

Detecting Fake Products in African E-Commerce Markets: An Artificial Intelligence Framework Using a Tanzania-Informed Simulation

Kenneth Longo Mlelwa (PhD)

Senior Lecturer - ICT Department the Mwalimu Nyerere Memorial Academy, Dar es Salaam,
Tanzania

DOI: <https://doi.org/10.51583/IJLTEMAS.2026.150900063>

Received: 13 September 2026; Accepted: 18 September 2026; Published: 10 October 2026

ABSTRACT

The rapid expansion of e-commerce in Africa has increased consumer access to goods while also creating opportunities for counterfeit and misleading product listings. This study proposes a multi-signal artificial intelligence framework for counterfeit-product risk detection and evaluates it through a reproducible Tanzania-informed simulation rather than through direct observation of Tanzanian marketplace transactions. A synthetic dataset of 1,000 product listings was generated with a 70:30 genuine-to-counterfeit class ratio. Five marketplace signals were modelled: price deviation, seller reputation, image similarity, review sentiment, and description quality. The revised simulation protocol specifies the class-conditional distributions, contamination assumptions, random seed, train-test split, cross-validation strategy, hyperparameter search, and evaluation metrics. Four supervised classifiers - logistic regression, support vector machine, random forest, and XGBoost - were compared. No raw product images were used to train a convolutional neural network; instead, image similarity was treated as a scalar signal that could in practice be produced by an upstream visual-comparison component. On a stratified 20% holdout set, logistic regression achieved 0.920 accuracy, 0.867 precision, 0.867 recall, 0.867 F1-score, and 0.950 ROC-AUC. Repeated stratified five-fold cross-validation produced mean accuracies of approximately 0.889-0.893 across the four models. The findings therefore support the technical feasibility of multi-signal screening under controlled simulated conditions, not operational validation in Tanzanian e-commerce. The paper concludes that future deployment should be validated with real African marketplace data, raw product imagery, network-level seller relationships, and explainable decision support for human moderators.

Keywords: African digital markets, artificial intelligence, counterfeit detection, e-commerce, machine learning, simulation, Tanzania

INTRODUCTION

Across Africa, e-commerce has become an increasingly important component of the digital economy as mobile connectivity, digital payments, and platform-based retail expand. In Tanzania, earlier work has documented the importance of trust, infrastructure, mobile technology, and institutional conditions to e-commerce development (Oreku, Mtenzi, & Ali, 2013; Makame, Kang, & Park, 2014; Kabanda & Brown, 2015; Mlelwa, Chachage, & Zaipuna, 2015). Mobile money has also reduced transaction frictions and reshaped digital exchange in the country (Economides & Jeziorski, 2017). These developments create opportunities for digital entrepreneurship, but they also increase the importance of reliable platform governance.

Counterfeit product listings are a platform-trust problem rather than only a brand-protection problem. A suspicious listing may combine several weak signals: an unusually low price, a recently created or poorly rated seller account, product imagery that resembles an authentic item but contains subtle inconsistencies, promotional reviews, or textual descriptions that differ from normal marketplace practice. None of these

signals is decisive in isolation. For that reason, a practical screening system should integrate multiple indicators and route high-risk cases for human review rather than rely on a single threshold.

Research on online deception has developed along several related paths. Early studies established the importance of reviews and trust in digital purchasing (Chen & Xie, 2008; Chiu, Wang, Fang, & Huang, 2014). Work on opinion spam then introduced linguistic, behavioral, and group-level techniques for identifying manipulated reviews and reviewers (Jindal & Liu, 2008; Ott, Choi, Cardie, & Hancock, 2011; Mukherjee, Liu, & Glance, 2012; Savage, Zhang, Yu, Chou, & Wang, 2015). More recent studies have used neural representations and comprehensive machine-learning pipelines for fake-review detection (Ren & Ji, 2017; Mewada & Dewang, 2022; Gupta, Jindal, & Kashyap, 2024). Image-oriented counterfeit research has meanwhile shown the value of fine-grained visual cues (Peng, Zou, He, & Zhu, 2022), and multimodal representation learning has expanded the technical possibilities for combining language and visual signals (Radford et al., 2021).

Despite this progress, two limitations remain relevant to African digital markets. First, publicly available labelled datasets that combine product, seller, review, and image-related signals are scarce. Second, evidence from adoption and trust studies should not be confused with empirical evidence of counterfeit detection performance. This paper therefore uses a deliberately limited claim: it evaluates a proposed detection architecture under a Tanzania-informed simulation. The simulation is motivated by documented features of the Tanzanian digital-commerce environment, but its numerical parameters are design assumptions for reproducible experimentation rather than estimates derived from a Tanzanian marketplace dataset.

The research question is: how effectively can conventional supervised machine-learning models discriminate between genuine and counterfeit product profiles when several marketplace signals are combined under a transparent synthetic design? The objective is not to demonstrate deployment-ready detection, but to test whether the proposed multi-signal architecture is technically coherent enough to justify subsequent validation on real marketplace data.

Contributions and Scope

This study makes four contributions. First, it frames counterfeit-product screening as a multi-signal classification problem that combines product, seller, review, and image-similarity information. Second, it replaces an under-specified synthetic experiment with a reproducible simulation protocol that states feature distributions, class proportions, overlap assumptions, contamination rates, and the random seed. Third, it evaluates four tabular classifiers with hyperparameter tuning, a stratified holdout test, repeated cross-validation, ROC-AUC, F1-score, confusion analysis, confidence intervals, and a non-parametric statistical comparison. Fourth, it explicitly separates proof-of-concept simulation results from real-world Tanzanian evidence.

The scope is intentionally constrained. The study does not scrape or observe live Tanzanian e-commerce platforms, does not use personal data, and does not train a convolutional neural network on raw product images. The image-similarity variable is a simulated score representing the type of signal that could be produced by an upstream image-comparison model. This distinction prevents the results from being interpreted as evidence of direct visual learning or operational marketplace performance.

LITERATURE REVIEW AND RESEARCH GAP

Trust, Reviews, and Deceptive Marketplace Behaviour

Online reviews affect consumer search and purchase decisions and have therefore become an important component of marketplace trust (Chen & Xie, 2008; Chiu et al., 2014). This same influence creates incentives for manipulation. Jindal and Liu (2008) formalized opinion-spam detection in product reviews, while Ott et al. (2011) demonstrated that deceptive reviews can be detected from linguistic cues. Mukherjee

et al. (2012) extended the problem to coordinated reviewer groups, showing that fraud can be relational rather than purely textual. Savage et al. (2015) further showed that anomalous rating behavior can reveal suspicious reviewers even without text analysis.

Later work moved toward learned representations and broader survey frameworks. Ren and Ji (2017) used neural models to learn document-level representations for deceptive opinion-spam detection. Mewada and Dewang (2022) synthesized false-review detection methods and emphasized persistent challenges in labelled data, feature design, and generalization. Gupta et al. (2024) similarly identified limitations related to training-data scarcity, concept drift, and inconsistency across datasets. These findings support a multi-signal approach because text or ratings alone cannot represent the full risk profile of a product listing.

Image-Based and Multimodal Counterfeit Detection

Counterfeit products frequently imitate authentic packaging, logos, textures, and visual layout. Peng et al. (2022) demonstrated that attention-based image models can capture fine-grained visual differences in luxury handbag authentication. Such studies justify visual analysis as a potentially valuable component of counterfeit detection, but they also illustrate why claims about image learning require actual image datasets and documented architectures.

Modern multimodal methods provide additional possibilities. Transformer-based language models such as BERT can produce contextual text representations (Devlin, Chang, Lee, & Toutanova, 2019), while visual-language models such as CLIP align image and text representations in a common embedding space (Radford et al., 2021). A future operational system could therefore compare product imagery against authentic exemplars while simultaneously analysing titles, descriptions, and reviews. The present study does not implement that full multimodal stack; instead, it treats image similarity as one input signal in a tabular proof of concept.

Graph-Based Fraud Detection and Explainability

Marketplace fraud also has a network structure. Sellers, reviewers, devices, payment accounts, and products may form repeated or coordinated relationships that are difficult to detect from isolated listings. Graph-neural-network approaches such as CARE-GNN have shown how relational information can improve fraud detection under camouflage behavior (Dou, Liu, Sun, Deng, Peng, & Yu, 2020). This suggests that seller-network analysis is an important extension for future African marketplace research.

Explainability is equally important because automated moderation affects sellers and consumers. Model-agnostic methods such as LIME (Ribeiro, Singh, & Guestrin, 2016) and SHAP (Lundberg & Lee, 2017) can provide local or feature-level explanations for predictions. In a production setting, an explanation such as 'high price deviation combined with weak seller reputation' is more actionable for moderators than a risk score alone. In this simulation, permutation importance is used as a global diagnostic, while local explanation methods are proposed for future deployment.

African and Tanzanian Context

Research on Tanzania has largely examined e-commerce adoption, infrastructure, trust, and digital-payment conditions rather than labelled counterfeit-product datasets. Oreku et al. (2013) emphasized implementation barriers and trust, Makame et al. (2014) examined adoption factors, Kabanda and Brown (2015) identified enablers and barriers among Tanzanian SMEs, and Mlelwa et al. (2015) described the development of e-commerce and associated security concerns. Economides and Jeziorski (2017) documented the importance of mobile money in Tanzania's digital economy. These studies justify Tanzania as a relevant contextual setting, but they do not provide empirical labels for counterfeit listings.

The resulting research gap is therefore not simply the absence of an algorithm. It is the absence of a transparent bridge between contextual African e-commerce research and reproducible counterfeit-detection

experimentation. This paper addresses that gap at the proof-of-concept stage by making the simulation assumptions explicit and by avoiding claims of direct field validation.

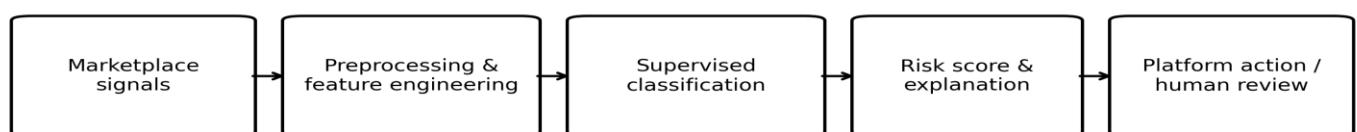
Table 1. Selected literature and the gap addressed by the present study

Study	Primary focus	Method / evidence	Implication for this study
Jindal & Liu (2008)	Opinion spam	Text/review features	Review signals matter but do not cover product authenticity.
Mukherjee et al. (2012)	Reviewer groups	Behavioral/network patterns	Fraud can be coordinated and relational.
Peng et al. (2022)	Counterfeit products	Fine-grained image model	Visual cues are useful when raw images are available.
Gupta et al. (2024)	Fake-review detection	Comprehensive review	Data scarcity and generalization remain major problems.
Dou et al. (2020)	Fraud networks	Graph neural network	Seller/reviewer relations are a future extension.
Oreku et al. (2013); Makame et al. (2014); Kabanda & Brown (2015)	Tanzanian e-commerce	Context/adoption studies	Support contextual framing, not counterfeit labels.

PROPOSED ARTIFICIAL INTELLIGENCE FRAMEWORK

The proposed framework is designed as a decision-support pipeline rather than an autonomous enforcement mechanism. Five signals enter the model: price deviation from a reference price, seller reputation, image similarity to authentic exemplars, review sentiment, and description quality. These variables are preprocessed into a feature vector and supplied to a supervised classifier that produces a counterfeit-risk probability. The probability can then be converted into a platform action such as no action, warning, temporary hold, or human moderation review.

Figure 1 summarizes the architecture. In an operational implementation, feature generation would be modular. Price and seller variables would come from marketplace transaction records; review and description features could be produced by transformer-based language models; image similarity could be produced by a visual encoder or visual-language model; and seller/reviewer relations could be incorporated through graph features. The present experiment evaluates only the tabular classification layer using simulated feature values.



Inputs: price deviation • seller reputation • image similarity • review sentiment • description quality

Figure 1. Proposed multi-signal counterfeit-product risk detection framework.

METHODOLOGY

Study Design and Classification Task

The study uses a controlled synthetic experiment. Each listing i is represented by a five-dimensional feature vector $X_i = [P_i, S_i, I_i, R_i, D_i]$, where P is price deviation, S is seller reputation, I is image similarity, R is review sentiment, and D is description quality. The target y_i takes the value 0 for a genuine listing and 1 for a counterfeit listing.

The dataset contains 1,000 listings with a 70% genuine and 30% counterfeit class distribution. The 70:30 ratio is a modelling assumption chosen to represent a marketplace in which counterfeit listings are material but not dominant; it is not an estimate of counterfeit prevalence in Tanzania. A fixed random seed of 42 is used to make the simulation reproducible.

Tanzania-Informed Simulation Protocol

The simulation is 'Tanzania-informed' only in its contextual motivation: the choice of price, seller reputation, reviews, and platform trust as core signals is consistent with the broader literature on Tanzanian digital commerce. The numerical distributions below are explicit experimental design assumptions used to create overlapping genuine and counterfeit profiles. They are not fitted to or inferred from Tanzanian marketplace transactions.

Continuous variables are bounded to operationally interpretable ranges. Price deviation is expressed as the absolute proportional difference from a reference market price. The remaining four variables are normalized to $[0,1]$, where higher values represent stronger seller reputation, greater visual similarity to an authentic exemplar, more positive review sentiment, and higher description quality. Beta distributions are used for bounded variables because they allow asymmetric but overlapping class profiles.

Conditional on class membership, the five features were initially sampled independently, and no cross-feature correlation structure was imposed before the contamination step. Class labels were assigned before feature generation rather than derived from a deterministic threshold rule.

Table 2. Feature-generation assumptions used in the revised simulation

Feature	Genuine-listing distribution	Counterfeit-listing distribution	Range / interpretation
Price deviation	Truncated Normal: mean 0.08, SD 0.07	Truncated Normal: mean 0.28, SD 0.14	0-0.50 (genuine), 0-0.75 (counterfeit); larger = greater deviation
Seller reputation	Beta(8, 2)	Beta(4, 4.5)	0-1; larger = stronger reputation
Image similarity	Beta(9, 2)	Beta(5, 3.5)	0-1; larger = closer to authentic exemplar
Review sentiment	Beta(7, 3)	Beta(5.5, 3.5)	0-1; larger = more positive sentiment
Description quality	Beta(7, 2.5)	Beta(4.5, 3.5)	0-1; larger = more complete/coherent description

To avoid making the classes unrealistically separable, overlap is introduced deliberately. Fifteen percent of counterfeit listings are treated as high-imitation cases and are resampled closer to the genuine distributions

across the five features. Seven percent of genuine listings are treated as low-quality or atypical genuine cases and are resampled toward the counterfeit distributions. This contamination step creates ambiguity that better reflects the screening nature of the problem.

The resulting dataset preserves the 70:30 class ratio shown in Figure 2. Because the experiment does not use real identities, transactions, or raw marketplace content, no personal or confidential data are involved.

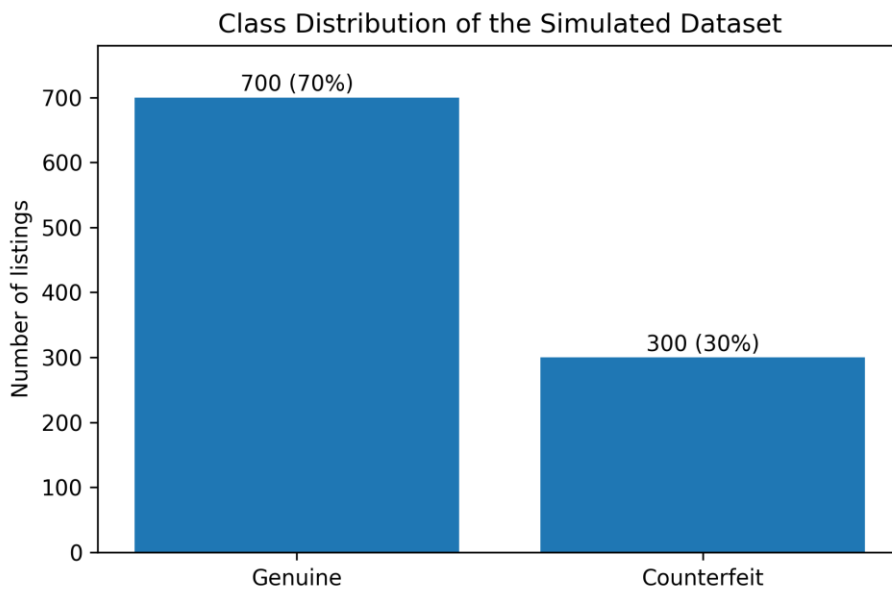


Figure 2. Class distribution of the 1,000-listing synthetic dataset.

Clarification of the Image-Similarity Variable

The original conceptual framework considered convolutional neural networks as a possible image-analysis component. However, the present dataset contains no raw product images, so training a CNN would not be methodologically defensible. The revised experiment therefore removes CNN as a candidate classifier.

Image similarity is instead represented as a normalized scalar feature. Conceptually, this value could be generated by an upstream image-matching system that compares a marketplace image with known authentic exemplars. A future empirical study with raw images should document image sources, sample counts, dimensions, preprocessing, augmentation, architecture, optimizer, learning rate, batch size, epochs, and train-validation-test separation before making claims about direct visual learning. Image-focused counterfeit work such as Peng et al. (2022) and multimodal visual-language models such as CLIP (Radford et al., 2021) provide appropriate directions for that extension.

Model Training, Hyperparameter Selection, and Evaluation

The simulated data are divided into an 80% training set and a 20% holdout test set using stratification to preserve the 70:30 class distribution. Hyperparameters are selected only from the training partition using five-fold GridSearchCV with ROC-AUC as the selection criterion. Logistic regression and SVM inputs are standardized. Logistic regression, SVM, and random forest use balanced class weights. XGBoost uses a positive-class weight equal to the training-set ratio of genuine to counterfeit observations.

For reproducibility, all stochastic data splits, cross-validation procedures, and model components supporting a random-state parameter were initialized with `random_state = 42`.

Four candidate models are evaluated: logistic regression, radial-basis-function SVM, random forest, and XGBoost. Logistic regression provides an interpretable linear baseline; SVM represents a nonlinear margin-

based classifier; random forest captures nonlinear interactions through bagged decision trees; and XGBoost provides a boosted-tree model (Chen & Guestrin, 2016).

Table 3. Selected model specifications after training-set grid search

Model	Selected hyperparameters	Class-imbalance treatment
Logistic regression	C = 1; standardized features	class_weight = balanced
SVM (RBF)	C = 1; gamma = 0.1; standardized features	class_weight = balanced
Random forest	200 trees; max_depth = 6; min_samples_leaf = 1	class_weight = balanced
XGBoost	150 trees; max_depth = 3; learning_rate = 0.03; subsample = 0.8	scale_pos_weight = genuine/counterfeit ratio

Holdout performance is reported with accuracy, precision, recall, F1-score, and ROC-AUC. Confusion matrices are used to inspect false positives and false negatives. To assess variability, the selected model specifications are also evaluated using repeated stratified five-fold cross-validation with five repetitions (25 validation folds). Mean scores and approximate 95% confidence intervals are reported from the fold-level variability.

A Friedman test is used to compare repeated cross-validation accuracy across the four models because the same folds are evaluated by each classifier. Pairwise Wilcoxon signed-rank tests with multiplicity adjustment are used only as a follow-up interpretation. Permutation importance on the holdout set is used as a global diagnostic of feature influence. This analysis does not replace local explanations such as LIME or SHAP, which would be preferable for explaining individual moderation decisions (Ribeiro et al., 2016; Lundberg & Lee, 2017).

5. RESULTS

5.1 Holdout Test Performance

Table 4. Performance on the stratified 20% holdout test set

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	0.920	0.867	0.867	0.867	0.950
SVM	0.915	0.852	0.867	0.860	0.948
Random Forest	0.900	0.833	0.833	0.833	0.943
XGBoost	0.910	0.839	0.867	0.852	0.939

All four models achieved useful discrimination under the simulated conditions. Logistic regression produced the strongest holdout accuracy (0.920) and ROC-AUC (0.950), while SVM produced very similar results. XGBoost achieved a slightly higher recall than its earlier unweighted configuration after class weighting, illustrating the importance of accounting for the 70:30 class distribution.

The similarity of the results is more important than a claim that one model is universally superior. The dataset was generated from five relatively low-dimensional signals, so a well-regularized linear model can perform competitively when class separation is approximately monotonic. More complex models may

become advantageous when real data introduce nonlinear interactions, high-dimensional text embeddings, image embeddings, and network features.

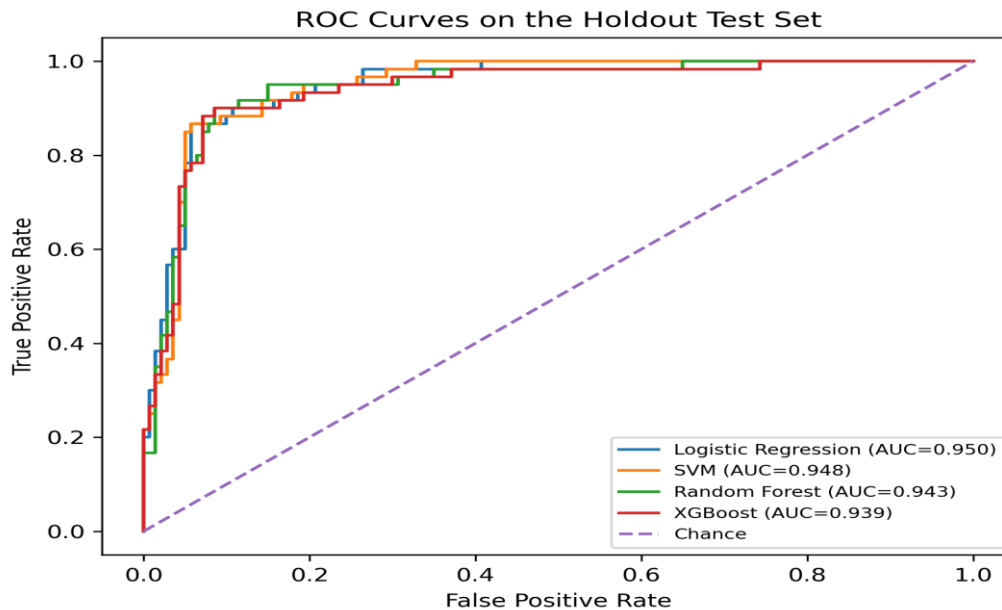


Figure 3. ROC curves for the four candidate classifiers on the holdout set.

5.2 Repeated Cross-Validation and Statistical Comparison

Table 5. Repeated stratified five-fold cross-validation results (5 repeats; 25 folds)

Model	Mean accuracy (95% CI)	Mean precision	Mean recall	Mean F1	Mean ROC-AUC
Logistic Regression	0.893 +/- 0.008	0.799	0.863	0.828	0.935
SVM	0.893 +/- 0.007	0.795	0.869	0.830	0.935
Random Forest	0.889 +/- 0.007	0.803	0.839	0.819	0.927
XGBoost	0.886 +/- 0.006	0.786	0.857	0.819	0.925

The repeated cross-validation results show that mean accuracy is tightly clustered between 0.886 and 0.893, and mean ROC-AUC lies between 0.925 and 0.935. The Friedman test detected an overall difference in fold-level accuracy among the four models (chi-square = 10.68, $p = 0.014$). However, pairwise Wilcoxon comparisons did not remain statistically significant after Holm correction at the 0.05 level. The practical interpretation is therefore that the four models are broadly comparable on this synthetic problem, with logistic regression and SVM showing a small but consistent numerical advantage rather than a decisive superiority.

Error Analysis

Using the holdout set, logistic regression correctly classified 132 of 140 genuine listings and 52 of 60 counterfeit listings. There were eight false positives and eight false negatives. In a real platform, the relative cost of these errors would need to be set through policy. False positives may unfairly delay legitimate sellers, while false negatives allow risky listings to remain visible. This trade-off supports the use of risk thresholds and human review rather than automatic removal.

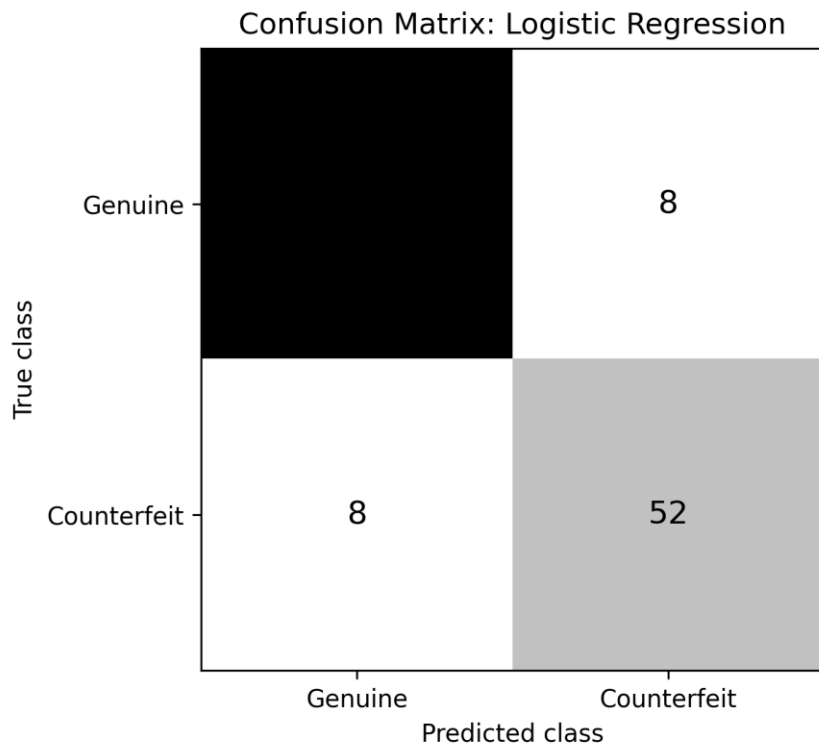


Figure 4. Confusion matrix for logistic regression on the holdout set.

Feature Importance

Permutation importance for the logistic-regression pipeline indicates that seller reputation and price deviation contribute most strongly to holdout ROC-AUC, followed by image similarity. Description quality has a smaller contribution, while review sentiment contributes little under the specific overlapping distributions used in this simulation. This pattern should not be interpreted as a general empirical ranking of marketplace signals. It is partly a consequence of the simulation parameters and is reported to make the model behavior transparent.

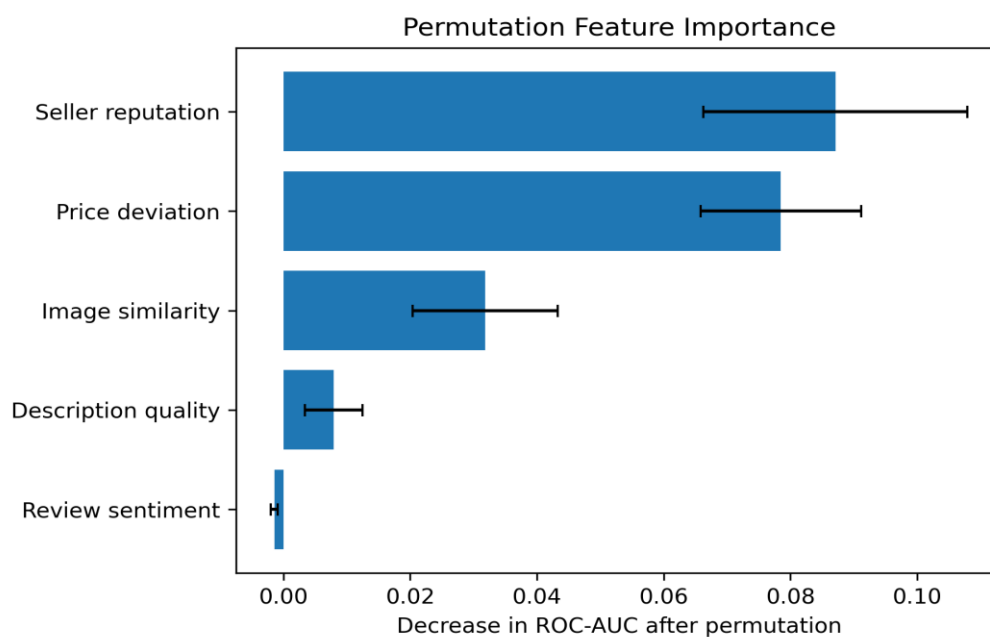


Figure 5. Permutation importance for the best holdout model. Error bars show variation across 30 permutations.

DISCUSSION

The revised experiment supports a narrower and more defensible conclusion than a direct 'evidence from Tanzania' claim. Under a transparent synthetic design, several standard classifiers can combine heterogeneous marketplace signals to discriminate between genuine and counterfeit profiles with useful accuracy. This demonstrates proof-of-concept feasibility, not field effectiveness.

The result that logistic regression and SVM perform competitively is informative. When a small number of normalized features already encode much of the class structure, complexity alone does not guarantee better performance. In operational systems, tree ensembles or more complex architectures may become more valuable as interactions, missing data, seasonal behavior, text embeddings, visual embeddings, and seller-network information are introduced. Conversely, interpretable baselines remain valuable because moderation decisions should be explainable and auditable.

The revised treatment of image information is also important. Image similarity remains conceptually relevant because counterfeit goods can mimic authentic packaging and branding, and image-based work has demonstrated that fine-grained visual analysis can detect subtle differences (Peng et al., 2022). However, a scalar similarity score is not equivalent to training a CNN. Future research should therefore evaluate raw images directly and may compare conventional CNNs with multimodal visual-language models such as CLIP (Radford et al., 2021).

Textual information can likewise be strengthened. The present review-sentiment and description-quality variables are simplified scalar proxies. Transformer encoders such as BERT can model contextual language patterns (Devlin et al., 2019), while recent fake-review research emphasizes the need to address data scarcity, concept drift, and cross-domain generalization (Mewada & Dewang, 2022; Gupta et al., 2024). A realistic platform deployment should therefore combine textual representations with seller behavior and visual evidence rather than treat any one modality as sufficient.

For Tanzania and other African markets, the main research priority is dataset development. Existing Tanzanian studies establish that trust, infrastructure, mobile technology, and payment conditions matter for e-commerce participation (Oreku et al., 2013; Makame et al., 2014; Kabanda & Brown, 2015; Mlelwa et al., 2015), but they do not provide labelled counterfeit records. Partnerships with marketplaces, consumer-protection agencies, brands, and logistics providers would be needed to build ethically governed benchmark data.

LIMITATIONS AND FUTURE RESEARCH

The first limitation is synthetic data. The distributions in Table 2 are explicit design assumptions and may not match the prevalence, covariance structure, fraud strategies, or data quality of real African e-commerce platforms. The reported metrics therefore measure performance against the simulation rules, not against real-world counterfeit cases.

Second, the sample contains only 1,000 listings and five engineered variables. Real marketplaces contain richer and noisier information, including categorical product attributes, seller histories, temporal behavior, raw text, raw images, device information, logistics events, and graph relationships. Performance may change substantially when these signals are introduced.

Third, the image component is represented by a precomputed similarity score rather than raw product imagery. No CNN or visual-language model is trained in this experiment. A future study should use documented image datasets and report architecture, preprocessing, augmentation, optimization settings, and image-level validation.

Fourth, Tanzania is used as a contextual motivation rather than as a source of labelled detection data. The framework should therefore not be interpreted as empirically validated for Tanzania or generalized to Africa as a whole. Future work should conduct multi-country validation and examine platform-specific differences.

Fifth, global feature importance is not sufficient for high-stakes moderation. Deployment should include local explanations, threshold governance, appeal mechanisms, and human oversight. LIME and SHAP provide useful starting points for explanation (Ribeiro et al., 2016; Lundberg & Lee, 2017), while graph-based fraud models may help identify coordinated seller or reviewer behavior (Dou et al., 2020).

CONCLUSION

This paper presented a revised artificial-intelligence framework for counterfeit-product screening in African e-commerce and evaluated it using a transparent Tanzania-informed simulation. The study intentionally distinguishes contextual motivation from empirical marketplace evidence. A reproducible synthetic dataset of 1,000 listings combined price deviation, seller reputation, image similarity, review sentiment, and description quality, with explicit class overlap and contamination assumptions.

Four supervised models were tuned and evaluated using a stratified holdout test and repeated cross-validation. Logistic regression achieved the strongest holdout results, while SVM, random forest, and XGBoost produced broadly comparable cross-validation performance. The results demonstrate that multi-signal screening is technically feasible under controlled simulated conditions, but they do not establish operational accuracy in Tanzania.

The next stage of research should therefore move from proof of concept to real-world validation. Priorities include collection of ethically governed African marketplace datasets, direct image and text modelling, graph-based seller analysis, explainable risk scoring, and human-in-the-loop moderation. Such validation is necessary before the framework can support consumer-protection or platform-governance decisions in practice.

REFERENCES

1. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. <https://doi.org/10.1145/2939672.2939785>
2. Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science*, 54(3), 477-491. <https://doi.org/10.1287/mnsc.1070.0810>
3. Chiu, C.-M., Wang, E. T. G., Fang, Y.-H., & Huang, H.-Y. (2014). Understanding customers' repeat purchase intentions in B2C e-commerce: The roles of utilitarian value, hedonic value and perceived risk. *Information Systems Journal*, 24(1), 85-114. <https://doi.org/10.1111/j.1365-2575.2012.00407.x>
4. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT 2019*, 4171-4186. <https://doi.org/10.18653/v1/N19-1423>
5. Dou, Y., Liu, Z., Sun, L., Deng, Y., Peng, H., & Yu, P. S. (2020). Enhancing graph neural network-based fraud detectors against camouflaged fraudsters. *Proceedings of the 29th ACM International Conference on Information & Knowledge Management (CIKM '20)*. <https://doi.org/10.1145/3340531.3411903>
6. Economides, N., & Jeziorski, P. (2017). Mobile money in Tanzania. *Marketing Science*, 36(6), 815-837. <https://doi.org/10.1287/mksc.2017.1027>
7. Gupta, R., Jindal, V., & Kashyap, I. (2024). Recent state-of-the-art of fake review detection: A comprehensive review. *The Knowledge Engineering Review*, 39, e8, 1-47. <https://doi.org/10.1017/S0269888924000067>
8. Jindal, N., & Liu, B. (2008). Opinion spam and analysis. *Proceedings of the 2008 International Conference on Web Search and Data Mining*, 219-230. <https://doi.org/10.1145/1341531.1341560>

9. Kabanda, S. K., & Brown, I. (2015). E-commerce enablers and barriers in Tanzanian small and medium enterprises. *The Electronic Journal of Information Systems in Developing Countries*, 67(1), 1-24. <https://doi.org/10.1002/j.1681-4835.2015.tb00485.x>
10. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30. https://papers.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
11. Makame, W. H., Kang, J., & Park, S. (2014). Factors influencing electronic commerce adoption in developing countries: The case of Tanzania. *South African Journal of Business Management*, 45(2), 83-96. <https://doi.org/10.4102/sajbm.v45i2.126>
12. Mewada, A., & Dewang, R. K. (2022). Research on false review detection methods: A state-of-the-art review. *Journal of King Saud University - Computer and Information Sciences*, 34(9), 7530-7546. <https://doi.org/10.1016/j.jksuci.2021.07.021>
13. Mlelwa, K. L., Chachage, B., & Zaipuna, Y. O. (2015). E-commerce trend in developing countries: A case study of Tanzania. *International Journal of Computer Applications*, 125(1), 27-33. <https://doi.org/10.5120/ijca2015905220>
14. Mukherjee, A., Liu, B., & Glance, N. (2012). Spotting fake reviewer groups in consumer reviews. *Proceedings of the 21st International Conference on World Wide Web*, 191-200. <https://doi.org/10.1145/2187836.2187863>
15. Oreku, G. S., Mtenzi, F. J., & Ali, A. D. (2013). A viewpoint of Tanzania e-commerce and implementation barriers. *Computer Science and Information Systems*, 10(1), 263-281. <https://doi.org/10.2298/CSIS110725002O>
16. Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 309-319. <https://aclanthology.org/P11-1032/>
17. Peng, J., Zou, B., He, X., & Zhu, C. (2022). Hybrid attention network with appraiser-guided loss for counterfeit luxury handbag detection. *Complex & Intelligent Systems*, 8, 2371-2381. <https://doi.org/10.1007/s40747-021-00633-1>
18. Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. *Proceedings of the 38th International Conference on Machine Learning*, 139, 8748-8763. <https://proceedings.mlr.press/v139/radford21a.html>
19. Ren, Y., & Ji, D. (2017). Neural networks for deceptive opinion spam detection: An empirical study. *Information Sciences*, 385-386, 213-224. <https://doi.org/10.1016/j.ins.2017.01.015>
20. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. <https://doi.org/10.1145/2939672.2939778>
21. Savage, D., Zhang, X., Yu, X., Chou, P., & Wang, Q. (2015). Detection of opinion spam based on anomalous rating deviation. *Expert Systems with Applications*, 42(22), 8650-8657. <https://doi.org/10.1016/j.eswa.2015.07.019>