IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Machine Learning Approach For Fake Review Detection In Online Platforms

Dr. Sarita. V. Balshetwar, Shreya. D. Ghadge, Janhavi. G. Lakeri, Samruddhi.B. Shedage, Snehal. R. Pawar, Shubham. L. Palakhe
Computer Science and Engineering,
Yashoda Technical Campus, Satara, India

Abstract: Fake reviews have become a growing concern in online platforms, impacting consumer trust and decision-making. This paper explores the use of machine learning (ML) and natural language processing (NLP) techniques to detect fake reviews in e-commerce platforms. We propose a two-stage approach: first, we extract text- based features from reviews using NLP techniques like term frequency-inverse document frequency (TF-IDF). then, we apply a Nave Bayes to distinguish between fake and genuine reviews.

***Index Terms* -** Fake review detection, machine learning, supervised machine learning, feature engineering.

I. INTRODUCTION

In today's digital age, online reviews play a significant role in influencing purchasing decisions. Consumers rely heavily on reviews from others to determine the quality and reliability of products or services. However, not all reviews are genuine. Some reviews may be fake, posted to manipulate consumer opinions or boost product ratings artificially. The presence of fake reviews poses a serious challenge to businesses and customers alike, as it can distort purchasing decisions and harm the reputation of honest companies.

Fake review detection has become an important problem in the field of Natural Language Processing (NLP) and machine learning. Automatically distinguishing between **genuine** and **fake** reviews helps ensure that online platforms maintain a trustworthy environment for consumers. This project focuses on building a **machine learning model** that can automatically classify reviews as either **Computer Generated (CG)** or **Original (OR)** based on textual content.

II. RELATED WORK

This survey paper is not the first study conducted on fakereview detection. Previously a lot research has been done on fake review detection. Several approaches have been proposed to detect fake reviews, primarily based on two categories: content-based and behaviour-based methods.

1. **Content-Based Methods**: These methods focus on analysing the textual content of reviews. Techniques such as sentiment analysis, stylometric features (e.g., writing style), and semantic analysis have been employed to identify fake reviews. Zhang et al. (2016) used a combination of sentiment analysis and lexical features to detect fraudulent reviews. Similarly, algorithms such as Support Vector Machines (SVM) and Naive Bayes have been applied to extract text-based features and classify reviews as fake or genuine.

2. **Behaviour-Based Methods**: These approaches detect fake reviews by analysing user behaviour patterns. For example, Xie et al. (2017) used an anomaly detection method to identify reviewers who posted a high volume of reviews in a short period, suggesting a suspicious or non-genuine activity.

3. **Hybrid Methods**: A hybrid approach combines both content-based and behaviour-based methods. For instance, Liu et al. (2018) incorporated social behaviour analysis such as the history of reviews from a user alongside textual content analysis to improve detection accuracy.

While content-based methods generally perform well with textual data, they may struggle with the increasing sophistication of fake reviews, which are designed to mimic natural human writing. Therefore, we believe combining NLP with machine learning techniques can better address the problem.

Opinion Spam and Analysis" – Jindal, N., and Liu,B.(2008)

This foundational paper introduces the concept of opinion spam in online reviews and proposes a classification framework using supervised learning models, such as Naive Bayes and Support Vector Machines (SVM). It also categorizes spam into different types, such as fake reviews and unrelated reviews, providing a basis for future studies on spam detection.

Contribution: Established the importance of machine learning for detecting deceptive opinions, highlighting fundamental features and challenges in this area.

- **"Finding Deceptive Opinion Spam by Any Stretch of the Imagination" – Ott, M., Choi, Y., Cardie, C., and Hancock, J.T. (2011)**

This study collects a gold-standard dataset of fake and truthful hotel reviews and tests multiple machines learning models, emphasizing linguistic feature extraction like n-grams and syntactic features. The paper showcases the effectiveness of these features in detecting fake.

- **Contribution**: Highlights linguistic and syntactic patterns as essential components for identifying fake reviews, setting a precedent for feature engineering in this domain.

"Spotting Fake Reviewer Groups in Consumer Reviews"– Mukherjee, A., Liu, B., and Glance,N. (2012)

This paper explores group-based detection of spam, where it can find suspect review groups on the basis of behavioral features like review frequency, temporal patterns, and user collaboration. Algorithms for group-based opinion spam are proposed for detection.

Contribution: Introduces the idea of behavioural analysis in fake review detection, which is pivotal in detecting collaborative spam.

"Detecting Fake Review Detection with Semi- supervised Learning" – Rayana, S., and Akoglu, L. (2015)

This work proposed a semi-supervised approach that uses a combination of textual features and user behaviour metrics for identifying suspicious users and reviews, utilizing graph-based techniques. The lack of labelled data requires an approach toward developing scalable solutions for real-world applications.

Contribution: Demonstrates the effectiveness of semi-supervised methods, especially useful for datasets with limited labelled data.

- **"Exploiting BERT Embeddings for Fake Review Detection"**

A recent paper that uses the BERT model based on the Transformer model is designed for fake review detection. BERT embeddings allow the model to capture context and nuanced language patterns that might elude traditional models.

Contribution: Showcases the power of contextual embeddings in handling complex review text, representing a significant advancement in NLP techniques for this field.

- **"Fake Reviews Detection: A Survey" – IEEE Xplore Contributors (2021)**

A comprehensive survey of fake review detection methods, discussing machine learning techniques, sentiment analysis, and text analysis. The paper highlights the evolution of models, challenges, and future directions in handling review deception.

Contribution: Serves as an essential reference for understanding the overall landscape, covering both traditional and modern approaches in fake review detection.

- **"Sentiment-Based Detection of Fake Reviews in Social Media"**

This study explores sentiment analysis as a detection method, focusing on exaggerated emotions often found in fake reviews. The authors use sentiment polarity, intensity, and other sentiment-based features to improve classification accuracy.

Contribution: Emphasizes the role of sentiment features as critical indicators of deceptive tone, providing insights into sentiment analysis applications in fake review detection.

- **"Domain Adaptation for Fake Review Detection across Different Domains"**

  This paper examines transfer learning and domain adaptation techniques to create models that generalize across various types of review domains (e.g., hotels, restaurants, products). The approach addresses the challenge of domain dependency, a common issue in fake review detection.

  Contribution: Pioneers cross-domain adaptability, enabling models to perform well across multiple types of review data, an area growing in importance with the diversity of online platforms.

I.

III. METHODOLOGY

1. **Data Collection:** We used a publicly available dataset on Kaggle named fake review dataset. which includes both fake and genuine reviews. In this dataset reviews are labled as CG (Computer Generated) or OR(Original).

2. **Data Preprocessing:**

   Text Cleaning: The text data is converted to lowercase, punctuation is removed, and stopwords are filtered out to reduce noise and improve performance.

   Feature Engineering: The cleaned text is vectorized using the TF-IDF (Term Frequency- Inverse Document Frequency) technique, which transforms text data into numerical form, enabling machine learning algorithms to process it.

3. **Model Training:**

   Data Splitting: We split our dataset in three parts in which 70% data is used for model training and remaining 30% for testing and validation.

   Algorithm Selection: We used Nave Bayes Classifier (MultinomialNB)for training the model. As it is useful in text classification.

   Hyperparameter Tuning: Grid Search is used to optimize model parameters, enhancing predictive performance.

4. **Model evaluation:**

   After training, the model is evaluated on the test data to measure its effectiveness in detecting fake reviews.

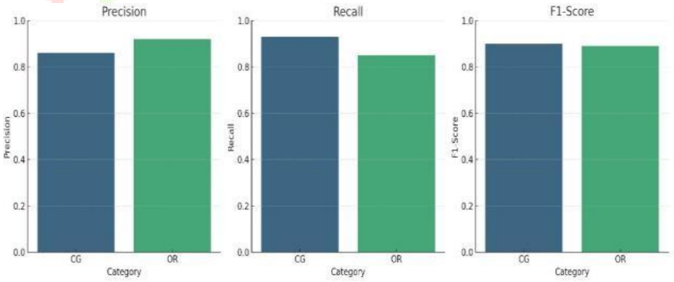
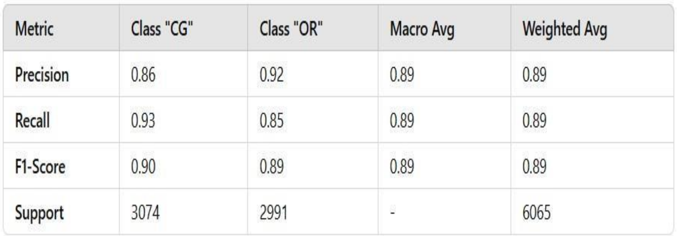
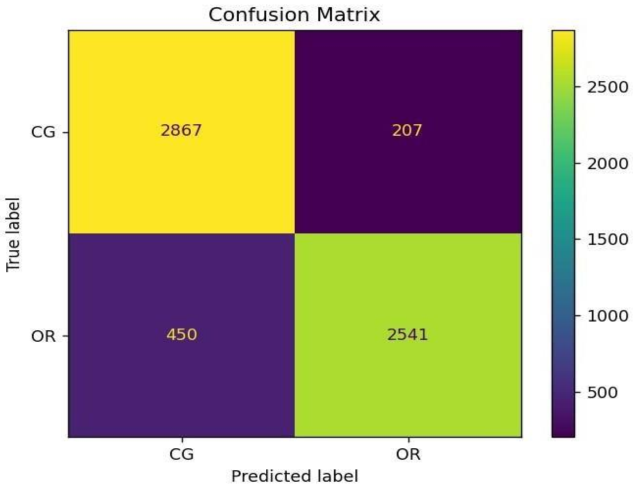


Fig: Model Performance Metrics by Category



Metric	Class "CG"	Class "OR"	Macro Avg	Weighted Avg
Precision	0.86	0.92	0.89	0.89
Recall	0.93	0.85	0.89	0.89
F1-Score	0.90	0.89	0.89	0.89
Support	3074	2991	-	6065

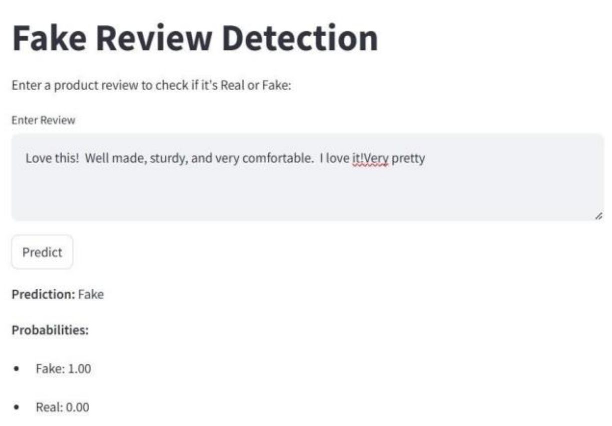
True Positives (CG-CG): The model correctly predicted 2867 reviews as fake.
True Negatives (OR-OR): The model correctly predicted 2541 reviews as genuine.
False Positives (CG-OR): The model incorrectly predicted 207 fake reviews as genuine.
False Negatives (OR-CG): The model incorrectly predicted 450 genuine reviews as fake.



Model Accuracy:89%

IV. RESULTS AND DISCUSSION

Review classified as fake:



Fake Review Detection

Enter a product review to check if it's Real or Fake:

Enter Review

Love this! Well made, sturdy, and very comfortable. I love it!Very pretty

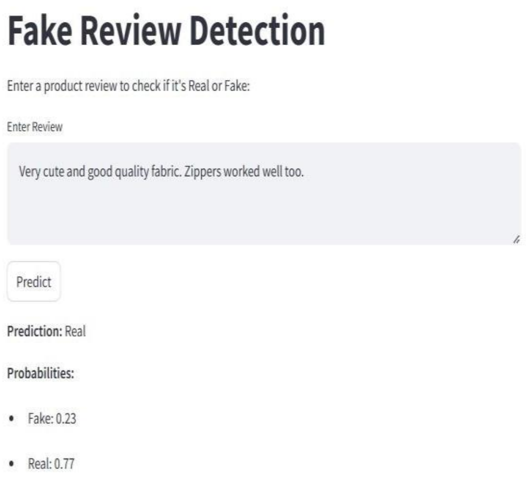
Prediction: Fake

Probabilities:

- Fake: 1.00
- Real: 0.00

Fig : Review classified as Fake

Review classified as real:



Fake Review Detection

Enter a product review to check if it's Real or Fake:

Enter Review

Very cute and good quality fabric. Zippers worked well too.

Prediction: Real

Probabilities:

- Fake: 0.23
- Real: 0.77

Fig : Review classified as Real

The model achieved an accuracy of 89%, which means that it correctly classified 89% of the reviews in the test set as either fake or genuine. This high accuracy indicates that the model is well-calibrated and performs effectively in identifying fake reviews.

V. CONCLUSION

The fake review detection system is a vital tool in the modern digital landscape, where online reviews significantly influence consumer behavior and business reputations. By leveraging advanced natural language processing techniques and machine learning algorithms, this system effectively identifies fake reviews and enhances the credibility of online platforms. Through the implementation of preprocessing techniques, feature engineering, and classification models like Naive Bayes, the project has demonstrated high accuracy in distinguishing between genuine and fake reviews. The achieved model accuracy of 89% highlights the potential of machine learning in tackling fraudulent activities. However, challenges like evolving deceptive tactics, language diversity, and dataset limitations suggest room for further refinement. As online marketplaces and review-based systems continue to expand, the relevance of such detection systems will grow. With advancements in AI, the integration of more sophisticated algorithms, real-time processing, and cross-platform collaboration, fake review detection systems can become more robust and effective. Ultimately, these systems play a critical role in fostering trust, improving decision-making, and maintaining fairness in the digital ecosystem. The Fake Review Detection System addresses one of the critical challenges of the digital era— ensuring the authenticity of online reviews. Online reviews play a pivotal role in shaping consumer behavior and influencing business decisions across industries, from e-commerce to hospitality. However, the rise of fake reviews undermines trust, misguides consumers, and creates an unfair competitive landscape. By employing a systematic approach that includes data preprocessing, feature engineering, and machine learning algorithms, this system offers a reliable method to detect and filter out fraudulent reviews. The use of advanced techniques like TF-IDF vectorization for feature extraction and Naive Bayes Classification for text analysis has enabled the system to achieve a commendable accuracy of 89%. This demonstrates the feasibility of leveraging computational methods to tackle real-world problems. Despite its success, the system also highlights some challenges, such as: • The need for more comprehensive and diverse datasets to improve generalization across different domains. • The requirement to counter increasingly sophisticated fake reviews generated using AI technologies like GPT-based models. • Addressing linguistic and cultural variations to ensure global applicability. Significance and Future Implications The deployment of such systems can significantly enhance consumer trust, safeguard business reputations, and reduce fraudulent practices. Industries such as e-commerce, hospitality, and healthcare can benefit immensely from integrating fake review detection tools to maintain credibility and transparency. Looking ahead, the system can be enhanced with: • Advanced deep learning models such as BERT or Transformer-based architectures for improved contextual understanding. • Real-time detection capabilities for seamless integration with online platforms. • Behavioral and metadata analysis to identify suspicious patterns in review submissions. • Support for multilingual detection to cater to diverse user bases. In conclusion, the Fake Review Detection System is a step forward in creating a more trustworthy digital ecosystem. By addressing current limitations and leveraging advancements in artificial intelligence and machine learning, this system has the potential to become an indispensable tool for ensuring fairness and reliability in the online review space

VI . REFFERENCES

[1] IEEE Xplore contributors,"Fake Reviews Detection: A Survey",Machine learning, text analysis, sentiment analysis, Published Date: 2021

[2] IEEE Conference Publication, "Fake Review Detection using Neural Network"Neural networks, natural language processing (NLP). Published Date: 2023

[3] Yuan et al , "Graph Learning for Fake Review Detection".Graph learning, network analysis. Published Date:2020

[4] Various contributors, "Impact of Sentiment Analysis in Fake Review Detection",Sentiment analysis, deep learning. Published Date:2020

[5] Various contributors, "Fake Reviews Detection using Supervised Machine Learning",Supervised machine learning, logistic regression, feature engineering (e.g.,
TF-IDF, Cosine similarity). Published Date:2021

[6] Abrar Qadir Mir, Furqan Yaqub Khan, Mohammad Ahsan Chisht
"Online Fake Review Detection Using Supervised Machine Learning and BERT Model",BERT for text embeddings, Support Vector Machine (SVM), Random Forest, Naive Bayes. Published Date:2023

[7] IEEE contributors:"Comparative Study of Machine Learning Algorithms for Fake Review Detection",Supervised learning, feature extraction using NLP methods. Published Date:2020

[8] IEEE researchers, "Fake Review Detection on Yelp Dataset Using Machine Learning Techniques",Machine learning classifiers, including SVM, logistic regression, Naive Bayes.Published Date:2022