WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

UPI Fraud Detection System Using Machine Learning Models

Nandhini Devi V[1], Sandhiya V[2], Reshma Mol R. S[3], T. Priya[4]

[1,2,3]Department of Computer Science, Kingston Engineering College, Vellore
[4]Guide, AP/CSE, Kingston Engineering College, Vellore tpriya.engineering@kingston.ac.in
Email: nandhinikamalesh18@gmail.com, Sandysandhiya1508@gmail.com , molreshma67@gmail.com

ABSTRACT

The introduction of Unified Payments Interface (UPI) has changed the digital transaction landscape in India, notwithstanding the possibility of several fraudulent operations. The use of machine learning (ML) techniques for UPI fraud detection is covered in brief here. In order to identify unusual patterns that might point to fraudulent activity, machine learning (ML) models examine transaction data using a combination of supervised, unsupervised, and semi-supervised learning techniques. As essential to optimizing the models' performance, effective feature selection and engineering techniques further improve the process. Additionally, combining anomaly detection algorithms with cooperative techniques improves the accuracy of fraud identification. The resilience of UPI security is increased, and quicker responses to new fraud techniques are made possible by the implementation of realtime monitoring mechanisms and adaptive learning strategies. Financial institutions can enhance the security of UPI transactions and preserve their integrity while fostering user trust by utilizing machine learning .

The system employs three different machine learning algorithms: Random Forest, Gradient Boosting, and AdaBoost, to predict whether a transaction is fraudulent or legitimate based on the transaction amount and the hour of the day. we aim to identify fraudulent patterns using transaction features such as amount, time, and transaction status. The model demonstrates its effectiveness in distinguishing legitimate transactions from fraudulent ones, ensuring a secure payment environment.

1. LITERATURE SURVEY

Fraud detection in online payment systems has emerged as a critical area of research, particularly with the widespread use of UPI (Unified Payments Interface) in India. UPI has revolutionized the digital payment landscape, but the increased usage also brings with it a rise in fraudulent activities. Detecting fraud in real-time is crucial to prevent financial loss, especially in a fast-paced transaction environment like UPI.

Machine learning has become a powerful tool for fraud detection, as it can model complex relationships in transaction data and make predictions based on historical patterns. Various algorithms, such as Random Forest, Gradient Boosting, and AdaBoost, have been explored for this purpose. These models can effectively classify transactions into "fraudulent" or "legitimate" categories by learning from features such as transaction amount, time, and user behavior.

Random Forest and Gradient Boosting have been widely used due to their high accuracy and ability to handle large, imbalanced datasets. AdaBoost, an ensemble learning algorithm, also has proven effectiveness, especially in boosting weak models to correct misclassifications.

Additionally, user interfaces designed with Tkinter provide an accessible way for users to interact with these machine learning models, making it easier for non-technical users to check transactions for fraud in real-time

2.PROPOSED SYSTEM

The proposed system is designed to predict fraudulent transactions in the UPI ecosystem using machine learning models. It aims to provide a real-time fraud detection solution that can be easily integrated into UPI transaction processing systems.

Key features of the proposed system include:

Data Collection:

Historical transaction data, including features like transaction ID, timestamp, sender and receiver details, amount, and transaction status.

Preprocessing:

Convert the timestamp into time-based features such as hour, day, and weekday

Clean the data to handle missing or inconsistent values.

Fraud Labeling:

- Label fraudulent transactions based on predefined conditions:
- Transactions with a Status of "FAILED.
- Transactions with amounts exceeding the 95th percentile.

Feature Engineering:

- Extract meaningful features from raw data (e.g., Amount (INR), hour).
- Create a binary fraud label (1 for fraud, 0 for legitimate).

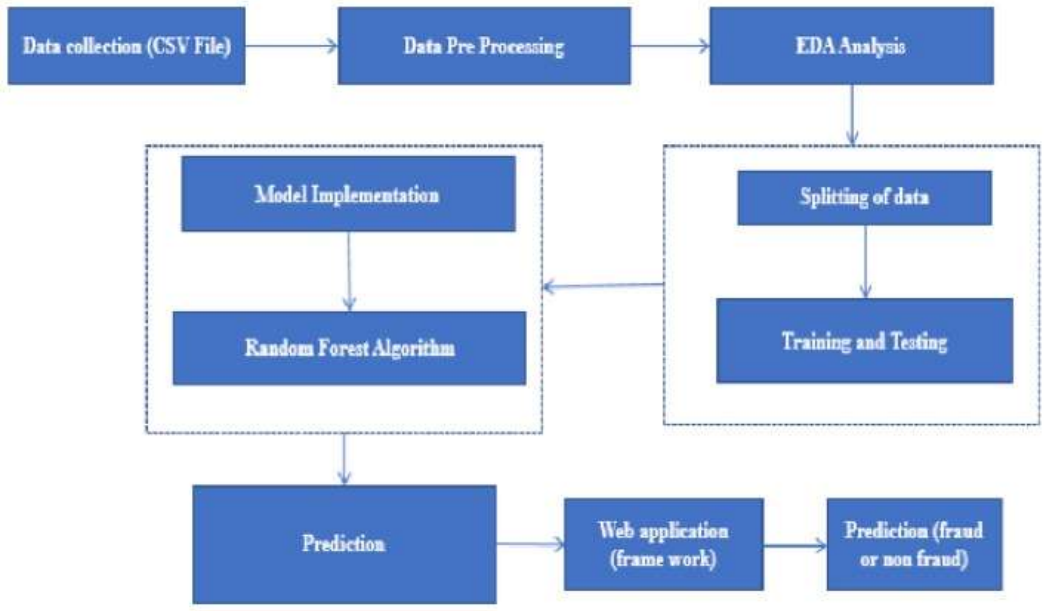
Multiple Machine Learning Models:

The system uses three machine learning algorithms: Random Forest, Gradient Boosting, and AdaBoost, to predict fraud. These models are trained on historical transaction data and have been selected for their high accuracy in binary classification tasks.

3.ARCHITECTURAL DESIGN FOR PROPOSED SYSTEM

The system architecture consists of the following components:

1. **Data Input Layer**:
    - The system accepts transaction data (amount, time, and status) as input.
    - Features are engineered (such as extracting the hour of the day) to prepare the data for model training.
2. **Model Training Layer**:
    - Machine learning models are trained using the processed data. Models like Random Forest are trained to predict fraud.
    - The models are saved for future predictions.
3. **Model Prediction Layer**:
    - The user inputs transaction details into the GUI. The system loads the selected model, processes the inputs, and predicts whether the transaction is fraudulent or legitimate..
4. **Visualization Layer**:
    - The system generates and displays graphs such as bar plots for feature importance and fraud statistics using Matplotlib and Seaborn.



4.ALGORITHMS/TECHNIQUES USED

 Random Forest Classifier: An ensemble learning technique that constructs multiple decision trees and combines their outputs to improve accuracy and robustness. It is used for classifying transactions into fraudulent or legitimate categories.

 Gradient Boosting Classifier: A boosting technique that combines weak learners (decision trees) to correct errors of previous models. It focuses on minimizing errors iteratively.

 AdaBoost Classifier: Another boosting algorithm that adjusts weights of misclassified instances to improve prediction accuracy. It is used to detect fraudulent transactions by boosting weak classifiers.

 Feature Engineering: Involves extracting useful features from raw data, such as the transaction amount and time, to improve model performance.

5. REFERENCES

 Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

 Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119-139.

 Raschka, S. (2015). *Python Machine Learning*. Packt Publishing.

 Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321-357.

 Dr. Virshree Tungare. (2019). A study on customer insight towards upi (unified payment interface) - an advancement of mobile payment system, International Journal of Science and Research (IJSR).

 Lakshmi K., Gupta H., and Ranjan J. (2019). Upi based mobile banking applications – security analysis and enhancements, in 2019 Amity International Conference on Artificial Intelligence (AICAI).

 Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 1189-1232.

 Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119-139.