MANIPAL UNIVERSITY JAIPUR

UPI Spam Detection Using Machine Learning

Synopsis

MCA - IV Sem

Submitted By

Student Name- Akash Saxena

Student Registration- 23FS20MCA00067

Faculty Coordinator

Dr. Vaibhav Bhatnagar

DEPARTMENT OF COMPUTER APPLICATIONS

2025

Introduction

Unified Payments Interface (UPI) has revolutionized digital transactions, making them fast and seamless. However, the increasing adoption of UPI has led to a surge in fraudulent transactions and spam messages. This project aims to build a Machine Learning (ML)-based UPI Spam Detection Model to classify transactions as legitimate or spam/fraudulent based on transaction details, metadata, and behavioural patterns.

Problem Statement

Detecting and preventing UPI-based spam and fraud is challenging due to:

- Anonymity of transactions
- Dynamic nature of fraud patterns
- Sophisticated scam techniques (phishing, social engineering)
- Lack of labelled datasets

Data Collection & Preprocessing

Dataset

The model will be trained using a dataset containing UPI transaction details, including:

- Transaction ID
- Sender & Receiver UPI ID
- Transaction Amount
- Timestamp

- Transaction Message/Text
- Transaction Category (e.g., P2P, Bill Payment, Merchant Payment, Loan, etc.)
- Device & IP Address
- Previous Fraud History

Data Cleaning & Feature Engineering

- Handling missing values
- Removing duplicate transactions
- Feature extraction from text messages (using TF-IDF, Word2Vec, BERT embeddings)
- Encoding categorical variables
- Creating time-based features (transaction frequency, hour of the day, etc.)
- Detecting anomalies in amount & recipient patterns

Machine Learning Models

Model Selection

Several supervised and unsupervised models will be tested:

1. Logistic Regression – Simple baseline classifier.
2. Random Forest – Handles imbalanced data well.
3. Gradient Boosting (XGBoost, LightGBM, CatBoost) – Best for structured data.
4. Neural Networks (LSTMs, Transformers) – For advanced text analysis.
5. Unsupervised Methods (Isolation Forest, DBSCAN, Autoencoders) – For anomaly detection.

Model Training & Evaluation

- Splitting data into training (80%) and testing (20%).
- Using cross-validation to improve generalization.
- Metrics for Evaluation:
    - Precision & Recall (to reduce false positives and negatives)
    - F1 Score
    - ROC-AUC Score
    - Confusion Matrix

Implementation & Deployment

Model Integration

- Develop an API (using Flask/FastAPI) to integrate with UPI systems.
- Real-time transaction monitoring for fraud detection.
- User alerts & flagging system for suspicious transactions.

Model Optimization

- Handling Imbalanced Data: Using SMOTE, weighted loss functions.
- Reducing False Positives: Using ensemble models, fine-tuning thresholds.
- Adversarial Training: Simulating real-world spam patterns for robustness.

Deployment Strategy

- Cloud Deployment (AWS, GCP, Azure) for scalability.
- Edge AI for real-time fraud detection in mobile banking apps.

Challenges & Future Scope

Challenges:

- Adapting to evolving fraud patterns
- Handling adversarial attacks on the model
- Balancing fraud detection & user experience (avoiding unnecessary transaction blocking)

Future Enhancements:

- Federated Learning to improve privacy in fraud detection.
- Explainable AI (XAI) to make fraud detection more interpretable.
- Blockchain-based fraud prevention for enhanced security.

Conclusion

This UPI Spam Detection Model aims to enhance transaction security and reduce financial fraud using advanced ML techniques. By leveraging structured transaction data, NLP-based text analysis, and real-time anomaly detection, the system can proactively identify fraudulent transactions and prevent financial losses.