Fraud Detection in UPI Transaction Using Machine Learning

Madhava J Kamarur[#1], Mr. Mohan H G [#2], Kishan S[#3], Navaneeth Y[#4], Prajwal K S[#5]

[#]*Computer Science and Engineering, Jawaharlal Nehru New College of Engineering, Shimoga.*

[1]mjkamarur06082003@gmail.com

[2]mohan@jnnce.ac.in [3]kishan.s.220803@gmail.com

[4]navaneethy2003@gmail.com [5]ksprajwal2003@gmail.com

ABSTRACT

The use of UPI has grown significantly in recent years. Fraud instances linked to UPI are increasing as it becomes the most widely used payment method for both regular and online purchases. In this research, we use a Convolutional Neural Network (CNN) to model the steps involved in processing UPI transactions and demonstrate how fraud detection may be done with it. Initially, a CNN is trained using a cardholder's typical behaviour. An inbound UPI transaction is deemed fraudulent if the trained CNN does not accept it with a high enough likelihood. At the same time, we work to prevent the rejection of legitimate transactions. To demonstrate the efficacy of our strategy and to contrast it with other methods found in the literature, we provide comprehensive experimental results.

I. INTRODUCTION

Thanks to widespread smartphone use and reasonably priced internet, online banking has become more convenient than old offline methods. People are choosing to use online banking over traditional banking services because of the falling cost of internet and the rise in smartphone usage. The implementation of the unified payment interface system is one example of a government initiative that has contributed to the rise in banking usage. Many customers have been encouraged to use remote banking by programs such as these. The security of banking has been exposed to several attacks as e-banking has grown, and it has been established that there is a significant vulnerability in online banking security. These days, two factor authentication is used for verification in the majority of banking applications [19]. Numerous studies have been conducted thus far to secure banks and banking services. For secure banking services, academics are suggesting blockchain, cryptography, biometrics, and secure transmission mechanisms. One essential component of civilisation is a bank. It is a financial organisation that encourages lending and deposits funds donated by the general population. The nation's economic growth and stability are significantly influenced by banks. It is a centralised system under state-specific government regulation.

II. METHODOLOGY

In order to prevent financial losses in real time, this methodology focusses on applying machine learning to detect fraud in UPI transactions. Data collection from synthetic and historical transaction records, including aspects like device information, user activity, and transaction details, is the first step in the process. While preprocessing entails cleaning and standardising the data, exploratory data analysis, or EDA, assists in identifying trends and connections. By adding risk and behavioural variables and then picking the most influential ones, feature

engineering improves model performance. Numerous methods are assessed, with hyperparameter adjustment guaranteeing peak performance, including Random Forest, Gradient Boosting, and Logistic Regression. Techniques like SMOTE are employed to correct class imbalance, and the model is evaluated using ROC-AUC, accuracy, precision, and recall measures. In order to manage changing fraud patterns, the system is implemented as a real- time microservice with ongoing upgrades and monitoring. In this strong architecture, scalability and dependability are guaranteed by tools like Flask, Scikit-learn, and Python.

1. System Architecture for Machine Learning-Based Fraud Detection in UPI Transactions A number of interconnected layers make up the suggested system architecture for identifying fraud in UPI transactions, which makes data processing, analysis, and decision-making smooth. By compiling transaction data from many sources, including banks, UPI payment gateways, and publicly accessible datasets, the Data Collection Layer acts as the framework. Important characteristics are recorded by this layer, including metadata (geolocation, device kind, and transaction mode), transaction information (amount, sender/receiver IDs, and timestamps), and user behaviour data (frequency and timing patterns).Data security and integrity are guaranteed by secure database connections and APIs. The Data Preprocessing Layer transforms the gathered raw data to guarantee its quality and analysis-suitability. This step entails normalising numerical data, extracting important properties like transaction intervals and high-value trends, and cleaning the data to handle noise and missing values. ensuring consistency, and employing methods such as One-Hot Encoding to encode categorical data into machine-readable representations.

2. Module for Feature Extraction By extracting important characteristics from transaction data, the feature extraction module is essential to the detection of fraudulent activity. These characteristics include sender-receiver relationships, transaction volume, frequency, device type, location, and temporal patterns like activity spikes and transaction intervals. Raw data is transformed into organised, usable input for the machine learning model using sophisticated statistical and data processing techniques. The system may successfully identify minute patterns that differentiate authentic transactions from fraudulent ones by concentrating on these extracted elements.

III. IMPLEMENTATION

Implementation of Fraud Detection in UPI Transactions Using Machine Learning

Application of Machine Learning for Fraud Detection in UPI Transactions The fraud detection system is implemented in a number of clearly defined steps, each of which is intended to guarantee the solution's efficacy and stability. A thorough description of the implementation procedure may be found below.

1. Gathering and Preparing Data Gathering an extensive dataset of UPI transactions, including both authentic and fraudulent examples, is the first stage in putting the system into place. Secure sources like banks, payment gateways, or publicly accessible datasets are used to gather transaction data. Preprocessing the data entails resolving missing information, deleting duplicates, and standardizing fields such as timestamps, transaction amounts, and user details.

2. Feature engineering and selection Feature engineering is the process of finding and extracting relevant attributes from the raw data, such as transaction amount, time, sender-receiver relationships, transaction frequency, location, and device type; temporal features, such as the intervals between transactions and unusual activity spikes, are also calculated

3. Model training and evaluation The foundation of the fraud detection system is machine learning algorithms, such as Random Forests, Gradient Boosting Machines (e.g., XGBoost, LightGBM), and Neural Networks. This preprocessed dataset is used to train networks. The data is divided into training, validation, and test sets during the training phase. To maximize the model's performance, hyperparameter tuning is done utilizing strategies like Grid Search and Random Search. The ability of the final model to generalize over unseen data is ensured by selecting it based on evaluation measures.

4. A system for detecting fraud in real time The UPI system incorporates the trained model to detect fraud in real time. To connect the model to the transaction processing system, a REST API is created. Upon initiating a transaction, the API retrieves pertinent information and inputs them into the model, which generates a probability score that indicates the possibility of fraud.

5. Development of User Interfaces To make using the fraud detection system easier, an intuitive user interface has been developed. The interface offers dashboards and visualizations to show performance indicators, suspicious activity patterns, and transactions that have been highlighted. To enhance the model, users—such as fraud analysts or payment processors—can modify detection criteria, examine transactions that have been identified, and offer input. Advanced search and filtering features enable in-depth analysis of particular cases.

6. Assessment and Examination Using criteria like accuracy, precision, recall, and F1- score, the model's performance is thoroughly assessed. Testing on various datasets is part of the evaluation process to gauge the system's generalization and resilience. To make sure the system can manage large transaction volumes without experiencing performance deterioration, stress testing is carried out.

7. Enhancement of Performance Optimizing the system for speed and scalability is necessary for real-time fraud detection. To improve computing performance, methods like hardware acceleration, quantization, and model pruning are used.

8. Privacy and Ethical Considerations: To guarantee adherence to moral and legal requirements, privacy-preserving methods including data encryption, anonymization, and safe storage are used. The technology protects user data and transaction details by

adhering to laws like the GDPR and India's Data Protection Bill. To make sure the system stays impartial and complies with ethical standards, audits are carried out on a regular basis.

9. Implementation Deploying the fraud detection system in an active UPI environment is the last step. To track system performance and adjust to new fraud trends, ongoing monitoring is put in place. The model is regularly retrained and updated using user feedback and fresh data from actual transactions to make sure it remains effective against changing threats.

IV. CONCLUSION

UPI fraud has grown to be a major worldwide issue, resulting in large financial losses and forcing UPI firms to spend money creating methods to identify and stop fraud. The goal of this research is to create algorithms that are effective, economical, and flexible in order to detect fraudulent transactions more quickly and accurately. To do this, a variety of machine learning techniques have been compared. The machine learning-based fraud detection system for UPI transactions offers a reliable and expandable way to deal with the growing risks of fraudulent activity in digital payment systems. To guarantee safe and easy transactions, this project skilfully combines machine learning algorithms, real-time processing capabilities, and advanced data analytics.

V. FUTURE SCOPE

1. Advanced Feature Engineering: To enable precise anomaly identification, the system gathers important features such transaction amount, frequency, temporal patterns, and sender-receiver relationships.

2. Machine Learning techniques: For reliable fraud detection, it makes use of cutting- edge techniques like Random Forests, Gradient Boosting Machines, and Neural Networks.

3. Real-Time Detection: by processing transactions in real-time, the system guarantees prompt detection and reaction to questionable activity.

4. Customisable Thresholds: To provide flexibility and adaptability, users can modify detection thresholds according to risk levels.

5. User-Friendly Interface: Stakeholders can examine trends, visualise flagged transactions, and offer comments via a dynamic dashboard.

REFERENCES

[1] Shaymaa Abdulla Al-Delayel, "Security Analysis of Mobile Banking Application in Qatar",2022.

[2] S.H. Projects and W. Lovo , ―JMU Scholarly Commons Detecting UPI fraud : An analysis of fraud detection techniques, ‖ 2020

[3] Tsai, C.; Su, P. The application of multi-server authentication scheme in internet banking transaction environments. Inf. Syst. e-Bus. Manag. 2021, 19, 77–105.

[4] Hammi,B.; Zeadally, S.; Adja, Y.C.E.; Giudice, M.D.; Nebhen, J. Blockchain-based solution for detecting and preventing fake check scams. IEEE Trans. Eng. Manag. 2022, 69, 3710–3725.

[5] AbdulRani, M.I.; Syed Mustapha Nazri, S.N.F.; Zolkaflil, S. A systematic literature review of money mule: Its roles, recruitment and awareness. J. Financ. Crime 2023. ahead-of-print.

[6] Ileberi, E.; Sun, Y.; Wang, Z. A machine learning based credit card fraud detection using the GA algorithm for feature selection. J.Big Data 2022, 9, 24.

[7] Zimba, A. ABayesian Attack-Network Modeling Approach to Mitigating Malware-Based Banking Cyberattacks. Int. J. Comput. Netw. Inf. Secur. (IJCNIS) 2022, 14, 25–39.

[8] Cauteruccio, F.; Terracina, G.; Ursino, D. Generalizing identity-based string comparison metrics: Framework and techniques. Knowl.-Based Syst. 2020, 187, 104820