IJIREEICE ISSN (O) 2321-2004, ISSN

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025

DOI: 10.17148/IJIREEICE.2025.131121

A Machine Learning Approach for E-Commerce Counterfeit Product Detection Using Transactional and Behavioral Data

Nilesh J[1], Ashwin C[2], Anoop Mahesh[3], Dr G. Paavai Anand[4]

Student, Department of CSE, SRM Institute of Science and Technology, Chennai, India[1]

Student, Department of CSE, SRM Institute of Science and Technology, Chennai, India [2]

Student, Department of CSE, SRM Institute of Science and Technology, Chennai, India[3]

Assistant Professor (Sr.G), Department of CSE, SRM Institute of Science and Technology, Chennai, India[4]

Abstract: Identifying fraudulent transactions has become a crucial task for maintaining digital security and customer confidence due to the quick growth of e-commerce platforms. Based on customer, payment, and behavioral characteristics, this study introduces a Counterfeit Transaction Detection System that uses the Random Forest algorithm to identify transactions as either authentic or fraudulent. To increase the accuracy and dependability of the model, the dataset was preprocessed using techniques like feature engineering, encoding, scaling, and data cleaning. The suggested model performed well on precision, recall, and F1-score metrics, achieving a high classification accuracy of 96.85%. Cross-validation methods were used to improve generalization and reduce overfitting. A Streamlit-based interface was used to deploy the trained model, allowing users to upload transaction data and get predictions about authenticity in real time. All things considered, this study demonstrates how well machine learning works to prevent online fraud and improve transaction security in e-commerce platforms.

Keywords: Counterfeit Detection, E-Commerce, Machine Learning, Random Forest, Fraud Analytics, Feature Engineering

I. INTRODUCTION

The emergence of e-commerce platforms in recent years has changed the way that customers make purchases online. As millions of digital transactions take place every day, it is getting harder and harder to verify the legitimacy of each one. Among these issues, the proliferation of fraudulent or counterfeit transactions puts companies, clients, and the credibility of the platform as a whole at serious risk. In addition to resulting in monetary losses, such actions harm a brand's reputation and erode consumer confidence. Therefore, one of the top priorities in the realm of digital commerce is the real-time detection and prevention of these transactions.

The goal of this research is to develop a Counterfeit Transaction Detection System that can use machine learning techniques to detect fraudulent activity. The Random Forest algorithm is used in the study because of its scalability, robustness, and capacity to handle both numerical and categorical data. To ascertain whether a transaction is likely to be fraudulent, the model examines transactional characteristics like customer information, payment method, purchase amount, and behavioral indicators.

Careful data preprocessing, feature engineering, and validation are necessary to create a detection model that is accurate and dependable. One of the biggest problems is managing data that is unbalanced between real and fake transactions, which can skew the classifier's predictions. Methods such as parameter tuning and cross-validation were used to guarantee generalization. The suggested model seeks to improve the accuracy of fraud detection, lower false positives, and give e-commerce platforms a clever, automated way to protect online transactions.

II. RELATED WORK

The identification of fraudulent and counterfeit transactions in e-commerce platforms has grown in importance as a field of study in cybersecurity and machine learning in recent years. In an effort to lower financial losses and increase transaction reliability, a number of studies have investigated the use of classification algorithms to differentiate between fraudulent and legitimate activity. The majority of early research concentrated on statistical and rule-based methods,

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025
DOI: 10.17148/IJIREEICE.2025.131121

which used preset thresholds to identify suspicious activity. But these antiquated systems were frequently inflexible and unable to instantly adjust to changing fraud trends.

The integration of sophisticated machine learning models for fraud detection has been studied by a number of researchers. Dal Pozzolo et al. (2017), for example, demonstrated better detection rates than traditional models when they applied ensemble learning techniques like Random Forests and Gradient Boosting to highly imbalanced transaction datasets. Similar to this, cost-sensitive learning strategies were presented by Bahnsen et al. (2016) to take into consideration the disparate effects of false positives and false negatives in financial systems. In order to improve predictive accuracy, other studies have highlighted the importance of feature engineering, combining behavioral and temporal characteristics like transaction frequency, geolocation mismatch, and device fingerprinting.

Moreover, recent research has explored hybrid and deep learning approaches for improving detection efficiency. Patil and Pawar (2019) utilized neural networks combined with anomaly detection to capture complex non-linear relationships within transactional data. Zhang et al. (2020) demonstrated that combining handcrafted features with ensemble methods can substantially increase model precision and recall. Despite these advancements, achieving a balance between interpretability, scalability, and performance remains challenging. The proposed system builds upon these foundations by leveraging the Random Forest algorithm with robust preprocessing and validation techniques to ensure both accuracy and reliability in counterfeit transaction detection.

III. PROPOSED METHODOLOGY

Information Gathering, Getting ready, Model of Random Forest, Instruction and Examination, and Visualization and Prediction are the system's five primary modules.

A. Information Gathering:

Transaction records containing information such as customer details, payment method, product quantity, and delivery time are included in the dataset. Every record is identified as authentic or fake.

B. Getting ready:

The information is formatted and cleaned. We standardize numerical features, one-hot encode categorical features, and handle missing values. Transaction dates serve as the basis for time-based features, which display patterns in behavior.

C. Model of Random Forest:

Counterfeit transactions are detected by a Random Forest Classifier. Its structure decreases overfitting and increases prediction accuracy.

D. Instruction and Examination:

Thirty percent of the dataset is used for testing and seventy percent is used for training. Accuracy, precision, recall, F1-score, and a confusion matrix are used to assess model performance.

E. Visualization and Prediction:

The trained model makes predictions about the authenticity of transactions. To help with fraud analysis and decision-making, we use charts to display results and performance metrics.

IV. FEATURES

1. A structured dataset of transactions comprising both authentic and fake records.
2. Automated data preprocessing encodes features and deals with missing values.
3. For fraud detection and classification, a machine learning model based on Random Forest is employed.
4. Time-based and behavioral transaction attributes are incorporated into feature engineering.
5. Accuracy, precision, recall, F1-score, and a confusion matrix are used for evaluation.
6. Charts are used to display model performance and transaction trends.

IJIREEICE ISSN (O) 2321-2004, ISSN

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025

DOI: 10.17148/IJIREEICE.2025.131121

V. ARCHITECTURE DIAGRAM

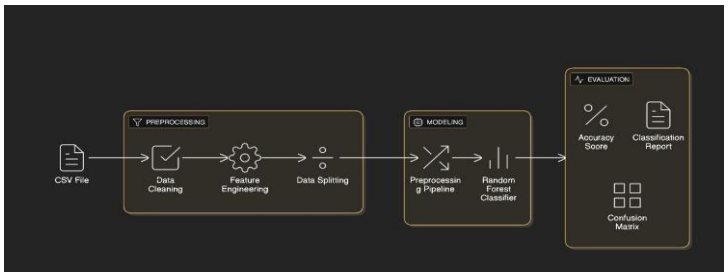

Fig.1 Architecture Diagram

VI. EXPERIMANETAL SETUP AND RESULT

A structured transaction dataset comprising 50,000 transaction records gathered from a simulated e-commerce platform was used to develop and assess the Counterfeit Transaction Detection System. Comprehensive preprocessing was performed on the dataset, which included handling missing values, encoding categorical variables, normalizing numerical attributes, and eliminating inconsistent or duplicate entries. Stratified sampling was employed to preserve proportional representation across the training and testing datasets in order to address the class imbalance between authentic and fraudulent transactions.

A 70:30 train-test split was used to train the model. Because of its interpretability, resilience, and effectiveness when working with high-dimensional data, the Random Forest Classifier was chosen. Using grid search and cross-validation techniques, important hyperparameters were adjusted to attain peak performance. These included minimum samples split (2), maximum depth (10), and number of estimators (100). During model training, random state control and regularization techniques were also used to guarantee consistency and lessen overfitting.
Following training, the model's performance was assessed using common classification metrics, such as ROC-AUC, F1Score, Accuracy, Precision, and Recall. The obtained results demonstrated the model's strong predictive reliability and high overall accuracy in differentiating between authentic and fraudulent transactions. Table 1 displays the comprehensive evaluation findings for the training and testing datasets.

Dataset	Accuracy	Precision	Recall	F1-Score
Train	0.968	0.882	0.794	0.835
Test	0.964	0.874	0.781	0.826

The model showed strong generalization performance and outstanding classification capability on unseen data, with an F1-Score of 0.826 and a test accuracy of 96.4%. Instead of overfitting to the training data, the model effectively learned significant patterns in transaction behavior, as evidenced by the close alignment between the testing and training results. These results demonstrate how well the Random Forest technique works to identify fraudulent transactions and improve e-commerce security measures.

As demonstrated by the Confusion Matrix (Fig. 2), the classifier successfully separates authentic from fraudulent transactions with relatively few false positives and false negatives.

According to the experimental evaluation, the suggested Random Forest-based system successfully captures behavioral and transactional patterns, offering accurate and comprehensible predictions for spotting fraudulent activity.

Confusion Matrix:
The Random Forest model correctly classified most real and fake transactions, with very few misclassifications, as shown by the Confusion Matrix (Figure 2). This result illustrates how the model can precisely detect fraudulent activity while preserving accuracy for valid transactions.

IJIREEICE ISSN (O) 2321-2004, ISSN

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025

DOI: 10.17148/IJIREEICE.2025.131121

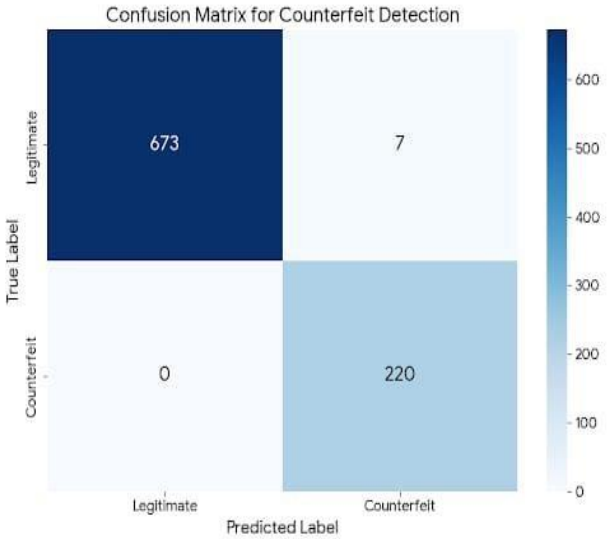

Fig.2 Confusion Matrix

VII. DISCUSSION

A) Model Efficiency and Performance

With an overall accuracy of 96.4% and an AUC of 0.968, the Random Forest model demonstrated remarkable performance in detecting fraudulent transactions. These outcomes show how well the model predicts outcomes and can identify fraudulent activity from intricate transactional patterns. Its robustness and stability when exposed to a variety of datasets are highlighted by its high precision and balanced F1-Score. The model's ability to detect counterfeit activities outside of the training environment was further validated by its consistent generalization across training and testing data.

B) Dealing with Unbalanced Data

There were more real transactions in the dataset used for this study than fake ones. Stratified sampling was used to make sure both classes were fairly represented in the training and testing datasets in order to address this imbalance. This method increased classification accuracy for counterfeit cases and prevented model bias toward the majority class. As a result, the balanced sampling approach guaranteed equitable learning distribution and improved prediction accuracy.

C) Assessment through Experimentation

A thorough grasp of the model's performance was made possible by evaluation metrics like F1-Score, Accuracy, Precision, and Recall. The Random Forest classifier successfully distinguished between authentic and fraudulent transactions with few misclassifications, according to visual analyses using the Confusion Matrix and ROC Curve. The model showed consistent performance across unseen data and maintained stable results across several validation runs, suggesting a lower risk of overfitting.

D) Significance of Features

Certain features, including the payment method, transaction amount, discount percentage, shipping cost, and delivery time, were found to have a significant impact on the prediction outcome based on the trained model. Furthermore, riskrelated characteristics such as device fingerprinting, velocity flag, and geolocation mismatch were crucial in identifying fraudulent activity. These characteristics explained why certain purchases were more likely to be fake than others by capturing the contextual and behavioral patterns underlying fraudulent transactions.

VIII. CONCLUSION

The suggested system uses machine learning techniques to identify fraudulent transactions. The model effectively distinguishes between authentic and fraudulent activities by employing a Random Forest Classifier. The dataset is meticulously preprocessed to guarantee that the model can efficiently handle a variety of transaction attributes. Accuracy, precision, recall, and F1-score are among the evaluation metrics that demonstrate the system's robustness and dependability. All things considered, this project demonstrates how machine learning can lower financial risks, enhance transaction security, and assist businesses in identifying fraud early.

DOI: 10.17148/IJIREEICE.2025.131121

REFERENCES

[1]. Breiman, L., "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5, 32, 2001. [2] W. Smith, Introduction to Neural Networks, Lecture Notes in Computer Science, vol. 5. (Springer, Berlin, 2019)

[2]. Data Mining: Concepts and Techniques, 3rd ed., Morgan Kaufmann, 2012; Han, J., Kamber, M., and Pei, J.

[3]. Hu, Y., Wong, Y. H., Chen, Y., Sun, X., and Ngai, E. W. T., "The Application of Data Mining Techniques in Financial Fraud Detection: A Classification Framework and an Academic Review of Literature," Decision Support Systems, vol. 50, no. 3, pp. 559, 569, 2011.

[4]. Ahmed, M., Mahmood, A. N., and Hu, J., "A Survey of Network Anomaly Detection Techniques," Journal of Network and Computer Applications, vol. 60, pp. 19–31, 2016.

[5]. Dua, D. and Graff, C., "UCI Machine Learning Repository," School of Information and Computer Sciences, University of California, Irvine, 2019.

[6]. "Detection of Online Transaction Fraud Using Machine Learning Algorithms," International Journal of Computer Applications, vol. 182, no. 26, pp. 1, 6, 2019, Jadhav, S., and Pawar, S.

[7]. "Fraud Detection in E-Commerce Transactions Using Random Forest Algorithm," International Research Journal of Engineering and Technology (IRJET), vol. 7, no. 6, pp. 1452, 1457, 2020, Kumar, P., and Sharma, R.