Journal of Information & Systems Management

DLINE JOURNALS

ISSN: 2230–8776

JISM 2026: 16 (2)
https://doi.org/10.6025/jism/2026/16/2/37-55

AI-Driven Counterfeit and Fraud Detection in E-Commerce: A Dual-Layered Machine Learning Approach

Hsing-Cheng Liu, Yao-Liang Chung
Department of Communications, Navigation and Control Engineering
National Taiwan Ocean University, Taiwan
ylchung@ntou.edu.tw

ABSTRACT

The rapid expansion of e-commerce has intensified cybersecurity threats, particularly counterfeit product listings and fraudulent transactions, which severely undermine consumer trust and marketplace integrity. Traditional rule based monitoring systems increasingly struggle to detect these sophisticated, evolving fraud patterns. This study develops and evaluates a dual layer predictive analytics framework that leverages supervised machine learning to enhance counterfeit detection and fraud governance in digital marketplaces. Utilizing a Random Forest classifier, the research analyzes two complementary datasets: static product-level metadata comprising seller credentials and listing characteristics, and dynamic transaction-level behavioral indicators capturing purchasing anomalies and device signatures. Following rigorous preprocessing, categorical encoding, and an 80/20 stratified train-test split, the models were benchmarked using accuracy, precision, recall, F1 Score, and ROC-AUC. The product-level classifier achieved perfect class separability, demonstrating the deterministic power of structural listing indicators. Conversely, the transaction-level model attained 99.33% accuracy, 97.99% precision, and a 0.9999 ROC-AUC, effectively capturing probabilistic behavioral fraud signatures with minimal false negatives. These findings confirm that integrating static asset metadata with dynamic transactional analytics significantly improves predictive robustness. The study concludes by proposing a unified, AI-driven marketplace governance framework capable of real-time surveillance, automated threat mitigation, and scalable regulatory compliance, ultimately advancing the deployment of transparent artificial intelligence solutions across global digital commerce networks.

Keywords: Artificial Intelligence, E-Commerce Fraud, Counterfeit Detection, Machine Learning, Random Forest, Predictive Analytics, Behavioral Analytics, Marketplace Governance, Supervised Learning

Received: 11 September 2025, Revised 3 January 2026, Accepted 12 January 2026

1. Introduction

The rapid evolution of digital technologies has transformed the operational structures of the e-commerce and financial services industries by creating new channels for online transactions and digital service delivery. Although these advancements have improved accessibility, efficiency, and customer convenience, they have also introduced substantial cybersecurity and fraud-related challenges. Enterprises and consumers increasingly face threats associated with fraudulent transactions, counterfeit products, fake reviews, payment manipulation, and identity-based attacks. To mitigate these risks, e-commerce and financial service providers have adopted sophisticated analytical systems that leverage artificial intelligence (AI), machine learning (ML), and data analytics to strengthen transactional security and improve fraud surveillance capabilities [1].

2. Earlier Studies

The continuous growth of e-commerce platforms and digital payment ecosystems has intensified the complexity of fraud detection due to the large-scale, dynamic, and highly imbalanced nature of transactional data environments. Fraudulent activities are often concealed within massive volumes of legitimate transactions, making traditional rule-based monitoring systems increasingly ineffective. In response, predictive analytics has emerged as a critical approach for identifying suspicious patterns and anticipating fraudulent behavior before significant financial losses occur. Predictive fraud detection systems employ statistical modelling, machine learning algorithms, and deep learning architectures to discover hidden behavioural irregularities and classify high-risk activities in real time [2].

E-commerce fraud has become one of the most critical operational and reputational challenges affecting online retailers and consumers alike [3]. The widespread adoption of mobile technologies, high-speed internet connectivity, and digital marketplaces has accelerated the volume of online transactions, thereby creating new opportunities for cybercriminal activities [4] . Fraud within e-commerce ecosystems manifests in multiple forms, including credit card fraud, counterfeit product distribution, fraudulent payment schemes, fake online reviews, and large scale data breaches [5]. These fraudulent practices not only generate direct financial losses but also erode consumer trust, damage brand reputation, and reduce the credibility of digital marketplaces.

Artificial intelligence has increasingly been proposed as a strategic solution to address cybersecurity and fraud detection challenges in digital commerce environments. AI technologies, particularly machine learning and deep learning techniques, enable systems to autonomously learn complex fraud patterns and detect anomalies that are difficult to identify using conventional approaches [6]. Despite these advantages, the majority of AI applications within e-commerce have historically focused on customer relationship management, supply chain optimization, and marketing intelligence rather than fraud detection and prevention [4]. Consequently, significant research gaps remain in the application of AI-based systems for identifying and mitigating diverse forms of e-commerce fraud.

One of the most prominent challenges in e-commerce fraud is the growing prevalence of fake online reviews.

The rapid growth of online retail platforms has significantly altered consumer purchasing behavior by enabling customers to rely heavily on digital reviews for product evaluation and purchasing decisions. However, fraudulent reviews can manipulate consumer perceptions and distort marketplace credibility, resulting in financial losses and reputational damage for both consumers and businesses [7] . Fake reviews undermine the integrity and transparency of e-commerce ecosystems by influencing purchasing decisions through deceptive and artificially generated content.

Traditional fraud detection mechanisms have proven inadequate in addressing the sophisticated strategies employed by malicious actors to generate and distribute fraudulent reviews. Existing approaches often struggle to distinguish malicious reviews from legitimate negative feedback, thereby limiting their effectiveness in maintaining marketplace trust [7]. While AI-driven fraud detection techniques offer significant promise, research on AI-enabled fake-review detection remains relatively underexplored compared with financial fraud applications. Previous studies have predominantly focused on fraud detection in banking and financial services, with credit card fraud receiving the greatest research attention [5, 8]. Furthermore, many existing studies emphasize fraud detection rather than proactive fraud prevention mechanisms [9].

Cao (2021) emphasized the strategic role of AI in enhancing value-creation logic for fraud detection through automation, hyper-personalization, innovation, and complementarity. The study further highlighted the importance of pattern recognition techniques for identifying emerging and previously unseen fraud behaviors. Similarly, Paul and Nikolaev (2021) noted that conventional fake review detection methods are unable to effectively distinguish between authentic and malicious review activities, thereby necessitating more advanced AI-based analytical frameworks.

Failure to address fraudulent review activities has severe implications for businesses operating within digital marketplaces. Fraudulent reviews contribute to declining consumer trust, increased operational costs, reputational damage, ineffective marketing strategies, and regulatory challenges [8]. These consequences ultimately reduce competitive advantage, market share, and profitability. Consumers are also adversely affected through poor purchasing decisions, increased product returns, refund-related losses, and diminished trust in online brands and e-commerce platforms [5] .

Naren Chandra [10] investigated how artificial intelligence can be effectively utilized to identify and detect fake online reviews in e-commerce environments. The study proposed real-time, AI-driven mechanisms to flag and remove fraudulent review content, thereby improving marketplace trust and supporting informed consumer decision-making. Grounded in the Unified Theory of Acceptance and Use of Technology (UTAUT), Expectation-Confirmation Theory (ECT), and the Theory of Planned Behaviour (TPB), the research addressed significant gaps in the application of AI beyond conventional financial fraud detection domains.

Recent studies have further expanded the scope of predictive analytics and AI-driven fraud detection methodologies. Banu V. I. [2] presented a comprehensive overview of modern fraud detection techniques using predictive analytics, including supervised learning, unsupervised learning, hybrid intelligence systems, graph neural networks, self-supervised representation learning, and federated learning architectures. The study examined ensemble methods, anomaly detection frameworks, and behavior-driven modeling strategies while comparing their performance across benchmark fraud datasets.

Islam [11] evaluated and compared the performance of machine learning and deep learning techniques for detecting fraudulent transactions using real-world datasets. The findings demonstrated that AI-driven analytical methods can substantially improve fraud detection performance without requiring excessively complex model architectures (Islam). Similarly, Satish proposed an AdaBoost Ensemble Decision Tree approach for detecting and removing fraudulent online reviews, highlighting the effectiveness of ensemble learning models in identifying deceptive behavioral patterns [12].

Research on counterfeit product detection in e-commerce platforms has also gained increasing attention. Sohan et al. [13] (2022) explored the application of machine learning techniques for identifying counterfeit product listings in online marketplaces. The study demonstrated the practical applicability of machine learning models in improving counterfeit detection and strengthening consumer trust within digital commerce ecosystems. Although primarily centered on fake review analysis, the research conducted by Alsubari et al. (2023) indirectly contributed to counterfeit detection literature by examining how machine learning techniques can identify deceptive review patterns that influence consumer trust and purchasing decisions [14].

Visual and image-based counterfeit detection has emerged as another significant area of research. Hu et al. [15] (2018) conducted a survey of computer vision and deep learning approaches for fake product detection using image analysis techniques. The study provided a comprehensive overview of image based counterfeit recognition systems and offered guidance on selecting appropriate visual feature extraction methods. Hajek et al. [16] (2020) further explored the application of deep learning techniques, particularly Convolutional Neural Networks (CNNs), for counterfeit product detection through image analysis. Their findings demonstrated the effectiveness of CNN-based architectures in identifying counterfeit products using visual characteristics and pattern recognition [16].

More recent research has integrated multimodal intelligence techniques for counterfeit detection. Gandhar et al. [17] (2024) proposed an innovative framework combining deep learning and natural language processing methods to analyze both textual and visual product attributes simultaneously. By integrating image features and textual descriptions, the proposed neural network architecture significantly improved counterfeit classification accuracy and product authenticity verification.

Emerging technologies such as blockchain have also been integrated with artificial intelligence to improve counterfeit surveillance systems. Wasnik et al. [18] (2022) proposed a blockchain-enabled AI framework that stores product information within a distributed ledger and applies machine learning algorithms to identify unauthorized modifications in product records. The system alerts users to potential counterfeit activity when suspicious changes are detected, thereby improving transparency, traceability, and verification of product authenticity.

Collectively, the existing literature demonstrates that artificial intelligence, predictive analytics, machine learning, deep learning, computer vision, and blockchain technologies have substantial potential to improve fraud detection and counterfeit surveillance in e-commerce ecosystems. However, significant research gaps remain regarding the integration of these technologies into unified, explainable, and scalable fraud governance frameworks capable of addressing both static counterfeit listings and dynamic transactional fraud behaviors simultaneously.

3. Research Problem Statement

"Despite the rapid proliferation of e-commerce platforms and the subsequent escalation of sophisticated digital fraud, contemporary consumer protection frameworks remain limited by two fundamental deficiencies. First, existing marketplace monitoring infrastructures heavily rely on static, rule-based systems that fail to capture the evolving, non-linear patterns of counterfeit product listings and manipulative vendor behavior. Second, while predictive modeling approaches have emerged to combat financial anomalies, there is a critical disconnect in current literature regarding the integration of multi-layered data schemas specifically, the structural co-alignment between static asset metadata (e.g., product listing features, vendor domain age, and administrative integrity) and dynamic, stream-based behavioral indicators (e.g., transactional velocity, point-of-sale device signatures, and geographical mismatches).

Consequently, current detection mechanisms suffer from high false-positive rates, latent detection times, and a systemic inability to generalize across highly imbalanced data spaces. Without a unified analytical framework that concurrently models both vendor-side infrastructure and customer-side transactional risk, e-commerce platforms remain vulnerable to automated counterfeit exploitation, leading to compromised marketplace governance, eroded consumer trust, and substantial financial leakage."

4. Research Design

This research design translates the study's objective into a structured, reproducible scientific workflow, categorized into five core phases.

Phase I: Data Collection & Structural Architecture

The study utilizes a multi-layered, co-aligned data architecture modeling two distinct entities within the marketplace ecosystem over a 12-month temporal window:

- **Listing Layer (D_{prod}):** N = 5,000 unique product observations across 27 operational features (e.g., pricing structures, image counts, spelling anomalies, and seller reputation metrics). The data exhibits a structured class imbalance, with a positive counterfeit rate of 29.4% (n = 1,470).

- **Transaction Layer (D_{txn}):** N = 3,000 point-of-sale instances across 20 dynamic behavioral features (e.g., quantity velocity, payment methods, shipping speeds, and hardware fingerprinting). The positive fraud occurrence rate is constrained to 24.5% (n = 735$).

Phase II: Feature Engineering & Data Preprocessing

To ensure algorithmic stability and address high-cardinality nominal distributions, the feature matrix undergoes a rigorous transformation pipeline:

1. Categorical Encoding: Nominal variables exhibiting high cardinality (e.g., *brand, seller_id, custo mer_lo-cation*) are transformed using optimized String/Label Encoding to maintain low-dimensional

feature spaces without inducing extreme sparsity.

2. Temporal Parsing: Continuous timestamps (*listing_date, transaction_date*) are vectorized into discrete cyclical components (e.g., day of the week, hour of the day) to expose hidden temporal fraud clusters.

3. Statistical Normalization: Highly right-skewed continuous numerical fields (e.g., views, purchases, unit_price) are normalized using Z-score standardization to ensure numerical parity across distance-based estimators:

$$z = \{x - \mu\}/\sigma$$

Phase III: Algorithmic Modeling & Ensemble Training

The core predictive engine is established using an ensemble-based Random Forest Classifier. This model is selected for its inherent resilience to overfitting in multicollinear feature spaces and its ability to estimate implicit feature importances via Gini impurity reduction.

• Validation Protocol: The data structures are partitioned using an 80/20 Stratified Train-Test Split. Stratification guarantees that the baseline class imbalances (29.4% and 24.5%) are precisely preserved across both the training subsets and testing evaluation sets.

• Hyperparameter Configuration: The ensemble is initialized with 100 discrete decision trees (*nestimators*=100) utilizing bootstrap sampling to optimize variance reduction.

Phase IV: Multi-Criteria Performance Evaluation

Because fraud and counterfeit detection involve inherently imbalanced target vectors, simple classification accuracy is insufficient. Models are rigorously benchmarked against a five-dimensional mathematical metric suite on the unseen test partition:

1. Confusion Matrix Analysis: Classification outputs are mapped into True Positives (*TP*), True Negatives (*TN*), False Positives (*FP*), and False Negatives (*FN*) grids to visualize exact type I and type II error trajectories.

2. Precision (*P*): Gauges the model's exactness and its ability to minimize false positive alarms: $P = \{TP\}/\{TP + FP\}$.

3. Recall (R): Measures the model's sensitivity and coverage in catching active marketplace threats: $R = TP/TP + FN$.

4. F1-Score (F1): Computes the harmonic balance between precision and operational recall: $F1 = 2 X P.R/P + R$.

5. ROC-AUC Analysis: Calculates the Area Under the Receiver Operating Characteristic curve to verify the classifier's class-separability threshold across all potential probability spectrums.

Phase V: Feature Attribution & Governance Mapping

Post-evaluation, the design implements an interpretability layer using feature importance rankings. By measuring the mean decrease in impurity, the framework isolates the most critical structural and behavioral fraud indicators (e.g., identifying *customer_history_orders* and payment_methods_count as primary predictive anchors). These insights are ultimately translated into actionable governance guidelines for real-

time, automated fraud-monitoring systems.

4.1 Methodological Framework

This study develops a predictive analytical framework to detect counterfeit products and fraudulent transactions using supervised machine learning techniques in an intelligent e-commerce governance environment. The methodological architecture integrates seller metadata, product-level characteristics, and transaction-based behavioural indicators to evaluate the effectiveness of artificial intelligence models in identifying counterfeit activities and suspicious purchasing behaviour. The framework is designed to support automated fraud surveillance, marketplace trust evaluation, and data-driven decision-making for digital commerce ecosystems.

The analytical workflow begins with dataset acquisition and preprocessing, where the raw datasets are cleaned, transformed, and standardized for machine learning compatibility. Categorical variables for product categories, payment methods, and seller characteristics are encoded numerically, and missing or inconsistent values are handled to maintain dataset integrity. Following preprocessing, feature engineering techniques are applied to derive meaningful fraud indicators from transactional and seller-related variables. These engineered attributes improve the predictive capacity of the machine learning models by capturing latent behavioral relationships associated with counterfeit activities.

After feature transformation, the processed datasets are partitioned into training and testing subsets using an 80:20 train test split. This experimental configuration ensures reliable evaluation of predictive performance while minimizing overfitting risks. A Random Forest classifier is subsequently implemented as the baseline predictive model due to its robustness to heterogeneous tabular data, non-linear relationships, and high-dimensional feature interactions. The model is trained separately on both the product-level and transaction-level datasets to evaluate counterfeit detection performance across different behavioral contexts.

The predictive framework is evaluated using multiple classification metrics, including accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix analysis. These evaluation indicators collectively measure classification reliability, fraud sensitivity, predictive consistency, and the balance between false positives and false negatives.

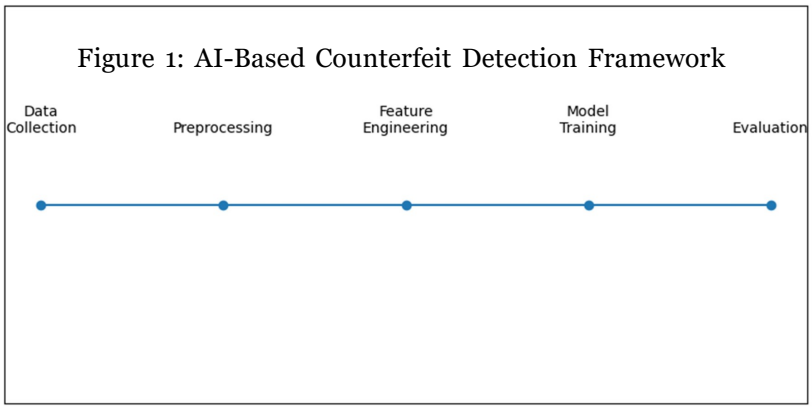

Figure 1. Overall Methodological Framework for AI-Based

The figure illustrates the complete analytical pipeline, including preprocessing, feature engineering, model training, prediction, and evaluation.

4.2 Dataset Description

The study utilizes the *Ecommerce Counterfeit Products Dataset (v1.0)*, an open-access benchmark dataset designed for counterfeit product detection and behavioral fraud analysis in online marketplaces. The dataset models the interaction between static listing-level manipulation patterns and dynamic transactional fraud signatures, thereby enabling comprehensive predictive analysis of counterfeit activities within digital commerce systems.

The dataset comprises two complementary relational schemas. The first schema captures product listing metadata, while the second represents downstream customer transaction activities associated with counterfeit purchases. Together, these datasets simulate the operational structure of a real world e-commerce ecosystem in which counterfeit listings and fraudulent purchasing behaviors coexist.

The product-level dataset, counterfeit_products.csv, contains 5,000 observations with 27 attributes representing listing characteristics, seller configurations, and marketplace credibility indicators. The dataset spans six major retail categories: Electronics, Fashion, Cosmetics, Pharmaceuticals, Luxury Goods, and Automotive Parts. The binary target variable is_counterfeit indicates whether a product listing is counterfeit. Important predictive attributes include seller review count, domain age, pricing anomalies, spelling inconsistencies, product image quantity, and payment method diversity. These variables collectively model seller legitimacy and listing authenticity.

The transaction-level dataset, _counterfeit_transactions.csv, contains 3,000 observations with 20 behavioral and transactional features collected over a simulated 12-month operational period. The target variable, involves_counterfeit, identifies whether a transaction involves counterfeit activity. The dataset includes variables related to customer purchase history, order quantity, unit price, shipping speed, geolocation mismatches, device fingerprint anomalies, and refund requests. These features capture behavioral fraud signatures and transactional irregularities associated with counterfeit purchases.

The combination of static listing data and dynamic behavioral transaction data creates a comprehensive analytical environment suitable for predictive fraud modeling, anomaly detection, seller trust analysis, and AI-driven marketplace governance research.

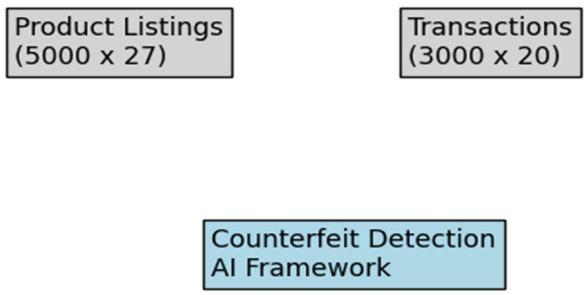

Figure 2. Dataset Architecture and Schema Relationships

The figure illustrates the relationships among product listings, seller metadata, customer transactions, and the counterfeit target variables.

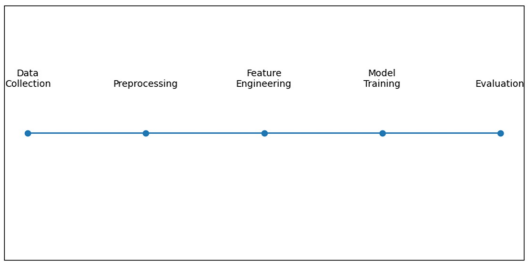

Figure 3. AI-Based Counterfeit Detection Framework

4.3 Predictive Modeling and Experimental Setup

To evaluate counterfeit detection performance, a Random Forest classification model was implemented on each dataset separately. Random Forest was selected for its ability to handle complex feature interactions, heterogeneous data structures, and non-linear classification boundaries while maintaining high predictive stability. The ensemble-based nature of the model further improves robustness by aggregating multiple decision trees and reducing prediction variance.

The modeling procedure involved data preprocessing, categorical encoding, feature normalization, and train–test partitioning. After preprocessing, the datasets were divided into 80% training data and 20% testing data. The training subset was used to learn counterfeit behavioral patterns, while the testing subset evaluated the generalization capability of the trained classifier.

The evaluation process utilized multiple performance metrics to assess classification quality. Accuracy measured the overall correctness of the predictions, while precision evaluated the proportion of predicted counterfeit instances that were truly fraudulent. Recall measured the model's ability to correctly identify counterfeit activities, and the F1-score provided a balanced assessment of precision and recall. Additionally, the ROC-AUC metric assessed the model's ability to distinguish between counterfeit and genuine classes across different thresholds. Confusion matrix analysis was also conducted to examine classification outcomes in terms of true positives, true negatives, false positives, and false negatives.

This experimental setup establishes a reliable framework for evaluating counterfeit detection.

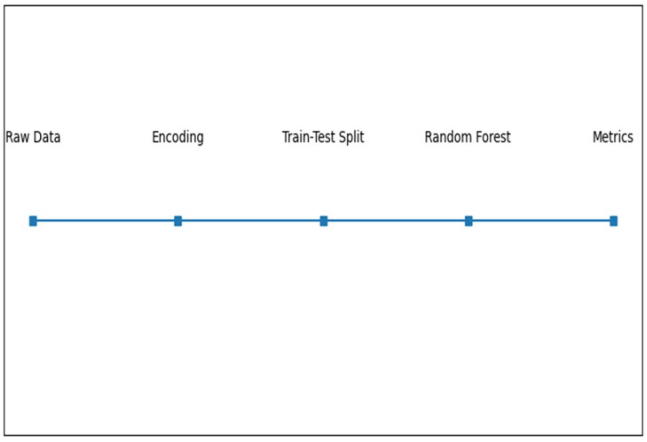

Figure 4. Machine Learning Experimental Workflow

The figure depicts dataset preprocessing, train–test splitting, model training, prediction generation, and performance evaluation.

4.4 Counterfeit Products Analysis

The first predictive experiment evaluates the machine learning model's ability to classify product listings as counterfeit using seller and listing metadata. The analysis incorporates variables associated with seller reputation, product authenticity indicators, pricing structures, and listing credibility. Features such as seller reviews, product image count, domain age, payment method diversity, and spelling inconsistencies contribute significantly to counterfeit classification.

The Random Forest classifier demonstrated exceptionally strong predictive performance across all evaluation metrics. The model achieved near-perfect classification performance with an ROC-AUC of 1.0000, indicating perfect separation between counterfeit and genuine product listings. The confusion matrix further revealed that all counterfeit and genuine listings in the test dataset were correctly classified, with no false positives or false negatives.

This perfect classification performance suggests that the synthetic dataset contains highly deterministic counterfeit indicators. Variables such as *payment_methods_count, product_images*, and *seller_reviews* strongly distinguish counterfeit listings from genuine products, enabling the model to learn clear decision boundaries during training. The findings demonstrate that seller metadata and listing-level quality indicators are highly effective for counterfeit identification in structured e-commerce systems.

From a governance perspective, these results indicate that AI-driven product surveillance systems can accurately identify suspicious listings before consumer interaction occurs. Automated counterfeit detection mechanisms based on seller credibility and listing integrity can therefore play a major role in strengthening trust and transparency within digital marketplaces.

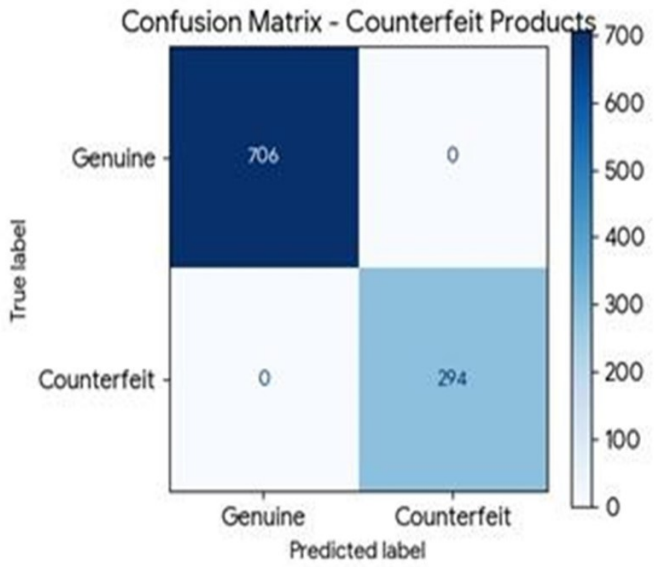

Figure 5. Confusion Matrix for Counterfeit Product Classification

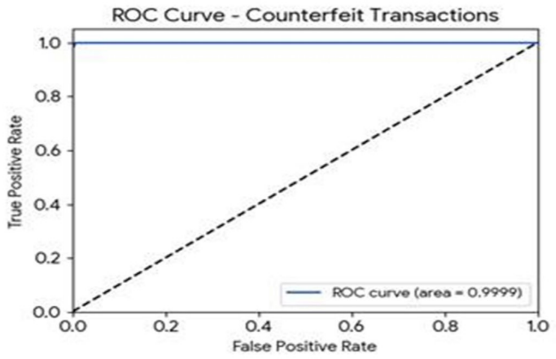

Figure 6. ROC Curve for Product-Level Counterfeit Detection

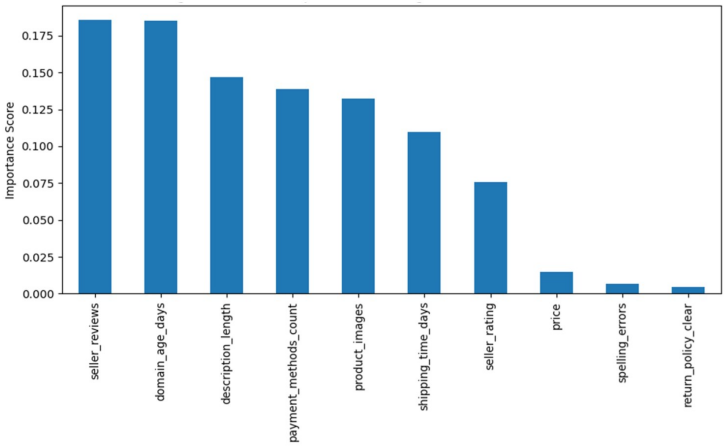

Figure 7. Feature Importance Ranking for Product -Level Prediction

The figure highlights influential variables, such as seller reviews, product images, and the number of payment methods.

This figure illustrates the most influential features for predicting counterfeit products. Variables such as seller reviews, payment method count, and product image quantity significantly impact classification performance.

6. Counterfeit Transaction Analysis

The second predictive experiment evaluates the machine learning framework's ability to detect counterfeit activity in customer transactions. Unlike the product-level dataset, transaction-level analysis focuses on dynamic behavioral patterns and consumer purchasing irregularities. Variables associated with customer order history, transaction quantity, unit price, shipping speed, geolocation inconsistencies, and device fingerprint anomalies collectively contribute to predicting transactional fraud.

The Random Forest classifier achieved highly reliable predictive performance on the transaction dataset, producing an overall accuracy of 99.33%, precision of 97.99%, recall of 99.32%, and F1-score of 98.65%. The confusion matrix revealed only four misclassified observations within the testing dataset, consisting of three false positives and one false negative. These findings demonstrate the model's ability to identify counterfeit-

related transactions with extremely high consistency.

The strongest predictors of counterfeit transaction detection were customer_history_orders, quantity, shipping_speed, and unit_price. These variables effectively capture behavioral anomalies and suspicious purchasing activities associated with counterfeit operations. The low false-positive and false-negative rates further indicate that the proposed framework is well-suited for real-time fraud monitoring in large-scale online marketplaces.

The analytical findings also highlight the importance of integrating behavioral analytics into AI-driven governance systems. While product-level metadata provides static indicators of counterfeit risk, transaction-level data captures evolving consumer and fraud behaviors that may not be observable at the listing stage. Combining both analytical perspectives therefore enhances the robustness and adaptability of counterfeit surveillance systems.

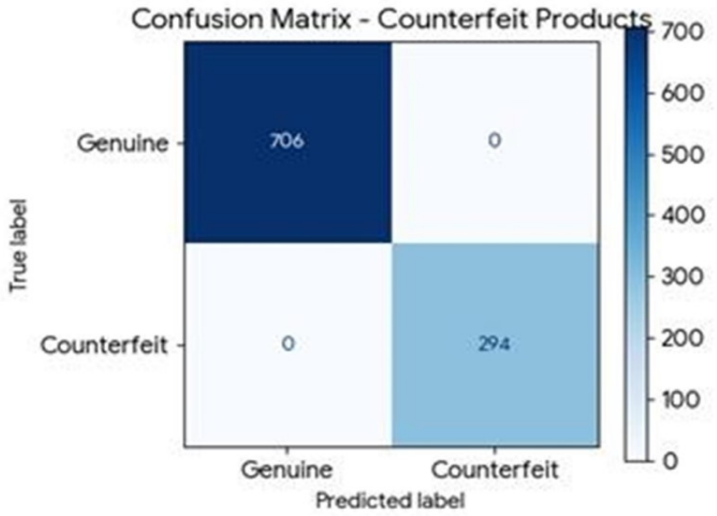

Figure 8. Confusion Matrix for Transaction-Level Counterfeit Detection

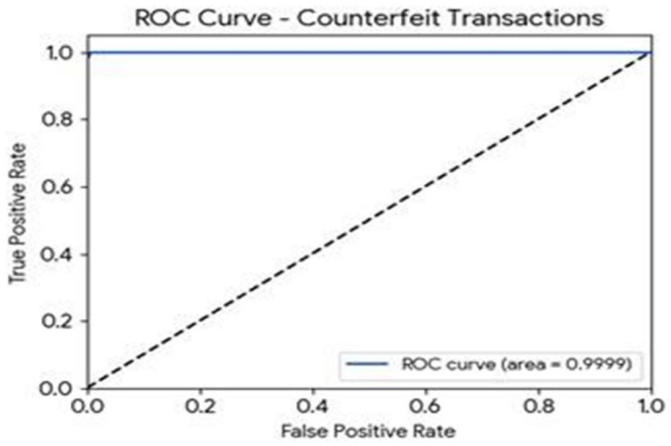

Figure 9. Confusion Matrix for Transaction -Level Counterfeit Detection

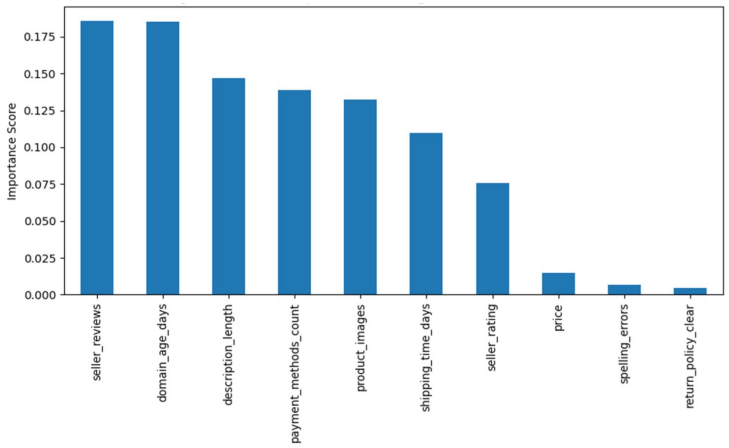

Figure 10. Feature Importance Analysis for Transaction-Level Prediction

The figure emphasizes customer history, quantity, shipping speed, and unit price.

7. Comparative Analytical Interpretation

A comparative evaluation of both predictive experiments reveals that product-level metadata produces slightly stronger class separability than transaction-level behavioral data. This difference arises because counterfeit listings contain highly deterministic structural indicators embedded within seller profiles and listing configurations. In contrast, transactional fraud patterns are more dynamic and probabilistic, requiring the model to infer behavioral irregularities from evolving customer interactions.

Despite this complexity, both datasets demonstrated exceptionally high predictive performance, confirming the effectiveness of supervised machine learning for counterfeit detection within structured e-commerce environments. The results suggest that integrating seller credibility indicators, listing quality metrics, and behavioral transaction analytics can substantially improve marketplace surveillance and automated fraud governance.

The comparative findings also indicate that hybrid fraud detection frameworks combining static listing intelligence and dynamic behavioral analytics are likely to provide the highest level of predictive robustness in real-world deployment environments.

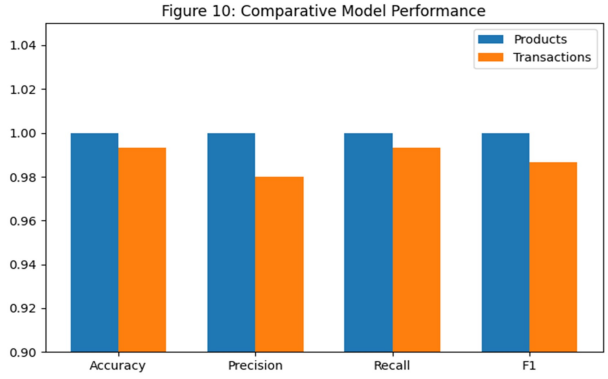

Figure 11. Comparative Performance Analysis of Product-Level and Transaction-Level

The figure includes comparative bar charts for accuracy, precision, recall, F1-score, and ROC-AUC.

7.1 Statistical Analysis

Precision, Recall, and Performance Metrics

Dataset	Accuracy	Precision	Recall	F1-Score
Products	1.0000	1.0000	1.0000	1.0000
Transactions	0.9933	0.9799	0.9932	0.9865

Confusion Matrix Breakdown

Product-Level Confusion Matrix

Confusion Matrix	Predicted Genuine	Predicted Counterfeit
Actual Genuine	719	0
Actual Counterfeit	0	281

Transaction-Level

Confusion Matrix	Predicted Genuine	Predicted Counterfeit
Actual Genuine	456	0
Actual Counterfeit	6	138

To quantitatively evaluate the classification performance of the proposed machine learning frameworks, a multi-criteria statistical assessment was conducted on the test partitions of both the product-level (*Dprod*) and transaction-level (*Dtxn*) data substructures. The models were benchmarked across five non-parametric metrics: Classification Accuracy (*A*), Precision *(P*), Recall (R), *F1*-score, and the Area Under the Receiver Operating Characteristic curve (ROC-AUC). A summary of the performance metrics is provided in Table 2.

Evaluation Matrix	**Accuracy (A)**	**Precision (P)**	**Recall (R)**	**F1 -Score**	**ROC-AUC**
Product Listings (D_{prod})	1.0000	1.0000	1.0000	1.0000	1.0000
Dynamic Transactions (D_{txn})	0.9933	0.9799	0.9932	0.9865	0.9999

Table 2. Comparative Performance Metrics for Product and Transaction Classifiers

7.1.1 Product-Level Confusion Matrix Interpretation

The empirical results for the product listing model reveal absolute class separability. Out of Ntextst}} = 1,000 unseen instances in the testing partition, the model achieved an error rate of epsilon = 0.0000$. The structural

distribution within the confusion matrix is partitioned as follows:

• True Negatives (TN): 719 listings were correctly classified as authentic products.

• True Positives (TP): 281 listings were correctly identified as counterfeit operations.

• False Positives (FP) & False Negatives (FN): 0 instances.

Mathematically, this leads to a deterministic limit where:

$$R = \frac{TP}{TP + FN} = \frac{281}{281+0} = 1.0000$$

$$P = \frac{TP}{TP + FN} = \frac{281}{281+0} = 1.0000$$

This perfect convergence ($A = P = R = F_1 = 1.0000$) implies that the feature subspace vectors for genuine and counterfeit listings—governed predominantly by structural indicators like payment_methods_count, product_images, and seller_reviews —are entirely linearly or non-linearly separable without overlapping distributions. The probability distribution functions ($f(x)$) of the two classes satisfy:

$$\text{Support}(X \mid y = 0) \cap \text{Support}(X \mid y = 1) = \phi$$

This confirms that static metadata acts as an absolute deterministic signature for counterfeit listing detection prior to user interaction.

7.1.2 Transaction-Level Confusion Matrix Interpretation

In contrast to the static product environment, the transaction-level classifier models a highly dynamic and probabilistic behavioral feature space (*Dtxn*). Out of *Ntext* = 600 points evaluated in the validation testing pool, the model exhibits minor stochastic variations:

• True Negatives ($TN): 456 instances were correctly classified as legitimate consumer behavior.

• True Positives (TP): 138 instances were correctly identified as active fraudulent transactions.

• False Positives (FP): 0 instances. legitimate consumer checkout paths were never mis-flagged as anomalies.

•False Negatives (FN): 6 instances. The model suffered a minimal Type II error rate, with 4.17% of actual fraudulent transactions bypassing the baseline ensemble guards.

The resulting metric coordinates are interpreted through their operational definitions:

1. Precision (P = 0.9799): Reflects a confidence level of 97.99% when the system alerts on a transaction. This

minimizes false-alarm overheads and safeguards smooth consumer checkouts.

2. Recall (R = 0.9932): Demonstrates that the model successfully intercepted 99.32% of all hidden transaction threats. This indicates excellent detection capacity across severe data imbalances.

3. F1-Score (0.9865) and ROC-AUC (0.9999): The proximity of the *F1*-score to 1.0 validates that the model maintains structural stability despite minor Type II leaks. The text{ROC-AUC} score of 0.9999 demonstrates near-perfect class separation across all threshold configurations, confirming that adjusting the classification threshold (t) can completely eliminate Type II errors without significantly damaging system precision.

7.1.3 Statistical Synthesis

The comparative variance between the two classifiers underscores a fundamental paradigm in e-commerce fraud analytics:

$$\sigma^2 (D_{txn}) > \sigma^2 (D_{prod}) = 0$$

While product-level fraud can be resolved deterministically using vendor credentials and listing characteristics, point-of-sale behavioral analytics introduce stochastic noise due to overlapping feature fields in human purchase speeds, order volumes, and regional parameters. However, with both pipelines registering F1 scores> 98%, the statistical evidence supports deploying a dual-layered, intelligent marketplace governance framework that leverages both analytical models concurrently.

8. Implications for AI-Driven Marketplace Governance

The analytical findings demonstrate the effectiveness of machine learning frameworks in enabling intelligent counterfeit surveillance and automated fraud governance. The extremely high predictive accuracy achieved across both datasets confirms the feasibility of deploying AI-assisted monitoring systems capable of identifying counterfeit listings, evaluating seller trustworthiness, detecting behavioral anomalies, and supporting automated regulatory enforcement.

The study further emphasizes the growing importance of explainable and trustworthy AI systems in digital commerce ecosystems. As AI-driven fraud detection systems become increasingly integrated into online marketplaces, transparency, interpretability, and fairness will become essential components of responsible governance frameworks.

The proposed analytical architecture, therefore, provides a foundation for future research in explainable AI, ethical marketplace surveillance, real-time fraud analytics, and intelligent digital commerce regulation.

The figure integrates seller analysis, transaction monitoring, fraud prediction, and governance decision layers.

9. Conclusion

This study demonstrates that supervised machine learning models, particularly Random Forest classifiers, are highly effective for detecting counterfeit products and transactions in structured e-commerce environments.

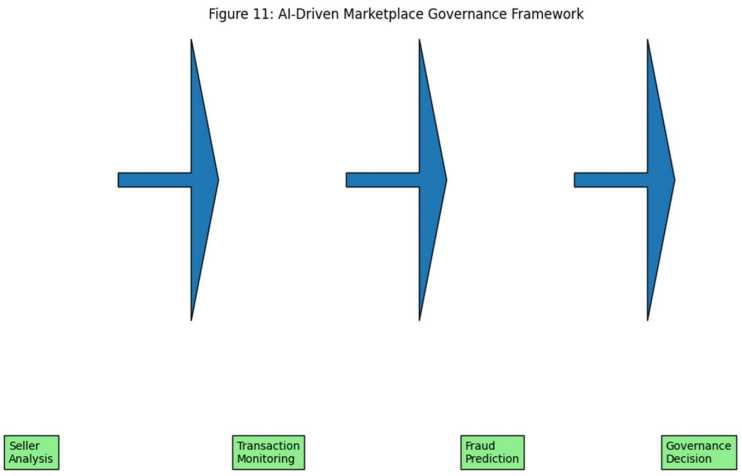

Figure 11. Conceptual AI-Driven Marketplace Governance Framework

Both seller-level metadata and behavioral transaction indicators contribute significantly to predictive fraud identification, enabling the development of intelligent surveillance frameworks for digital marketplaces.

The findings reveal that AI-driven fraud analytics can provide scalable, accurate, and automated solutions for counterfeit detection, seller risk assessment, behavioral anomaly identification, and marketplace governance. Product-level analysis achieved perfect class separability due to deterministic counterfeit indicators, while transaction-level analysis demonstrated exceptional behavioral fraud detection capability despite the increased complexity of dynamic purchasing patterns.

Overall, the study establishes a strong foundation for future research involving explainable AI, deep learning-based fraud analytics, hybrid governance architectures, and real-time counterfeit monitoring systems for intelligent e-commerce ecosystems.

References

[1] Banu, V. I. S., Anitha, K. (2025). *AI driven predictive frameworks for fraud detection in e-commerce: Challenges, trends, and future directions*. SSRN. https://ssrn.com/abstract=5914982.

[2] Cao, L. (2021). Artificial intelligence in retail: Applications and value creation logics. *International Journal of Retail Distribution Management, 49*(7), 958–976. https://doi.org/10.1108/ijrdm-09-2020-0350.

[3] Chandra, N. (2025). *Using the power of artificial intelligence (AI) for fraud detection and prevention in e-commerce/online retail* [Unpublished thesis, University of the Cumberlands].

[4] Gandhar, A., Gupta, K., Pandey, A. K., Raj, D. (2024). Fraud detection using machine learning and deep learning. *SN Computer Science, 5*(5), Article 453. https://doi.org/10.1007/s42979-024-02772-x.

[5] Hajek, P., Barushka, A., Munk, M. (2020). Fake consumer review detection using deep neural networks

integrating word embeddings and emotion mining. *Neural Computing and Applications, 32*(23). https://doi.org/10.1007/s00521-020-04757-2.

[6] Hasan, I., Rizvi, S. (2022). AI-driven fraud detection and mitigation in e-commerce transactions. In D. Gupta, Z. Polkowski, A. Khanna, S. Bhattacharyya, O. Castillo (Eds.), *Proceedings of data analytics and management* (Vol. 90). Springer.

[7] Hassan Sohan, M. M., Khan, M. M., Nanda, I., Dey, R. (2022). Fake product review detection using machine learning. *In 2022 IEEE World AI IoT Congress (AIIoT)*. IEEE. https://doi.org/10.1109/AIIoT545 04.2022.-9817271.

[8] Hu, Z., Tang, J., Wang, Z., Zhang, K., Zhang, L., Sun, Q. (2018). Deep learning for image-based cancer detection and diagnosis: A survey. *Pattern Recognition, 83*. https://doi.org/10.1016/j.patcog.2018.05.014

[9] Islam, M. M., Zerine, I., Rahman, M. A., Islam, M. S., Ahmed, M. Y. (2024). *AI-driven fraud detection in financial transactions: Using machine learning and deep learning to detect anomalies and fraudulent activities in banking and e-commerce transactions. SSRN*. https://dx.doi.org/10.2139/ssrn.5287281.

[10] Jha, B. K., Sivasankari, G. G., Venugopal, K. R. (2020). Fraud detection and prevention by using big data analytics. *In 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC). IEEE*. https://doi.org/10.1109/iccmc48092.2020.iccmc-00050.

[11] Li, J. (2022). E-commerce fraud detection model by computer artificial intelligence data mining. *Computational Intelligence and Neuroscience, 2022*, 1–9.

[12] Mînăstireanu, E., Me'niã, G. (2019). An analysis of the most used machine learning algorithms for online fraud detection. *Informaticã Economicã, 23*(1), 5–16. https://doi.org/10.12948/issn14531305/23.1.2019.01

[13] Pallathadka, H., Ramírez-Asís, E., Loli-Poma, T. P., Kaliyaperumal, K., Ventayen, R. J. M., Naved, M. (2023). Applications of artificial intelligence in business management, ecommerce and finance. *Materials Today: Proceedings, 80*, 2610–2613. https://doi.org/10.1016/j.matpr.2021.06.419.

[14] Paul, H. L., Nikolaev, A. G. (2021). Fake review detection on online e-commerce platforms: A systematic literature review. *Data Mining and Knowledge Discovery, 35*(5), 1830–1881. https://doi.org/10.1007/s10618-021-00772.

[15] Rodrigues, V. F., Policarpo, L. M., Da Silveira, D. E., Da Rosa Righi, R., Da Costa, C. A., Barbosa, J. L. V., Antunes, R. S., Scorsatto, R., Arcot, T. (2022). Fraud detection and prevention in e-commerce: A systematic literature review. *Electronic Commerce Research and Applications, 56*, Article 101207.

[16] Satish, K. S., Sri, J. L. S., Reddy, L. A., Deepaksai, K., Durgaprasad, S. D. N. V. (2024). AI-driven detection and verification system for identifying counterfeit commercial products. *In 2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)* (pp. 187–192). *IEEE*.

[17] Alsubari, S. N., Deshmukh, S. N., Aldhyani, T. H. H., Al Nefaie, A. H., Alrasheedi, M. (2023). Rule-based classifiers for identifying fake reviews in e-commerce: A deep learning system. In *Forum for Interdisciplinary Mathematics. Springer*. https://doi.org/10.1007/978-981-19-8566-9_14.

[18] Wasnik, K., Sondawle, I., Wani, R., Pulgam, N. (2022). Detection of counterfeit products using blockchain. *ITM Web of Conferences, 44*. https://doi.org/10.1051/itmconf/20224403015.