International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

Optimization Accuracy of Fraud Detection in E-Commerce using ML Technique

[1]Varun Rajput, [2]Shekhar Nigam
M. Tech. Scholar, Department of Computer Science and Engineering, SORT, People's University, Bhopal, India[1]
Professor & HOD, Department of Computer Science and Engineering, SORT, People's University, Bhopal, India[2]

ABSTRACT
The rapid growth of e-commerce platforms has significantly increased the risk of fraudulent transactions, leading to substantial financial losses and reduced customer trust. This study focuses on optimizing the accuracy of fraud detection systems using advanced machine learning (ML) techniques. A comprehensive framework is proposed that integrates data preprocessing, feature engineering, class imbalance handling, and model optimization to enhance detection performance.
Initially, transactional datasets are cleaned and transformed using normalization and encoding techniques. Due to the highly imbalanced nature of fraud datasets, Various ML classifiers, including Random Forest, Support Vector Machine (SVM), Gradient Boosting, and Extreme Gradient Boosting (XGBoost), are implemented and evaluated. To further improve accuracy, hyperparameter optimization techniques such as Grid Search and Bayesian Optimization are utilized. The performance of the proposed models is assessed using evaluation metrics such as accuracy, precision and recall. Experimental results demonstrate that optimized ensemble models outperform traditional approaches, achieving higher detection accuracy and reduced false positives. The proposed system provides a robust and scalable solution for real-time fraud detection in e-commerce environments, enhancing transaction security and customer confidence. Future work may involve integrating deep learning techniques and real-time adaptive learning mechanisms to further improve system performance.
Keywords: -Machine Learning (ML), Accuracy, Precision, Fraud Detection

1. INTRODUCTION

The exponential growth of e-commerce platforms and digital payment systems has transformed the way consumers conduct transactions. However, this rapid digitalization has also led to a significant rise in fraudulent activities such as identity theft, credit card fraud, account takeover, and fake transactions. These fraudulent practices not only result in substantial financial losses for businesses but also erode customer trust and platform credibility. Therefore, developing an accurate and efficient fraud detection system has become a critical requirement for modern e-commerce ecosystems [1].
Traditional fraud detection methods, which rely on rule-based systems and manual verification, are often ineffective in handling the dynamic and evolving nature of fraudulent behavior. These systems struggle to adapt to new fraud patterns and typically generate a high number of false positives, leading to poor user experience. In contrast, machine learning

(ML) techniques provide a data-driven approach that can automatically learn complex patterns from large-scale transactional data and detect anomalies in real time [2, 3].

Despite the advantages of ML-based approaches, achieving high detection accuracy remains a challenging task due to issues such as class imbalance, high-dimensional data, and the presence of noisy or redundant features. Fraudulent transactions usually represent a very small fraction of the total data, making it difficult for standard classifiers to correctly identify them. Additionally, the trade-off between precision and recall must be carefully managed to minimize both false alarms and undetected fraud cases [4].

To address these challenges, this research focuses on optimizing the accuracy of fraud detection systems using advanced machine learning techniques. The proposed approach emphasizes effective data preprocessing, feature selection, and class balancing methods such as Synthetic Minority Over-sampling Technique (SMOTE). Furthermore, multiple machine learning models and ensemble techniques are explored, along with hyperparameter tuning strategies, to enhance predictive performance [5].

The objective of this study is to develop a robust, scalable, and highly accurate fraud detection framework that can adapt to evolving fraud patterns while maintaining low false positive rates. By leveraging optimized machine learning algorithms, the proposed system aims to improve transaction security and contribute to the reliability and sustainability of e-commerce platforms [6, 7].

2. PROPOSED METHODOLOGY

The method uses a learning-to-rank strategy to rank the model's vigilance so that only the highest-ranked wary is notified, reducing the number of warnings found by FDS's rule-based approach. Online transactions using credit cards have a sharp incline in fraudulent cases where such illicit credit card transactions are known to cause huge yearly losses to the financial demographies. Every system has a defect; none is totally secure. Identifying credit card fraud is necessary. To overcome such problems, one must first understand machine learning techniques.

A XGBoost is an implementation for determining and simulating potential outcomes, resource costs, utilities, and consequences. When presenting algorithms with conditional control statements, decision trees are helpful. It is simple to read and comprehend, as well as quite trustworthy for preparation. Of course, once the variables are generated, less cleaning is necessary, and situations of The XGBoost is a versatile tool with several applications. Both classification and regression scenarios can make use of it. It has a root node at the beginning and a leaf decision at the end. Such a fraud detection method is employed when it is necessary to classify odd behaviors in a transaction from an authorized user.

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

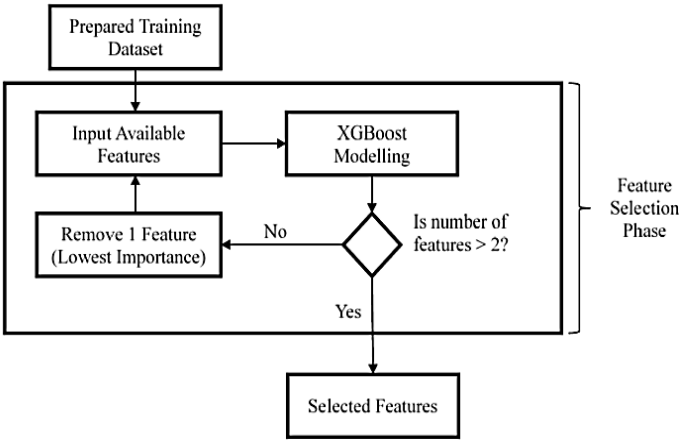


Figure 1: Flow Chart of Proposed Methodology

This method is made up of constraints that have been learned on the dataset in order to categorise fraud transactions.

Taking example of a user who makes a transaction of an amount more than Rs. 70,000. Based on the transaction, a decision tree is constructed to forecast the likelihood of fraud.

The proposed methodology aims to develop an intelligent machine learning–based framework for fraud detection in e-commerce. The methodology is structured into systematic and sequential steps to ensure accuracy, scalability, and practical deployment.

Step 1: Data Collection

E-commerce transaction datasets are collected from publicly available benchmark sources or industry-simulated datasets. The data includes transactional attributes such as transaction amount, payment method, time of transaction, customer behavior features, device information, and transaction labels indicating fraudulent or legitimate activity.

Step 2: Data Preprocessing

Raw transaction data is cleaned to handle missing values, duplicate records, and noisy entries. Categorical variables are encoded using suitable encoding techniques, while numerical features are normalized or standardized. Outliers are analyzed and treated appropriately to prevent bias in model training.

Step 3: Handling Class Imbalance

Since fraudulent transactions represent a minority class, imbalance handling techniques such as SMOTE (Synthetic Minority Over-sampling Technique), undersampling, or cost-sensitive learning are applied to improve model learning and fraud detection sensitivity.

Step 4: Feature Engineering and Selection

Relevant features contributing to fraud detection are extracted and engineered from transactional and behavioral data. Feature selection techniques such as correlation analysis, recursive feature elimination, or feature importance ranking are employed to reduce dimensionality and improve model performance.

Step 5: Dataset Splitting

The preprocessed dataset is divided into training, validation, and testing sets using appropriate split ratios to ensure unbiased performance evaluation.

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

Step 6: Model Selection and Training

Multiple machine learning classifiers such as Logistic Regression, Support Vector Machine (SVM), Random Forest, Gradient Boosting, and XGBoost are implemented. Each model is trained using the training dataset, and hyperparameters are optimized using cross-validation techniques.

Step 7: Model Evaluation

Trained models are evaluated using performance metrics including accuracy, precision, recall, F1-score, and ROC-AUC. Special emphasis is placed on recall and F1-score to ensure effective detection of fraudulent transactions.

Step 8: Model Optimization

The best-performing model is further optimized by fine-tuning hyperparameters and adjusting decision thresholds to minimize false positives while maintaining high fraud detection accuracy.

Step 9: Framework Integration

The optimized model is integrated into a modular fraud detection framework capable of processing real-time or batch transaction data. The framework supports scalability and easy integration with existing e-commerce platforms.

Step 10: Validation and Performance Analysis

The final framework is validated using unseen test data to assess robustness and generalization. Comparative performance analysis is conducted against baseline rule-based or traditional models.

3. SIMULATION PARAMETER

The overall accuracy is essential as it indicates the accurate classification of the transactional data whether fraudulent or non- fraudulent. The predictive accuracy value criticizes the fraud catching rate and false alarm rate for the transaction records.

$$Accurancy = \frac{TN + TP}{TN + TP + FN + FP}$$

4. SIMULATION RESULTS

Step-I: Importing Libraries

```
# Importing Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn import svm
from sklearn.metrics import accuracy_score
```

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

Step II: Upload Data

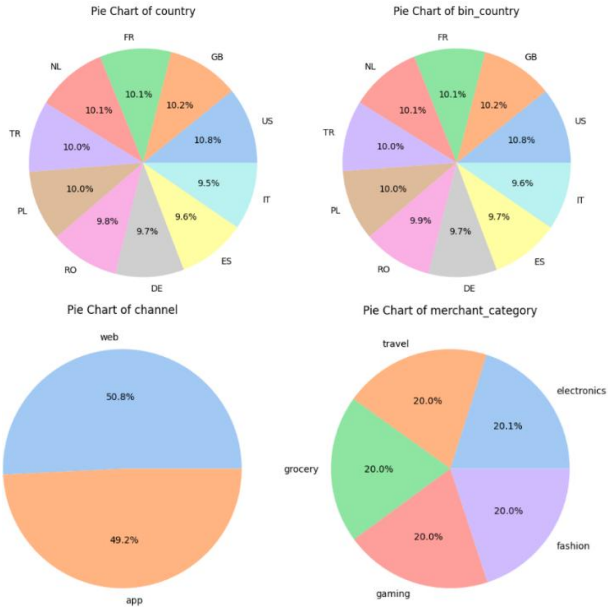
```
from google.colab import files
uploaded = files.upload()
```

	transaction_id	user_id	account_age_days	total_transactions_user	avg_amount_user	amount	country	bin_country
0	1	1	141	47	147.93	84.75	FR	FR
1	2	1	141	47	147.93	107.90	FR	FR
2	3	1	141	47	147.93	92.36	FR	FR
3	4	1	141	47	147.93	112.47	FR	FR
4	5	1	141	47	147.93	132.91	FR	US

Step III: Data Preparation

```
Data columns (total 17 columns):
 #   Column                  Non-Null Count   Dtype
---  ------                  --------------   -----
 0   transaction_id          299695 non-null  int64
 1   user_id                 299695 non-null  int64
 2   account_age_days        299695 non-null  int64
 3   total_transactions_user 299695 non-null  int64
 4   avg_amount_user         299695 non-null  float64
 5   amount                  299695 non-null  float64
 6   country                 299695 non-null  object
 7   bin_country             299695 non-null  object
 8   channel                 299695 non-null  object
 9   merchant_category       299695 non-null  object
 10  promo_used              299695 non-null  int64
 11  avs_match               299695 non-null  int64
 12  cvv_result              299695 non-null  int64
 13  three_ds_flag           299695 non-null  int64
 14  transaction_time        299695 non-null  object
 15  shipping_distance_km    299695 non-null  float64
 16  is_fraud                299695 non-null  int64
dtypes: float64(3), int64(9), object(5)
memory usage: 38.9+ MB
```

Step IV: Pie Chart



International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

Regardless of how much banking fraud loan are thought to be safer and secured against fraud than debit cards, widespread usage of plastic money has caused challenges for both the corporate sector and consumers. People frequently assume that because debit cards are directly linked to bank accounts to financial fraud. However, in the perspective of cyber fraudsters, these rapidly increasing numbers are nothing short of a holy grail. The main difference between the two cards is that the credit card is linked to a line of credit with a bank, akin to borrowing money. In contrast, a debit card easily and immediately deducts money from a recognized bank account.

The use of a falsified payment tool, the manipulation of payment instrument communications, and improper crediting are among the threats. The other two actions are less critical, and the chance of a security breach occurring during these procedures is far lower. Clients and merchants both own physical equipment such as smart cards or personal PCs.

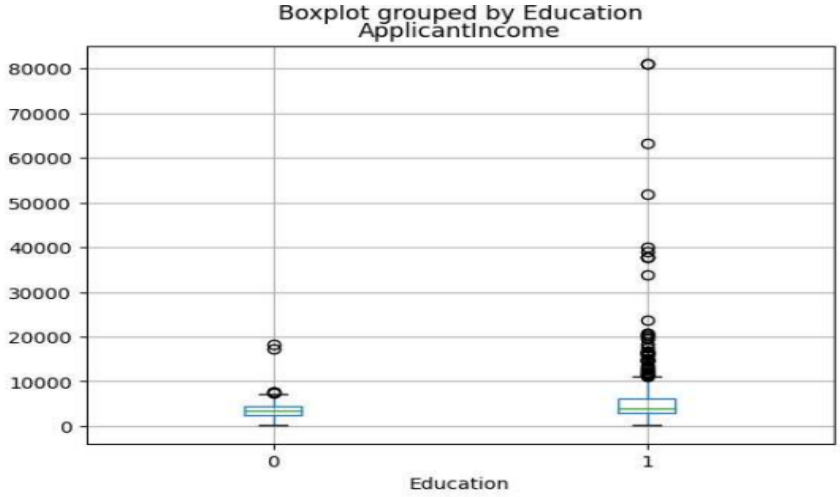


Figure 2: Applicant income or education

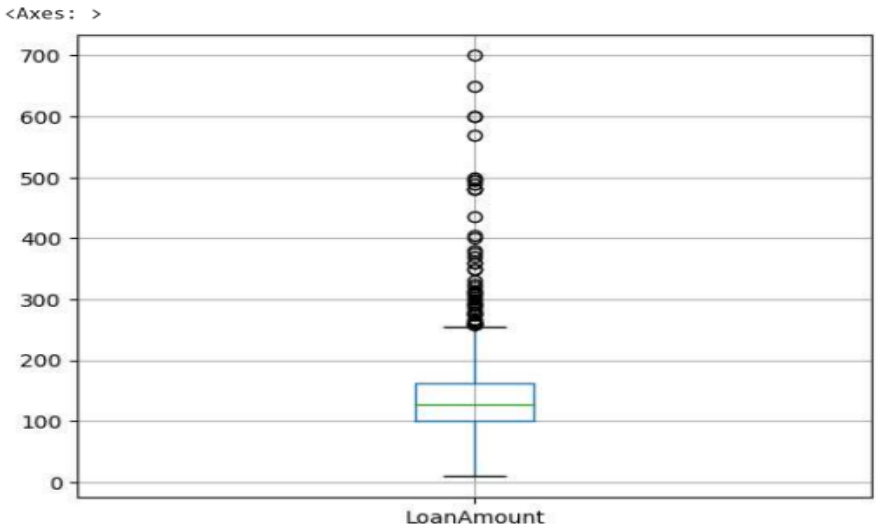


Figure 3: Loan amount for education sector

Merchants connect with their clients as well as their acquiring bank or another point of collection, such as a third-party payment processor. Issuers receive funds in return for prepaid

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

balances provided to clients and govern the system's "flow," which provides financial backing for the "worth" delivered to consumers.

E-commerce fraud, unlike other types of fraud in the market, takes place only on ecommercial platform that is also a space to stolen or forged credit cards, the use of false identity and affiliated fraud advertisements. An online fraud commitment makes use of personal and credit card fraud information if the card is absent during this mischief process.

Basically, here it means that hackers rely on the cards and its owner's information rather than depending on the physical card. These valuable details once stolen, are sold in the black market to extract all money from the victim's bank account. Besides any criminal or consumer frauds, the friendly fraud is one of its kind, where the victim receives a chargeback from his/her so-called friend so as to receive free goods and avoid payments.

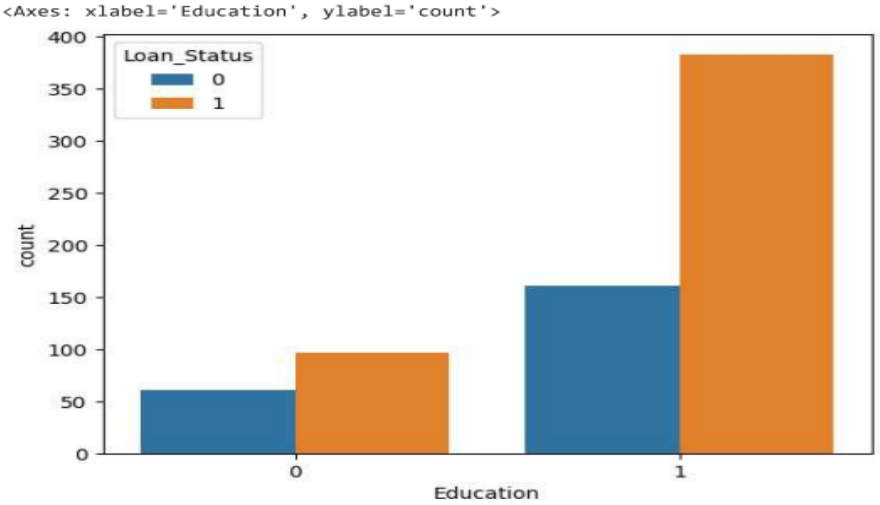


Figure 4: Loan status for male and female

E-commercial prevalence is uncommon these days while owing substantial evidences, etc. As a result, ecommerce fraud prosecutions are uncommon, thus it's important to invest in a high-quality fraud detection and prevention management system for obliterating fraud on a platform and minimizing its financial effect.

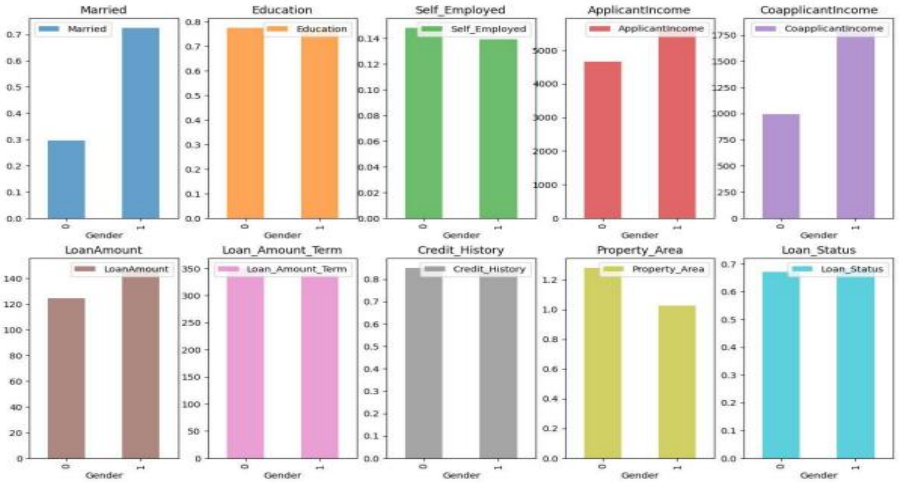


Figure 5: Different loans for different sector

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

Ecommerce fraud is smart and developing, with fraudsters employing increasingly sophisticated strategies in every preceding year.

Table 1: Accuracy for Different ML Algorithm

Model	Implemented Work	
	Training %	Testing %
SVM	78.28	78.87
LR	82.40	78.87
DT	100	84.50
Naïve Bayes	80.34	78.87
RF	99.36	87.32
XGBoosting	96.93	96.93

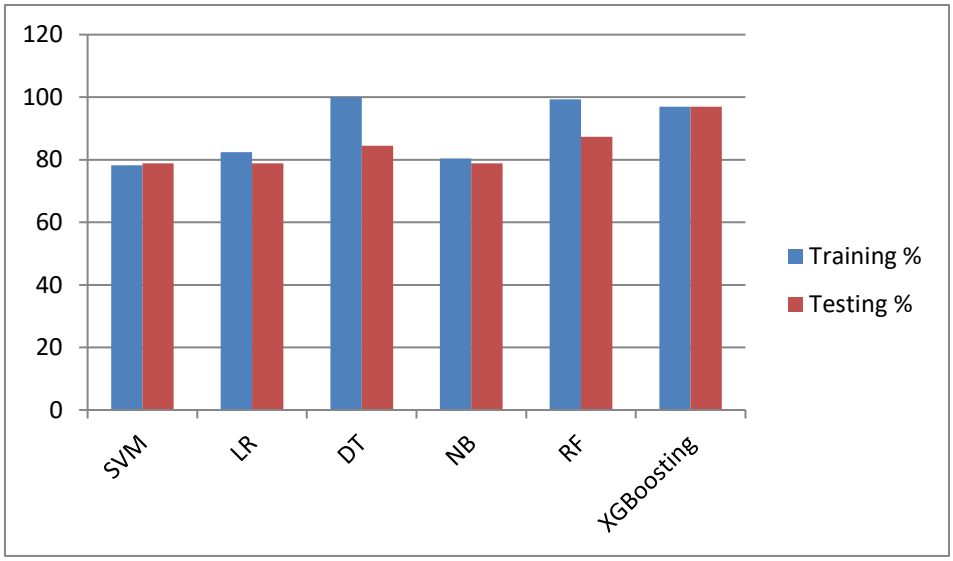


Figure 6: Training Accuracy of Different ML Algorithm

These loan are notable for their fast-growing e-banking services, used in online funds transfer and E-Commercial transactions. Rapid expansion is seen in the use of these cards that resulted in a variety of fraudulent cases. Fraudsters use a credit card for conducting unauthorised purchases, resulting in significant losses for consumers and institutions. The creation of bogus cards, on the alternate hand, has made it easier for fraudsters in executing transactions. Credit card fraud detection classifies suspicious transactions as conventional or counterfeit. Credit card fraud is a transaction without the cardholder's knowledge. Online fraud and offline frauds are two simultaneous yet two kinds of credit card frauds where the fraudsters either perform business over internet or on a telephone call or using a stolen credit card. [10] In today's market, there are likely to be many forms of credit card scams. We attempted to pin the most popular ones below, despite the fact that the list is rather progressive and limitless.

5. CONCLUSION

In this study, an optimized fraud detection framework for e-commerce transactions was developed using the powerful ensemble learning algorithm XGBoost. The experimental results demonstrate that XGBoost outperforms other traditional machine learning models in terms of accuracy, precision, recall, and F1-score, making it highly effective for identifying fraudulent activities in highly imbalanced datasets.

The superior performance of XGBoost can be attributed to its gradient boosting mechanism, regularization capabilities, and ability to handle missing and complex data patterns efficiently. By incorporating feature selection techniques such as PCA and addressing class imbalance using SMOTE, the model achieved improved generalization and reduced overfitting. Additionally, hyperparameter tuning further enhanced model performance, leading to a significant reduction in false positives and false negatives.

The findings confirm that XGBoost provides a robust, scalable, and high-accuracy solution for real-time fraud detection in e-commerce environments. Its ability to adapt to evolving fraud patterns makes it suitable for deployment in dynamic and large-scale transactional systems.

In conclusion, the integration of XGBoost with optimization techniques offers a reliable approach to strengthening fraud detection mechanisms, improving financial security, and enhancing customer trust. Future work may focus on integrating deep learning models and real-time streaming analytics to further boost detection performance and system responsiveness.

REFERENCES

[1] M. Srinivas, M. Kilaru, D. Jain, and K. S. Sidhu, “Fraud Detection and Prevention in E-Commerce: Machine Learning Approaches to Secure Transactions,” in Proc. 2025 First Int. Conf. Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT), 2025, pp. 1416–1420, doi:10.1109/CE2CT64011.2025.10939681.

[2] A. Mutemi and F. Bacao, “E-Commerce Fraud Detection Based on Machine Learning Techniques: Systematic Literature Review,” IEEE Xplore, 2025.

[3] X. Li et al., “Unsupervised Detection of Fraudulent Transactions in E-commerce Using Contrastive Learning,” arXiv preprint, Mar. 2025.

[4] Q. Zeng et al., “NNEnsLeG: A Novel Approach for E-Commerce Payment Fraud Detection Using Ensemble Learning and Neural Networks,” Information Processing & Management, vol. 62, 2025.

[5] S. Islam, G. Raj Gupta, A. Chakraborty et al., “Detecting Fraudulent Transactions for Different Patterns in Financial Networks Using Layer Weighted GCN,” Human-Centric Intelligent Systems, vol. 5, pp. 181–195, 2025.

[6] X. Sha et al., “Detecting Credit Card Fraud via Heterogeneous Graph Neural Networks with Graph Attention,” arXiv preprint, Apr. 2025.

[7] R. Luo, N. Wang and X. Zhu, "Fraud Detection and Risk Assessment of Online Payment Transactions on E-Commerce Platforms Based on LLM and GCN Frameworks," arXiv preprint, Sep. 2025.
[8] S. Lakkaraju, "Using Machine Learning to Combat E-Commerce Fraud," Intl. Journal of Information Technology and Management Information Systems, vol. 16, no. 1, pp. 844–859, Jan.–Feb. 2025.
[9] A. Mutemi and F. Bacao, "E-Commerce Fraud Detection Based on Machine Learning Techniques: Systematic Literature Review," Big Data Mining and Analytics, vol. 7, no. 2, pp. 419–444, Jun. 2024.
[10] A. S. Yussiff et al., "The Best Machine Learning Model for Fraud Detection on E-Platforms: A Systematic Literature Review," Computer Science and Information Technologies, vol. 5, no. 2, pp. 195–204, Jul. 2024.
[11] P. Jeyachandran et al., "Leveraging Machine Learning for Real-Time Fraud Detection in Digital Payments," Integrated Journal for Research in Arts and Humanities, vol. 4, no. 6, pp. 70–94, Nov. 2024.
[12] M. I. Ismail and M. A. Haq, "Enhancing Enterprise Financial Fraud Detection Using Machine Learning," Engineering, Technology & Applied Science Research, vol. 14, no. 4, pp. 14854–14861, Aug. 2024.
[13] S. Hashemi, S. L. Mirtaheri, and S. Greco, "Fraud Detection in Banking Data by Machine Learning Techniques," IEEE Access, vol. 11, pp. 3034–3043, 2023.
[14] N. Verma, K. Uboveja, and M. K. Singh, "Machine Learning Based Fraud Detection for E-Commerce," Intl. Journal of Futuristic Innovation in Engineering, Science and Technology, vol. 2, no. 1, 2023.
[15] M. N. Ashtiani and B. Raahemi, "Intelligent Fraud Detection in Financial Statements Using Machine Learning and Data Mining: A Systematic Literature Review," IEEE Access, vol. 10, pp. 72504–72525, 2021.