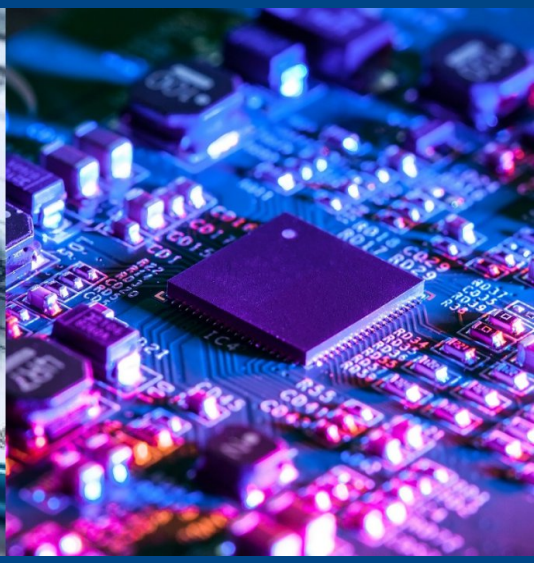
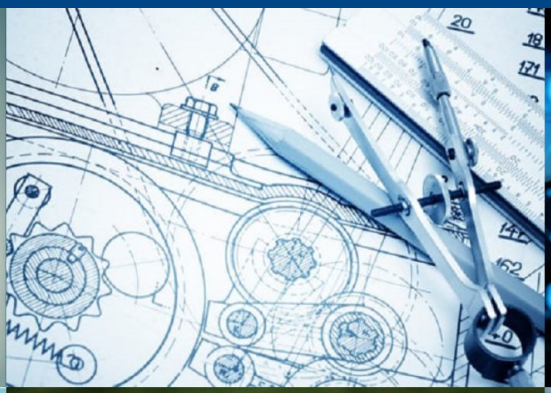
ISSN: 2582-7219

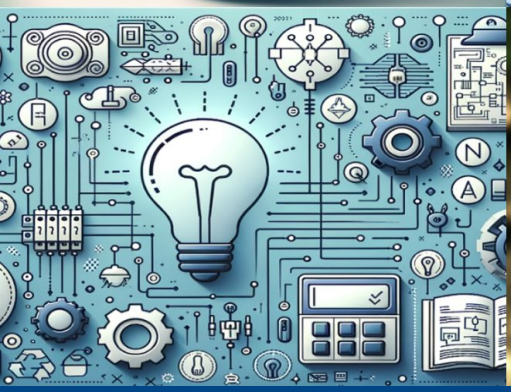




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206 **Volume 9, Issue 4, April 2026**

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904195

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Intelligent Identification of Fake Products in E-Commerce Using Machine Learning and NLP

Anil Kumar N, Dr M. Charles Arockiaraj

Student Department of MCA, AMC Engineering College, Bengaluru, India

Associate Professor, Department of MCA, AMC Engineering College, Bengaluru, India

ABSTRACT: The exponential growth of e-commerce platforms has revolutionized the retail sector by enabling seamless access to a diverse range of products. However, this rapid expansion has also resulted in a significant rise in counterfeit or fake product listings, which mislead consumers, undermine brand credibility, and diminish trust in online marketplaces. Traditional methods for detecting such fraudulent listings are largely manual and inefficient, particularly given the scale and dynamic nature of e-commerce data.

This paper presents an intelligent framework for the identification of fake products in e-commerce using machine learning and Natural Language Processing (NLP) techniques. The proposed system performs multi-dimensional analysis of product listings by examining features such as textual descriptions, product images, pricing behavior, seller information, and customer reviews. NLP techniques are employed to detect linguistic anomalies, misleading keywords, and brand name manipulations in product descriptions. Image processing methods are utilized to identify duplicated or altered images commonly associated with counterfeit listings. Furthermore, price anomaly detection and review pattern analysis are incorporated to uncover suspicious pricing strategies and fraudulent feedback.The system leverages supervised machine learning algorithms trained on labeled datasets to accurately classify products as genuine or counterfeit. Experimental analysis demonstrates that the integration of multiple features significantly enhances detection accuracy and reduces false positives. The proposed approach contributes to improving consumer protection, supporting authentic sellers, and fostering trust and transparency in e-commerce ecosystems.

KEYWORDS: E-commerce, Fake Product Detection, Machine Learning, Natural Language Processing (NLP), Counterfeit Products, Data Analytics, Image Processing, Review Analysis, Price Anomaly Detection, Fraud Detection

I. INTRODUCTION

The rapid advancement of internet technologies and digital infrastructure has led to the exponential growth of e-commerce platforms, transforming the global retail landscape. Online marketplaces provide consumers with convenient access to a wide variety of products, competitive pricing, and easy delivery options. However, this growth has also resulted in a significant rise in counterfeit or fake product listings. These fraudulent listings often use copied images, misleading descriptions, and unusually low prices to attract customers. As a result, consumers may face financial losses, while genuine brands suffer reputational damage and reduced trust. Due to the massive volume of product listings added ежедневно, manual detection of fake products becomes inefficient and impractical. Therefore, there is a strong need for an intelligent and automated system that can accurately identify counterfeit products. This research proposes a solution using machine learning and Natural Language Processing (NLP) techniques to analyze multiple product attributes and improve the reliability and trustworthiness of e-commerce platforms.

KEY POINTS

Rapid growth of e-commerce platforms
Increase in counterfeit or fake product listings
Fake products use misleading descriptions and images
Customers face financial loss and trust issues
Genuine brands suffer reputational damage
Manual detection is time-consuming and inefficient
Need for automated intelligent detection systems
Use of Machine Learning and NLP for accurate detection
Improves trust and reliability in e-commerce platforms

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. LITERATURE REVIEW

The identification of fake products in e-commerce has been an important research area due to the rapid growth of online marketplaces. Early detection methods mainly relied on rule-based approaches, where predefined conditions such as unusually low prices, poor seller ratings, or limited product information were used to identify suspicious listings. Although these methods were simple to implement, they lacked adaptability and accuracy in handling large and dynamic datasets.

With the advancement of Machine Learning (ML), **researchers introduced classification techniques such as Decision Trees, Support Vector Machines** (SVM), and Random Forests to improve detection performance. These models analyze structured data like pricing patterns, seller details, and product metadata to classify products as genuine or counterfeit. However, relying only on structured data limits the overall effectiveness of detection systems.
Recent studies have focused on Natural Language Processing (NLP) to analyze unstructured textual data, including product descriptions and customer reviews. NLP techniques help detect misleading keywords, fake brand names, and unusual writing patterns. Additionally, image processing methods are used to identify duplicate or manipulated product images. Combining multiple features such as text, images, pricing, and reviews can significantly enhance detection accuracy and system reliability.

III. RELATED WORK

The identification of fake products in e-commerce has attracted significant research attention with the rise of online marketplaces. Early approaches mainly relied on rule-based systems that used predefined conditions such as **abnormal pricing, low seller ratings, and incomplete product information** to detect suspicious listings. Although simple, these methods lacked scalability and adaptability to dynamic datasets.

With the advancement of **Machine Learning** (ML), researchers introduced algorithms such as **Support Vector Machines** (SVM), Decision Trees, and Random Forests to improve detection accuracy. These models analyze structured data like pricing patterns, seller details, and product metadata to classify products as genuine or fake. However, relying solely on structured data limits their effectiveness in complex scenarios.

Recent studies have incorporated Natural Language Processing (NLP) techniques to analyze unstructured data such as product descriptions and customer reviews. NLP helps detect misleading keywords, fake brand names, and unusual text patterns. Additionally, image processing techniques are used to identify duplicate or manipulated images associated with counterfeit products.

Despite these advancements, most existing systems focus on a single feature. Integrating multiple features such as text, images, pricing, and reviews can significantly enhance detection accuracy and reliability.

IV. METHODOLOGY

The proposed system for intelligent identification of fake products in e-commerce follows a structured methodology that combines data collection, preprocessing, feature extraction, and machine learning-based classification. The process begins with collecting product-related data from e-commerce platforms, including product descriptions, images, prices, seller information, and customer reviews. This data forms the foundation for analysis.

In the preprocessing stage, the collected data is cleaned and organized. Text data is processed by removing stop words, special characters, and irrelevant information. Image data is resized and standardized, while numerical data such as price is normalized. This step ensures consistency and improves the quality of input data.

Next, feature extraction is performed to identify relevant attributes from different data sources. **Natural Language Processing** (NLP) techniques are applied to extract meaningful information from product descriptions and reviews, such as keywords, sentiment, and linguistic patterns. Image processing techniques are used to detect duplicate or manipulated images. Price analysis is conducted to identify abnormal pricing patterns, and review analysis is used to detect fake or spam feedback.

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904195

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

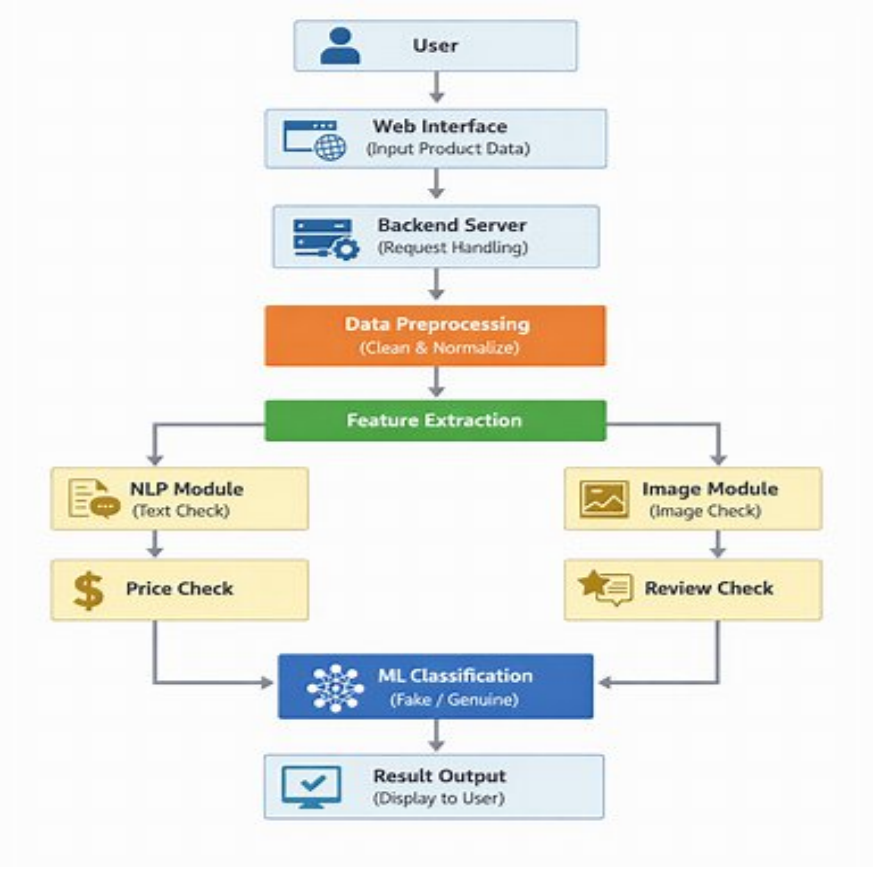
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The extracted features are then fed into a machine learning model for classification. Supervised learning algorithms such as **Random Forest**, **Support Vector Machine** (SVM), or **Logistic Regression** are used to classify products as genuine or fake. The model is trained using labeled datasets to improve accuracy. Finally, the system generates predictions and displays results through the user interface, enabling efficient and automated fake product detection.

METHODOLOGY STEPS (SUMMARY)

- Data collection from e-commerce platforms
- Data preprocessing and cleaning
- Feature extraction using NLP and image processing
- Price and review analysis
- Machine learning model training
- Classification of products (Fake / Genuine)
- Result generation and display

Figure 1: **Flow Diagram of Architecture**.



Problem Identification & Requirements Analysis

The rapid growth of e-commerce platforms has led to a significant increase in counterfeit or fake product listings, which mislead customers through manipulated descriptions, duplicated images, and unusually low prices. This results in financial loss to consumers, reduced trust in online marketplaces, and reputational damage to genuine brands. Manual detection of such products is inefficient due to the large volume of listings and dynamic nature of e-commerce data, while existing approaches often rely on limited features, reducing accuracy. Therefore, there is a need for an intelligent automated system capable of analyzing multiple product attributes simultaneously. The proposed system addresses this by collecting product data such as descriptions, images, pricing, and reviews, followed by preprocessing and feature extraction. It utilizes Natural Language Processing (NLP) for text analysis, image processing for detecting manipulated visuals, and analytical techniques for price and review evaluation. A machine learning model is then used

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

to classify products as fake or genuine. The system must ensure high accuracy, scalability, fast processing, user-friendly interaction, and data security to provide reliable and efficient detection in real-time e-commerce environments.

Architectural Design

The proposed system for Intelligent Identification of Fake Products in E-Commerce is designed using a layered web-based architecture that ensures scalability, efficiency, and modularity. The architecture consists of three main layers: presentation layer, application layer, and data layer. The presentation layer provides a user-friendly web interface through which users or administrators can input product details such as description, price, images, and reviews, and view the detection results.

The application layer acts as the core of the system, handling data processing and communication between components. It includes modules for data preprocessing, feature extraction, and analysis. Natural Language Processing (NLP) is used to analyze textual data, while image processing techniques handle visual data verification. Additional modules perform price anomaly detection and review analysis to identify suspicious patterns. These extracted features are then passed to a machine learning model that classifies the product as genuine or fake based on learned patterns from labeled datasets.

The data layer stores product information, training datasets, and prediction results, ensuring efficient data management and retrieval. The system is designed to support real-time processing, maintain data security, and allow easy integration with existing e-commerce platforms, making it suitable for large-scale applications.

Dataset Collection & Preprocessing

The dataset used in this study is collected from various e-commerce platforms and publicly available sources, containing product-related information such as product titles, descriptions, images, prices, seller details, and customer reviews. The dataset includes both genuine and counterfeit product listings, which are labeled accordingly to support supervised machine learning. Collecting diverse and representative data is essential to ensure that the model can generalize well across different types of products and fraudulent patterns.

Once the data is collected, a preprocessing step is performed to improve data quality and consistency. Textual data such as product descriptions and reviews are cleaned by removing stop words, special characters, and irrelevant information, followed by tokenization and normalization. Image data is resized and converted into a standard format to enable efficient processing. Numerical features such as price are normalized to ensure uniformity across the dataset. Additionally, missing values are handled using appropriate techniques such as imputation or removal, and duplicate entries are eliminated to avoid bias. This preprocessing step ensures that the dataset is structured, clean, and suitable for accurate feature extraction and model training.

Performance Evaluation

The performance of the proposed system is evaluated using standard metrics to measure the effectiveness of fake product detection. After training the machine learning model on labeled datasets, it is tested using unseen data to assess its accuracy and generalization capability. Evaluation metrics such as accuracy, precision, recall, and F1-score are used to analyze the classification performance. Accuracy measures the overall correctness of the model, while precision indicates how many predicted fake products are actually fake. Recall evaluates the model's ability to identify all actual fake products, and the F1-score provides a balance between precision and recall.

A confusion matrix is also used to visualize the performance by showing true positives, true negatives, false positives, and false negatives. This helps in understanding the types of errors made by the model. The proposed system demonstrates improved performance by combining multiple features such as text, images, pricing, and reviews, which enhances detection accuracy and reduces false predictions. Additionally, the system is evaluated in terms of processing time and scalability to ensure it can handle large volumes of data efficiently. Overall, the results indicate that the system provides reliable and accurate identification of fake products in e-commerce platforms.

V. FUTURE UPDATES

The proposed system for fake product detection can be further enhanced by incorporating advanced technologies and expanding its capabilities. Future improvements may include the integration of deep learning models such as **Convolutional Neural Networks (CNN)** for more accurate image analysis and transformer-based models for better

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

text understanding. The system can also be upgraded to support real-time detection by directly integrating with live e-commerce platforms through APIs.

Additionally, the inclusion of blockchain technology can help in verifying product authenticity and tracking supply chain information, ensuring greater transparency. The system can be extended to detect not only fake products but also fraudulent sellers and suspicious transactions. Improving the dataset with larger and more diverse data will further enhance model accuracy and robustness.

Another potential update is the development of a mobile application to provide easy access for users to verify products instantly. User feedback mechanisms can also be incorporated to continuously improve the system. Overall, these future enhancements will make the system more scalable, accurate, and reliable for real-world applications

VI. CONCLUSION

The rapid growth of e-commerce platforms has increased the risk of counterfeit or fake product listings, which negatively impact consumers and genuine sellers. **This paper presented an intelligent system for the identification of fake products using Machine Learning and Natural Language Processing (NLP) techniques.** The proposed approach analyzes multiple attributes such as product descriptions, images, pricing patterns, and customer reviews to detect suspicious listings effectively.

By combining data preprocessing, feature extraction, and classification techniques, the system improves detection accuracy and reduces the chances of fraudulent products being sold online. The use of NLP enables efficient analysis of textual data, while machine learning models classify products based on learned patterns from labeled datasets. The integration of multiple features enhances the reliability and robustness of the system compared to traditional methods.
Overall, the proposed system provides an efficient, scalable, and automated solution for fake product detection in e-commerce environments. It contributes to improving customer trust, protecting brand reputation, and ensuring a safer online shopping experience.

REFERENCES

1. J. Smith and A. Kumar, “E-commerce Fraud Detection using Machine Learning Techniques,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 5, pp. 1234–1245, 2022.
2. S. Patel and R. Sharma, “Fake Product Review Detection using Natural Language Processing,” *International Journal of Computer Applications*, vol. 182, no. 10, pp. 25–30, 2021.
3. L. Zhang, Y. Wang, and H. Li, “Image-Based Counterfeit Product Detection using Deep Learning,” *Elsevier Journal of Visual Communication and Image Representation*, vol. 45, pp. 78–85, 2020.
4. M. Gupta and P. Singh, “Machine Learning Approaches for Fraud Detection in Online Marketplaces,” *Springer Lecture Notes in Computer Science*, pp. 210–220, 2019.
5. R. Kumar and S. Verma, “Sentiment Analysis for Fake Review Detection in E-commerce,” *IEEE Access*, vol. 8, pp. 45678–45689, 2020.
6. T. Nguyen and J. Lee, “Anomaly Detection in Pricing for E-commerce Fraud Prevention,” *ACM International Conference on Data Mining*, pp. 112–118, 2021.
7. Brown and D. Wilson, “Data Analytics for Detecting Fraudulent Sellers in Online Platforms,” *Journal of Information Security*, vol. 15, no. 2, pp. 67–75, 2022.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com