Fraud Detection in Financial Transaction using Machine Learning

Mahinda [#1], Usha Devi[*2],

[#]*Department of Computer Science, Rathinam College of Arts and Science*

(Autonomous), Coimbatore, Tamilnadu, India

mahinda2006r@gmail.com, *usha.devi145@gmail.com*

Abstract *- The increasing use of digital financial transactions has led to a rise in fraudulent activities, making effective fraud detection essential. This project proposes a machine learning-based system to detect fraudulent transactions in real time. It uses algorithms such as Random Forest, Gradient Boosting, and Logistic Regression to analyze transaction data and classify them as legitimate or fraudulent. The system is implemented as a web application using FastAPI and React, enabling efficient processing and user interaction. It also incorporates Explainable AI (SHAP) to provide transparency in predictions. The proposed system improves accuracy, reduces manual effort, and adapts to evolving fraud patterns, making it a reliable and scalable solution for modern financial systems.*

Keywords *- Fraud Detection, Machine Learning, Financial Transactions, Random Forest, Gradient Boosting, Logistic Regression, Real-Time Detection, Explainable AI, SHAP, Data Preprocessing, Anomaly Detection, Web Application*

1. INTRODUCTION

The rapid growth of digital banking, online payments, and e-commerce has made financial transactions faster and more convenient. At the same time, it has increased the risk of fraudulent activities such as unauthorized payments, credit card fraud, identity theft, and money laundering. Traditional fraud detection methods based on manual checking and fixed rules are often slow, costly, and less effective in identifying new fraud patterns. Therefore, there is a need for an intelligent system that can detect fraud quickly and accurately.

The project titled Fraud Detection in Financial Transaction Using ML focuses on using machine learning techniques to identify suspicious transactions from large volumes of financial data. ML models learn patterns from previous transaction records and classify whether a transaction is genuine or fraudulent based on factors such as amount, time, location, and user behavior. This system helps banks and financial organizations reduce financial losses, improve customer trust, and enhance the overall security of online transactions.

2. RELATED WORKS

Many earlier fraud detection systems used rule-based methods in which predefined rules were applied to identify suspicious financial transactions. These systems were simple and easy to implement, but they often failed to detect new fraud techniques and produced high false alarms. To improve accuracy, researchers introduced machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and Naive Bayes. These models learn patterns from historical transaction data and classify transactions as genuine or fraudulent more effectively than traditional methods.

Recent studies focus on advanced techniques such as ensemble learning, deep learning, and anomaly detection for better fraud identification. Algorithms like Random Forest,

XGBoost, and Neural Networks have shown high performance in detecting complex fraud patterns. Real-time fraud detection systems are also being developed to analyze transactions instantly and prevent fraud before completion. Hybrid approaches that combine rule-based systems with machine learning models are widely used to achieve better accuracy, faster response time, and reduced false positive rates.

3.System Design

3.1.Data Collection
The system begins by collecting financial transaction data from various sources such as user input forms, databases, or CSV files. The collected data includes details like transaction ID, amount, sender and receiver information, time, location, and device type. This data serves as the foundation for fraud detection.

3.2.Data Preprocessing
In this step, the collected data is cleaned and prepared for analysis. Missing values are handled, outliers are removed, and categorical data is converted into numerical form. Feature scaling and normalization are applied to ensure consistency. This step improves data quality and enhances model performance.

3.3.Model Training
The preprocessed data is used to train machine learning models such as Random Forest, Gradient Boosting, and Logistic Regression. The models learn patterns and relationships between transaction features to distinguish between legitimate and fraudulent transactions.

3.4.Model Evaluation
The trained models are evaluated using metrics such as accuracy, precision, recall, and F1-score. This step ensures that the model performs well and can reliably detect fraud while minimizing false positives and false negatives.

3.5.Backend Development
The backend is developed using FastAPI, which handles API requests and integrates the machine learning model. It processes incoming transaction data, sends it to the model for prediction, and returns the results to the frontend.

3.6.Database Design
A database is created to store transaction data, user details, and prediction results. Tables are designed with proper relationships to ensure efficient storage, retrieval, and security of data.

3.7.Frontend Development
The frontend is developed using React to provide a user-friendly interface. Users can input transaction details, view fraud detection results, and access dashboards and reports. The interface is designed to be interactive and easy to use.

3.8.Real-Time Prediction
When a new transaction is entered, it is sent to the backend through APIs. The machine learning model processes the data in real time and classifies it as legitimate or fraudulent, along with a risk score.

3.9.Alert and Notification System
If a transaction is detected as suspicious, the system generates alerts and notifications. These alerts help users or administrators take immediate action to prevent fraud.

3.10.Deployment and Maintenance
Finally, the system is deployed on a local server or cloud platform. Regular monitoring, updates, and maintenance are performed to ensure system performance, security, and accuracy over time.

3.11.Algoritnm Selection
The selection of appropriate algorithms is a crucial step in developing an effective fraud detection system. Since financial transaction data is mostly structured and involves binary classification (fraud or legitimate), machine learning algorithms that perform well on classification tasks and can handle large datasets are chosen. The selected algorithms for this project include Random Forest, Gradient Boosting, and Logistic Regression, as they

provide high accuracy, efficiency, and reliability in detecting fraudulent transactions.

3.11.1. Random Forest

Random Forest is chosen because it is an ensemble learning algorithm that builds multiple decision trees and combines their outputs to improve prediction accuracy. It is highly effective in handling large datasets, reducing overfitting, and managing imbalanced data. This makes it suitable for fraud detection, where fraudulent transactions are rare compared to legitimate ones.

3.11.2.Gradient Boosting

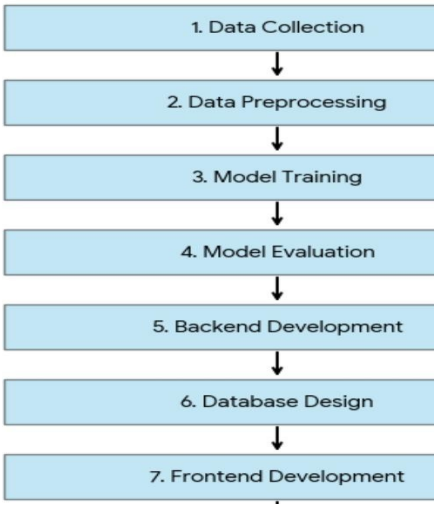
Gradient Boosting is selected due to its ability to improve model performance by iteratively correcting errors made by previous models. It is particularly useful for detecting complex and subtle fraud patterns that may not be captured by simpler algorithms. Its high accuracy and ability to reduce both false positives and false negatives make it a strong choice for this system.

3.11.3.Logistic Regression

Logistic Regression is included as a baseline model because of its simplicity, speed, and interpretability. It provides probability-based outputs, making it easy to understand how different features influence fraud detection. Although it is less complex than ensemble methods, it performs well for linearly separable data and helps in comparing model performance.

4. METHODOLOGY

The methodology of the fraud detection system begins with data collection and preprocessing. Transaction data is gathered from sources such as databases, CSV files, or user inputs, including attributes like transaction amount, time, location, and user details. The collected data is then cleaned and prepared by handling missing values, removing outliers, encoding categorical variables, and normalizing numerical features. Since fraud datasets are typically imbalanced, resampling techniques are applied to ensure balanced learning. The processed data is then split into training and testing sets for model development. In the next phase, machine learning models such as Random Forest, Gradient Boosting, and Logistic Regression are trained using the prepared data to identify patterns of fraudulent behavior. The models are evaluated using performance metrics like accuracy, precision, recall, and F1-score to ensure reliability. Once validated, the model is integrated into the system for real time prediction through a FastAPI backend and a React-based frontend. The system classifies transactions and generates alerts for suspicious activities, while continuous monitoring and updates ensure improved performance and adaptability to new fraud patterns.O



5. CONCLUSION

The project Fraud Detection in Financial Transaction Using ML successfully demonstrates how machine learning can be used to identify fraudulent activities quickly and accurately in financial systems. By analyzing transaction patterns and customer behavior, the system can classify transactions as genuine or suspicious, helping organizations reduce financial losses and improve security. Compared to traditional rule-based methods, the proposed approach offers better accuracy, faster detection, and adaptability to new fraud patterns. This project highlights the importance of intelligent technologies in creating

safer and more reliable digital payment environments.

6. ACKNOWLEDGEMENT

This project titled **Fraud Detection in Financial Transaction Using ML** is the outcome of research work carried out in the Department of Computer Science under the guidance of faculty members. The authors are grateful to the Department of Computer Science for providing the necessary facilities, valuable guidance, and continuous support throughout the project work. The authors also express sincere thanks to all staff members, friends, and well-wishers who directly or indirectly helped in the successful completion of this project.

7. REFERENCES

1.V. Bhusari and S. Patil, "Study of Machine Learning Algorithms for Credit Card Fraud Detection," International Journal of Computer Applications, vol. 182, no. 44, pp. 1–5, 2019.

2. A. Dal Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, "Calibrating Probability with Undersampling for Unbalanced Classification," in Proceedings of IEEE Symposium on Computational Intelligence and Data Mining, 2015.

3. S. Jurgovsky et al., "Sequence Classification for Credit Card Fraud Detection," Expert Systems with Applications, vol. 100, pp. 234–245, 2018.

4. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, 2016.

5. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.

6. L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

7. F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

8. S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in Advances in Neural Information Processing Systems (NeurIPS), 2017.

9. J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed., Morgan Kaufmann, 2012.

10. A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, O'Reilly Media, 2019.