iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** X **Month of publication:** October 2024

DOI: https://doi.org/10.22214/ijraset.2024.64825

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue X Oct 2024- Available at www.ijraset.com

Survey on Fake Review Detection System

Suchitra Deokute[1], Vaishnavi Gaikwad[2], Alina Tejani[3], Anuja Mhetre[4]
Department of Computer Engineering, Dhole Patil College of Engineering, Pune

Abstract: ***In today's business landscape, online reviews play a crucial role in shaping commerce. A large portion of purchase decisions for online products is driven by customer feedback. Consequently, some individuals or groups try to manipulate product reviews to their advantage. Fake online reviews have a considerable effect on the experiences of online consumers, sellers, and e-commerce platforms. Although there has been academic research focused on identifying fake reviews, there is still a need for studies that thoroughly examine and summarize their origins and impacts. This work introduces a combination of semi-supervised and supervised text mining models to detect fake reviews, comparing their performance using datasets from reviews.***

Keywords: Online Reviews, Fake Reviews, Semi-supervised Learning, Supervised Learning, Review Datasets, machine learning.

I. INTRODUCTION

When making a purchase online, it is common to first examine the ratings and reviews left by previous buyers. If the ratings are high and the reviews align with your expectations, it can encourage a quick decision to buy. However, the issue arises when the authenticity of the reviewer is unclear. Determining whether a review comes from a genuine buyer is nearly impossible. Some reviews may be paid promotions, or posted by acquaintances of the seller to artificially boost the product's reputation. This increases the risk of misunderstanding product quality and being misled by deceptive advertising. Customers often rely on seller descriptions and user reviews on e-commerce platforms, but false reviews can skew purchasing decisions. Fraudulent reviews can significantly impact consumers' choices, as studies suggest that buyers are highly sensitive to both positive and negative feedback. Reviews not only shape sales, but can also create considerable financial benefits, which in turn fuels the spread of fake reviews. Given the massive volume of review data, manually identifying fake reviews is impractical, making automated detection a valuable research area. With millions of online users, it is important to develop methods to identify and classify fake and genuine reviews, ensuring that consumers can make well-informed decisions and choose quality products.

As the internet economy grows, new forms of online deception continue to evolve, with fraudulent verification techniques becoming more sophisticated. Research in detecting fake reviews must therefore continue to advance. Genuine reviews tend to include a mix of nouns, adjectives, prepositions, determiners, and coordinating conjunctions to describe clear, sensory attributes. On the other hand, fake reviews often feature more verbs, adverbs, and pronouns. By analyzing the distribution of parts of speech in review texts, it becomes possible to identify fraudulent reviews. Many fake reviewers engage in "brushing" practices, where they provide false ratings, whether positive or negative, for products they haven’t used. As such, their reviews often have distinctive traits. Reviewer behavior can be analyzed based on factors such as activity patterns, the maximum number of reviews in a day, total number of reviews, and the proportion of positive or misleading reviews. All of this information, derived from metadata, can help identify patterns indicative of fake reviews after careful analysis.

II. PROBLEM STATEMENT

In the age of e-commerce, online reviews have become a critical factor influencing consumer purchasing decisions. However, the increasing prevalence of fake reviews poses a significant challenge to both consumers and businesses. These deceptive reviews, often posted by individuals with ulterior motives or automated systems, mislead potential buyers about product quality, resulting in skewed ratings and unreliable feedback. As online markets expand, the ability to distinguish between genuine and fake reviews has become crucial for maintaining trust in digital platforms. Manual detection of fraudulent reviews is impractical due to the sheer volume of data, creating a pressing need for automated solutions that can accurately detect and filter fake reviews. The lack of effective, scalable tools to address this issue not only erodes consumer confidence but also negatively impacts sellers, leading to unjust competition and distorted market dynamics. Therefore, there is an urgent need for research and development of robust detection systems capable of identifying fake reviews and preserving the integrity of online reviews, ensuring that consumers make informed purchasing decisions based on authentic feedback.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue X Oct 2024- Available at www.ijraset.com

III. SCOPE AND OBJECTIVE

The primary focus of this project is to develop an automated system that detects fake reviews on online platforms using machine learning techniques. As online reviews significantly influence consumer decisions, it is essential to identify and eliminate fraudulent reviews that distort product ratings and mislead buyers. This system will analyze review data from various e-commerce platforms, utilizing machine learning models to classify reviews as genuine or fake. By incorporating both supervised and semi-supervised learning approaches, the system aims to improve the accuracy of fake review detection. The project will also involve evaluating different machine learning algorithms on real-world datasets, such as hotel reviews, to determine the most effective method. Ultimately, the system seeks to enhance the reliability of online reviews, providing a scalable solution that helps consumers make informed purchasing decisions and contributes to a deeper understanding of fake review patterns and their impact on the digital marketplace.

IV. LITERATURE SURVEY

Sr no	Appearance (in Time New Roman or Times)		
	Author	Year	Description
1	Wesam Hameed Asaad, Ragheed Allami, Yossra Hussain Ali.	2023	Online customer reviews have become an increasingly influential tool in shaping purchasing decisions. However, the growing impact of these reviews has led to a surge in the publication and promotion of fake reviews by some businesses, either to enhance their own product's reputation or to undermine their competitors. These counterfeit reviews can have an especially detrimental impact on small businesses, with even a single negative fake word removal, and stemming. Subsequently, features were extracted using TFIDF techniques to leverage the benefits of sentiment analysis and to ascertain the presence of spam comments in the feature extraction approach.
2	Maysam Jalal Abd , Mohsin Hasan Hussein	2024	Most consumers on the internet rely on reviews to help them make decisions about what to buy because they provide a reliable way to read other people's opinions about a specific product. The creation of fake reviews that influence consumers' purchase decisions is a persistent and detrimental problem. Therefore, developing techniques to help companies and customers to distinguish between genuine and fraudulent reviews are still an important but challenging task.
3	Prof. Shekhar Patle, Gaurav Pawar, Shrinivas Pawar, Onkar Davkare, Rutuja Javalekar .	2023	This research paper presents a novel approach to address this challenge by leveraging supervised machine learning techniques for the detection of fake reviews. The proposed system begins by constructing a comprehensive dataset consisting of genuine and fake reviews, along with relevant features such as review text, reviewer information, and rating patterns.

©IJRASET: All Rights are Reserved

V. METHODOLOGY

The methodology for this research on fake review detection encompasses several structured steps, including data acquisition, preprocessing, model development, and evaluation.

1) Data Acquisition:

Datasets containing both authentic and fraudulent reviews will be sourced from various online platforms, such as Amazon, Yelp, and hotel review sites. These datasets will be selected to represent a diverse array of products and review styles, ensuring a comprehensive analysis. The data will include multiple attributes, such as review text, user ratings, timestamps, and product information.

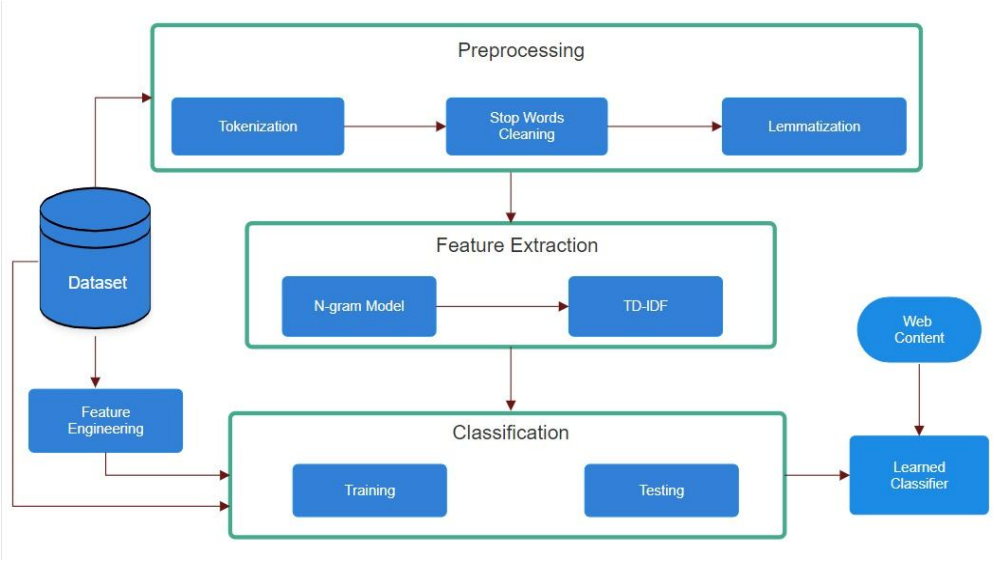
2) Data Preprocessing

The collected data will undergo a thorough preprocessing phase, which includes:

Data Cleaning: Removing irrelevant entries, duplicates, and handling missing values to ensure high-quality data for analysis.

Text Preprocessing: This includes converting all text to lowercase, removing special characters, and employing techniques like stemming or lemmatization to standardize the language used in reviews.

Feature Engineering: Key features will be extracted from the review text, such as sentiment scores, word counts, and specific lexical features, as well as metadata like user review frequency and overall ratings.



Fig(1): System Architecture

3) Model Development

A combination of supervised and semi-supervised machine learning algorithms will be implemented for detecting fake reviews. The primary models will include:

Support Vector Machines (SVM): Chosen for its effectiveness in classification tasks, particularly in high-dimensional spaces.

Logistic Regression: Used as a baseline model due to its simplicity and interpretability.

Ensemble Methods: Such as Random Forest, which combines multiple decision trees to improve prediction accuracy.

Deep Learning Approaches: Future work will incorporate advanced techniques like Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) to capture intricate patterns in textual data.

4) Training and Testing

The dataset will be divided into training and testing subsets, typically using an 80/20 split. The training set will be utilized to fit the machine learning models, while the testing set will evaluate their performance. Techniques such as cross-validation will be applied to enhance model robustness and mitigate overfitting.

5) Model Evaluation

The performance of each model will be assessed using multiple metrics:
Accuracy: The ratio of correctly identified reviews to the total number of reviews.
Precision: The proportion of true positive reviews among all reviews predicted as fake.
Recall: The proportion of true positive reviews among all actual fake reviews.
F1 Score: The harmonic mean of precision and recall, providing a balanced view of performance.
Confusion Matrix: To visualize the performance of the classification model and understand the misclassifications.

VI. CONCLUSION

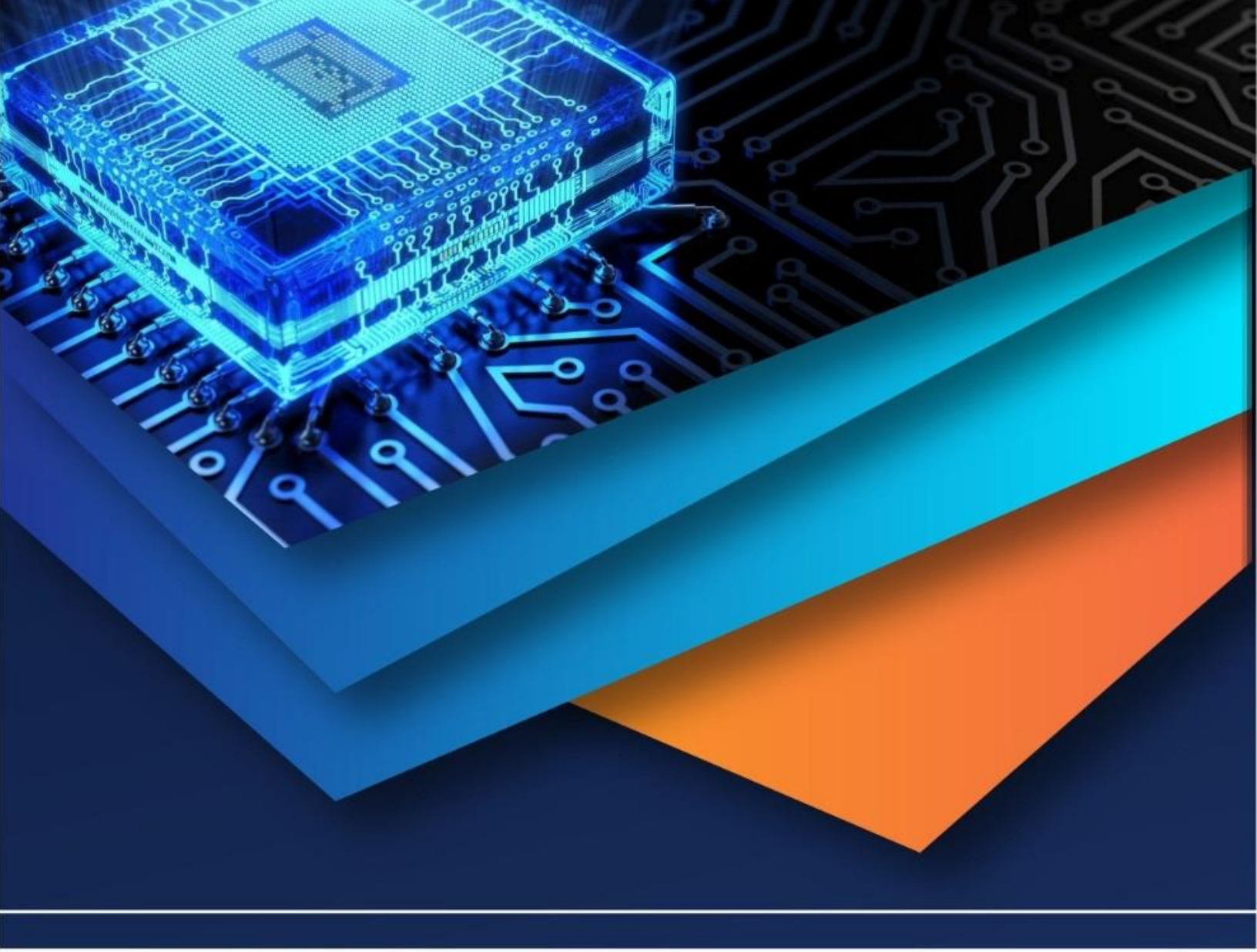
In this paper, we propose a model for detecting fake reviews using machine learning algorithms, specifically Support Vector Machines (SVM). Our model demonstrates a high level of accuracy in identifying fraudulent reviews. Fake review detection is an emerging area of research, particularly due to the limited availability of open datasets. Through this project, our goal is to not only achieve high accuracy but also minimize the time required to identify fake reviews. Additionally, the model is designed to detect multiple fake reviews, making it a practical solution for real-world applications.

VII. FUTURE SCOPE

Future research directions include several key improvements. First, a more detailed investigation could incorporate performance and time-based analysis to determine if certain individuals are posting reviews at an unusually high frequency. Additionally, expanding the range of datasets used, such as incorporating the Amazon dataset, would enrich the study. Future studies should also explore the use of deep learning techniques like LSTM, CNN, and RNN. A hybrid approach combining feature selection with sentiment analysis is strongly recommended to improve accuracy. Finally, to better recognize user behavior, there should be a focus on developing chatbots that can accurately interpret and respond to human emotions.

REFERENCES

[1] Chengai Sun, Qiaolin Du and Gang Tian, “Exploiting Product Related Review Features for Fake Review Detection,” Mathematical Problems in Engineering, 2016.
[2] A. Heydari, M. A. Tavakoli, N. Salim, and Z. Heydari, ”Detection of review spam: a survey”, Expert Systems with Applications, vol. 42, no. 7, pp. 3634–3642, 2015.
[3] 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: Human Language Technologies (ACL-HLT), vol. 1, pp. 309–319, Association for Computational Linguistics, Portland, Ore, USA, June 2011.
[4] J. W. Pennebaker, M. E. Francis, and R. J. Booth, ”Linguistic Inquiry and Word Count: Liwc,” vol. 71, 2001.
[5] S. Feng, R. Banerjee, and Y. Choi, “Syntactic stylometry for deception detection,” in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers, Vol. 2, 2012.
[6] J. Li, M. Ott, C. Cardie, and E. Hovy, “Towards a general rule for identifying deceptive opinion spam,” in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL), 2014.
[7] E. P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, “Detecting product review spammers using rating behaviors,” in Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM), 2010.
[8] J. K. Rout, A. Dalmia, and K.-K. R. Choo, “Revisiting semi-supervised learning for online deceptive review detection,” IEEE Access, Vol. 5, pp. 1319–1327, 2017
[9] Beutel A, Murray K, Faloutsos C, Smola AJ (2014) CoBaFi - Collaborative Bayesian filtering. In: Proceedings of 23rd international conference on world wide web, pp 97–108
[10] Cao Q, Sirivianos M, Yang X, Pregueiro T (2012) Aiding the detection of fake accounts in large scale social online services. In: Proceedings of 9th USENIX symposium on networked systems design and implementation, pp 197–210



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)