Machine Learning Based Upi Fraud Detection System

Dr.J.Jensy Rajakumari[1], Ms. K. Jeyapriya[2]
[1]*Professor, Department. of CSE, Jayaraj Annapackiam CSI College of Engineering, Nazareth, India*
[2]*PG Scholar, Department. of CSE, Jayaraj Annapackiam, CSI College of Engineering, Nazareth, India*

***Abstract*—In recent years, online payment methods such as UPI, mobile banking, and digital wallets have become very popular because they make transactions faster and more convenient. At the same time, the number of fraud cases, including phishing, unauthorized access, and suspicious transactions, has also increased.**

This project focuses on developing a machine learning-based system to detect fraudulent UPI transactions. In our approach, we use machine learning algorithms like Random Forest, XGBoost, and Logistic Regression to classify whether a transaction is genuine or fraudulent.

Before training the model, the transaction data is processed through several steps such as data cleaning, normalization, and feature engineering. We also apply anomaly detection techniques to identify unusual patterns. The performance of the model is evaluated using metrics like accuracy, precision, recall, F1-score, and ROC-AUC.

The proposed system can detect fraud in real-time and provides a scalable and efficient solution to improve the security of digital payments.

***Index Terms*—UPI Fraud Detection, Machine Learning, Online Payments, Cyber Security, Random Forest, XGBoost**

I. INTRODUCTION

This project presents a system that uses machine learning techniques to detect fraudulent transactions in UPI and other online payment platforms. The main goal of this system is to provide a secure, automated, and real-time solution for fraud detection.

Nowadays, digital transactions are increasing rapidly, and along with that, fraud activities such as OTP scams and phishing attacks are also becoming more common. To address this problem, our system analyzes transaction details like amount, time, and location to identify unusual patterns and detect fraud.

II. LITERATURE REVIEW

The proposed approach in this project focuses on developing an effective machine learning-based system for detecting fraudulent transactions in UPI and other online payment platforms. With the rapid growth of digital payment systems such as mobile banking, UPI, and digital wallets, financial transactions have become faster and more convenient. However, this convenience has also led to an increase in fraudulent activities such as phishing attacks, unauthorized transactions, and identity theft. To address these challenges, this study introduces a data-driven solution that analyzes transaction patterns and identifies anomalies using machine learning techniques. The system utilizes algorithms such as Random Forest, XGBoost, and Logistic Regression to classify transactions as either legitimate or fraudulent. Among these models, Random Forest is given priority due to its robustness, ability to handle large datasets, and improved accuracy in classification tasks.

In this study, transaction data is carefully processed through multiple stages, including data cleaning, normalization, and feature engineering, to ensure high-quality input for the model. Important features such as transaction amount, time, location, device information, and user behavior are considered for detecting unusual patterns. The dataset is divided into training and testing sets to evaluate the performance of the model effectively. Various evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC are used to measure the efficiency of the system. These metrics help in understanding how well the model can detect fraudulent transactions while minimizing false positives.

III. METHODOLOGY

The proposed methodology focuses on developing a machine learning-based system to detect fraudulent transactions in UPI and online payment platforms through a structured and automated process. Initially, transaction data is collected from reliable sources such as datasets or real-time systems. This data includes important attributes like transaction ID, user ID, amount, time, location, device information, and transaction status. Since raw data may contain noise, missing values, or inconsistencies, it undergoes preprocessing steps such as data cleaning, normalization, and encoding to improve data quality and ensure consistency.

During the training phase, the model learns relationships between input features and fraud patterns. Once trained, the model is evaluated using performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to measure its effectiveness. Based on these evaluation results, the model is further optimized to improve its prediction capability. The final system is capable of assigning a fraud probability score to each transaction, enabling real-time fraud detection. Due to its scalability, efficiency, and ability to handle large volumes of transaction data, the proposed method provides a reliable solution for enhancing security in digital payment systems.

3.1. Proposed Method

The proposed method involves analyzing transaction data using machine learning techniques to identify fraudulent activities effectively. Initially, the preprocessed transaction data is fed into the model, where important features such as transaction amount, time, location, and user behavior are considered.

The learning process involves training the model on historical transaction data, where it learns to differentiate between normal and fraudulent transactions. The trained model generates probability scores for each transaction, indicating the likelihood of fraud. Based on these scores, the system classifies transactions accordingly. This approach ensures high accuracy and efficiency in fraud detection while maintaining interpretability and scalability for real-world applications.

3.2. Algorithm Used

The selection of appropriate machine learning algorithms plays a crucial role in the effectiveness of the fraud detection system. In this project, algorithms such as Random Forest, XGBoost, and Logistic Regression are used to achieve accurate and reliable predictions. Among these, Random Forest is selected as the primary model due to its ability to handle large datasets, reduce overfitting, and provide high accuracy.

Random Forest works by creating multiple decision trees and combining their outputs to make a final prediction, which improves overall performance. XGBoost is also used because of its efficiency and ability to handle complex relationships in data, while Logistic Regression provides a simple and interpretable baseline model. The combination of these algorithms helps in improving the robustness and reliability of the system, ensuring better detection of fraudulent transactions.

3.3. Dataset

The dataset used in this project consists of transaction records that include various attributes such as transaction ID, user ID, amount, time, location, device ID, merchant details, and fraud status. These records represent both legitimate and fraudulent transactions, allowing the model to learn patterns associated with each category. The dataset is carefully prepared to ensure consistency and reliability during model training.

Before training, the data undergoes preprocessing steps such as removing missing values, eliminating duplicates, and normalizing numerical features. The dataset is then divided into training and testing sets, typically in an 80:20 ratio, to evaluate the performance of the model effectively. To improve model accuracy and prevent overfitting, techniques such as data balancing and feature selection are applied. This ensures that the model performs well even when dealing with imbalanced data, where fraudulent transactions are relatively rare compared to normal ones.

3.4. Preprocessing

Data preprocessing is an essential step in the fraud detection system, as it ensures that the input data is

clean, consistent, and suitable for machine learning algorithms. Raw transaction data often contains missing values, noise, and inconsistencies that can affect model performance. To address these issues, preprocessing techniques such as data cleaning, normalization, and encoding are applied.

Missing values are handled by either removing incomplete records or filling them with appropriate values. Duplicate entries are removed to avoid redundancy, and numerical features are normalized to bring them to a common scale. Categorical variables, such as transaction type or location, are converted into numerical format using encoding techniques. These preprocessing steps help improve the accuracy and efficiency of the model by providing high-quality input data for feature extraction and classification.

IV. IMPLEMENTATION

The implementation phase focuses on developing a practical system for detecting fraudulent transactions in UPI and online payment platforms using machine learning techniques. In this stage, all the modules designed in the methodology, including data preprocessing, feature extraction, model training, prediction, and result visualization, are integrated into a complete working application. The system begins with a data input module, where transaction data is collected either from a dataset or through a user interface. The input data is first verified for completeness and correctness before being processed further.

The processed data is then used to train machine learning models such as Random Forest, XGBoost, and Logistic Regression. During training, the models learn the relationship between transaction features and fraud patterns. After training, the models are evaluated using performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to measure their effectiveness. Based on these evaluation results, model tuning and optimization techniques are applied to improve performance.

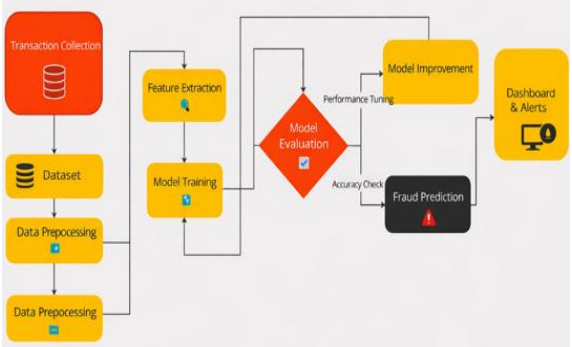

Figure 1: System Architecture for UPI Fraud Detection

V. CONCLUSION

The proposed project, "Machine Learning Based UPI Fraud Detection System," successfully develops an automated system capable of identifying fraudulent transactions in digital payment platforms. By using machine learning algorithms such as Random Forest, XGBoost, and Logistic Regression, the system is able to analyze transaction patterns and detect suspicious activities with high accuracy. Among these models, Random Forest provides better performance due to its robustness and ability to handle complex data.

Furthermore, the implementation of a real-time prediction system with a user-friendly interface makes the solution practical and efficient for modern financial environments. Overall, this project highlights the importance of machine learning in enhancing cybersecurity and reducing financial fraud in digital transactions. The proposed system provides a scalable, accurate, and efficient solution that can be further extended with advanced techniques to handle evolving fraud patterns in the future.

REFERENCES

[1] Kumar V., Patil V., Ingle D.R. (2021), "A Novel Approach for Fraud Detection Using Machine Learning Techniques," International Conference on Emerging Technologies.
[2] Singh R. et al. (2019), "Machine Learning Approaches for Fraud Detection in Financial Transactions," Journal of Cyber Security.
[3] Wang Y. et al. (2022), "Online Payment Fraud Detection Using Data Mining Techniques," IEEE Conference Proceedings.

[4] Arefinia A., Geethanjali S.G. (2023), "A Survey on Fraud Detection Using Machine Learning," International Journal of Computer Science.
[5] Lalinia P. et al. (2024), "Advanced Fraud Detection Techniques in Digital Payments," Journal of Artificial Intelligence Research.
[6] Scikit-learn Documentation
[7] XGBoost Documentation