



# Fake Review Detection in E-Commerce Using Machine Learning and NLP: A Comparative Study

Ajeetha G

Department of Computer Science, EGS Pillay Engineering College, Nagapattinam, India,  
Anna University Affiliated, Chennai, India

**Abstract:** The proliferation of fake reviews in e-commerce platforms has become a critical challenge affecting consumer trust and purchasing decisions. This paper presents a comparative study of machine learning approaches for automatic fake review detection using Natural Language Processing (NLP) techniques. We evaluate three classification algorithms — Logistic Regression, Naive Bayes, and Support Vector Machine (SVM) — on a dataset of 8,087 reviews using TF-IDF feature extraction. Experimental results demonstrate that SVM achieves the highest accuracy of 86.60% with precision and recall of 0.87, outperforming Logistic Regression (86.01%) and Naive Bayes (84.37%). The proposed system effectively distinguishes between genuine (CG) and fake (OR) reviews, providing a reliable foundation for trustworthy product recommendation systems.

**Keywords:** Fake review detection, Natural Language Processing, TF-IDF, Support Vector Machine, E-commerce, Opinion spam.

## I. INTRODUCTION

The rapid growth of e-commerce platforms such as Amazon, Flipkart, and eBay has made online reviews a primary source of information for consumers making purchasing decisions. However, the increasing prevalence of fake and deceptive reviews has severely undermined the credibility of these platforms. Studies estimate that nearly 30-40% of online reviews are fabricated, either by competitors attempting to damage rival products or by sellers trying to artificially boost their own ratings.

Fake review detection has therefore emerged as a critical research problem at the intersection of Natural Language Processing (NLP) and machine learning. Traditional manual detection methods are time-consuming and ineffective at scale. Automated approaches using machine learning offer a scalable and reliable solution.

In this paper, we propose a comparative study of three machine learning algorithms — Logistic Regression, Naive Bayes, and Support Vector Machine — combined with TF-IDF feature extraction for automatic fake review detection. Our system is evaluated on a real-world dataset of 8,087 reviews and achieves a best accuracy of 86.60% using SVM.

The remainder of this paper is organized as follows: Section II reviews related work, Section III describes the proposed methodology, Section IV presents experimental results, and Section V concludes the paper.

## II. LITERATURE REVIEW

Detecting fake reviews has been an active area of research over the past decade. Ott et al. [1] first introduced a benchmark dataset of deceptive hotel reviews and demonstrated that Support Vector Machines with unigram features could achieve reasonable detection accuracy. Their work established the foundation for subsequent NLP-based approaches.

Mukherjee et al. [2] proposed a behavioral analysis approach to identify fake reviewers on Yelp, focusing on reviewer activity patterns rather than review content alone. Their study highlighted that combining behavioral and linguistic features improves detection performance.

Li et al. [3] explored deep learning approaches for opinion spam detection, demonstrating that neural network models could capture semantic patterns in reviews more effectively than traditional bag-of-words methods.

Rashid et al. [4] applied TF-IDF vectorization combined with Logistic Regression and Naive Bayes classifiers for fake review classification, reporting accuracies in the range of 80-85%. Our work extends this by including SVM and achieving superior results of 86.60%.

Recent studies have explored the use of BERT and transformer-based models for fake review detection, achieving higher accuracies but requiring significantly more computational resources. Our proposed approach focuses on lightweight, interpretable machine learning models suitable for real-world deployment.



### III. METHODOLOGY

#### A. Dataset

In this study, we used the Fake Reviews Dataset obtained from Kaggle, containing 40,432 reviews collected from the Amazon e-commerce platform. The dataset consists of two classes — Computer Generated fake reviews (OR) and Original genuine reviews (CG). For our experiments, 8,087 reviews were used in the test set. The dataset covers multiple product categories including Home and Kitchen, Electronics, and Clothing.

#### B. Data Preprocessing

The raw review text was preprocessed using the following steps: (1) Removal of null and missing values, (2) Text tokenization, (3) Removal of English stop words, and (4) TF-IDF feature extraction with maximum 5,000 features.

#### C. Feature Extraction

We applied Term Frequency-Inverse Document Frequency (TF-IDF) vectorization to convert raw text into numerical feature vectors. TF-IDF assigns higher weights to words that are frequent in a document but rare across the entire corpus, making it highly effective for distinguishing fake from genuine reviews.

#### D. Classification Models

Three machine learning classifiers were evaluated: (1) Logistic Regression — A linear classifier that models the probability of a review belonging to the fake or genuine class. (2) Naive Bayes — A probabilistic classifier based on Bayes theorem, widely used for text classification tasks. (3) Support Vector Machine (SVM) — A powerful classifier that finds the optimal hyperplane separating fake and genuine reviews in high-dimensional feature space.

#### E. Experimental Setup

The dataset was split into 80% training and 20% testing sets using stratified random sampling with random state 42 to ensure reproducibility.

### IV. RESULTS AND DISCUSSION

#### A. Performance Comparison

Table I presents the performance comparison of the three machine learning models evaluated on the fake reviews dataset.

TABLE I. Comparison of Classification Models

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	86.01%	0.86	0.86	0.86
Naive Bayes	84.37%	0.85	0.84	0.84
SVM (Proposed)	<b>86.60%</b>	0.87	0.87	0.87

#### B. Discussion

Among the three classifiers, Support Vector Machine achieved the highest accuracy of 86.60% with precision and recall of 0.87. This demonstrates that SVM is highly effective for fake review detection due to its ability to handle high-dimensional TF-IDF feature spaces.

Logistic Regression performed comparably with 86.01% accuracy, making it a strong alternative when interpretability is required. Naive Bayes achieved 84.37% accuracy, which is slightly lower but computationally faster and suitable for real-time applications.

All three models successfully classified both Computer Generated fake reviews (OR) and Original genuine reviews (CG) with balanced precision and recall scores, indicating no significant class bias in the predictions.

#### C. Comparison with Existing Work

Our proposed SVM model achieving 86.60% accuracy compares favorably with existing studies. Rashid et al. [4] reported accuracies of 80-85% using similar TF-IDF based approaches. Our results demonstrate an improvement of approximately 1-2% over previously reported baselines, validating the effectiveness of our comparative approach.



## V. CONCLUSION

This paper presented a comparative study of machine learning approaches for fake review detection in e-commerce platforms. We evaluated three classifiers — Logistic Regression, Naive Bayes, and Support Vector Machine — using TF-IDF feature extraction on a real-world dataset of 8,087 reviews.

Experimental results demonstrate that SVM achieves the best performance with 86.60% accuracy, 0.87 precision, and 0.87 recall. All three models successfully distinguished between Computer Generated fake reviews and Original genuine reviews with balanced performance metrics.

The proposed approach provides a lightweight, interpretable, and scalable solution for fake review detection suitable for real-world e-commerce deployment. Future work includes incorporating BERT and transformer-based models, multilingual fake review detection, Graph Neural Network based reviewer behaviour analysis, and real-time fake review detection API integration.

## REFERENCES

- [1] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," in Proc. ACL, 2011, pp. 309-319.
- [2] A. Mukherjee, B. Liu, and N. Glance, "Spotting fake reviewer groups in consumer reviews," in Proc. WWW, 2013, pp. 191-200.
- [3] J. Li, M. Ott, C. Cardie, and E. Hovy, "Towards a general rule for identifying deceptive opinion spam," in Proc. ACL, 2015, pp. 1566-1576.
- [4] J. Rashid, S. M. A. Shah, and A. Irtaza, "Fake review detection using machine learning," IEEE Access, 2019.
- [5] N. Jindal and B. Liu, "Opinion spam and analysis," in Proc. WSDM, 2008, pp. 219-230.
- [6] G. Wang, S. Xie, B. Liu, and S. Y. Philip, "Review graph based online store review spammer detection," in Proc. ICDM, 2011, pp. 1242-1247.
- [7] E. P. Lim, V. A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers using rating behaviors," in Proc. CIKM, 2010, pp. 939-948.