interface (UPI) has become one of the most widely used platforms for real-time financial transactions. While UPI offers convenience, speed, and accessibility, it has also become a target for fraudulent activities such as unauthorized transactions, phishing attacks, identity theft, and account takeovers. These increasing fraud incidents highlight the need for an intelligent and automated fraud detection system to ensure secure digital transactions. This project presents a UPI Fraud Detection System using Machine Learning techniques to identify and prevent fraudulent transactions in real time. The proposed system analyzes historical transaction data and user behavior patterns to distinguish between legitimate and fraudulent activities. Key features such as transaction amount, transaction frequency, time of transaction, location variance, device usage, and transaction velocity are extracted and processed for model training. Machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM) are applied to classify transactions as either genuine or fraudulent.

Keywords: UPI FRUAD , Machine learning , historical transaction,random forest, fraudulent

I. INTRODUCTION

In recent years, digital payment systems have revolutionized the way financial transactions are conducted across the world. In India, the introduction of the Unified Payments Interface (UPI) by the National Payments Corporation of India (NPCI) has significantly transformed the digital payment ecosystem. UPI enables instant, real-time fund transfers between bank accounts using mobile devices, eliminating the need for cash, debit cards, or net banking credentials. Its simplicity, interoperability, and 24×7 availability have led to widespread adoption across urban and rural regions.

UPI supports various transaction types such as peer-to-peer transfers, merchant payments, bill payments, and online purchases. Applications such as Google Pay, PhonePe, Paytm, and BHIM have further boosted its popularity. As a result, millions of transactions are processed daily, involving large volumes of sensitive financial data. While UPI has improved financial inclusion and economic efficiency, the rapid growth in digital transactions has also increased the risk of fraudulent activities.

II. BACKGROUND AND MOTIVATION

The rise of digital payment platforms has revolutionized financial transactions worldwide, making payments faster, more convenient, and accessible from anywhere. In India, the Unified Payments Interface (UPI) has become one of the most widely used digital payment methods, enabling instant money transfers across banks through mobile devices. UPI transactions are highly interoperable, secure, and user-friendly, contributing to the growth of a cashless economy.

However, the growing adoption of UPI has also attracted malicious actors who exploit vulnerabilities in mobile devices, applications, and user behavior to commit fraudulent activities. Common forms of UPI fraud include phishing attacks, fake customer care calls, QR code scams, social engineering attacks, and unauthorized account access. These fraudulent activities can lead to financial losses, erode trust in digital payment systems, and impact user confidence.

Motivation:

The primary motivation for this project arises from the increasing incidence of UPI fraud and the limitations of existing detection systems. Some key motivating factors include:

Rising Digital Payment Adoption: The exponential growth of UPI transactions makes real-time monitoring essential for fraud prevention.

Complex and Evolving Fraud Patterns: Fraudsters continuously develop new techniques that traditional systems cannot easily detect.

Need for Automation: Manual verification is slow, costly, and prone to human error; automated solutions are necessary.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 12, December 2025

DOI: 10.17148/IJIREEICE.2025.131213

Enhancing Security and Trust: Effective fraud detection improves user confidence in digital payments and reduces financial risks.

Applicability of Machine Learning: Machine learning algorithms can learn from historical transaction data to identify suspicious behavior, making them well-suited for adaptive fraud detection.

III. LITERATURE REVIEW

Financial fraud detection has been widely studied, especially in contexts like credit card and online banking fraud. Traditional systems often rely on rule-based engines and manual reviews, which fail to capture adaptive fraud patterns due to rigid thresholds and static rules. With the arrival of machine learning, researchers have focused on automated and intelligent techniques capable of learning from data. Early studies highlight the effectiveness of classification models such as Logistic Regression, Decision Trees, and Support Vector Machines in identifying anomalies in transaction data.

For example, Bhattacharyya et al. (2011) investigated supervised machine learning models for credit card fraud detection. They found that Random Forest and SVM achieve robust performance when handling high-dimensional transaction features and imbalanced datasets. The study emphasized feature engineering and class imbalance handling as crucial steps for practical performance improvements.

In recent years, with the massive adoption of UPI in India, researchers have started focusing on fraud detection specific to UPI transactions. Unlike traditional banking systems, UPI traffic demonstrates unique challenges related to mobile usage patterns, rapid peer-to-peer transfers, and varying transaction contexts.

One such study by Kumar and Singh (2020) applied supervised learning algorithms to mobile payment datasets. They compared the performance of Logistic Regression, Decision Trees, Random Forests, and Gradient Boosting Machines. Their results showed that ensemble models like Random Forest and XGBoost significantly outperformed standard classifiers by learning complex non-linear relationships in transaction features. The author also stressed the need for real-time detection capabilities due to the high volume of UPI traffic.

IV. METHODOLOGY

The proposed system aims to detect fraudulent UPI transactions using machine learning techniques. The methodology involves collecting transaction data, preprocessing the data, extracting relevant features, training machine learning models, and evaluating their performance. The system is designed to analyze transaction patterns and classify each transaction as either legitimate or fraudulent.

The overall workflow ensures accuracy, efficiency, and adaptability to changing fraud patterns while maintaining real-time processing capability.

Data Collection:

The dataset used in this project consists of historical UPI transaction records. Each transaction includes attributes such as transaction ID, transaction amount, transaction time, sender and receiver details, device information, location, transaction frequency, and transaction status (fraud or non-fraud).

Due to privacy and security concerns, publicly available or simulated datasets are used that closely resemble real-world UPI transactions. The dataset is labeled to support supervised machine learning techniques.

Data Preprocessing:

Data preprocessing is a critical step to improve model performance. The following preprocessing techniques are applied:

Removal of duplicate and irrelevant records.

Handling missing and inconsistent values.

Encoding categorical variables into numerical format.

Normalization or standardization of numerical features.

Handling class imbalance using resampling techniques such as SMOTE.

Preprocessing ensures clean and structured data suitable for machine learning algorithms.

Feature Extraction and Selection:

Relevant features are extracted from the transaction dataset to represent user behavior and transaction characteristics. Key features include:

Transaction amount

Transaction frequency

Time-based features (hour, day, unusual time)

Location variance

Device or IP change

Transaction velocity

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 12, December 2025

DOI: 10.17148/IJIREEICE.2025.131213

Feature selection techniques are applied to identify the most significant attributes, reducing dimensionality and improving computational efficiency.

Model Selection and Training:

Multiple supervised machine learning algorithms are implemented and compared, including:

Logistic Regression

Decision Tree

Random Forest

Support Vector Machine (SVM)

The dataset is split into training and testing sets. Models are trained using the training data, and hyperparameters are tuned to optimize performance. Ensemble models are preferred due to their ability to handle non-linear relationships and reduce overfitting.

Model Evaluation:

The trained models are evaluated using standard performance metrics:

Accuracy

Precision

Recall

F1-score

Confusion Matrix

These metrics help assess the effectiveness of each model in detecting fraudulent transactions while minimizing false positives.

Fraud Detection and Alert Generation:

Once trained, the best-performing model is used to classify new UPI transactions in real time. Transactions identified as suspicious are flagged, and alerts can be generated for further verification or user confirmation.

This step helps prevent unauthorized transactions and reduces financial losses.

V. IMPLEMENTATION

The implementation process involves the following steps:

Data Preprocessing: Duplicate and irrelevant records are removed, missing values are handled, categorical features are encoded numerically, and numerical features are normalized. SMOTE is applied to balance the dataset for better model learning.

Feature Engineering: Relevant features representing user behavior and transaction characteristics (amount, frequency, time, location, velocity) are extracted.

Model Training: Supervised machine learning algorithms including Logistic Regression, Decision Tree, Random Forest, and SVM are trained on the preprocessed data. Hyperparameter tuning is performed to optimize performance.

Evaluation and Prediction: Models are evaluated using accuracy, precision, recall, F1-score, and confusion matrix. The best-performing model (Random Forest) is used to classify new transactions as fraudulent or legitimate, and alerts are generated for suspicious activity.

The implemented system efficiently detects fraud while reducing false positives, providing a scalable and automated solution to enhance UPI transaction security.

VI. ANALYSIS AND DISCUSSION

The performance of the UPI Fraud Detection System is evaluated by analyzing the results obtained from multiple machine learning models, including Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM). The system uses metrics such as accuracy, precision, recall, F1-score, and confusion matrix to assess the effectiveness of each model in detecting fraudulent transactions.

From the experimental results, it is observed that ensemble-based models, particularly Random Forest, outperform other algorithms in terms of all evaluation metrics. Random Forest effectively captures complex, non-linear patterns in transaction data and reduces overfitting, leading to better fraud detection accuracy. Logistic Regression and SVM perform reasonably well but show lower recall and F1-score, indicating they may miss some fraudulent transactions. Decision Tree performs moderately but is prone to overfitting on training data.

The confusion matrix analysis demonstrates that the Random Forest model correctly identifies most fraudulent and legitimate transactions, minimizing false positives and false negatives. Handling the class imbalance using SMOTE significantly improves the model's sensitivity towards the minority class (fraudulent transactions), which is critical in fraud detection.

Overall, the analysis shows that machine learning provides an effective solution for UPI fraud detection. Feature engineering and preprocessing steps, such as encoding, normalization, and balancing, play a vital role in improving the

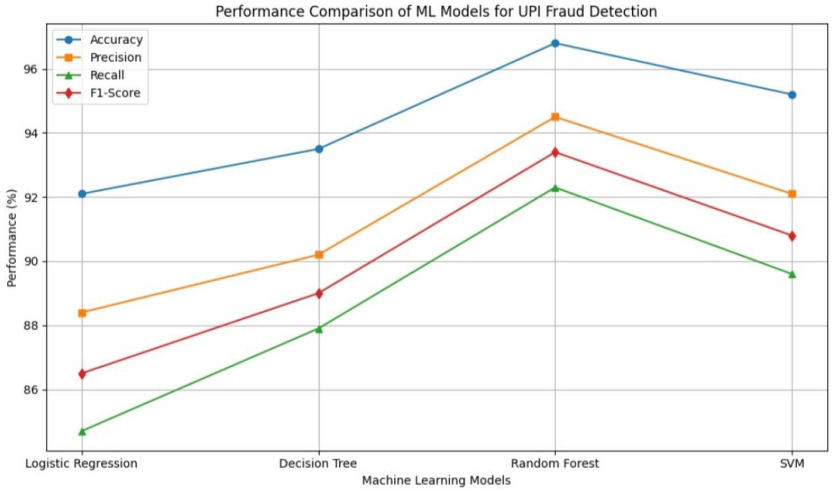
IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 12, December 2025

DOI: 10.17148/IJIREEICE.2025.131213

system's performance. The proposed system not only identifies fraudulent transactions with high accuracy but also reduces the risk of unnecessary alerts for genuine users, thereby improving the reliability and trustworthiness of digital payment platforms.



VII. CONCLUSION

The rapid growth of digital payment systems, particularly the Unified Payments Interface (UPI), has transformed financial transactions by providing convenience, speed, and accessibility. However, this growth has also led to an increase in fraudulent activities, posing serious challenges to users and financial institutions. The primary objective of this project was to design and implement an effective fraud detection system for UPI transactions using machine learning techniques.

In this project, a machine learning-based approach was successfully developed to identify fraudulent UPI transactions by analyzing transaction behavior and user patterns. The system incorporated essential stages such as data collection, preprocessing, feature engineering, model training, and performance evaluation. Multiple supervised machine learning algorithms, including Logistic Regression, Decision Tree, Support Vector Machine (SVM), and Random Forest, were implemented and compared to identify the most suitable model for fraud detection.

The experimental results demonstrated that ensemble-based algorithms, particularly Random Forest, achieved superior performance in terms of accuracy, precision, recall, and F1-score. This indicates the effectiveness of machine learning in capturing complex and non-linear fraud patterns that traditional rule-based systems fail to detect. Data preprocessing and class balancing techniques played a crucial role in improving detection accuracy and reducing false positives.

The proposed system is capable of identifying suspicious transactions and generating alerts, thereby minimizing financial losses and enhancing the security of UPI payment platforms. Overall, this project validates the practical applicability of machine learning in strengthening fraud detection mechanisms and improving trust in digital payment systems.

REFERENCES

[1]. Bhattacharyya, S., Jha, S., Tharakunnel, K., and Westland, J. C.,"Data Mining for Credit Card Fraud: A Comparative Study,"Decision Support Systems, vol. 50, no. 3, pp. 602–613, 2011.

[2]. Dal Pozzolo, A., Bontempi, G., Snoeck, M., and Snoeck, D.,"Handling Imbalanced Data in Fraud Detection," Neurocomputing, vol. 150, pp. 213–220, 2015.

[3]. Kumar, A., and Singh, R.,"Fraud Detection in Digital Payment Systems Using Machine Learning," International Journal of Computer Applications, vol. 176, no. 29, pp. 1–6, 2020.

[4]. Sharma, N., Verma, S., and Gupta, R.,"UPI Transaction Fraud Detection Using Machine Learning Techniques," International Journal of Engineering Research and Technology (IJERT), vol. 10, no. 5, pp. 120–125, 2021.

[5]. Patel, J., and Mehta, D.,"Feature Selection and Classification Techniques for Fraud Detection," International Journal of Computer Science and Information Security, vol. 17, no. 4, pp. 45–50, 2019.

[6]. Zhou, C., Paffenroth, R. C., and Tracy, K.,"Anomaly Detection with Robust Deep Autoencoders," Proceedings of the 23rd ACM SIGKDD Conference, pp. 665–674, 2020.

[7]. National Payments Corporation of India (NPCI),"Unified Payments Interface (UPI) – Overview and Guidelines,"

[8]. NPCI Official Documentation.

DOI: 10.17148/IJIREEICE.2025.131213

[9]. Pedregosa, F., et al.,"Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[10]. Géron, A.,Hands-On Machine Learning with Scikit-Learn and TensorFlow,O'Reilly Media, 2019.
[11]. Brownlee, J.,Machine Learning Mastery with Python,Machine Learning Mastery, 2018.