 ISSN 2579-7069

Machine Learning-Based Fraud Detection in E-Commerce Transactions

Evelyn[1,*] Adi Suryaputra Paramita[2]

[1]*School of Business Management Petra Christian University, Indonesia*
[2]*Information Systems Department, School of Information Technology, Universitas Ciputra Surabaya, Indonesia*

(Received April 16, 2025; Revised August 20, 2025; Accepted November 25, 2025; Available online January 31, 2026)

Abstract

The rapid growth of e-commerce has heightened fraud risks, demanding advanced fraud detection solutions. This study evaluates five machine learning models Logistic Regression, SVM, KNN, Random Forest, and Gradient Boosting for detecting fraudulent transactions in e-commerce environments. The models were assessed based on accuracy, precision, recall, F1-score, ROC-AUC, and error-related indicators. Results indicate that ensemble-based models, particularly Gradient Boosting and Random Forest, consistently outperform linear models like Logistic Regression, achieving superior balance between precision and recall. Gradient Boosting emerged as the top performer, with the highest accuracy (0.9763), F1-score (0.9765), and ROC-AUC (0.9880), while maintaining a low false negative rate (4.38%). These findings suggest that machine learning models, particularly ensemble methods, provide robust and efficient fraud detection frameworks. The study emphasizes the importance of using recall and F1-score as primary metrics to balance fraud detection sensitivity and operational efficiency.

Keywords: Fraud Detection, E-Commerce, Machine Learning, Gradient Boosting, Recall, Ensemble Models

1. Introduction

The rapid expansion of e-commerce has transformed global commercial activities, enabling high-volume and high-speed digital transactions across diverse sectors, including Business-to-Business (B2B) environments. However, this growth has also intensified exposure to fraud risks, as increasing transaction complexity and scale create broader attack surfaces for cybercriminals. Prior studies highlight that the surge in e-commerce adoption is directly associated with escalating cybersecurity threats, particularly fraudulent transactions that undermine financial stability and operational integrity [1]. As transaction volumes grow, fraud schemes become increasingly sophisticated, posing serious challenges to both businesses and consumers while eroding trust in digital marketplaces [2].

The prevalence of fraudulent transactions has significant financial and reputational consequences. Empirical evidence indicates that organizations across various sectors suffer substantial economic losses annually due to weaknesses in online payment systems and digital transaction infrastructures [3], [4]. Beyond financial damage, fraud incidents severely impact consumer confidence, as users become increasingly concerned about privacy, data protection, and transaction security in online platforms [5], [6]. Consequently, ensuring effective fraud detection and prevention mechanisms has become a critical requirement for sustaining trust and long-term growth in e-commerce ecosystems.

Traditional fraud prevention approaches often struggle to cope with the dynamic and adaptive nature of fraudulent activities. Static rule-based systems and conventional statistical techniques lack flexibility and frequently fail to detect emerging fraud patterns in real time. This limitation has driven increasing interest in advanced analytical solutions based on Artificial Intelligence (AI), particularly machine learning, which offers adaptive and data-driven capabilities for fraud detection [7]. Complementary technologies such as blockchain have also been explored for securing transaction processes by enhancing transparency and tamper resistance; however, these approaches are typically preventive rather than predictive and require integration with analytical detection mechanisms to be fully effective [8].

***Corresponding author: Evelyn (evelyn@petra.ac.id)**
DOI: https://doi.org/10.47738/ijiis.v9i1.295

Machine learning has emerged as a pivotal tool for identifying anomalous transaction patterns indicative of fraud. By analyzing large-scale transactional data, machine learning models can uncover subtle and complex relationships that are difficult to detect using traditional methods. Recent studies demonstrate that supervised and unsupervised learning techniques, including anomaly detection networks and ensemble-based approaches, are effective in capturing unusual purchasing behaviors, abnormal transaction frequencies, and atypical monetary patterns associated with fraud [9]. Furthermore, advances in neural network architectures, such as recurrent and convolutional neural networks, have enabled near real-time fraud detection, allowing organizations to respond promptly to suspicious activities [10].

Despite these advances, detecting fraudulent transactions remains a challenging task. One of the most prominent challenges is achieving an appropriate balance between accuracy and sensitivity. While high recall is essential to minimize missed fraud cases, overly sensitive models often generate excessive false positives, leading to operational inefficiencies and increased investigation costs [10]. Additionally, fraud detection datasets are typically imbalanced, with legitimate transactions vastly outnumbering fraudulent ones, which can bias supervised learning models and degrade classification performance [4], [11], [12]. These challenges underscore the need for systematic evaluation of machine learning models using appropriate metrics and carefully engineered features.

In this context, supervised machine learning approaches remain foundational for fraud detection due to their ability to learn from labeled transaction data and recognize known fraud patterns [13]. However, the evolving nature of fraud schemes necessitates continuous assessment and refinement of these models to ensure robustness and adaptability [14]. Prior studies report that models such as Logistic Regression, Random Forest, and Support Vector Machines (SVM) can achieve high detection accuracy, with ensemble-based methods often demonstrating superior performance across diverse datasets [15]. Nevertheless, performance outcomes are highly dependent on feature representation, dataset characteristics, and evaluation strategies.

Therefore, this study aims to empirically evaluate the performance of selected supervised machine learning models for fraud detection in e-commerce transactions, with particular attention to both numerical and categorical transaction features. The research systematically assesses commonly used classifiers, including Logistic Regression, Random Forests, and SVM, using evaluation metrics that are critical in fraud detection contexts, namely accuracy, precision, recall, and F1-score. By analyzing model performance and feature contributions, this study seeks to provide practical insights into optimizing fraud detection systems that balance detection sensitivity with the minimization of false positives. Ultimately, the findings are expected to contribute to the development of more effective and reliable machine learning-based fraud detection frameworks capable of addressing the evolving challenges of e-commerce security.

2. Literature Review

2.1. Fraud Detection in E-Commerce

Traditional fraud detection approaches in e-commerce have predominantly relied on rule-based systems that apply predefined thresholds and expert-defined rules to identify suspicious transaction patterns. While such methods are effective in detecting known fraud behaviors, they suffer from limited adaptability and require frequent manual updates to address emerging fraud tactics. As fraud schemes become increasingly sophisticated and dynamic, these static systems struggle to maintain effectiveness, resulting in higher vulnerability within digital transaction environments and reduced detection accuracy over time [16].

In contrast, machine learning-based fraud detection methods offer adaptive and data-driven capabilities that enable continuous learning from historical transaction data. Supervised learning techniques, in particular, utilize labeled datasets to classify transactions as fraudulent or legitimate, allowing models to capture complex and non-linear patterns that traditional methods often overlook [13], [17]. Empirical studies have demonstrated that ensemble-based models, such as Random Forests and XGBoost, consistently achieve higher detection accuracy while reducing false positives compared to individual classifiers [9], [18]. Furthermore, deep learning approaches, including Convolutional Neural Networks and Autoencoders, have shown strong potential in detecting subtle anomalies within high-dimensional transaction data, further enhancing fraud detection performance in e-commerce systems [19].

Recent empirical research reinforces the effectiveness of supervised machine learning approaches in transaction fraud detection. Comparative studies indicate that machine learning models significantly outperform traditional rule-based systems across various e-commerce scenarios, particularly when combined with effective feature engineering strategies that integrate both numerical and categorical transaction attributes [13], [20]. Moreover, ensemble learning frameworks have been shown to provide superior robustness and generalization capabilities in handling evolving fraud patterns and imbalanced datasets commonly found in e-commerce environments [16], [20]. These findings collectively highlight the growing consensus that machine learning-based approaches constitute a more responsive and effective foundation for securing modern e-commerce transactions.

2.2. Machine Learning Models for Fraud Detection

Machine learning models play a central role in e-commerce fraud detection due to their ability to analyze large-scale transactional data and capture complex fraud patterns. Commonly adopted supervised learning models include Logistic Regression, SVM, Random Forest, and Gradient Boosting, each offering distinct advantages depending on data characteristics and detection objectives. Logistic Regression is frequently used as a baseline model because of its simplicity, computational efficiency, and interpretability in binary classification tasks. However, its reliance on linear decision boundaries limits its effectiveness when addressing non-linear relationships and complex feature interactions often present in fraud datasets, necessitating additional feature engineering to maintain performance in high-dimensional settings [21].

SVM are particularly effective in high-dimensional feature spaces and can model non-linear decision boundaries through kernel functions, making them suitable for fraud detection scenarios involving complex transaction patterns. Their robustness against overfitting enables reliable performance on moderately sized datasets; however, scalability remains a challenge when applied to large transaction volumes due to computational complexity and sensitivity to hyperparameter selection [22]. In contrast, ensemble-based models such as Random Forest and Gradient Boosting have demonstrated superior performance in fraud detection tasks by leveraging multiple decision trees to improve generalization. Random Forest effectively reduces variance and handles heterogeneous numerical and categorical features, although its ensemble structure can limit interpretability [23].

Gradient Boosting techniques, including advanced implementations such as XGBoost, further enhance detection performance by sequentially optimizing model errors and capturing intricate data relationships. These models are widely recognized for their high predictive accuracy in structured transactional data but require careful hyperparameter tuning to mitigate sensitivity to noise and overfitting risks [24]. Overall, existing studies emphasize that no single model is universally optimal for fraud detection; rather, model selection should consider dataset properties, computational constraints, and interpretability requirements. Consequently, hybrid and ensemble-based approaches, alongside continuous performance evaluation, are increasingly advocated to enhance the robustness and adaptability of fraud detection systems in evolving e-commerce environments [25].

2.3. Evaluation Metrics in Fraud Detection

Evaluating fraud detection models requires carefully selected performance metrics that reflect the asymmetric costs of classification errors. Commonly used metrics include accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC). While accuracy measures the overall proportion of correctly classified transactions, it can be misleading in fraud detection scenarios due to severe class imbalance, where legitimate transactions vastly outnumber fraudulent ones. In such contexts, high accuracy may simply indicate that the model favors the majority class while failing to detect fraud effectively [11]. Consequently, relying solely on accuracy does not provide a reliable assessment of model performance in real-world fraud detection systems.

Precision, recall, and F1-score offer more informative insights into fraud detection effectiveness by explicitly accounting for classification errors. Precision reflects the reliability of fraud alerts by measuring the proportion of correctly identified fraudulent transactions among all predicted fraud cases, making it particularly relevant when false positives incur high operational costs [26]. Recall, or sensitivity, quantifies the model's ability to identify actual fraudulent transactions and is widely regarded as a critical metric in fraud detection, as false negatives missed fraud cases can lead to substantial financial losses and erosion of customer trust [27]. The F1-score balances precision and

recall, providing a single metric that captures trade-offs between false positives and false negatives, which is especially valuable in evaluating models trained on imbalanced datasets [28].

AUC-ROC further complements these metrics by evaluating a model's discriminative ability across all classification thresholds, offering a threshold-independent measure of performance [29]. In practical fraud detection systems, recall is often prioritized over precision and overall accuracy, as the cost of failing to detect fraudulent transactions typically outweighs the inconvenience caused by false alarms. Empirical studies demonstrate that optimizing recall and minimizing false negatives often through sampling strategies, feature enrichment, or model tuning significantly enhances fraud detection effectiveness in operational environments [12]. Accordingly, a comprehensive evaluation framework that emphasizes recall while managing trade-offs with precision is essential for deploying reliable machine learning-based fraud detection systems in e-commerce and financial services.

3. Methodology

3.1. Dataset Description

The dataset used in this study consists of 46,046 e-commerce transaction records, each represented by 13 features, including numerical and encoded categorical attributes. The dataset has been fully preprocessed prior to modeling, ensuring that all features are in numerical format and suitable for machine learning algorithms. No missing values were observed across the dataset, indicating high data completeness and reliability for experimental evaluation.

The target variable is defined as Is Fraudulent, a binary class label indicating whether a transaction is fraudulent (1) or non-fraudulent (0). This variable serves as the primary outcome for supervised learning models tasked with identifying fraudulent transaction patterns based on historical data.

Regarding class distribution, the dataset exhibits a relatively balanced composition, with approximately 51.3% fraudulent transactions and 48.7% non-fraudulent transactions. Although the dataset does not suffer from severe class imbalance, the cost-sensitive nature of fraud detection necessitates careful evaluation using metrics beyond accuracy, such as recall and F1-score, to ensure effective identification of fraudulent cases. This balanced yet realistic distribution provides a suitable foundation for evaluating the performance and robustness of machine learning models in fraud detection scenarios.

3.2. Feature Description

The dataset comprises a combination of numerical and categorical features that capture key characteristics of e-commerce transactions relevant to fraud detection. These features are designed to represent both transactional behavior and contextual information associated with customers and products, enabling machine learning models to identify patterns indicative of fraudulent activities.

The numerical features include transaction amount, customer age, and account age (in days). Transaction amount reflects the monetary value of each transaction and is a critical indicator in fraud detection, as fraudulent activities often involve unusually high or irregular spending patterns. Customer age provides demographic context that may correlate with purchasing behavior, while account age represents the duration since account creation, which is particularly relevant as newly created accounts are frequently associated with higher fraud risk. All numerical features have been scaled to ensure consistent value ranges and to prevent dominance of features with larger magnitudes during model training.

The categorical features describe transaction-related attributes, including payment method and product category. Payment methods encompass options such as credit card, debit card, bank transfer, and digital wallets, which may exhibit different fraud risk profiles. Product categories, such as electronics, clothing, home and garden, health and beauty, and toys and games, provide additional contextual information, as certain product types are more commonly targeted in fraudulent transactions. These categorical variables have been transformed into numerical representations using encoding techniques, enabling their integration into machine learning models while preserving their discriminatory information.

3.3. Data Preprocessing

Data preprocessing was conducted to ensure that the dataset was suitable for effective machine learning model training and evaluation. Given the heterogeneous nature of transaction data, preprocessing focused on standardizing feature representations, transforming categorical variables, and preparing the dataset for supervised learning without introducing information leakage.

Feature scaling and normalization were applied to all numerical attributes to ensure comparable value ranges across features. Standardization was employed to rescale numerical variables to a common scale with zero mean and unit variance. This step is particularly important for distance-based and margin-based algorithms, such as SVM and Logistic Regression, as it prevents features with larger numerical ranges from disproportionately influencing the learning process.

Categorical variables were transformed into numerical representations using encoding techniques suitable for machine learning algorithms. Payment method and product category attributes were encoded into binary indicator variables, allowing the models to capture categorical distinctions without imposing ordinal relationships. The encoded categorical features were subsequently aligned with the scaled numerical features to form a unified feature space, ensuring compatibility across all evaluated models.

Finally, the prepared dataset was validated for completeness and consistency prior to model training. All features were consolidated into a fully numerical matrix, free from missing values and redundant attributes. The target variable was clearly defined as a binary class label, enabling straightforward application of supervised learning techniques. This preprocessing pipeline ensured that the dataset was model-ready and supported fair and reproducible comparisons across different machine learning algorithms in the subsequent experimental stages.

3.4. Experimental Framework

The experimental framework was designed to systematically evaluate the performance of machine learning models for fraud detection in e-commerce transactions. The overall workflow begins with the input of the preprocessed transaction dataset, followed by feature–target separation, model training, and performance evaluation. After preprocessing, the dataset was divided into input features and a binary target variable indicating fraudulent and non-fraudulent transactions. Multiple supervised machine learning models were then trained using the same feature set to ensure fair and consistent comparisons across algorithms.

To assess model performance and generalization capability, a structured evaluation strategy was employed. The dataset was initially partitioned into training and testing subsets using a stratified split to preserve the original class distribution. The training set was used to fit the models, while the testing set served as an independent hold-out dataset for final performance evaluation. This approach ensures that the reported results reflect the models' ability to generalize to unseen transaction data.

In addition to the train–test split, stratified k-fold cross-validation was applied to the training data to obtain robust performance estimates and reduce variance caused by random data partitioning. Cross-validation enabled each model to be evaluated across multiple data folds, ensuring stability and reliability of the results. Performance metrics derived from cross-validation and testing phases were subsequently analyzed and compared to identify the most effective machine learning approach for fraud detection.

3.5. Machine Learning Models

This study employs several widely used supervised machine learning models for fraud detection, selected based on their proven effectiveness, interpretability, and suitability for transactional data. The models represent a combination of linear, kernel-based, and ensemble learning approaches, enabling a comprehensive performance comparison across different learning paradigms.

Logistic Regression was used as a baseline classifier due to its simplicity and strong interpretability in binary classification tasks. Logistic Regression models the probability of a transaction being fraudulent using a linear decision boundary, making it effective for identifying straightforward fraud patterns. Despite its limitations in capturing

complex non-linear relationships, Logistic Regression remains valuable as a benchmark model for evaluating the relative performance of more advanced techniques.

SVM were employed to handle complex and non-linear decision boundaries in high-dimensional feature spaces. By utilizing a Radial Basis Function (RBF) kernel, SVM can effectively separate fraudulent and legitimate transactions that are not linearly separable. This capability makes SVM suitable for capturing subtle fraud patterns, although its performance is sensitive to kernel and regularization parameters.

Random Forest, an ensemble-based method, was selected for its ability to handle heterogeneous numerical and categorical features while reducing overfitting through the aggregation of multiple decision trees. Random Forest models are particularly effective in fraud detection due to their robustness to noise and capacity to capture non-linear feature interactions. However, their ensemble structure may reduce interpretability compared to linear models.

Gradient Boosting was included to further evaluate the effectiveness of sequential ensemble learning. Gradient Boosting models iteratively build decision trees by minimizing classification errors from previous iterations, enabling high predictive accuracy on structured transaction data. While highly effective, these models require careful parameter tuning to prevent overfitting, particularly in noisy datasets.

Regarding general parameter settings, standard configurations were applied to all models to ensure fair comparison. Default or commonly accepted parameter values were initially used, with constraints applied to control model complexity and prevent overfitting. No aggressive hyperparameter optimization was performed at this stage, allowing the evaluation to focus on the intrinsic learning capability of each model rather than extensive parameter fine-tuning.

3.6. Evaluation Metrics

The performance of the machine learning models was evaluated using several standard classification metrics commonly adopted in fraud detection research, including accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC). These metrics provide complementary perspectives on model performance and enable a comprehensive assessment of classification effectiveness under varying error conditions.

Accuracy measures the proportion of correctly classified transactions among all observations. While it provides a general overview of model performance, accuracy alone is insufficient in fraud detection scenarios due to the asymmetric cost of misclassification. Precision quantifies the proportion of correctly identified fraudulent transactions among all transactions predicted as fraud, reflecting the reliability of fraud alerts. Recall (sensitivity) measures the proportion of actual fraudulent transactions that are correctly detected by the model, while the F1-score, as the harmonic mean of precision and recall, captures the trade-off between false positives and false negatives. AUC-ROC evaluates the discriminative capability of a model across different decision thresholds, offering a threshold-independent measure of performance.

Among these metrics, recall and F1-score were emphasized as key indicators due to their critical relevance in fraud detection contexts. High recall is essential to minimize false negatives, as undetected fraudulent transactions can result in substantial financial losses and reputational damage. However, maximizing recall alone may lead to excessive false positives, which increase operational costs. Therefore, the F1-score was used to balance recall and precision, ensuring that models achieve effective fraud detection performance without disproportionately increasing false alarms. This metric selection aligns with real-world fraud detection objectives, where capturing fraudulent activities accurately while maintaining operational efficiency is of primary importance.

4. Results and Discussion

4.1. Model Performance Comparison

To provide a comprehensive comparison of model performance, this study evaluates five supervised machine learning models using multiple metrics derived from both cross-validation and test datasets. Table 1 summarizes the test-set performance of all models, including accuracy, precision, recall, F1-score, ROC-AUC, and error-related indicators. Overall, ensemble-based models consistently outperformed linear and distance-based approaches across all evaluation metrics.

 ISSN 2579-7069

Table 1. Performance Comparison of Machine Learning Models on the Test Dataset

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC	FPR	FNR
Logistic Regression	0.8109	0.8098	0.8253	0.8175	0.8969	0.2043	0.1747
SVM (RBF)	0.9749	0.9991	0.9520	0.9750	0.9779	0.0009	0.0480
KNN	0.9710	0.9991	0.9444	0.9710	0.9834	0.0009	0.0556
Random Forest	0.9756	0.9954	0.9568	0.9757	0.9859	0.0047	0.0432
Gradient Boosting	0.9763	0.9976	0.9562	0.9765	0.9880	0.0025	0.0438

During cross-validation, Gradient Boosting achieved the highest mean accuracy (0.9759), F1-score (0.9760), and ROC-AUC (0.9875), closely followed by Random Forest, which exhibited comparable recall (0.9559) and F1-score (0.9752). These results indicate strong generalization capability and stability, as reflected by the low standard deviation values across folds. This consistency is critical for evaluating the robustness of fraud detection models in dynamic e-commerce environments.

On the test dataset, similar performance trends were observed, confirming the robustness of the models across both validation and real-world data. Gradient Boosting emerged as the best-performing model, achieving the highest accuracy (0.9763), F1-score (0.9765), and ROC-AUC (0.9880), while maintaining a high recall of 0.9562. Random Forest demonstrated nearly identical recall (0.9568) but slightly lower precision and ROC-AUC. In contrast, KNN and SVM (RBF) achieved high precision values exceeding 0.999 but exhibited lower recall, indicating a higher tendency to miss fraudulent transactions. Logistic Regression, despite being interpretable, showed substantially lower performance across all metrics, with a test accuracy of 0.8109 and ROC-AUC of 0.8969. This highlights its limitations in capturing the complex, non-linear relationships inherent in transactional data, even after preprocessing and feature scaling.

Based on the comprehensive evaluation of accuracy, recall, F1-score, and ROC-AUC, Gradient Boosting was identified as the best-performing model for fraud detection in this study, offering the most balanced trade-off between fraud detection sensitivity and classification reliability.

Figure 1 illustrates the confusion matrix of the Gradient Boosting model on the test dataset. The model correctly classified 4,472 legitimate transactions (true negatives) and 4,520 fraudulent transactions (true positives). Only 207 fraudulent transactions were misclassified as legitimate (false negatives), resulting in a false negative rate of approximately 4.38%, while false positives remain extremely low (11 cases). This result demonstrates the model's strong ability to detect fraud while minimizing missed fraudulent transactions, which is critical in operational e-commerce environments where detecting fraud accurately is essential for financial security.

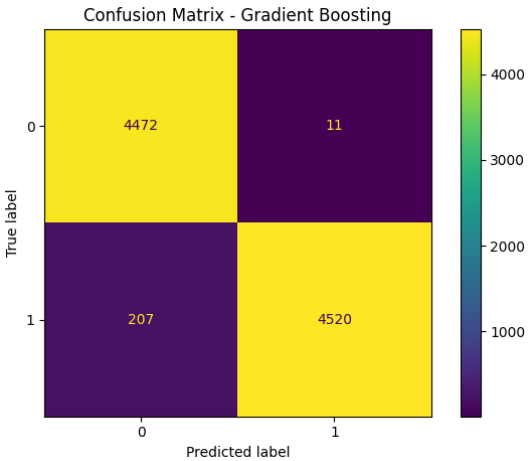

Figure 1. Confusion Matrix of the Gradient Boosting Model on the Test Dataset

Next, Figure 2 presents the ROC curves for all evaluated models on the test dataset. Gradient Boosting achieves the highest ROC-AUC value (0.988), indicating superior discriminative capability across all classification thresholds. Random Forest and KNN follow closely, with AUC values of 0.986 and 0.983, respectively. In contrast, Logistic Regression shows substantially weaker separation between classes, with an AUC of 0.897. This ROC analysis

 ISSN 2579-7069

highlights that ensemble-based models consistently outperform simpler classifiers in distinguishing fraudulent from non-fraudulent transactions.

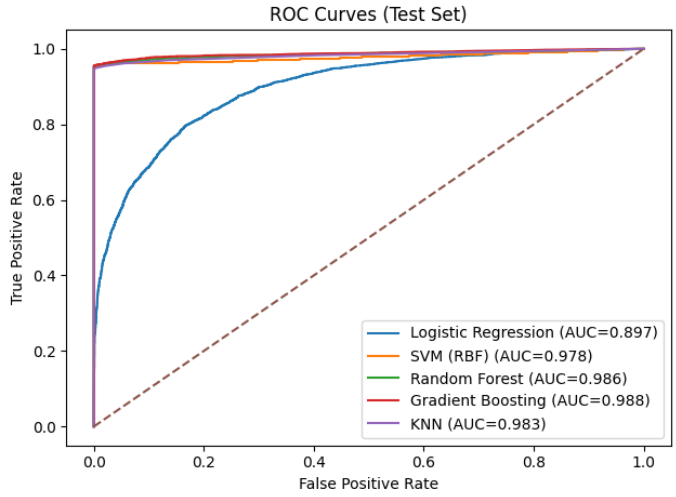

Figure 2. ROC Curves for All Evaluated Models on the Test Dataset

Figure 3 shows the Precision–Recall (PR) curves, which are particularly informative in fraud detection due to class imbalance and asymmetric error costs. Gradient Boosting achieves the highest Average Precision (AP = 0.992), maintaining near-perfect precision across a wide range of recall values. Although SVM and KNN exhibit extremely high precision, their recall drops more sharply, indicating a higher tendency to miss fraudulent transactions. Logistic Regression displays the weakest PR curve, reinforcing its limited suitability for high-risk fraud detection scenarios where recall is crucial for minimizing missed fraud cases.

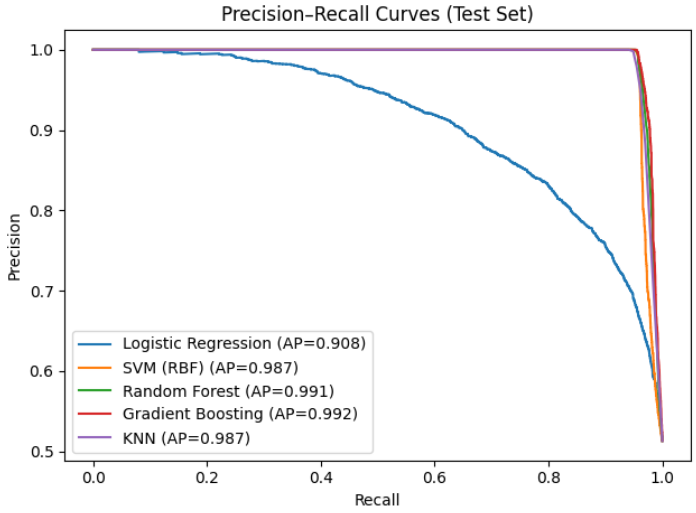

Figure 3. Precision–Recall (PR) Curves

Figure 4 depicts the top features identified by the Gradient Boosting model. The most influential feature is product category (electronics), followed by credit card payment method and home & garden product category. This suggests that fraud patterns are strongly associated with specific product types and payment channels. Numerical attributes such as transaction amount and account age contribute to a lesser extent, indicating that categorical transaction context plays a dominant role in fraud detection for this dataset. This aligns with prior research indicating that non-linear relationships between categorical and numerical features are crucial in detecting fraud effectively [13], [17].

 ISSN 2579-7069

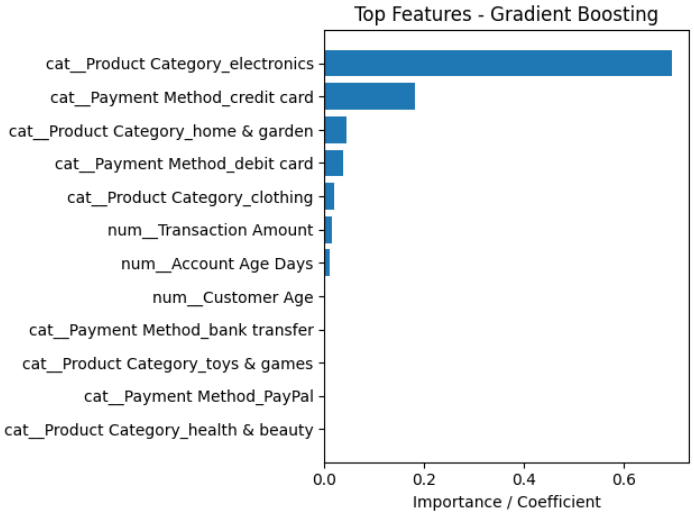

Figure 4. Top Features Identified by the Gradient Boosting Model

Figure 5 illustrates the most influential features based on Logistic Regression coefficients. While similar categorical features appear relevant, their linear influence is insufficient to model complex fraud behavior effectively. The contrast between Figures 4 and 5 highlights the limitation of linear models in capturing non-linear interactions that ensemble-based methods, such as Gradient Boosting, can successfully exploit. This reinforces the importance of selecting models that can capture the underlying complexities of transaction data.

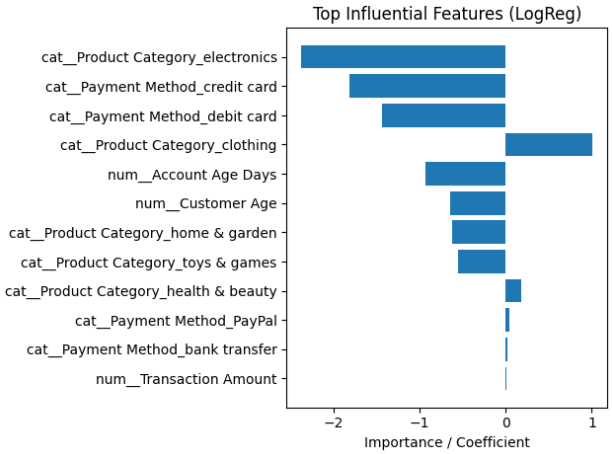

Figure 5. Most Influential Features Based on Logistic Regression Coefficients

Collectively, the quantitative results and visual analyses demonstrate that Gradient Boosting provides the most balanced and reliable performance for e-commerce fraud detection in this study. Its superior recall, low false negative rate, and strong discriminative power make it particularly suitable for real-world deployment, where undetected fraud poses significant financial and reputational risks.

4.2. Discussion

The experimental results demonstrate that ensemble-based learning methods, particularly Gradient Boosting and Random Forest, consistently outperform linear and instance-based models in detecting fraudulent e-commerce transactions. This finding aligns with prior studies reporting that ensemble approaches provide superior robustness and generalization by aggregating multiple decision boundaries and capturing complex feature interactions inherent in transactional data [9], [18]. The superior performance of Gradient Boosting can be attributed to its sequential learning mechanism, which iteratively emphasizes misclassified instances, enabling the model to learn subtle, non-linear fraud patterns that traditional rule-based systems and linear classifiers often fail to capture [13], [17], [19]. This confirms the

growing consensus that adaptive, data-driven models offer a more effective foundation for modern fraud detection than static rule-based approaches [16].

Models such as KNN and SVM achieved exceptionally high precision, indicating that their fraud predictions were highly reliable when triggered. However, their comparatively lower recall and higher false negative rates suggest a conservative decision boundary that prioritizes certainty over coverage. Similar trade-offs have been observed in previous studies, where kernel-based and distance-based models tend to favor precision at the expense of sensitivity, particularly in complex and evolving fraud environments [22]. From a practical perspective, this behavior can be problematic, as undetected fraudulent transactions represent direct financial losses and reputational risks. Logistic Regression, despite its interpretability and moderate recall, exhibited high false positive and false negative rates, reinforcing existing findings that linear models struggle to adequately represent non-linear relationships and intricate feature dependencies in fraud datasets, even after preprocessing and feature scaling [21].

From an operational standpoint, these results emphasize the importance of selecting evaluation metrics and models that reflect the asymmetric cost structure of fraud detection. Consistent with prior research, recall and false negative rate emerge as more critical indicators than overall accuracy or precision, as missed fraud cases can lead to substantial financial damage and erosion of consumer trust [27], [12]. The Gradient Boosting model achieved the lowest false negative rate (4.38%) among all evaluated models while maintaining near-perfect precision, supporting empirical evidence that ensemble-based methods offer an optimal balance between detection sensitivity and operational efficiency [30]. These findings suggest that e-commerce organizations should prioritize ensemble-driven machine learning frameworks, combined with continuous evaluation and feature refinement, to enhance the resilience of fraud detection systems against evolving and increasingly sophisticated fraud threats [25].

5. Conclusion

This study demonstrates that ensemble-based machine learning models, particularly Gradient Boosting, are highly effective for fraud detection in e-commerce transactions. The results underscore the superiority of Gradient Boosting and Random Forest over simpler models like Logistic Regression and SVM, as they successfully balance high accuracy and recall, which is crucial in fraud detection. These findings align with existing literature, where ensemble models have been shown to outperform traditional fraud detection methods due to their ability to model complex, non-linear relationships inherent in transactional data.

The Gradient Boosting model, with its superior recall and low false negative rate, offers a practical solution for real-world deployment, where detecting fraud without missing critical instances is paramount. Moreover, the model's ability to maintain precision while maximizing recall ensures operational efficiency, a critical factor for organizations dealing with large-scale transactional data in real-time.

In conclusion, this study highlights the importance of incorporating ensemble-based machine learning techniques in fraud detection systems, emphasizing recall as a key performance metric. Future research should explore advanced hybrid models and optimization techniques to further improve detection accuracy and efficiency, particularly in dynamic, high-volume e-commerce environments.

6. Declarations

6.1. Author Contributions

Author Contributions: Conceptualization E. and A.S.P.; Methodology, E. and A.S.P.; Software, E.; Validation, E.; Formal Analysis, E.; Investigation, A.S.P.; Resources, E.; Data Curation, A.S.P.; Writing Original Draft Preparation, E.; Writing Review and Editing, E. and A.S.P.; Visualization, A.S.P. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J. V Ngwira and J. Phiri, "Contextual Antecedents of E-Commerce Adoption for Supply Chain Management by Retail and Consumer Goods Traders in Developing Countries," *Open J. Bus. Manag.*, vol. 12, no. 01, pp. 472–489, 2024, doi: 10.4236/ojbm.2024.121029.

[2] A. Faraz and K. J. Simpao, "E-Commerce Fraud Perceptions in Pakistan and Impacts of Risks and Preventive Measures," *ITQAN J. Islam. Econ. Manag. Financ.*, vol. 4, no. 1, pp. 31–43, 2024, doi: 10.57053/itqan.v4i1.61.

[3] A. A. Alhashmi, A. M. Alashjaee, A. A. Darem, A. F. Alanazi, and R. Effghi, "An Ensemble-Based Fraud Detection Model for Financial Transaction Cyber Threat Classification and Countermeasures," *Eng. Technol. Appl. Sci. Res.*, vol. 13, no. 6, pp. 12433–12439, 2023, doi: 10.48084/etasr.6401.

[4] M. R. Baker, Z. N. Mahmood, and E. H. Shaker, "Ensemble Learning with Supervised Machine Learning Models to Predict Credit Card Fraud Transactions," *Rev. d'Intelligence Artif.*, vol. 36, no. 4, pp. 509–518, 2022, doi: 10.18280/ria.360401.

[5] M. Alhashem, M. K. Nabi, R. Pant, A. Inkesar, N. Khan, and M. A. Khan, "Exploring the Factors Affecting Online Trust in B2C E-Commerce Transactions in India: An Empirical Study," *Int. J. Prof. Bus. Rev.*, vol. 8, no. 12, pp. e03945, 2023, doi: 10.26668/businessreview/2023.v8i12.3945.

[6] M. Hossain, S. Islam, M. M. Rahman, and N. Arif, "Impact of Online Payment Systems on Customer Trust and Loyalty in E-Commerce Analyzing Security and Convenience," *Acad. J. Sci. Technol. Eng. Math. Educ.*, vol. 4, no. 3, pp. 1–15, 2024, doi: 10.69593/ajsteme.v4i03.85.

[7] P. J. Khan and SMT. M. Vani, "'The Multi-Perspective Fraud Detection Method for Multi-Participant for E-Commerce Transactions,'" *J. Eng. Sci.*, vol. 16, no. 04, pp. 174–179, 2025, doi: 10.36893/JES.2025.V16I04.029.

[8] I. Yuhertiana dan A. Hadi Amin, "Artificial Intelligence Driven Approaches for Financial Fraud Detection: A Systematic Literature Review," KnE Soc. Sci., vol. 9, no. 20, hal. 448–468, 2024, doi: 10.18502/kss.v9i20.16551.

[9] M. D. V. Prasad, "Multi-Entity Real-Time Fraud Detection System using Machine Learning: Improving Fraud Detection Efficiency using FROST-Enhanced Oversampling," *J. Electr. Syst.*, vol. 20, no. 7s, pp. 1380–1394, 2024, doi: 10.52783/jes.3710.

[10] M. Maashi, B. Alabduallah, and F. Kouki, "Sustainable Financial Fraud Detection Using Garra Rufa Fish Optimization Algorithm with Ensemble Deep Learning," *Sustainability*, vol. 15, no. 18, pp. 13301, 2023, doi: 10.3390/su151813301.

[11] P. Verma and P. Tyagi, "Analysis of Supervised Machine Learning Algorithms in the Context of Fraud Detection," *ECS Trans.*, vol. 107, no. 1, pp. 7189–7200, 2022, doi: 10.1149/10701.7189ecst.

[12] T. Albalawi and S. Dardouri, "Enhancing Credit Card Fraud Detection Using Traditional and Deep Learning Models with Class Imbalance Mitigation," *Front. Artif. Intell.*, vol. 8, no. 1, pp. 1–10, Okt 2025, doi: 10.3389/frai.2025.1643292.

[13] S. Cho, "Fraud Detection in Malaysian Financial Institutions Using Data Mining and Machine Learning," *J. Inf. Technol.*, vol. 7, no. 1, pp. 13–21, 2023, doi: 10.53819/81018102t4152.

[14] M. I. Shabiya and R. R. Chandrika, "LSO-RVAE: Latent Space Optimization with Sparrow Search and Residual Variational Autoencoder for Credit Card Fraud Detection," *Eng. Res. Express*, vol. 7, no. 3, pp. 35244, 2025, doi: 10.1088/2631-8695/adf52a.

[15] A. Alhchaimi, "Cloud-Based Transaction Fraud Detection: An In-Depth Analysis of ML Algorithms," *Wasit J. Comput. Math. Sci*., vol. 3, no. 2, pp. 19–31, 2024, doi: 10.31185/wjcms.253.

[16] Z. Miao, "Financial Fraud Detection and Prevention," *J. Organ. End User Comput*., vol. 36, no. 1, pp. 1–27, 2024, doi: 10.4018/joeuc.354411.

[17] S. Siddique, "Fraud Detection in E-Commerce – A Machine Learning Approach," *Interantional J. Sci. Res. Eng. Manag*., vol. 09, no. 06, pp. 1–9, 2025, doi: 10.55041/ijsrem50949.

[18] L. Yangyan and C. Tingting, "CRAFIC Framework: Multi-Account Collaborative Fraud Detection, Efficient Feature Extraction and Relationship Modelling Combined with CNN-LSTM and Graph Attention Network," *IET Commun*., vol. 19, no. 1, pp. 1–15, 2025, doi: 10.1049/cmu2.70014.

[19] A. Tanikonda, "Deep Learning for Anomaly Detection in E-Commerce and Financial Transactions: Enhancing Fraud Prevention and Cybersecurity," *J. Inf. Syst. Eng. Manag*., vol. 10, no. 30s, pp. 70–77, 2025, doi: 10.52783/jisem.v10i30s.4776.

[20] A. Skármeta, "Artificial Intelligence in E-Commerce Fraud Detection: A Paradigm Shift in Digital Security," *Int. Sci. J. Eng. Manag.*, vol. 04, no. 05, pp. 1–9, 2025, doi: 10.55041/isjem03834.

[21] Mr. Rumit Raval and Dr. Pallavi Devendra Tawde, "Fraud Detection in Online Transactions Using Machine Learning and Data Analytics.," *Int. J. Adv. Res. Sci. Commun. Technol*., vol. 5, no. 2, pp. 284–289, 2025, doi: 10.48175/IJARSCT-23343.

[22] D. Hemalatha, "Online Payment Fraud Detection Using Machine Learning," *Interantional J. Sci. Res. Eng. Manag*., vol. 09, no. 05, pp. 1–9, 2025, doi: 10.55041/ijsrem48218.

[23] M. Hassan and N. Kaabouch, "Impact of Feature Selection Techniques on the Performance of Machine Learning Models for Depression Detection Using EEG Data," *Appl. Sci*., vol. 14, no. 22, pp. 10532, 2024, doi: 10.3390/app142210532.

[24] L. Li, "A Comparative Study on Supervised Machine Learning Algorithms for Credit Card Transaction Fraud Detection," *Appl. Comput. Eng*., vol. 179, no. 1, pp. 74–80, 2025, doi: 10.54254/2755-2721/2025.ld26477.

[25] M. Koppula, "Predictive Maintenance to Reduce Machine Downtime in Factories Using Machine Learning Algorithms," *Int. J. Adv. Res. Comput. Sci*., vol. 16, no. 2, pp. 71–77, 2025, doi: 10.26483/ijarcs.v16i2.7224.

[26] S. Rahmadani, A. Subekti, and M. Haris, "Improving Classification Performance on Imbalanced Medical Data Using Generative Adversarial Network," *J. Ilmu Komput. dan Inf*., vol. 17, no. 1, pp. 9–17, 2024, doi: 10.21609/jiki.v17i1.1177.

[27] N. V Thieu, "PerMetrics: A Framework of Performance Metrics for Machine Learning Models," *J. Open Source Softw*., vol. 9, no. 95, pp. 6143, 2024, doi: 10.21105/joss.06143.

[28] M. Woo, J. Lee, H. Kim, and S. Park, "Subgroup Evaluation to Understand Performance Gaps in Deep Learning-Based Classification of Regions of Interest on Mammography," *Plos Digit. Heal*., vol. 4, no. 4, pp. e0000811, 2025, doi: 10.1371/journal.pdig.0000811.

[29] E. M. Al-dahasi, R. K. Alsheikh, F. A. Khan, and G. Jeon, "Optimizing Fraud Detection in Financial Transactions with Machine Learning and Imbalance Mitigation," Expert Syst., vol. 42, no. 2, pp. e13682, 2025, doi: 10.1111/exsy.13682.

[30] C. Idemudia, E. E. Agu, and S. Obeng, "Analysis of Machine Learning Techniques in Detecting and Preventing E-Commerce Fraud Effectively," *Int. J. Front. Sci. Technol. Res*., vol. 7, no. 1, pp. 25–34, 2024, doi: 10.53294/ijfstr.2024.7.1.0046.