International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI:10.55041/ISJEM05829
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

A Framework for Identifying False Reviews in E-Commerce Platforms using Machine Learning

Dr. K. Satyam[1], Cheerla Vanitha[2]

[1]Associate Professor, Department of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, Andhra Pradesh, India.

[2] Post Graduate, Department of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, Andhra Pradesh, India.

Abstract

Online reviews now have a much greater impact on consumer purchase decisions due to the quick development of e-commerce platforms. However, the existence of fraudulent or misleading evaluations has grown to be a significant problem, impacting the dependability and credibility of online marketplaces. By examining textual review data and related metadata, this study suggests a machine learning-based method for identifying fraudulent reviews on e-commerce sites. The suggested approach uses supervised machine learning methods for classification, text vectorization techniques for feature extraction, and data preprocessing. Users can input bulk review datasets using CSV files or examine individual reviews using the system's web-based application. To increase forecast reliability, the system also incorporates confidence score analysis and duplicate review detection. The model successfully divides reviews into authentic, fraudulent, and neutral categories while offering comprehensive insights into review patterns, according to experimental data. The created system can help e-commerce sites detect false reviews and enhance online feedback systems' reliability**.**

Keywords

Fake Review Detection, Machine Learning, Text Classification, E-Commerce Reviews, Natural Language Processing, TF-IDF Vectorization, Review Analysis, Duplicate Review Detection

I.Introduction

Customer reviews are becoming a significant factor influencing purchasing decisions due to the quick expansion of online shopping platforms. Reviews are frequently used by customers to assess the dependability, quality, and general level of satisfaction of a product. But as e-commerce platforms have grown in popularity, false or misleading evaluations have also become more common. These reviews are produced on purpose to alter product ratings and sway consumers.

Fake reviews have the potential to deceive consumers, undermine consumer confidence, and harm online marketplaces' reputation. Companies may make bad remarks to hurt rivals or use fake reviews to advertise their goods. Because of this, e-commerce companies now face a significant issue in locating and eliminating fraudulent reviews.

By examining language patterns, review activity, and sentiment traits, machine learning approaches offer efficient ways to identify bogus reviews. Machine learning models can automatically differentiate between authentic and fraudulent reviews by using text preparation, feature extraction, and classification algorithms.

In order to categorize reviews as authentic, fraudulent, or neutral, a machine learning-based false review detection method is suggested in this study. In order to predict review authenticity, the system preprocesses data, uses vectorization techniques to transform textual input into numerical features, and then applies a trained machine learning model. Through

a web-based interface, the proposed system also facilitates prediction confidence evaluation, duplicate review detection, and bulk CSV analysis.

II. Proposed System / Methodology

The suggested approach uses text analysis and machine learning techniques to identify fraudulent reviews on e-commerce sites. In order to categorize user reviews as authentic, fraudulent, or neutral, the system analyzes user reviews, extracts relevant elements from the text, and then applies a trained machine learning model. Users can upload large review datasets for analysis or examine individual reviews thanks to the system's web-based implementation.

Data Collection

E-commerce product reviews with review text, ratings, reviewer information, and other metadata make up the dataset used in this study. The machine learning model is trained and assessed using the dataset, which consists of about 25,000 reviews. Positive, negative, and neutral reviews of various kinds are included in the statistics.

Data Preprocessing

The dataset goes through a number of preparation procedures to enhance data quality before the machine learning model is trained. These procedures involve resolving missing values, removing superfluous symbols, changing text to lowercase, and getting rid of stop words. This procedure aids in getting the review text ready for feature extraction and categorization.

Feature Extraction

Following preprocessing, TF-IDF (Term Frequency–Inverse Document Frequency) vectorization is used to transform the textual input into numerical characteristics. By using this method, written reviews are converted into a structured numerical representation that machine learning algorithms can process. Important terms that aid in identifying fraudulent reviews are identified by TF-IDF.

Machine Learning Model

To determine if a review is authentic or fraudulent, a supervised machine learning classification model is employed. The machine learns patterns that differentiate authentic reviews from fraudulent ones after being trained on labeled review data. New reviews that are submitted via the system are then categorized using the learned model.

Duplicate Review Detection

The algorithm also detects duplicate reviews to improve the detection of fraudulent reviews. Reviews created by bots or spam are frequently indicated by duplicate content. To help detect suspicious activity, the system examines review texts and highlights reviews that are repeated or extremely similar.

III. Implementation

The Python programming language is used in conjunction with a number of machine learning and web development technologies to construct the suggested fake review detection system. Python was selected because it offers strong libraries for machine learning, text processing, and data analysis. In this system, the dataset is handled and preprocessed using libraries like Pandas and NumPy, and the machine learning model is constructed and trained using Scikit-learn. To increase the quality of the review text data, stop words are removed, superfluous symbols are eliminated, and the text is converted to lowercase.

Following preprocessing, textual reviews are transformed into numerical feature vectors that the machine learning algorithm may handle using TF-IDF (Term Frequency–Inverse Document Frequency) vectorization. The trained classification algorithm can determine if a particular review is authentic, fraudulent, or neutral by identifying patterns in the labeled dataset.

The Flask framework is used to incorporate the trained model into a web application to give users an interactive platform. Users can upload CSV files with several reviews for bulk analysis or enter individual reviews into the web application. The system applies the trained machine learning model, processes the reviews, and produces prediction results and confidence scores after the data is presented. The system also has a duplicate review identification technique that finds reviews in the dataset that are repeated or extremely similar, which might be a sign of spam or automated review activity. A user-friendly interface presents the results and offers analytical insights like the quantity of authentic, fraudulent, and neutral reviews that were found throughout the analysis.

IV. Results and Discussion

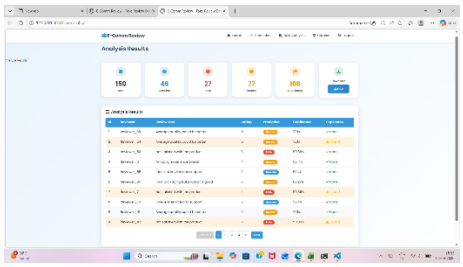

Figure 1: Dataset Analysis Dashboard

An overview of the review dataset used to train and assess the false review detection algorithm is given by the dataset analysis dashboard. About 25,000 reviews—textual comments from users across various product categories and e-commerce platforms—are processed by the system. To highlight the features of the dataset, the dashboard shows key information including model accuracy, average rating, and average reaction time. Positive (39.9%), negative (39.7%), and neutral (20.3%) reviews are represented by percentage values in the dataset's sentiment distribution. The machine learning model is able to identify significant patterns across several sentiment categories because to this balanced distribution.

Furthermore, the dashboard classifies reviews according to platforms, geographical areas, and product categories. Groceries, cars, books, travel, fashion, technology, and cosmetics are among the product categories. Reviews gathered from various online marketplaces, including Nykaa, Snapdeal, Myntra, and Flipkart, are highlighted in platform distribution. The distribution of reviews in the West, South, East, North, and Central regions is further demonstrated by regional statistics. These discoveries aid in enhancing model reliability and comprehending dataset diversity.

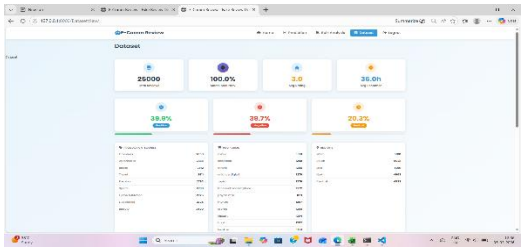

Figure 2: Bulk Review Analysis Results

When processing a batch of reviews via CSV upload, the analysis results page shows how the suggested fake review detection system works. In this instance, the system uses the trained machine learning model to assess 150 reviews and categorize them into several groups.

According to the findings, 46 reviews are recognized as authentic, 27 as fraudulent, and 77 as neutral.

Additionally, the machine finds 108 duplicate reviews, which can point to spam or recurring content in the dataset. Because repeated evaluations are frequently linked to automated spam or bot-generated feedback, duplicate identification is a crucial feature.

Each review entry in the table includes several attributes such as reviewer ID, review text, rating, predicted class label, prediction confidence, and duplicate status. The prediction confidence value indicates how strongly the machine learning model supports its classification decision.

V. Conclusion

A machine learning-based method for identifying fraudulent reviews on e-commerce sites is presented in this paper. The suggested method converts text into useful numerical representations by analyzing textual review data utilizing preprocessing strategies and feature extraction techniques like TF-IDF vectorization. The reviews are then categorized as authentic, fraudulent, or neutral using a supervised machine learning model. Both individual review prediction and bulk review analysis via CSV uploads are supported by the system, which is designed as a web-based application. Furthermore, prediction confidence scoring and duplicate review identification improve the analysis's dependability. The suggested system can successfully detect questionable reviews and offer helpful insights into review patterns, according to experimental results. Overall, the technique helps make online review systems in e-commerce settings more reliable and credible.

References

[1] N. Jindal and B. Liu, “Opinion spam and analysis,” in Proceedings of the International Conference on Web Search and Data Mining (WSDM), New York, NY, USA, 2008, pp. 219–230.

[2] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, “Finding deceptive opinion spam by any stretch of the imagination,” in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL), Portland, OR, USA, 2011, pp. 309–319.

[3] A. Mukherjee, B. Liu, and N. Glance, “Spotting fake reviewer groups in consumer reviews,” in Proceedings of the 21st International World Wide Web Conference (WWW), Lyon, France, 2012, pp. 191–200.

[4] S. Rayana and L. Akoglu, “Collective opinion spam detection: Bridging review networks and metadata,” in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, Australia, 2015.

[5] F. Li, M. Huang, Y. Yang, and X. Zhu, “Learning to identify review spam,” in Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), Barcelona, Spain, 2011.