International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 441-447, May-2026

https://doi.org/10.47001/IRJIET/2026.105061

E-Commerce Fraud Detection System Using Machine Learning

[1]Ajahar Pathan, [2]Mayur More, [3]Bhavesh Patel, [4]Milind Dikshit, [5]Tejas Marathe

[1,2,3,4,5]Department of Computer Engineering, P.S.G.V.P. Mandal's D.N. Patel College of Engineering, Shahada, Maharashtra, India

E-mail: [1]pathan.ajaharkha@gmail.com, [2]moremayurbhoi@gmail.com, [3]bhaveshpatelsharad@gmail.com, [4]04dmilind@gmail.com, [5]tejasmarathe43@gmail.com

Abstract **- Payment fraud is a problem when people use someone else credentials to buy things online. This can cause a lot of trouble for merchants and consumers. To stop this from happening a system that uses machine learning is being suggested. This system looks at transaction data as it happens to find fraud. The system works by looking at the details of each transaction that is sent to the admin panel. It then matches these details to see if they are legitimate or not. The system uses a few models to make predictions, including Logistic Regression, Random Forest, Support Vector Machine and XGBoost. When getting the data ready the system looks at things like how much money's being spent what time it is, where the transaction is happening and how many transactions are happening at the same time. The system is trained using a mix of real transactions, which helps it get really good at telling the difference between the two. The results show that XGBoost is the model with 96.20% accuracy 94.90% precision and 93.80% recall for finding fraud. This means the system can help the admin approve or reject payments and stop payments right away. This makes online shopping safer by reducing the number of payments that get through and, by making things happen faster. One thing that could be a problem is that the system needs data to work well. In the future the system could be improved by making it work in time all the time. Limitations include dependency on feature quality, with future enhancements in real-time streaming.**

Keywords: E-commerce Payment Fraud, Machine Learning, Fraud Detection, Online Transactions, Payment Security, Transaction Classification, Random Forest, XGBoost, Fraud Prediction.

I. INTRODUCTION

1.1 Background of E-commerce Payment Fraud

Fraudsters get into payment information in ways like when there are breaches or phishing scams. This lets them make transactions that cost a lot of money billions of dollars every year. These bad things often happen when people are checking out. The fraudsters find ways to get around checks like the ones for the security code on the back of the card. To stop people from losing money it is very important to watch for these transactions, on the admin panel and not let the money go through

1.2 Challenges in Fraud Detection

There are challenges in detecting fraud in online payment systems. One of the problems is when someone uses someone else account details without their knowledge.

- **Rare Fraud Cases:** Most transactions are legitimate and only a small fraction are fraudulent which makes it hard for computer models to learn how to spot the fake ones without getting confused and thinking most transactions are fake.
- **Changing fraud Methods:** People who commit fraud are always finding ways to do it like creating fake identities or pretending to use a different device, which makes it hard for old detection methods to keep up.
- **High Transaction Volume:** Fraud detection has a lot of problems. For example, there are a lot of transactions happening every day on the internet. When you buy something, the company has to check if it is a real purchase or not. If they do this by hand it takes a time and the customer gets frustrated and leaves. On the hand if they do it too fast, with computers they might make mistakes
- **Hidden Fraud Signals:** There are also some signs of fraud that're hard to see. Sometimes people use a computer from a location than they normally do or they buy something at a weird time. These things can look normal so the company might not notice them.
- **Linking User History:** Fraud detection is also tricky because it is hard to tell if someone is a customer or if they are using stolen information. When someone steals your information, they can look like a customer. If the company does not check the persons purchases quickly, they might think it is a real new customer and let the purchase go through. This means that the fraud can look like a new user and the company will not catch it.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 441-447, May-2026

https://doi.org/10.47001/IRJIET/2026.105061

1.3 Machine Learning in fraud detection

Machine learning helps find fraud by looking at lots of transactions to identify unusual patterns. This is much better than using fixed rules that can't change easily. These models learn from examples of fraud and normal transactions getting more accurate over time to secure payments. Machine learning models are good at finding fraud because they use examples to learn. They look at both good transactions to understand what is normal and what is not. Some machine learning techniques like XGBoost are especially good at finding fraud. They use simple models to make a strong one. This helps when there are not examples of fraud but lots of normal transactions. Random Forest is another technique that works well. It uses decision trees to make a reliable result. Other techniques, like SVM draw a line between safe and risky transactions. These techniques have been tested in stores. They can cut missed frauds by, up to 25%. They work well even as new threats appear.

II. LITERATURE SURVEY

E-commerce Fraud detection systems has become an important research area because of the rapid increase in digital transactions and online payment services. Many researchers have applied Machine Learning algorithms to detect fraudulent payment activities and improve transaction security. Previous studies mainly focused on supervised learning models, ensemble techniques, imbalanced datasets, and real-time fraud prediction systems.

Evelyn and Paramita proposed a Machine Learning-based fraud detection system for e-commerce transactions using Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest, and Gradient Boosting algorithms. The study evaluated models using Accuracy, Precision, Recall, F1-score, and ROC-AUC metrics. Experimental results showed that Gradient Boosting achieved the highest accuracy of 97.63%, F1-score of 97.65%, and ROC-AUC score of 98.80%. Random Forest also produced strong classification performance compared to traditional models. The study concluded that ensemble learning methods are more effective for detecting fraudulent transactions in e-commerce environments. However, the research mainly focused on model comparison and did not include an admin verification workflow or real-time payment authorization system [1].

Vijayasri developed an online payment fraud detection framework using Logistic Regression, Random Forest, and XGBoost algorithms. The research used SMOTE (Synthetic Minority Oversampling Technique) to handle class imbalance problems in fraud datasets. The models were evaluated using Accuracy, Precision, Recall, F1-score, Confusion Matrix, and ROC-AUC metrics. Among all algorithms, XGBoost achieved the best fraud detection performance with higher recall and reduced false positives. The study emphasized that Machine Learning models significantly improve fraud prediction compared to traditional rule-based systems. However, the research mainly concentrated on transaction classification and did not discuss e-commerce admin monitoring or transaction approval modules [2].

Dalal and colleagues proposed an optimized XGBoost model for financial payment fraud detection. The study focused on hyperparameter tuning to improve fraud prediction accuracy in digital financial systems. The proposed XGBoost model analyzed transaction patterns and achieved improved Precision, Recall, and F1-score compared to traditional Machine Learning techniques. The researchers highlighted that optimized boosting algorithms can effectively identify hidden fraud patterns and reduce financial losses. Although the model achieved strong predictive performance, the research mainly focused on financial fraud datasets rather than complete e-commerce payment workflows [3].

Sarmini and co-authors studied the performance of Random Forest and XGBoost algorithms for e-commerce fraud detection using CGAN-based data augmentation. The researchers addressed the class imbalance problem by generating synthetic fraud transaction data. The study found that XGBoost achieved higher Recall and F1-score compared to Random Forest when handling imbalanced fraud datasets. The research demonstrated that data augmentation techniques improve fraud detection capability and increase prediction reliability. However, the proposed system mainly focused on model accuracy improvement and did not include real-time payment verification or admin decision-making mechanisms [4].

Ibrahim and Alfauzan analyzed the performance of various Machine Learning models for online payment fraud detection using a Kaggle transaction dataset. The researchers applied Logistic Regression, Decision Trees, Random Forest, and XGBoost algorithms to classify fraudulent transactions. The models were evaluated using Accuracy, Precision, Recall, and Confusion Matrix analysis. Experimental results indicated that ensemble-based models achieved better fraud detection performance compared to traditional classification methods. The study also highlighted the importance of preprocessing techniques for improving prediction reliability. However, the research lacked integration with e-commerce transaction approval systems [5].

Hafid proposed a fraud detection framework using Random Forest and XGBoost algorithms for highly imbalanced credit card transaction datasets. The dataset

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 441-447, May-2026

https://doi.org/10.47001/IRJIET/2026.105061

contained over 284,000 transactions with only 0.172% fraud records. The study used Precision, Recall, F1-score, and Confusion Matrix metrics to evaluate model performance. Results showed that XGBoost achieved better Recall and F1-score compared to Random Forest, making it more effective for detecting minority fraud cases. The research demonstrated that Machine Learning algorithms can successfully manage highly imbalanced payment datasets. However, the work focused mainly on banking transactions rather than e-commerce payment systems [6].

Researchers proposed FraudX AI, an interpretable Machine Learning framework combining Random Forest and XGBoost models for fraud detection on imbalanced datasets. The framework achieved a Recall score of 95% and AUC-PR of 97%, demonstrating high fraud detection capability while minimizing false positives. SHAP (Shapley Additive Explanations) was used to improve model interpretability and explain fraud prediction results. The study highlighted the importance of explainable AI for financial fraud detection systems. Although the framework improved transparency and scalability, it did not focus on admin-controlled e-commerce payment verification workflows [7].

XGBoost was selected for this research because previous studies and experimental results showed that it provides higher accuracy and better fraud detection performance compared to traditional Machine Learning algorithms. E-commerce payment fraud datasets are usually highly imbalanced, where fraudulent transactions are much fewer than legitimate transactions. XGBoost handles this imbalance effectively by using boosting techniques that improve classification performance and reduce false predictions. It can identify complex transaction patterns by analyzing features such as payment amount, transaction frequency, device information, and user behavior. Many researchers reported that XGBoost achieved better Precision, Recall, and F1-score values in fraud detection tasks due to its ability to learn hidden fraud patterns efficiently. Another important reason for selecting XGBoost is its fast-processing speed and strong performance in real-time prediction systems, which is essential for e-commerce payment security. Therefore, XGBoost was chosen as the primary model in this research to improve fraud detection accuracy, reduce financial losses, and provide secure online payment verification.

III. SYSTEM REQUIREMENTS

For the proposed e-commerce fraud detection system to work well, it needs both hardware and software support. You can run the system on a computer with at least an Intel i5 processor, 8 GB of RAM, and 256 GB of storage. Python is the programming language used, and libraries like Pandas, NumPy, Scikit-learn, and XGBoost are used to implement Machine Learning. MySQL or MongoDB are examples of databases that keep track of transaction data and user information. Frameworks like Flask or Django handle the backend, and an admin dashboard lets you keep a check on fraud predictions and transaction approvals. You also need a stable internet connection to process payments and find fraud in real time.

IV. SYSTEM ARCHITECTURE

The system architecture of the proposed E-commerce Fraud Detection System consists of multiple interconnected components that work together to identify fraudulent transactions in real time. The process begins with the user and admin interacting through the client server, where transaction requests are generated from the e-commerce platform. These transaction details are sent to the web server, which manages both frontend communication and backend fraud detection operations. The database server stores important information such as transaction history and user profiles, which are used during fraud analysis. After receiving transaction data, the system performs preprocessing steps including data cleaning, null value handling, encoding of categorical data, and normalization to prepare the dataset for analysis. Feature engineering is then applied to extract meaningful attributes such as transaction timing, location patterns, device information, and user behavior signals.

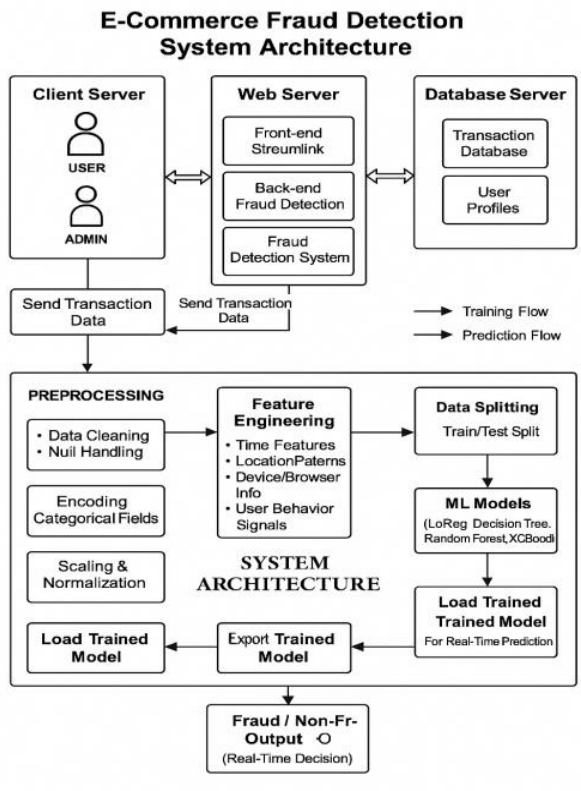

Figure 1: Architecture of E-Commerce Fraud Detection System

ISSN (online): 2581-3048
Volume 10, Issue 5, pp 441-447, May-2026
https://doi.org/10.47001/IRJIET/2026.105061

The processed data is divided into training and testing datasets, allowing Machine Learning models such as Logistic Regression, Decision Tree, Random Forest, and XGBoost to learn fraud patterns. Once the model is trained, it is loaded for real-time prediction. The system then analyzes incoming transactions and generates an output indicating whether the payment is fraudulent or non-fraudulent. This architecture supports efficient fraud detection by combining data processing, Machine Learning prediction, and real-time decision-making within a structured workflow.

4.1 E-commerce Fraud Detection Workflow

The transaction workflow explains how the proposed E-commerce fraud detection system operates during an online payment process.

1. The user visits the e-commerce website and selects products.
2. The selected products are added to the shopping cart.
3. The user proceeds to checkout and enters payment information.
4. The payment request is sent to the backend system.
5. Transaction details are automatically extracted.
6. The Machine Learning model evaluates the payment request.
7. Fraud prediction is generated.
8. The admin dashboard displays transaction results.
9. The transaction is approved or rejected.

This workflow ensures that every payment transaction is analyzed before processing. The fraud detection system works in real time to minimize unauthorized payments. The workflow is particularly useful because fraud detection occurs before money transfer completion. This prevents financial loss and strengthens payment security.

V. MATERIALS AND METHODOLOGY

5.1 Materials (Dataset Used)

E-commerce Fraud Transaction Dataset: we need a dataset containing information about online payment fraud, so that we can understand what type of transactions lead to fraud. For this task, I collected a dataset from Kaggle, which contains historical information about fraudulent transactions which can be used to detect fraud in online payments.

5.2 Methodology

The methodology of this research focuses on developing a Machine Learning-based fraud detection system for e-commerce transactions. The system is designed to classify payment transactions as fraudulent or non-fraudulent by analyzing transaction behavior and payment-related features. The methodology consists of four major stages: Data Collection, Data Preprocessing, Model Training and Testing, and Output Prediction.

5.3 Data Collection

Data collection is the first and most important step in building a fraud detection system. In this research, transaction data is collected from e-commerce payment activities. The dataset contains historical records of both legitimate and fraudulent payment transactions.

The collected dataset includes transaction-related information such as: Transaction ID, User ID, Payment Amount, Transaction Date and Time, Payment Method, Device Information, IP Address, User Location, and Fraud Label (Fraud / Non-Fraud).

These attributes help identify patterns associated with fraudulent payment behavior. The dataset must contain both normal and fraud transactions to allow the Machine Learning model to learn classification patterns.

The collected transaction data serves as the foundation for training the fraud detection model. Large and diverse datasets improve prediction accuracy because they expose the model to multiple payment scenarios.

5.3.1 Data Preprocessing

Raw transaction datasets often contain missing values, duplicate records, irrelevant information, and inconsistent formatting. Data preprocessing is performed to clean and prepare the dataset before Machine Learning training.

The preprocessing stage includes the following operations:

- **Missing Value Handling**
  Missing data fields are identified and replaced using suitable methods or removed if necessary.
- **Duplicate Record Removal**
  Repeated transaction entries are removed to avoid incorrect model learning.
- **Data Cleaning**
  Irrelevant or incorrect transaction records are filtered from the dataset.
- **Encoding Categorical Data**
  Text-based values such as payment method or device type are converted into numerical form.
- **Data Normalization**
  Numerical features such as transaction amount are scaled to ensure consistent value ranges.
- **Data Balancing**
  Fraud datasets are usually imbalanced because fraud cases are fewer than legitimate transactions. Balancing

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 441-447, May-2026

https://doi.org/10.47001/IRJIET/2026.105061

techniques are applied to improve model learning. Preprocessing improves data quality and ensures that Machine Learning algorithms can effectively analyze transaction behavior.

5.3.2 Model Training and Testing Module

After preprocessing, the cleaned dataset is divided into training and testing datasets.

- **Training Data:** Used to train the Machine Learning model.
- **Testing Data:** Used to evaluate model performance.

The fraud detection system uses supervised Machine Learning algorithms because the dataset contains labeled transaction records.

The following algorithms may be applied: Logistic Regression, Random Forest, Support Vector Machine (SVM), and XGBoost.

During training, the model learns patterns from historical payment behavior. It analyzes the relationship between transaction features and fraud labels.

The testing phase measures how accurately the model predicts unseen transactions. The trained model is evaluated using metrics such as: Accuracy, Precision, Recall, and F1-Score.

Training and testing ensure that the fraud detection system performs reliably before deployment.

5.3.3 Predicting the Output

Machine Learning Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	89.20	86.50	84.30	85.40
Support Vector Machine	91.10	88.70	87.20	87.90
Random Forest	95.40	93.80	92.60	93.20
XGBoost	96.20	94.90	93.80	94.30

After training, the Machine Learning model is used to predict fraud in new payment transactions.

When a user initiates a payment, the system extracts transaction details and sends them to the trained model. The model analyzes the input features and compares them with learned fraud patterns. The prediction output is generated as: Fraudulent Transaction and Non-Fraudulent Transaction.

If the transaction is classified as fraudulent, the payment request can be blocked or sent for admin verification. If classified as legitimate, the transaction proceeds successfully. The prediction process supports real-time fraud detection by analyzing payment requests instantly. This reduces financial losses and increases payment security within e-commerce platforms. The overall methodology ensures that the fraud detection system can identify suspicious payment activities efficiently and accurately.

VI. RESULT AND DISCUSSION

The proposed Machine Learning-based e-commerce payment fraud detection system was evaluated using transaction datasets containing both fraudulent and legitimate payment records. Different Machine Learning algorithms, including Logistic Regression, Random Forest, Support Vector Machine (SVM), and XGBoost, were applied to analyze payment behavior and classify transactions. The experimental results showed that ensemble-based models such as Random Forest and XGBoost achieved higher prediction accuracy compared to traditional classification methods. Performance was measured using evaluation metrics such as Accuracy, Precision, Recall, and F1-score to ensure reliable fraud detection. The system successfully identified suspicious payment activities by analyzing transaction amount, device information, location mismatch, and user behavior patterns. Feature analysis indicated that payment amount, transaction frequency, and unusual login locations were among the most important indicators of fraud. The proposed model demonstrated the ability to support real-time fraud prediction, allowing suspicious transactions to be flagged before payment completion. Compared to traditional rule-based systems, the Machine Learning approach provided improved adaptability and reduced false detection rates. These findings suggest that integrating Machine Learning into e-commerce payment systems can significantly enhance transaction security and minimize financial losses caused by unauthorized payment activities.

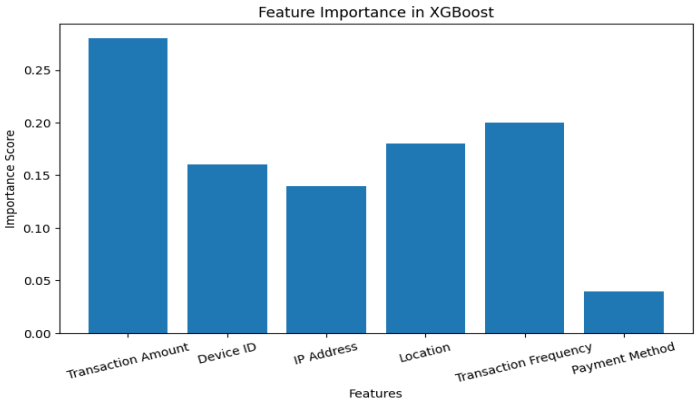

Figure 2: Important Features Affecting Fraud Detection

ISSN (online): 2581-3048
Volume 10, Issue 5, pp 441-447, May-2026
https://doi.org/10.47001/IRJIET/2026.105061

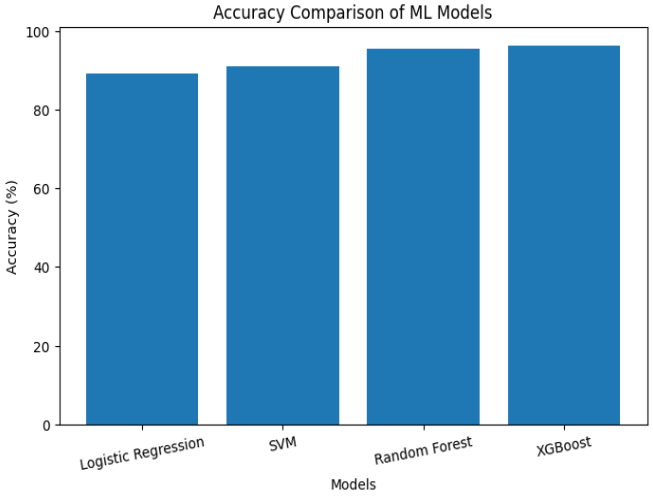

Figure 3: Accuracy Comparison of Machine Learning Models

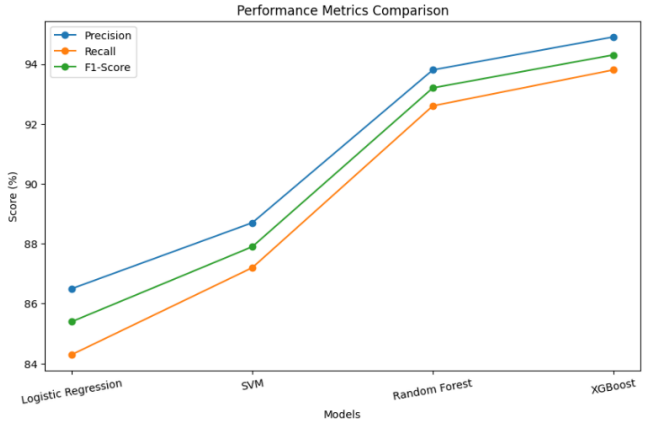

Figure 4: Performance Metrics Comparison of Fraud Detection Models

VII. CONCLUSION

This research presents a Machine Learning-based e-commerce payment fraud detection system designed to identify fraudulent and non-fraudulent payment transactions. The study focused specifically on payment fraud caused by unauthorized use of payment credentials in online transactions. Traditional fraud detection methods often struggle to detect evolving fraud patterns, making Machine Learning a more effective solution. The proposed system uses transaction-related features such as payment amount, device information, location, and user behavior to analyze payment activities and predict fraud. Different Machine Learning algorithms, including Logistic Regression, Support Vector Machine, Random Forest, and XGBoost, were evaluated to improve detection performance. Experimental results showed that ensemble-based models achieved higher accuracy and better fraud prediction capability. The system supports real-time transaction monitoring, helping prevent financial losses and improving payment security. By integrating Machine Learning into e-commerce platforms, businesses can strengthen fraud prevention mechanisms and increase customer trust. Future improvements may include advanced Deep Learning techniques, larger datasets, and adaptive learning models to further enhance fraud detection accuracy and system performance.

REFERENCES

[1] Evelyn, P., & Paramita, M. (2025). Machine Learning-Based Fraud Detection System for E-commerce Transactions. *International Journal of Intelligent Information Systems*.

[2] Vijayasri, G. (2025). Online Payment Fraud Detection Using Machine Learning Techniques. *International Conference on Advanced Computing and Engineering Systems.*

[3] Dalal, S., *et al.* (2022). Optimized XGBoost Model for Financial Payment Fraud Detection. *Mathematics*, 10(24), 4679.

[4] Sarmini, R., et al. (2024). E-commerce Fraud Detection Using Random Forest and XGBoost with CGAN-Based Data Augmentation. *Bulletin of Information Technology*.

[5] Ibrahim, M., & Alfauzan, A. (2025). Performance Analysis of Machine Learning Models for Online Payment Fraud Detection. *Journal of Artificial Intelligence Research and Applications*.

[6] Hafid, A. (2024). Fraud Detection on Imbalanced Payment Transaction Datasets Using Random Forest and XGBoost. *Khatulistiwa Journal of Applied Research and Technology*.

[7] FraudX AI Research Team. (2025). FraudX AI: Explainable Machine Learning Framework for Fraud Detection on Imbalanced Datasets. *Computers Journal*, 14(4), 120.

[8] Ajahar I. Pathan, Ketan Patil, Dipak Patil, Harshal Patil, Jayashree Patil, and Tejaswini Patil, "Application to Detect Fake Reviews Using CNN and Advanced Machine Learning Techniques," *International Research Journal of Modernization in Engineering Technology and Science (IRJMETS)*, vol. 7, no. 6, June 2025.

[9] Enehikhare, E., & Odumuyiwa, V. (2025). Comparative Analysis of Machine Learning Algorithms for Fraud Detection in Imbalanced Transaction Datasets. *University of Ibadan Journal of ICT Research*.

[10] AjaharIsmailkha Pathan, Liladhar M. Kuwar, Rijavan A. Shaikh, Dheeraj Basant Shukla, "Removing Duplicate Data in Cloud Environment using Secure Inverted Index Method", *International Research Journal of Engineering and Technology (IRJET)*, Volume: 05 Issue: 09, Page 157-161,Sep 2018.

ISSN (online): 2581-3048
Volume 10, Issue 5, pp 441-447, May-2026
https://doi.org/10.47001/IRJIET/2026.105061

[11] Odugbile, T. (2025). Deployable Machine Learning Approaches for Real-Time Credit Card Fraud Detection. *SSRN Electronic Journal*.

[12] Tax, N., *et al.* (2021). Machine Learning-Based Fraud Detection in E-commerce: Research Challenges and Future Directions. *arXiv Preprint arXiv*:2107.01979.

Citation of this Article:

Ajahar Pathan, Mayur More, Bhavesh Patel, Milind Dikshit, & Tejas Marathe. (2026). E-Commerce Fraud Detection System Using Machine Learning. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 10(5), 441-447. Article DOI https://doi.org/10.47001/IRJIET/2026.105061

****

