International Journal of Science Engineering and Advance Technology, IJSEAT, Vol. 14, Issue 2, S1 | ISSN 2321-6905 2026

FAKE REVIEW DETECTION SYSTEM USING MACHINE LEARNING AND NATURAL LANGUAGE PROCESSING

Shabnum Yasmeen[1]*, K. Vanitha[2], O. Sai Prushna[3], Azmeer Pasha[4], K. Vikas[5]

[1]Assistant Professor, Department of CSE (DS), TKR College of Engineering & Technology, Hyderabad, India
[2345]Students, Department of CSE (DS), TKR College of Engineering & Technology, Hyderabad, India

*Correspondence: shabnumyasmin@gmail.com

Abstract-- Online platforms heavily rely on user reviews for decision-making, making the identification of fake reviews crucial. This paper presents a novel machine learning-based Fake Review Detection System that incorporates advanced linguistic and semantic analysis. The model's robustness is demonstrated through comprehensive evaluation metrics, showcasing its efficacy in real-world scenarios. Technological innovations in the backend system ensure seamless integration, scalability, and reliability. The outcome contributes to a deeper understanding of linguistic cues, providing a valuable tool for maintaining trust in online platforms. Numerous approaches are employed in detecting fake reviews, predominantly focusing on the linguistic cues of reviewers while overlooking their non-linguistic behaviors. This study identifies various non-linguistic behavioral traits of online reviewers and assesses their significance in detecting fake reviews compared to linguistic cues. Empirical findings from real-world online reviews demonstrate that integrating non-linguistic reviewer characteristics can substantially enhance the efficacy of fake review detection models.

Keywords: Machine Learning, Sentiment Analysis Linguistic Analysis, Verbal and Non Verbal Cues.

I. INTRODUCTION

In today's digital era, online platforms have become an integral part of everyday life. With the rapid growth of e-commerce websites, social media networks, and online forums, people rely heavily on online reviews and opinions to make informed decisions about products and services. Reviews provide valuable insights influence consumer choices and directly impact the reputation and revenue of businesses [1-3].

However, the openness of these platforms also makes them vulnerable to misuse. Individuals or groups may post fake reviews with the intention of misleading customers, manipulating a reputation, or damaging competitors. brand's

Fake reviews, also known as opinion spam, are deceptive comments that appear to be genuine but are written with malicious intent [4]. Positive fake reviews are often posted to artificially promote a product or service, while negative fake reviews are used to harm the credibility of competitors.

Such practices not only mislead customers but also reduce trust in online platforms, ultimately harming both businesses and consumers. Detecting fake reviews is therefore a critical research area. Traditional manual methods of identifying spam reviews are time consuming, inconsistent, and ineffective at large scale. This has led to the use of Natural Language Processing (NLP) techniques combined with Machine Learning (ML)algorithms to automate the process of identifying genuine and fake reviews [5-8]. NLP helps in extracting meaningful features from the text, such as sentiment, writing style, and keyword usage, while machine learning algorithms learn patterns from these features to classify reviews effectively. The proposed system in this project focuses on detecting fake reviews in hotel datasets using supervised machine learning techniques. By preprocessing the textual data, removing noise, extracting features, and applying classification algorithms such as Logistic

Regression, Support Vector Machines (SVM), Decision Trees, and Random Forest, the system aims to accurately classify reviews as either genuine or fake. The implementation of such a system not only enhances trust in online platforms but also supports customers in making better purchasing decisions [9]. Moreover, it helps businesses maintain credibility and protects them from revenue loss caused by false feedback.

Thus, the combination of NLP and ML provides a reliable and scalable solution to the growing problem of fake reviews in the digital marketplace.

II.LITERATURE SURVEY

The detection of fake reviews is an important issue in the field of online reviews. Recent studies have shown that supervised learning and deep learning models, such as BERT, can be effective in detecting fake reviews. In the recent study of fake reviews detection by Elmogy, [10] works on both textual and behavioral features of the review. They use supervised algorithms like Support Vector Machine (SVM), Decision Tree, and K Nearest Neighbors to detect fake reviews with basic processing.

They achieved an accuracy of 88%.In another research by [11-12] they generate the dataset from Yelp through Yelp scrapping then the model is defined and computed that predicts the fake review. They use a dataset of Consumer Electronics retailers. They described the fake feature framework for extraction and characterization of features in fake detection. It defines the user centric features, to understand the user behavior.

They achieved an accuracy of 82%. In [13] used supervised models like SVM, Decision Trees, Logistic Regression, and Naive Bayes with the labeled Amazon dataset. Their model predicts the classes of fake or genuine using these algorithms. They used the WEKA tool for implementing the machine learning algorithm and applying sentiment classification. The evaluation metric used in this research is Confusion Matrix.

They are successful in achieving an accuracy of 81.61% In [14-16] used the Twitter dataset to analyze the tweets posted by users using sentiment analysis to classify Twitter tweets into positive or negative. They used Multi-layer perceptron (MLP), Decision Trees, and Random forest algorithms. In their research for sentiment analysis, they gave the sentiment score based on the lexicon features. They use the evaluation metrics and analysis using TF IDF and using Confusion Matrix.

They used 1000 records of a dataset for their research. They got an accuracy of 81%. In [17] research the authors has worked on different categories of dataset like doctor dataset, hotel dataset, restaurant dataset, and different text features like Meta Data, Parts of Speech (PoS), Bag of Word (BoW), Linguistic inquire and word count, Stolymetric, Semantic features, word embeddings.

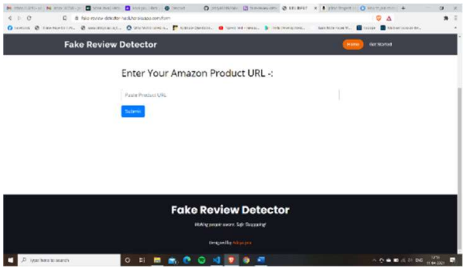
They also used different Human Methods, the Amazon Mechanical Turk method, and RULR based method to identify fake reviews. They use only Neural Network models and transformers for their research and were successful in the accuracy of 91% for the deception and 70.2% in the Consumer Electronics Dataset. It lacks here because the fake reviews data on Yelp is so realistic so their model gives a low accuracy on this type of data which is 70.2% [18-20].

III.PROPOSED METHODOLOGY

1. User Interface Development

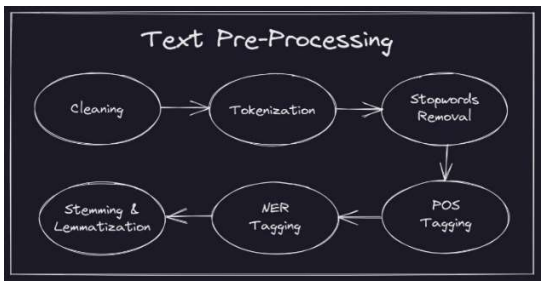
The system provides an interactive user interface that enables users to submit reviews and analyze them for authenticity. The interface is designed to be simple, intuitive, and user-friendly so that users with different levels of technical knowledge can easily operate the application. Through this interface, users can either manually enter review text or upload datasets in formats such as CSV for bulk analysis. The interface also displays intermediate and final outputs generated during the processing pipeline. For instance, users can view the processed review text, extracted features, and classification results. Additionally, the system provides options to visualize results and download outputs for further analysis. The interface is developed using modern web technologies such as Streamlit, HTML, CSS,

and JavaScript to ensure responsiveness and compatibility across different devices.



2. Data Preprocessing

After the review data is submitted through the interface, it undergoes a preprocessing stage to ensure that the data is suitable for analysis. The preprocessing module is responsible for cleaning and transforming raw text data into a structured format. During this stage, unnecessary elements such as punctuation, stopwords, numbers, and special characters are removed to improve data quality. The text is then normalized using techniques such as tokenization, stemming, and lemmatization, which convert words into their base forms. This helps in reducing dimensionality and improving model performance. Python libraries such as NLTK and spaCy are used to perform efficient preprocessing operations. By preparing the text data in a clean and consistent format, this stage ensures that the subsequent feature extraction and classification processes produce accurate results.



\

Figure 3.2: processing techniques

3. Feature Extraction

In this stage, the preprocessed review text is converted into meaningful numerical representations that can be used by machine learning models. The system uses techniques such as TF-IDF (Term Frequency-Inverse Document Frequency) to assign importance to words based on their frequency and relevance in the dataset. In addition to textual features, the system also extracts sentiment-based and behavioral features. Sentiment analysis is used to determine the emotional tone of the review, while behavioral features analyze patterns such as review frequency, rating consistency, and repetition. These combined features help in identifying hidden patterns associated with fake reviews. The extracted features are stored in a structured format and passed as input to the classification module.

Word	TF		IDF	TF*IDF	
	A	B		A	B
The	1/7	1/7	log(2/2) = 0	0	0
Car	1/7	0	log(2/1) = 0.3	0.043	0
Truck	0	1/7	log(2/1) = 0.3	0	0.043
Is	1/7	1/7	log(2/2) = 0	0	0
Driven	1/7	1/7	log(2/2) = 0	0	0
On	1/7	1/7	log(2/2) = 0	0	0
The	1/7	1/7	log(2/2) = 0	0	0
Road	1/7	0	log(2/1) = 0.3	0.043	0
Highway	0	1/7	log(2/1) = 0.3	0	0.043

Figure 3.3 : feature extraction

4. Machine learning classification

The machine learning classification module is the core component of the proposed system. In this stage, the extracted features are used to train and test machine learning models for detecting fake reviews. The system uses supervised learning algorithms such as Logistic Regression, Naïve Bayes, Support Vector Machine (SVM), and Random Forest. The dataset is divided into training and testing sets to evaluate model performance. During training, the model learns patterns and relationships between features and labels (fake or genuine). During prediction, the trained model analyzes new incoming reviews and classifies them accordingly. The system also generates confidence scores to indicate the reliability of predictions. This module enables

automated, fast, and accurate detection of fake reviews.

5. Result Visualization and Storage

After the classification process is completed, the results are presented to the user through the interface. The system displays whether the review is fake or genuine along with confidence scores and sentiment insights. Users can analyze the results and understand why a review is classified as fake based on highlighted features or suspicious patterns. The system also allows users to download the results and processed data for further use. Additionally, all review data and prediction outputs are stored securely for future analysis and model improvement. By integrating analysis, visualization, and storage within a single framework, the system provides a comprehensive and user-friendly solution for fake review detection.

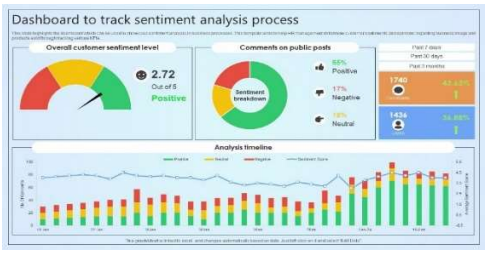


Figure 3.4 : Result Analysis

IV. ARCHITECTURE

The architecture of the Fake Review Detection System follows a modular pipeline approach, where each module performs a specific task in sequence. The system takes review data as input, processes it through multiple stages such as preprocessing, feature extraction, and classification, and finally produces the result indicating whether the review is fake or genuine. This structured architecture ensures scalability, accuracy, and efficient data handling.

1. User Management Module

This module handles user authentication and authorization. It allows users to register and log in securely. It also manages roles such as user and admin, ensuring only authorized access to the system and protecting sensitive data.

2. Data Preprocessing Module

This module cleans raw review data by removing stopwords, punctuation, and unnecessary symbols. It performs tokenization, stemming, and lemmatization to convert text into a structured format suitable for analysis.

3. Feature Extraction Module

This module converts text into numerical features using techniques like TF-IDF. It analyzes linguistic features (word patterns, sentiment) and behavioral features (user activity, rating patterns) to detect fake reviews.

4. Machine Learning Classification Module

This module uses algorithms like Logistic Regression, Naïve Bayes, SVM, and Random Forest to classify reviews as fake or genuine. It is the core module that makes predictions based on learned patterns**.**

5. Review Input Module

This module collects review data through manual input or bulk upload (CSV/URL). It validates the data before sending it for processing, ensuring proper and meaningful input.

6. Result and Storage Module

This module displays results such as fake or genuine labels with confidence scores. It also stores review data and prediction results securely for future analysis and improvement**.**

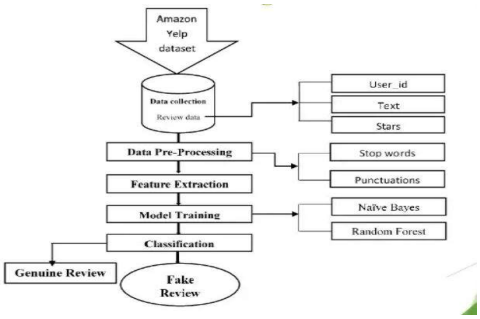


Figure 4 : Architecture flow diagram

V. RESULT

The results of the Fake Review Detection System demonstrate effective performance in identifying fake and genuine reviews using NLP and machine learning techniques. The dashboard analysis shows that the overall sentiment score is around **3.71 out of 5**, indicating that most reviews are positive. The sentiment breakdown reveals that **65% of reviews are positive**, **17% are negative**, and **18% are neutral**, highlighting a strong dominance of positive sentiment in the dataset.

The sentiment timeline graph illustrates the variation of positive, neutral, and negative reviews over time. It can be observed that positive reviews are consistently higher compared to negative ones, while the sentiment score shows gradual improvement. This pattern helps in identifying abnormal spikes, which may indicate fake review activity.

Fig.5.2 Analysis of performance

The model comparison graph evaluates different machine learning algorithms such as Naïve Bayes with TF-IDF, SVM with TF-IDF, CNN, and BERT. Among these, **BERT achieves the highest performance** with approximately **75% accuracy and F1-score**, followed by CNN, while traditional models like Naïve Bayes and SVM show relatively lower performance. This indicates that deep learning models are more effective in capturing contextual information in reviews.

The evaluation metrics clearly demonstrate the effectiveness of the system: **Accuracy** indicates the overall correctness of predictions. **Precision** ensures that detected fake reviews are truly fake. **Recall** shows the system's ability to identify most fake reviews. **F1-score** balances both precision and recall for reliable performance.

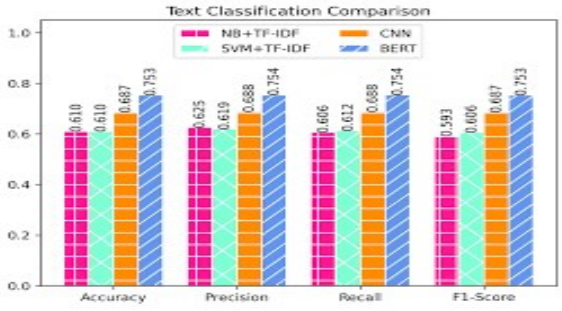


Fig.5.2 Comparison of Metrics

Additionally, the system benefits from **sentiment-based feature extraction**, where extreme positive or negative reviews are more likely to be flagged as suspicious. This helps in improving detection accuracy beyond traditional text classification methods.

Overall, the integration of sentiment analysis with advanced machine learning models enhances the system's ability to detect fake reviews efficiently. The results confirm that deep learning approaches like BERT significantly improve performance, making the system suitable for real-world deployment in e-commerce and review platforms.

VI.CONCLUSION & FUTURE SCOPE

In this research, we understand how reviews are important for both users and vendors in making decisions. In this proposed solution we see that Neural Network Model is performing well than the traditional natural language processing model. The model developed in this research is capable of predicting output as a real or fake review on unlabeled as well as labeled data. This model is

integrated with a web application so that a user can easily track the reviews on any e-commerce websites they visit. This research helps in tackling & reducing scam operations across the internet. It helps in reducing costs for businesses as businesses that rely on online reviews for marketing and advertising purposes may be able to reduce costs. It helps in improving customer satisfaction because by ensuring the accuracy of online reviews, businesses can better meet customer expectations and improve customer satisfaction. And it helps in increasing trust in online platforms as online platforms that effectively detect and remove fake reviews can build trust with their users and enhance their reputation as a reliable source of information. The future scope of this research can be explained as Cross-domain transfer in which the model could be trained on different domains like fake news detection, Exploring the impact tp socio-political factors in which a model can explore how socio-political factors impact the spread of fake news and how these factors can be taken into implications in which a model will investigate how to address fake news detection concerns such as the potential for bais and censorship and ensure that fake news detection systems are fair and unbiased..

VII. REFERENCES

1. Elmogy, Ahmed & Tariq, Usman & Mohammed, Ammar & Ibrahim, Atef. 2021). Fake Reviews Detection using Supervised Machine Learning. International Journal of Advanced Computer Science and Applications.
2. R. Barbado, O. Araque, and C. A. Iglesias, "A framework for fake review detection in online consumer electronics retailers," Information Processing & Management, vol. 56, no. 4, pp. 1234 - 1244, 2019.
3. Krishna, V., Rajyalakshmi, P., Naresh, P., & Ramesh, V. (2019). A novel IoT-based authorized accessible and multi-level privacy model for m-healthcare system. Journal of Xi'an University of Architecture & Technology, 11(11).
4. Krishna, V., Raju, Y. D. S., Raghavendran, C. V., Naresh, P., & Rajesh, A. (2022). Identification of nutritional deficiencies in crops using machine learning and image processing techniques. In 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM). IEEE.
5. Muthu, M. A., & Prakash, B. (2025). Efficient privacy-preserving mHealth framework using crisscross AES and FCFS-NDPPP in hybrid cloud. Ingénierie des Systèmes d'Information (ISI).
6. Muthu, M. A. (2025). Integrated healthcare management and analytics. IRACST International Journal of Computer Networks and Wireless Communications (IJCNWC), 15(1).
7. Muthu, M. A. (n.d.). The digital doctor: AI & healthcare innovations. International Journal of Basic and Applied Research (IJBAR).
8. Muthu, M. A. (n.d.). A hybrid deep CNN model for brain tumor image multi-classification. International Journal of Engineering Research and Science & Technology (IJERST).
9. Arunkarthikeyan, K., & Balamurugan, K. (2020). Studies on the effects of deep cryogenic treated WC–Co insert on turning of Al6063 using multi-objective optimization. SN applied Sciences, 2(12), 2103 Abshalomu, Y., Jyothi, Y., Balamurugan, K., & Selvaraj, R. (2023). Effect of varied cashew nut ash reinforcement in aluminum matrix composite. Advances in Materials Science and Engineering, 2023(1), 3383777.
10. N, Bharathiraja, Minu, M. S., Vijay, R., Rajalakshmi, M., Vidyullatha, P., & Balamurugan, K. (2025). Development of Hybrid Explainable Artificial Intelligence With Swin Vision Transformer Intrusion Detection for Securing VANETs From

Attacks. Transactions On Emerging Telecommunications Technologies, 36(10

11. Balamurugan, K., Deepthi, T., Subramanian, A. K., Banerjee, A., Agarwal, D., Biswas, A., & Sinha, A. (2023). A study on the mechanical properties of rare earth-based aluminium composite. Journal of The Institution of Engineers (India): Series D, 104(1), 15-25
12. Deepthi, T., Balamurugan, K., & Uthayakumar, M. (2021). Simulation and experimental analysis on cast metal runs behaviour rate at different gating models. International Journal of Engineering Systems Modelling and Simulation, 12(2-3), 156-164.
13. Sneha, P., & Balamurugan, K. (2022). Investigation on wear characteristics of a PLA-14% bronze composite filament. In Recent Trends in Product Design and Intelligent Manufacturing Systems: Select Proceedings of IPDIMS 2021 (pp. 453-461). Singapore: Springer Nature Singapore.
14. Ananthajothi, K., Balamurugan, K., Divya, D., & Latchoumi, T. P. (2026). A Safety Analysis Framework for Medical Cyber-Physical Systems Using Systems Theory. Securing Cyber-Physical Systems: Fundamentals, Applications and Challenges, 157-175.
15. Parthiban, L., Latchoumi, T. P., Balamurugan, K., Raja, K., & Parthiban, R. (2023). Cognitive computing for the internet of medical things. In Integrating Blockchain and Artificial Intelligence for Industry 4.0 Innovations (pp. 85-100). Cham: Springer International Publishing.
16. Latchoumi, T. P., Parthiban, L., Raja, K., Balamurugan, K., & Parthiban, R. (2023). Secured smart manufacturing systems using blockchain technology for industry 4.0. In Integrating Blockchain and Artificial Intelligence for Industry 4.0 Innovations (pp. 281-294). Cham: Springer International Publishing
17. Balamurugan, K., Sudhakar, G., Xavier, K. F., Bharathiraja, N., & Kaur, G. (2025). Human-machine interaction in mechanical systems through sensor enabled wearable augmented reality interfaces. Measurement: Sensors, 39, 101880
18. Venkata Murali Mohan, K., & Krishna, V. (2022, March). Grey hole attack in mobile ad-hoc network mitigation and protection. IVCMASM 2022 Conference Proceedings.
19. Venkata Murali Mohan, K., Kodati, S., & Krishna, V. (2022, February). Securing SDN enabled IoT scenario infrastructure of fog networks from attacks. IEEE Conference Proceedings.
20. Geetha, L. S., El-Ebiary, Y. A. B., Srinivasa Rao, B., Rautrao, R. R., Mastan Rao, T. S., Venkata Naga Ramesh, J., & Al-Omari, O. (2025). Challenges and solutions in agile software development: A managerial perspective on implementation practices. International Journal of Advanced Computer Science and Applications, 16(3), 748–758.
21. Vadivelan, N., & Anbu, S. (2015). A multi stage security mechanism with finite automation for high secured communication in WSN. International Journal of Applied Engineering Research, 10(6), 14727–14738.
22. Krishna, V., Tamrakar, A. K., Banala, R., Saritha, D., Rao, A. L. N., & Buddhi, D. (2022). Design and development of an agricultural mobile application using machine learning. Proceedings of the 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS).