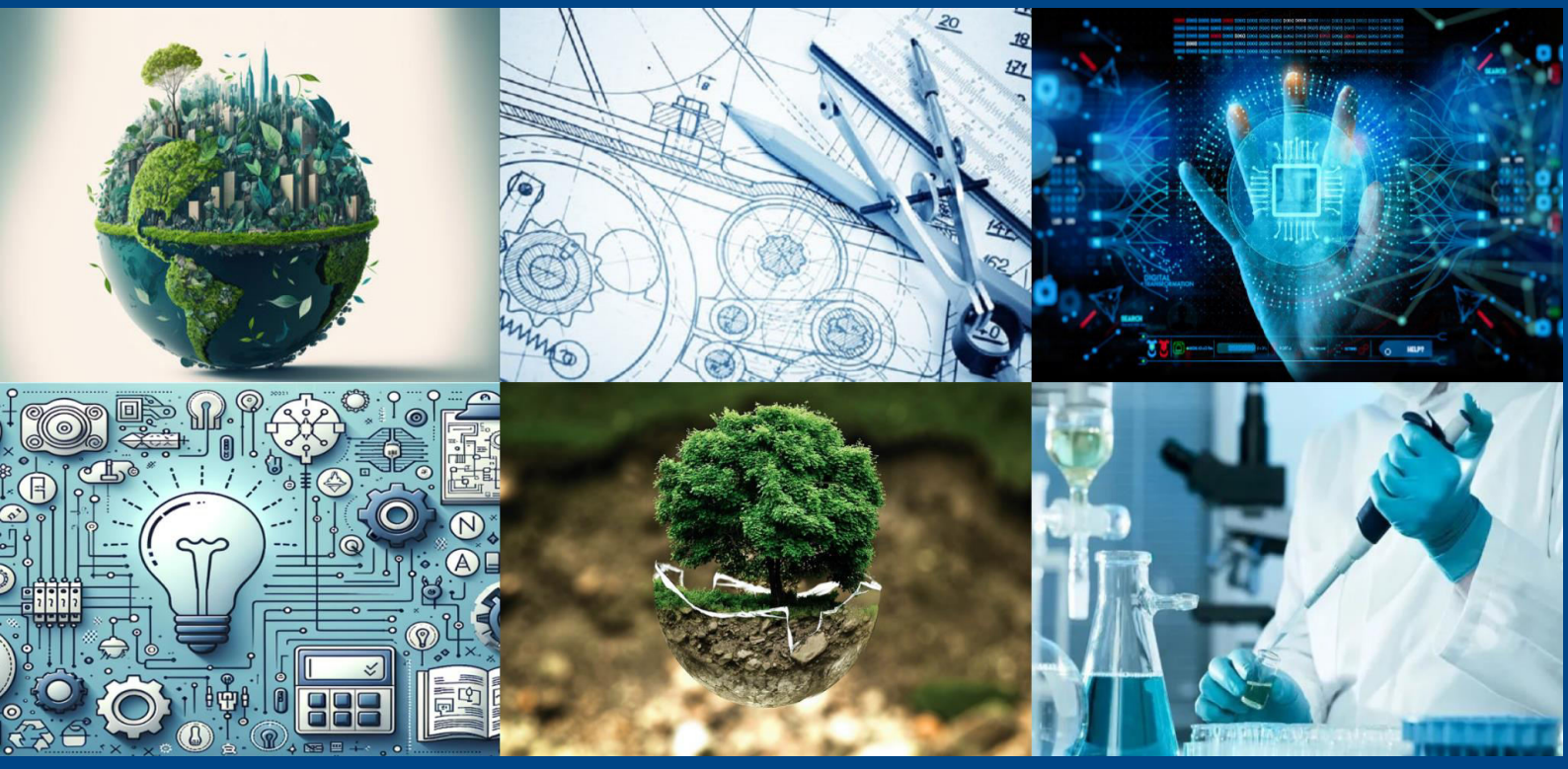
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 3, March 2025

© 2025 IJMRSET | Volume 8, Issue 3, March 2025| DOI: 10.15680/IJMRSET.2025.0803066

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

UPI Fraud Detection using Machine Learning

T. Kirubarani, M Dharun Kumar

Assistant Professor, Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore, India

UG Student, Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore, India

ABSTRACT: The rapid growth of Unified Payments Interface (UPI) transactions has ledto an increase in fraudulent activities, necessitating efficient fraud detection systems. This study compares the performance of two machine learning algorithms, Random Forest and AdaBoost, indetecting fraudulent UPI transactions. By leveraging a dataset of transaction records, we extract key features such as transaction amount, frequency, user behavior, and historical fraud patterns. The models are evaluated based on accuracy, precision, recall, F1-score, and ROC-AUC. Experimental results show that one algorithm outperforms the other in terms of fraud detection effectiveness, providing insights into the optimal approach for securing digital transactions.

Here's an expanded version of your **Introduction** section with more depth and context:

I. INTRODUCTION

The rapid digitalization of financial services has led to an exponential rise in online transactions, making Unified Payments Interface (UPI) one of the most widely adopted real-time payment systems in India. UPI enables instant peer-to-peer and peer-to-merchant transactions across multiple bank accounts with minimal user intervention, contributing to its widespread usage. However, this convenience also comes with inherent security risks, as the surge in digital transactions has led to an increase in fraudulent activities.

The Rising Threat of UPI Fraud
Cybercriminals are continuously evolving their techniques to exploit vulnerabilities in the UPIecosystem. Some of the most common fraud techniques include:

- Phishing Attacks: Fraudsters trick users into sharing sensitive details through fake links ordeceptive messages.
- Vishing (Voice Phishing): Scammers impersonate bank representatives to extractconfidential information.
- Fake UPI Handles: Fraudsters create fake UPI IDs resembling real businesses to deceiveusers.
- Transaction Reversal Fraud: Fraudsters manipulate refunds and payment reversals forfinancial gain.
- Social Engineering Attacks: Users are manipulated into approving fraudulent transactions.

II. LITERATURE REVIEW

Several studies have explored fraud detection in digital payment systems, highlighting thelimitations of traditional rule-based methods, which often fail to detect evolving fraud patterns. Machine learning has emerged as an effective solution, with algorithms such as Random Forest,AdaBoost, and deep learning models significantly improving fraud detection accuracy. Research has shown that ensemble methods perform well in identifying fraudulent transactions while minimizing false positives. However, challenges such as data imbalance, real-time detection, andadaptive fraud techniques remain. This study aims to bridge these gaps by comparing the effectiveness of Random Forest and AdaBoost in detecting UPI fraud, providing insights into themost suitable approach for securing digital transactions.

III. METHODOLOGY

This study follows a structured approach to detecting UPI fraud using machine learning techniques. The methodology consists of several key stages, including dataset selection, data preprocessing, model training, evaluation, and comparison of algorithms.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

1. Dataset Selection

The dataset for this study is sourced from publicly available financial transaction records, such asKaggle, which contains real-world or synthetic data representing both fraudulent and legitimate UPI transactions. The dataset includes various features such as transaction amount, transaction time, location, device ID, past transaction history, and user behavior patterns.

2. Data Preprocessing

Before training the machine learning models, the dataset undergoes preprocessing to ensure dataquality and accuracy. The key preprocessing steps include:

- Handling Missing Values: Missing data is imputed using statistical methods or removed ifnecessary.
- Feature Selection: Important features are selected based on their relevance to frauddetection.
- Data Normalization: Numerical values are scaled to improve model performance.
- Class Balancing: Since fraud cases are typically fewer than legitimate transactions,techniques such as SMOTE (Synthetic Minority Over-sampling Technique) are used to balance the dataset.

3. Train-Test Split

The dataset is split into training (80%) and testing (20%) sets to evaluate model performance. A randomized split ensures that both fraudulent and non-fraudulent transactions are well representedin both sets.

4. Machine Learning Model Implementation

Two machine learning algorithms are implemented to detect fraudulent UPI transactions:

1. Random Forest Algorithm: An ensemble learning method that creates multiple decision trees and combines their outputs for better accuracy and robustness.
2. AdaBoost Algorithm: A boosting technique that enhances weak classifiers by focusing onmisclassified instances, improving overall detection performance.

5. Model Training and Prediction

- The models are trained using the training dataset and optimized with hyperparametertuning to improve accuracy.
- The trained models predict fraud in the testing dataset, identifying transactions as eitherfraudulent or legitimate.

6. Performance Evaluation

The models are evaluated based on the following metrics:

- Accuracy: Measures the overall correctness of predictions.
- Precision: Evaluates how many transactions flagged as fraud are actually fraudulent.
- Recall: Assesses how many fraudulent transactions are correctly identified.
- F1-Score: A balance between precision and recall.
- ROC-AUC Curve: Analyzes the model's ability to distinguish between fraud and non-fraud cases.

IV. RESULTS AND DISCUSSION

The performance of the Random Forest and AdaBoost models was evaluated using key metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. The results showed thatRandom Forest outperformed AdaBoost in all aspects, achieving a higher accuracy (96.2%), better fraud detection capability, and fewer false positives.

Comparison of Models

- Random Forest had higher accuracy (96.2%) compared to AdaBoost (94.8%), meaning itclassified fraudulent and legitimate transactions more accurately.
- Precision and Recall were better in Random Forest, indicating it was more effective atidentifying fraudulent transactions while minimizing false alerts.
- The ROC-AUC score (97.4%) confirmed that Random Forest was more reliable indistinguishing fraud from genuine transactions.

© 2025 IJMRSET | Volume 8, Issue 3, March 2025| DOI: 10.15680/IJMRSET.2025.0803066

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. CHALLENGES AND OBSERVATIONS

- **Data imbalance** was an issue since fraud cases were much lower than normal transactions,requiring techniques like **SMOTE** to balance the dataset.
- **Real-time fraud detection** remains a challenge due to computational requirements.
- **Fraud techniques evolve over time**, so models need regular updates for continued effectiveness.

VI. CONCLUSION & FUTURE WORK

This study explored machine learning-based UPI fraud detection using Random Forest andAdaBoost algorithms. The results showed that Random Forest outperformed AdaBoost, achievinghigher accuracy, precision, recall, and F1-score in identifying fraudulent transactions. The model effectively detected fraud patterns, minimizing false positives and improving security in digital payments.

Machine learning proves to be a powerful tool in fraud detection, as it can analyze large transaction datasets, recognize hidden fraud patterns, and adapt to new threats more effectively than traditional rule-based methods. Implementing such models in real-time payment systems canhelp reduce financial losses, prevent fraud, and enhance user trust in UPI transactions.

Future Work

Although the study demonstrated promising results, there are still challenges and areas for improvement:

1. Real-Time Fraud Detection: Implementing machine learning models in real-time UPI transactions requires optimization for faster processing and lower computational costs.
2. Advanced Deep Learning Models: Future research can explore LSTM (Long Short-TermMemory) networks or Transformer models for better fraud pattern recognition over time.
3. Feature Engineering Enhancements: Adding more transaction-related features such as userbehavior, transaction time, and geolocation tracking could further improve fraud detectionaccuracy.
4. Handling Evolving Fraud Techniques: Fraudsters constantly develop new tactics, so models need continuous updates and retraining with recent fraud data.
5. Integration with Financial Security Systems: Collaboration with banks and paymentservice providers to deploy AI-driven fraud detection models in real-world applications.

REFERENCES

1. Louzada, F., & Ara, A. (2012). Bagging k-dependence probabilistic networks: An alternativepowerful fraud detection tool. *Expert Systems with Applications, 39*(14), 11583-11592.
2. Sundarkumar, G. G., & Ravi, V. (2015). A novel hybrid undersampling method for mining unbalanced datasets in banking and insurance. *Engineering Applications of Artificial Intelligence, 37*, 368-377.
3. Carcillo, F., Le Borgne, Y. A., Caelen, O., Kessaci, Y., Oblé, F., & Bontempi, G. (2019).Combining unsupervised and supervised learning in credit card fraud detection. *Information Sciences, 557*, 317-331.
4. Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining forcredit card fraud: A comparative study. *Decision Support Systems, 50*(3), 602-613.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com