ML-POWERED UPI FRAUD DETECTION

[1] A.Arockia Helen Sushma, [2] S.Madhu Bala, [3] J.Margrate Sneka, [4] N.Pavithra, [5]M.Pooja
[1]Assistant Professor, Department of Electronics and Communication Engineering, SSM Institute of Engineering and Technology, Dindigul – 624 002, Tamil Nadu, India
[2,3,4,5]Department of Electronics and Communication Engineering, SSM Institute of Engineering and Technology, Dindigul – 624 002, Tamil Nadu, India

[1]helensushma@gmail.com,[3]margratesnekaj@gmail.com, [4]pavinaga0507@gmail.com, [5]srip6403@gmail.com

Abstract: This project aims to develop a real-time fraud detection system for UPI transactions using Machine Learning, specifically SMOTE and XGBoost. Since fraudulent transactions are rare compared to legitimate ones, SMOTE is used to balance the dataset by generating synthetic fraud samples, improving model accuracy. XGBoost, a powerful gradient boosting algorithm, is implemented to classify transactions as fraudulent or legitimate based on patterns in user behavior and transaction details. The system is designed to detect fraud instantly, helping prevent unauthorized payments by triggering alerts or blocking suspicious transactions. By continuously learning from new fraud patterns, this approach enhances the security of digital payments and reduces financial fraud risks effectively.

Keywords: UPI, Financial Fraud, SMOTE, XGBoost, Machine Learning, Real-Time Detection, Data Imbalance

Corresponding Author: *A.Arockia Helen Sushma*
Assistant Professor, Department of Electronics and Communication Engineering, SSM Institute of Engineering and Technology, Dindigul – 624 002
Mail: karthickkiwi@gmail.com

I. INTRODUCTION

As digital payments, particularly UPI, gain popularity in India, they become prime targets for fraudulent activities. Traditional rule-based fraud detection systems struggle to detect new, evolving patterns. To address this, machine learning-based systems offer dynamic, data-driven models capable of real-time analysis and adaptive learning.

The objective of implementing a machine learning algorithm for real-time financial fraud detection and prevention in UPI transactions using SMOTE and XGBoost is to enhance security and minimize fraudulent activities.

Since financial fraud is a critical issue with highly imbalanced datasets, SMOTE (Synthetic Minority Over-sampling Technique) is employed to balance the data by generating synthetic fraud samples, improving model training and accuracy.

XGBoost (Extreme Gradient Boosting) is chosen for its efficiency, scalability, and ability to handle large datasets while preventing overfitting. The model is designed to analyze transactional patterns, detect anomalies, and classify transactions as either legitimate or fraudulent in real time.

Upon detecting fraudulent transactions, the system can trigger alerts, block unauthorized payments, and apply additional verification measures to prevent losses.

II. LITERATURE SURVEY

Existing systems rely on static rules and basic verification methods, which result in high false positives and negatives. Recent studies suggest leveraging algorithms like Random Forest, SVM, and Neural Networks can provide more adaptable fraud detection. These studies also highlight the importance of addressing class imbalance in financial datasets.

III. PROBLEM DEFINITION

Rule-based fraud detection systems lack adaptability, generate high false alarm rates, and cannot detect unseen fraud types. Moreover, the scarcity of fraud examples in datasets limits traditional models from learning complex fraudulent behavior patterns.

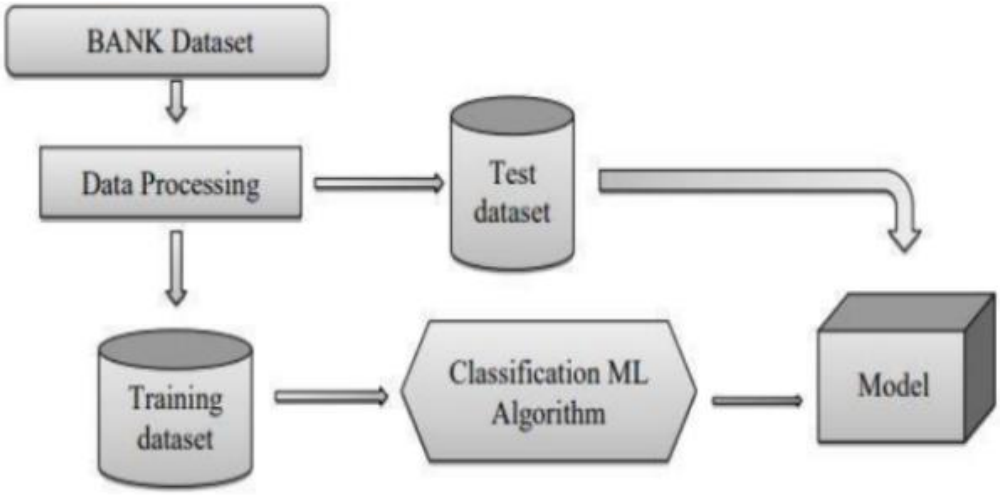
IV. PROPOSED SYSTEM

The proposed model uses:

SMOTE (Synthetic Minority Over-sampling Technique): to artificially balance the training data by generating synthetic samples of the minority class (fraud).

XGBoost (Extreme Gradient Boosting): to accurately classify transactions by learning from transaction patterns and user behavior.

Additional algorithms like Random Forest and SVM are considered for comparative performance evaluation.



Methodology for UPI Fraud Detection using Machine Learning

1. Data Collection

Sources: Collect transactional data, including features like transaction amount, timestamp, location, merchant ID, device information, account history, and transaction status (legitimate or fraud).

Data Privacy: Ensure anonymization of sensitive data to comply with data protection regulations.

2. Data Preprocessing

Cleaning: Handle missing, inconsistent, or duplicate entries.

Encoding: Convert categorical variables (e.g., device type, transaction type) into numerical values using one-hot encoding or label encoding.

Feature Engineering: Create new features such as:

Transaction frequency

Average transaction amount

Time since last transaction

Geolocation deviations.

Normalization/Scaling: Normalize numerical features to bring them to a common scale.

3. Exploratory Data Analysis (EDA)

Pattern Detection: Analyze typical behavior versus fraudulent behavior.

Imbalance Handling: Since fraud transactions are rare, use techniques like:

Oversampling (SMOTE)

Undersampling

Synthetic data generation

4. Model Selection

Supervised Learning Models: Choose appropriate models such as:

Logistic Regression

Decision Trees

Random Forest

XGBoost

Support Vector Machines (SVM)

Neural Networks (especially for large datasets)

Unsupervised Learning Models (for anomaly detection if labeled data is unavailable):

Isolation Forest

One-Class SVM

Autoencoders

5. Model Training and Validation

Train-Test Split: Split the dataset (e.g., 70% training, 30% testing).

Cross-Validation: Apply k-fold cross-validation to validate model performance.

Hyperparameter Tuning: Use Grid Search or Random Search to optimize model parameters.

6. Evaluation Metrics

Because fraud detection is a class-imbalance problem, use:

Precision

Recall

F1-Score

Area Under the ROC Curve (AUC-ROC)

Confusion Matrix (to visualize TP, FP, FN, TN)

7. Model Deployment

Real-time Scoring: Integrate the model with UPI transaction systems to analyze transactions in real-time.

Alert System: Trigger alerts or block suspicious transactions immediately.

8. Continuous Learning and Updating

Regularly retrain models with fresh data to capture new fraud patterns.

Implement active learning (where the model queries uncertain examples for human review).

9. Security and Privacy Considerations

Secure model APIs.

Protect user transaction data.

Follow regulatory and compliance requirements (e.g., GDPR, RBI guidelines in India).

5. System Architecture

The architecture includes:

Data Collection and Pre-processing

Feature Extraction (amount, time, frequency, etc.)

Data Balancing using SMOTE

Model Training (XGBoost, Random Forest, SVM)

Real-time Classification and Alerting System

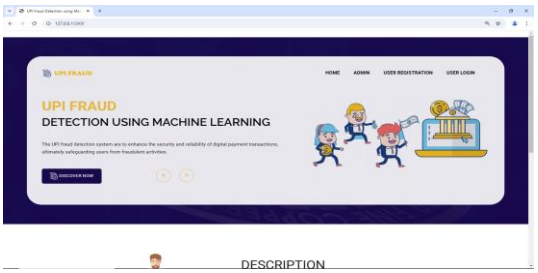
6. Dataset

The dataset comprises labeled UPI transaction records with details like transaction amount, type, time, and user ID. Pre-processing steps include normalization, missing value handling, and categorical encoding.

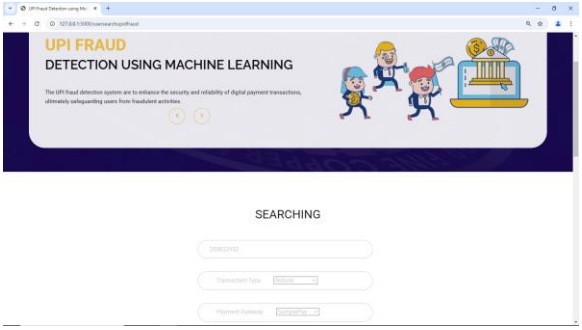
7. Results and Discussion

The system successfully classifies transactions in real time, demonstrating high precision and recall. Performance metrics confirm that XGBoost, when combined with SMOTE, significantly reduces false positives and improves fraud detection rates compared to traditional methods.

Home page



Searching Details



IV. CONCLUSION

This research confirms the efficacy of machine learning in enhancing fraud detection for digital payments. By utilizing SMOTE and XGBoost, the model adapts to evolving fraud patterns and ensures timely prevention, contributing to safer financial ecosystems.

Future Scope:

The future scope for UPI (Unified Payments Interface) fraud detection is vast and evolving, as digital payments continue to grow rapidly in India and around the world. With the increasing volume of transactions, fraud detection systems will need to become more sophisticated and proactive to ensure the security and trust of UPI users. Some potential areas of development and trends for UPI fraud detection include:

1**. AI and Machine Learning Integration**

Predictive Analytics: Machine learning algorithms can analyze transaction patterns in real-time, detect anomalies, and flag potentially fraudulent activities. Predictive models can help identify unusual behavior before it leads to fraud.

Behavioral Biometrics: Using AI to track how users interact with their devices (such as typing speed, swipe patterns, and device orientation) could help detect abnormal activity and potential fraud.

Deep Learning: More advanced neural networks can continuously learn from historical data and improve fraud detection systems over time. These systems would become more adept at distinguishing legitimate transactions from fraudulent ones.

2. Biometric Authentication

Facial Recognition: Integrating facial recognition technology with UPI transactions for more secure authentication could reduce the risk of fraud.

Fingerprint Scanning: Biometrics like fingerprint or retina scans, combined with UPI, can provide a higher level of transaction authentication, ensuring that only the rightful user can authorize payments.

3. Multi-Factor Authentication (MFA)

Incorporating multi-factor authentication (MFA) can further reduce the risk of fraud. For instance, requiring both biometric authentication and a one-time password (OTP) could significantly raise the security bar.

Contextual Authentication: Authentication methods can become dynamic, asking for more security measures based on factors such as the transaction amount, location, or frequency of transactions.

4. Real-Time Monitoring and Analytics

Real-Time Fraud Detection: By using real-time data analysis, UPI platforms can flag potentially fraudulent transactions instantly, limiting the damage before the transaction is completed.

Transaction Velocity Checks: Monitoring the speed and volume of transactions can help detect patterns consistent with fraud. For instance, multiple rapid transactions from the same account in a short period can raise red flags.

5. Cross-Platform Fraud Detection

Fraud detection systems could become more sophisticated by collaborating with other platforms like banks, fintechs, and even third-party services. Shared data across ecosystems could improve the accuracy of fraud detection by understanding a broader context.

This could include leveraging global fraud detection networks that share data across multiple payment systems to prevent cross-platform fraud.

REFERENCES:

1. Yash Patil et al., UPI Fraud Detection Using Machine Learning, IJSREM, 2024.
2. S. Jagadeesan et al., UPI Fraud Detection Using ML, IJARCCE, 2024.
3. Miss Sayalee S. Bodade, Review on UPI Fraud Detection, IJNRD, 2023.
4. Vitthal B. Kamble et al., Enhancing UPI Fraud Detection, IJRASET, 2025.
5. G. D. Clifford et al., "AF classification from a short single lead ECG recording: the PhysioNet/Computing in Cardiology Challenge 2017," in 2017 Computing in Cardiology (CinC), Rennes, France, pp. 1–4.
6. O. Yildirim, U. B. Baloglu, R. S. Tan, E. J. Ciaccio, and U. R. Acharya, "A new approach for arrhythmia classification using deep coded features and LSTM networks," Computer Methods and Programs in Biomedicine, vol. 176, pp. 121–133, 2019.
7. G. B. Moody and R. G. Mark, "The impact of the MIT-BIH Arrhythmia Database," IEEE Eng. Med. Biol. Mag., vol. 20, no. 3, pp. 45–50, 2001.
8. A. Y. Hannun et al., "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," Nature Medicine, vol. 25, no. 1, pp. 65–69, 2019.
9. J. Xie, R. Zhu, Y. Xu, and Y. Zhou, "A robust arrhythmia classification method based on LSTM and wavelet denoising," IEEE Access, vol. 7, pp. 184001–184012, 2019.
10. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, pp. 770–778, 2016.
11. A. A. Alsharabi, H. Wang, and L. Ye, "A deep learning-based approach for ECG signal denoising and arrhythmia classification," Computer Methods and Programs in Biomedicine, vol. 213, p. 106523, 2022.
12. R. K. Tripathy, U. R. Acharya, and D. Bhattacharyya, "Use of features from RR-time series and ECG signal for automated classification of cardiac arrhythmias using decision tree," Journal of Medical Systems, vol. 41, no. 11, pp. 1–13, 2017.
13. D. Banerjee, H. E. Michelis, and R. R. Khanduja, "Real-time arrhythmia detection using machine learning and cloud computing," in 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC), pp. 690–695.
14. H. Chen, J. Wu, and X. Lin, "ECG heartbeat classification based on ensemble deep learning approach," in 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 2715–2719.
15. Z. Zhao, M. Zhang, and Y. Zhou, "ECG feature extraction and classification for arrhythmia detection using wavelet transform and hybrid neural networks," IEEE Access, vol. 7, pp. 104078–104088, 2019.
16. P. de Chazal and R. B. Reilly, "A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features," IEEE Transactions on Biomedical Engineering, vol. 53, no. 12, pp. 2535–2543, 2006.