IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

Fake Product Review Detection and Sentiment Analysis using Machine Learning

Prof. Shravani M Kolawale[1], Shruti kharat[2], Sejal Kannake[3], Shruti Patil[4], Satwik Jadhav[5]
Guide, Department of Computer Science[1]
Students, Department of Computer Science[2,3,4,5]
Shrimati Kashibai Navale College of Engineering, Pune, India
Sejalkannake151@gmail.com, shrutipatil0316@gmail.com

Abstract: *Online reviews frequently change the minds of potential customers and inform their shopping decisions. However, many online platforms are dealing with many fake reviews, and this is creating a lot of issues. These issues include lost, reduced trust and misinformation. Text data is generally complex and even the structured data has a lot of variations. In addition, circumstances of spams are present which make it even more complex. The sophisticated work of sentiment analysis and fake reviews detection is expected of this nature. This work provides a framework of research that utilizes machine learning to support pairs of fake review detection and sentiment analysis of these reviews. We use and employ a flexible solution to normalize the text data that includes tokenization and stopword removal and also stems it with the Porter Stemming Algorithm. Text representation techniques are. More specifically, to make the classification pair, we implement classification token pairs and also pair classification support to this pair classification sets. To utilize both the pair classification sets, two support machines are built and therefore claim that the pairs are to be in support. The ANN model captures complex nonlinear textual features. In addition, it is posited that the Random Forest classifier improves the pair classification of the two sets by means of the use pair classification sets. We use pair classification sets to build the two support machines. More specifically, to utilize both the pair classification sets, two support machines are built and therefore claim that the pairs are to be in support. The Random Forest model pair classification of the two sets not only provides more accuracy than the Support Vector and the ANN model combined, in a pair classification of over 5.88%, but also provides more comprehensive bounds than the Support Vector and the ANN model. The Random Forest model provides an overall classification of 97.94% accuracy, greater than pair classification of over 5.88% of both the Support Vector and the ANN model combined. Results show that approaches that employ ensemble learning techniques are very productive for detecting fake reviews. The system that is proposed increases its capacity of providing functional classes of dishonest reviews combined with its ability to perform sentiment analysis. This research builds on existing work towards the improvement of online review systems by providing effective and scalable solutions on real e-commerce systems.*

Keywords: Fake reviews, Sentiment analysis, Machine Learning, Deep Learning, Opinion mining, Natural Language Processing

I. INTRODUCTION

E-commerce platforms are changing how people see products or services. In digital marketplaces, reviews are a big part of how customers make decisions. Because of this, reviews are manipulated. Intentional fake reviews rise and are made to increase or decrease a product's rating and are intended to mislead customers and jury products. These fake reviews can ruin a platform's credibility, and serious people/companies worry about this due to the deterioration of trust and the money lost. Finding fake reviews is very inefficient and unscalable, because the review finder depends on a human being.

DOI: 10.48175/IJARSCT-34392

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

Luckily, the future looks bright, and many believe it is due to the advancement of machine learning. These unique models analyze digital and written data. Sentiment analysis is a very important element and even more important when determining a review's true and false emotional essence. Research has been done on review fraud and several approaches to the machine learning models have been made involving support vector machines, naive Bayes, etc. Unfortunately, although many find the models hopeful, need order and control. For instance, poorly organized data, specifically which data is defined most relevant, and flexibility of the model.

This research focuses on designing a system that accurately and efficiently detects fake product reviews and perform simultaneous sentiment analysis. The framework includes pre-processing of removing stop words and performing stemming to enhance the quality of data. The machine learning models of Artificial Neutral Networks (ANN) and Random Forest (RF) are adopted to classify reviews in terms of originality and sentiment polarity. This research helps to (i) address the need of creating a sophisticated pre-processing in the system, (ii) offer an analysis of RF and ANN classifies as well comment on the results and (iii) justify the system. The results obtained indicate the successful implementation of ensemble learning to enhance the quality of classification.

II. LITERATURE SURVEY

Alzahrani et al. [1] use deep learning to design a system to analyze reviews of e-commerce products. The neural networks used can understand the structure and meaning of the text in the reviews. The authors show that their system achieves a better level of accuracy compared to the traditional machine learning systems. The role of deep learning in managing large datasets of reviews is confirmed by the authors.

Alsubari et al. [2] present the first framework that employs supervised learning to detect fake reviews using the analytics of big data. The framework focuses on the selection of features and the preprocessing of data in order to reach an acceptable classification level. Several machine learning systems are used in finalizes order to detect the so-called "deceptive» patterns in the reviews. The experiments prove that the supervised models of reviews, in a reliable manner, reveal fraudulent reviews.

Marwat et al. [3] address the SentiDeceptive system, a cross-disciplinary approach that integrates sentiment analysis and deception detection systems. The analyzed reviews are misleading and are therefore proxy reviews. SentiDeceptive improves the level of detection of deceptive rating reviews on social networking systems. The authors remark that SentiDeceptive is the first system that integrates the analysis of sentiment reviews and the analyses of authenticity.

Elmogy et al. [4] describe machine learning systems of supervised type in order to detect reviews that are not real. Among these systems are Decision Trees and a system of Support Vector Machines. The authors prove the high level of effectiveness given the adequate selection of features and a rigorous data preprocessing. The authors come to the conclusion that, in review analysis, the supervised systems are characterized by the high level of effectiveness in revealing reviews that are not real.

Tabany and Gueffal [5] uses SVM to conduct sentiment analysis and classify fake reviews in the Amazon datasets while trying to improve the representation of features and the accuracy of classification. This means that the SVM proves to be efficient in the separation of true reviews from fake reviews and is the optimal choice for structured datasets.

Alsubari et al. [6] extends their work on fake reviews by further detecting them via supervised learning and data analytics. The study inspects various classification methods and the significance of feature engineering. Results verify artificial intelligence to be a great tool to identify untruthful reviews and firmly suggest a major role of feature engineering.

Bathla et al. [7] have created a deep learning-based format that include detail analysis of the components to identify fake reviews. It allows for the identification of the review content to improve the precision of detection. The study strongly supports deep learning due to having the ability to breakdown review content.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 18, April 2026 Impact Factor: 8.2

Ennaouri and Zellou [8] provide a detailed review of the machine learning-based models for fake review detection aprovided by other researchers. They describe the challenges such as data imbalance and sparse features. Finally, they assert that combined and cross-layer approaches provide the best solution.

Bhattacharya [9] describes the approach that helps in monitor and eliminate the fake reviews from products by using machine learning. The study aims for a real-time detection and the removal of instantly performed deception. It strongly advocates that the automated approach to improve the respect of online marketplaces.

Qazi et al. [10] perform a systematic literature review that covers the use of machine learning for detecting opinion spam. The study reviews the key problems that arise when building proprietary models, supervised, and hybrid model spam detection systems. It outlines the problems of scalability and feature selection. The review outlines the gaps for future research.

Park and Chai [11] design a machine-learning model for fake news detection based on a user-centered approach. While their primary focus is fake news, the methods these authors discuss are applicable for the detection of fake reviews. The study focuses on user behavior and content attributes. The results show that various classification techniques increased detection accuracy.

Alshehri [12] describes an online system for the detection of fake reviews that employs multiple machine learning techniques. The study seeks to find the best model by comparing multiple systems. The study reports that systems that rely on ensemble methodologies yield the best results and that the system is ready for implementation for real-time use.

Jaiswal and Javale [13]deliver a fake product review monitoring system. The system encompasses various methods of detection to analyze and categorize reviews. The study demonstrates the significance of integrating systems and shows accurate results from the system.

Mutemi and Bacao [14] systematically review fraud detection on various e-commerce platforms. Detection of fake reviews is just one of the methods that the study elaborates on. The study covers an array of techniques, including big data fraud detection. The authors note the limitations of their study in relation to the real-time processing of large data.

Qazi et al. [15] build upon their opinion spam detection review by examining the recent progress made in the machine learning landscape. The study recognizes the need for proper consideration of hybrid models and deep learning methodologies. It pinpointed some of the missing pieces when it comes to big data. The results will help to make the next steps in the evolution of the fake review detection systems.

III. METHODOLOGY

The system enables online Review prediction through text data analysis. The system and real time text data, that can be processed by any application, can be assessed as an integrated ecosystem. The system includes phases of both training and testing for classification. The system can classify text data through symbolic analysis, such as predicting and discerning Review data forgery. The system employs machine learning algorithms to discern predictive language in Reviews. Review the system in terms of accuracy and false ratio.

There are two machine learning models for classification: Artificial Neural Network (ANN) and Random Forest (RF). ANN is highly capable at capturing nonlinear patterns in data and uses hidden layers. These are trained using a technique called backpropagation. The Random Forest classifier is slightly different. It uses decision trees and creates an ensemble of them. While this ensemble is working, each decision tree is trained on random subsets of data and features. The predictions for the class are accepted after the decision tree ensemble votes. Because of this voting mechanism, Random Forests are less likely to overfit the training data and captures the real data pattern more reliably.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

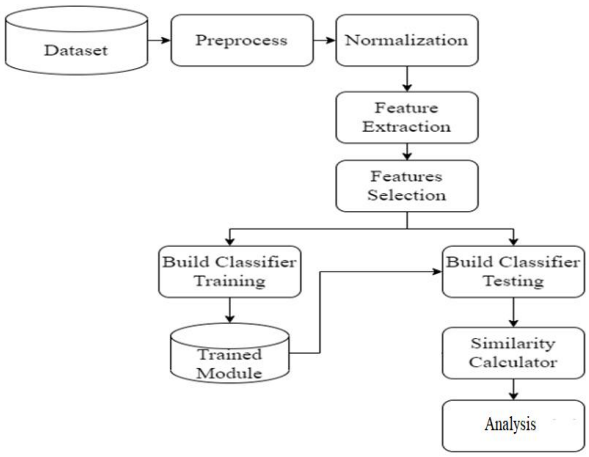

Figure 1: System Architecture

We're breaking down our fake review detection system into a few easy steps to keep things neat! We've got some awesome engineering coming together with NLP (Natural Language Processing) and ML (Machine Learning) that we think will help with this problem a lot!

1. Data Collection

For this first step we need to pull together some product reviews from the internet. This could be from e-commerce sites or social media. Buying reviews is common, so we need both real and fake reviews, and ideally from a diverse range of sources.

2. Data Preprocessing

Text data is really messy, so we need to clean this up. We remove words that aren't as important, we smell and lemmatize, and we do some tokenizing as well. After this, we'll be left with a lot less noise, and much much better data quality.

3. Feature Extraction and Feature Selection

We can use ML models to assess the reviews, but first we need to create a feature list. We'll do this by using Bag of Words. We can't use every feature we list. We need to use our judgment to pick the best of the best. This is important to help the system learn to spot which reviews are fake or real!

5. Training Models

Using our feature list and data, we can train some models. We plan to use ANN and Random Forest models, as well as some other different types of models.

6. Model Testing

The trained model is validated on unseen testing data. Model performance metrics are defined as accuracy, precision, recall, and F1-score to validate the system's effectiveness.

7. Classification and Prediction

In real time and as part of batch processing, the system classifies reviews as either fake or genuine. Sentiment analysis to detect review polarity may also be performed. The analysis is performed, and results are displayed in the form of a graph or a report as an improvement to existing methods in the proposal.

IV. RESULTS

Here, we evaluate the effectiveness of both components using the proposed classification method to measure the bitcoin implementation efficiency and the accuracy of false news detection. We use an Intel(R) Core(TM) i3-2328M CPU 2.20GHz, with 4 GB of RAM. The system is built using a Java-based 3-tier analytics platform with distributed architecture.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

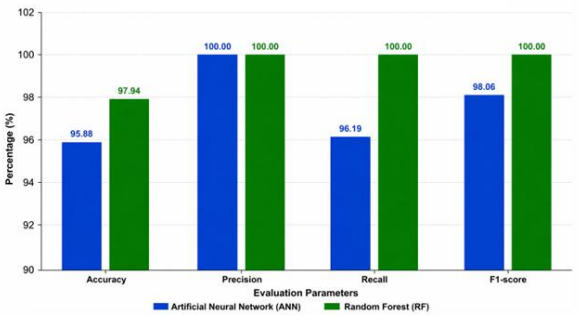

Figure 2: Performance Comparison of ANN and Random Forest

Figure 2 on the obtained experimental results from the implementation of Java, the models developed with the Artificial Neural Network (ANN) and Random Forest (RF) classifiers proved usefulness when it comes to the detection of fake product reviews and the accomplishment of sentiment analysis. Concerning the ANN model, output of the experiment showed the following values: an accuracy of 95.88%, 100% precision, and 96.19% recall, with an F1-score of 98.06. The values show that the ANN model displayed the capability to detect reviews and classify them with minimal/very few false positive results. Based on the results of the experiment, the RF model performed the detection task of classifying fake reviews with an accuracy of 97.94% and also, perfect precision, recall, and F1-score of 100% to the ANN model. The results show that the RF model proved to be more effective as it was able to identify true positive instances with zero false positive results. With the implementation of Java, all the steps of the model, from preparation to classification, and the obtained results showed that all the steps of the model would yield consistent results. The results showed that through the use of ensemble in Random Forest, the performance on fake product reviews detection using Random Forest and ANN was more positive**.**

Table 1: Performance Comparison of ANN and Random Forest

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ANN	95.88	100	96.19	98.06
Random Forest (RF)	97.94	100	100	100

The table 1 study shows that both ANN and Random Forest models work well at detecting fraudulent reviews. Both models were able to achieve 100% precision. However, the Random Forest Classifier was better compared to the ANN Classifiers in the accuracy, recall, and F1-score measures. For Random Forest Classifiers, recall (100%) and accuracy (97.94%) were higher than ANN (95.88%) concerning these measures. The Random Forest Classifier was also more robust and better performance proven by its higher F1-score. Random Forest Classifiers, with their more powerful basing, give the ANN Classifiers better accurate and trustworthy results**.**

V. CONCLUSION

This study provides a good example of a machine-learning based system that not only identifies fake product reviews but also performs sentiment analysis. It mainly provides product reviews and other relevant user-generated content. The focus of the study is the implementation of a machine learning system that uses a structured pipeline, consisting of data preprocessing, feature extraction, and classification. This pipeline aims to address the challenges presented by the nature of the data, which in this case is mostly in the form of noise and lacks actual content. Preprocessing techniques such as tokenization, the removal of stop words, and stemming not only improve the quality of the data but also reduce the data dimension by removing the irrelevant portions. The representation of the textual data is also further improved by the application of TF-IDF in feature extraction, which apportions weights to the key terms in the data. To evaluate the performance of the developed system in detecting fake reviews and performing sentiment analysis, artificial neural network (ANN) and random forest (RF) classifiers were utilized. The experiments proved both models to be precise and robust. The Random Forest Classifier proved to be better than the ANN model based on accuracy, recall, and the

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

F1-score, thus achieving about full performance. The ANN model was primarily based on the reliance of the ensemble learning techniques as well as the properties of the high-dimensional data. The detection of fake reviews and sentiment analysis provides users with a deeper insight into the opinions of other users and promotes better decision-making both for consumers and for the businesses. The system is designed to be scalable and provides a basis for real-time applications after further optimization. The system can also be optimized by investing in better datasets with ample computing resources. These improvements can address the system's limitations and be the focus of the upcoming work. This research study provides a valuable solution for tracking fake reviews and sentiments. By doing so, it improves trust and transparency in e-commerce systems.

REFERENCES

[1]. Alzahrani, Mohammad Eid, et al. "Developing an Intelligent System with Deep Learning Algorithms for Sentiment Analysis of E-Commerce Product Reviews." Computational Intelligence and Neuroscience 2022.1 (2022): 3840071.

[2]. Alsubari, S. Nagi, et al. "Data analytics for the identification of fake reviews using supervised learning." Computers, Materials & Continua 70.2 (2022): 3189-3204.

[3]. Marwat, M. Irfan, et al. "Sentiment Analysis of Product Reviews to Identify Deceptive Rating Information in Social Media: A SentiDeceptive Approach." KSII Transactions on Internet & Information Systems 16.3 (2022).

[4]. Elmogy, Ahmed M., et al. "Fake reviews detection using supervised machine learning." International Journal of Advanced Computer Science and Applications 12.1 (2021).

[5]. Tabany, Myasar, and Meriem Gueffal. "Sentiment analysis and fake amazon reviews classification using SVM supervised machine learning model." Journal of Advances in Information Technology 15.1 (2024): 49-58.

[6]. Alsubari, S. Nagi, et al. "Data analytics for the identification of fake reviews using supervised learning." Computers, Materials & Continua 70.2 (2022): 3189-3204.

[7]. Bathla, Gourav, et al. "Intelligent fake reviews detection based on aspect extraction and analysis using deep learning." Neural Computing and Applications 34.22 (2022): 20213-20229.

[8]. Ennaouri, Mohammed, and Ahmed Zellou. "Machine learning approaches for fake reviews detection: A systematic literature review." Journal of Web Engineering 22.5 (2023): 821-848.

[9]. Bhattacharya, Swarnajit. "Monitoring and Removal of Fake Product Review Using Machine Learning (Ml)." (2023).

[10]. Qazi, Atika, et al. "Machine learning-based opinion spam detection: A systematic literature review." IEEE Access 12 (2024): 143485-143499.

[11]. Park, Minjung, and Sangmi Chai. "Constructing a user-centered fake news detection model by using classification algorithms in machine learning techniques." IEEE Access 11 (2023): 71517-71527.

[12]. Alshehri, Asma Hassan. "An Online Fake Review Detection Approach Using Famous Machine Learning Algorithms." Computers, Materials & Continua 78.2 (2024).

[13]. Jaiswal, Mohini, and Deepali Javale. "Fake Product Review Monitoring System." 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI). Vol. 2. IEEE, 2024.

[14]. Mutemi, Abed, and Fernando Bacao. "E-commerce fraud detection based on machine learning techniques: Systematic literature review." Big Data Mining and Analytics 7.2 (2024): 419-444.

[15]. Qazi, Atika, et al. "Machine learning-based opinion spam detection: A systematic literature review." IEEE Access 12 (2024): 143485-143499..