E-Commerce Fraud Detection Using Machine Learning

Prof. V.A Bijalpure[*1], Shravani Jadhav[*2], Neharika Deore[*3], Joveriya Kazi[*4], Ashra Shaikh[*5]

[*1]Prof, Department Of Computer Technology, K.K. Wagh Polytechnic, Nashik, Maharashtra, India.

[*2,3,4,5]Student, Department Of Computer Technology, K.K. Wagh Polytechnic, Nashik, Maharashtra, India.

ABSTRACT

In the rapidly growing landscape of online shopping, e-commerce platforms face a significant

rise in fraudulent activities such as fake orders, stolen credit card usage, and account takeovers. These fraudulent transactions not only lead to substantial financial losses for companies but also damage customer trust and brand reputation. This project, "E-Commerce Fraud Detection Based on Machine Learning", aims to address these challenges by implementing a machine learning–based solution to accurately detect and prevent fraudulent transactions. The proposed system analyzes historical transaction data to identify behavioral patterns and anomalies that indicate potential fraud. Key transaction features—such as order amount, payment method, IP address, location, time of purchase, and device type are processed and evaluated using Python and libraries. Multiple algorithms, including Logistic Regression, Decision Tree, Random Forest, and XGBoost, are applied and compared to determine the most accurate and efficient model for fraud detection.The system classifies transactions as genuine or suspicious and can generate alerts for potentially fraudulent activities, enabling businesses to take immediate action. This solution is adaptable for e-commerce platforms like Amazon and Flipkart, payment gateways such as Paytm and Razorpay, and subscription services like Netflix. In the future, the system can be enhanced with real-time detection, dashboard integration, and deep learning techniques to further improve accuracy and efficiency, thereby safeguarding both businesses and customers.

Keywords E-commerce, Fraud detection, Machine learning, Online transactions, Anomaly detection, Transaction analysis, Cybersecurity, Python, Scikit-learn, Random Forest, Logistic Regression, Decision Tree, XGBoost, Customer behavior analysis, Fake orders, Credit card fraud, Account takeover, Data analytics, Predictive modeling

I. INTRODUCTION

The rise of e-commerce has transformed shopping with convenience and variety but also

increased online fraud like fake orders, stolen cards, and identity theft. These issues cause major financial and trust losses for businesses and customers. Traditional rule-based detection methods are no longer effective against evolving cyber threats.

Machine Learning (ML) offers a powerful solution by analyzing transaction data to detect

unusual patterns in real time. The project **"E-Commerce Fraud Detection Based on Machine Learning"** aims to build a Python-based system using tools like pandas, scikit-learn, and XGBoost to classify transactions as genuine or fraudulent. Features such as order value, IP address, payment method, and device type are analyzed to flag suspicious activity.

This system can be integrated into e-commerce platforms and payment gateways, enhancing

security and customer trust. Future improvements may include real-time detection, visual

dashboards, and deep learning models for even greater accuracy — ensuring a safer digital

shopping experience.

Goals

1. To develop a machine learning–based system capable of detecting fraudulent e-commerce transactions with high accuracy.

2. To analyze transaction features such as order amount, IP address, location, device type, payment method, and time of purchase for identifying suspicious patterns.

3. To compare and evaluate multiple ML algorithms like Logistic Regression, Decision Tree, Random Forest, and XGBoost to determine the most effective model for fraud detection.

4. To implement an alert mechanism that notifies the concerned authority when a transaction is predicted to be fraudulent.

5. To create a scalable and adaptable fraud detection solution that can be integrated into various e-commerce platforms, payment gateways, and subscription-based services.

II. METHODOLOGY

1. Data Collection Module

- Responsible for gathering historical transaction data containing both genuine and fraudulent records.
- Data includes features such as order amount, payment method, IP address, location, device type, and time of purchase.

2. Data Preprocessing Module

- Cleans and prepares the dataset by handling missing values, removing duplicates, and encoding categorical variables.
- Performs feature scaling to normalize the data and ensure better model performance.

3. Model Training and Testing Module

- Implements multiple machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, and XGBoost.
- Splits the dataset into training and testing sets to build and evaluate the models.

4. Model Evaluation Module

- Compares the performance of all models using metrics like accuracy, precision, recall, F1-score, and confusion matrix.
- Selects the most suitable algorithm for fraud detection.

5. Fraud Detection and Alert Module

- Uses the selected model to classify new transactions as genuine or fraudulent.
- Generates alerts when a transaction is flagged as suspicious, enabling timely preventive action.

6. **Integration and Future Enhancement Module**

- Ensures compatibility with e-commerce platforms, payment gateways, and subscription-based services.

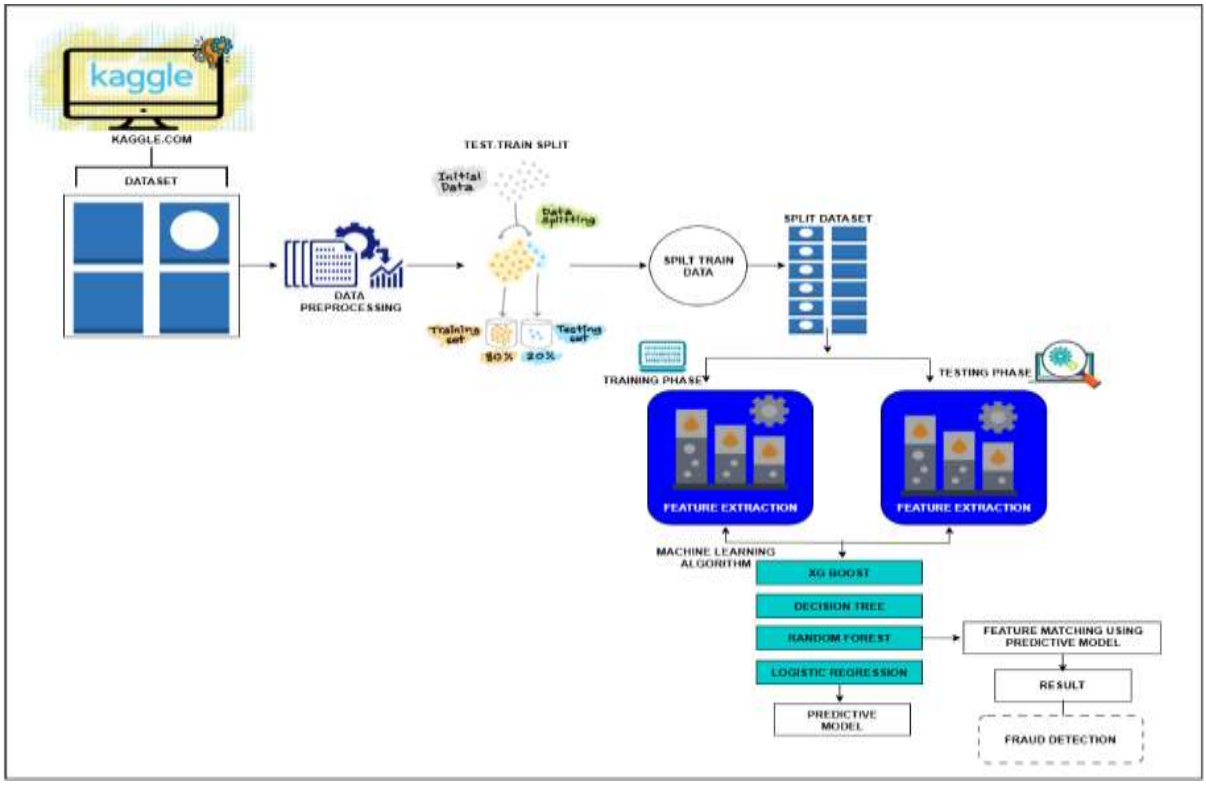
Provides scope for real-time fraud detection, dashboard integration, and deep learning–based improvements.

IV. SYSTEM ARCHITECTURE

The figure illustrates the overall architecture of the proposed fraud detection system. The process begins with dataset acquisition from a publicly available source, followed by data preprocessing to clean, transform, and normalize the raw transactional data.

The pre-processed dataset is then divided into training and testing sets using a standard train–test split strategy (typically 80% for training and 20% for testing). During the training phase, relevant feature extraction techniques are applied to identify significant attributes influencing fraudulent behaviour.

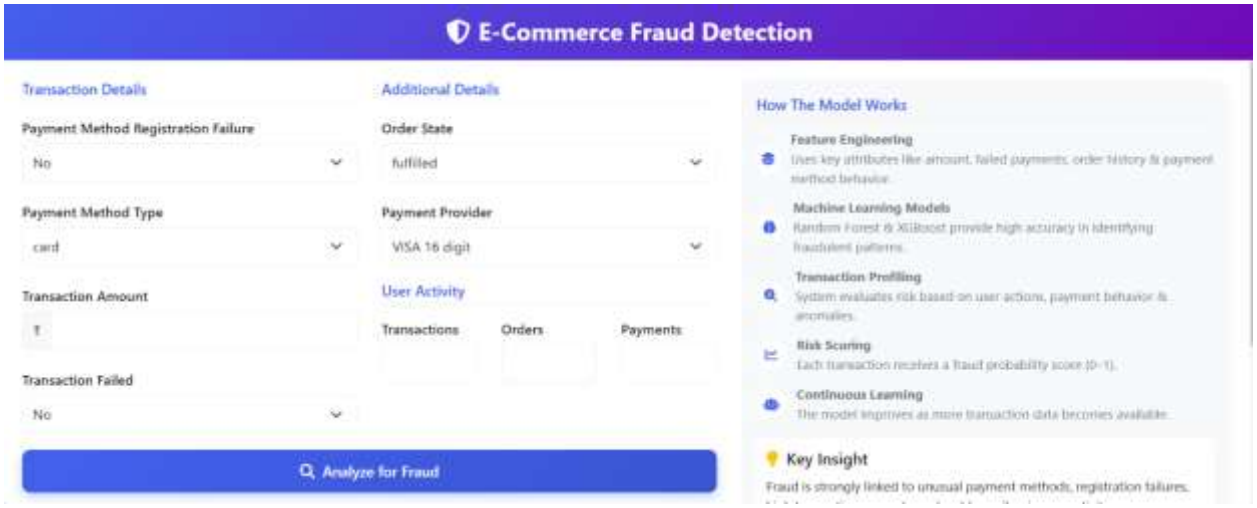
Multiple machine learning algorithms, including XGBoost, Decision Tree, Random Forest, and Logistic Regression, are



implemented to build predictive models. These models are trained on the training dataset and evaluated using the testing dataset.

In the testing phase, the trained model performs feature matching and classification to determine whether a transaction is legitimate or fraudulent. The final output of the system is a fraud detection decision, enabling timely identification of suspicious activities.

Filling details about fraud



After Response

V. RESULTS AND DISCUSSION

The proposed E-Commerce Fraud Detection System was implemented and tested using a labeled transaction dataset. The dataset was preprocessed and divided into training and testing sets in an 80:20 ratio to ensure fair model evaluation. Logistic Regression, Decision Tree, Random Forest, and XGBoost algorithms were trained and compared.

A. Performance Evaluation

The models were evaluated using standard metrics such as Accuracy, Precision, Recall, and F1-Score. The results show that ensemble methods, especially Random Forest and XGBoost, performed better than individual models. Higher recall values were achieved, which is important in fraud detection to reduce missed fraudulent transactions.

B. Feature Analysis

Important transaction features such as transaction amount, payment method, device type, IP address, time, and location played a key role in identifying fraud. Unusual spending patterns and location mismatches were found to be strong indicators of fraudulent activity.

C. System Effectiveness

The system was able to classify transactions in near real-time with minimal delay. Suspicious transactions were successfully flagged, allowing quick preventive actions. Compared to traditional rule-based systems, the proposed machine learning approach adapts better to changing fraud patterns.

D. Discussion

The results indicate that the proposed system is effective and scalable for real-world e-commerce applications. However, performance may be affected by imbalanced data and new fraud techniques. Future improvements may include advanced deep learning models and real-time data processing to further enhance detection accuracy.

VI. CONCLUSION

In conclusion,the project "E-Commerce Fraud Detection Based on Machine Learning" successfully demonstrates how advanced data analysis and predictive modeling can be applied to enhance the security of online transactions. By leveraging algorithms such as Logistic Regression, Decision Tree, Random Forest, and XGBoost, the system can effectively identify suspicious patterns and classify transactions as genuine or fraudulent. This approach not only helps businesses prevent financial losses but also builds customer trust by ensuring a secure shopping experience. While the current model provides promising results, its performance can be further improved with real-time detection, deep learning integration, and larger datasets. Overall, the system offers a scalable, adaptable, and practical solution for tackling fraud in the rapidly growing e-commerce industry.

VII. REFERENCES

1. **Monteith, M. et al. (2021).** *Increasing cybercrime since the pandemic: Concerns for psychiatry.* — Because it shows the broader social and psychological consequences of increased cybercrime since COVID-19, framing why fraud detection matters beyond pure technical or financial loss. PMC+1

2. **Kodate, S. et al. (2020).** *Detecting problematic transactions in a consumer-to-consumer e-commerce network.* — Very relevant for marketplace / peer-to-peer settings; uses network / graph techniques to spot anomalies.

3. **Ali, A. et al. (2022).** *Financial fraud detection based on machine learning: A systematic literature review.* — Helps you understand which ML methods (supervised, unsupervised, hybrid) perform well in fraud detection.