iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: https://doi.org/10.22214/ijraset.2024.59666

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 III Mar 2024- Available at www.ijraset.com

Fake Review Detection using BERT Transfer Learning Algorithm

P. Boobalan[1], Devi Sireesha Seru[2], Bhavana Madireddy[3], Madhan Kumar[4], Ganesh A[5]
Information Technology Department, Puducherry Technological University

Abstract: ***In this research, our main focus is on tackling the issue of fake online reviews, which has become increasingly prevalent due to the vast amount of content generated by users on various online platforms. These reviews have a big impact on what people decide to buy, emphasizing how important it is for them to be real and trustworthy. However, the rise in fake reviews has cast a shadow of doubt over the credibility of online platforms, making it imperative to find effective ways to discern the real from the fake. To address this challenge, we propose a strong model that utilizes the capabilities of advanced deep learning algorithms—BERT (Bidirectional Encoder Representations from Transformers) and Bidirectional Long Short-Term Memory Networks (BiLSTMs)—to accurately identify and filter out fake reviews. BERT excels in understanding the nuances of language and BiLSTMs effectively capture the order of words and phrases. Our plan involves training and validating this model using datasets sourced from reputable platform like Kaggle. By amalgamating these powerful algorithms and datasets, we aim to significantly enhance the accuracy and credibility of our fake review detection system. Ultimately, our goal is to restore consumer trust and empower individuals to make more informed choices in the expansive realm of online consumerism.***
Keywords: Fake reviews, ensemble model, Algorithms, deep learning, BERT, LSTMs, BiLSTMS.

I. INTRODUCTION

In today's digital landscape, the proliferation of online reviews has transformed the way consumers make decisions about purchases. These reviews provide a firsthand account of experiences with a product, service, or establishment. They offer a sense of community, empowering consumers with collective knowledge and influencing their choices. However, this immense influence has attracted a darker side – the rise of fake reviews. Fake reviews are strategically fabricated to either boost or undermine a product or service, often as a part of marketing strategies or to tarnish a competitor's reputation. Detecting fake reviews is a multifaceted challenge that requires sophisticated technological solutions. These reviews are becoming increasingly sophisticated, mimicking genuine ones to evade detection. Cutting-edge technologies like artificial intelligence, natural language processing, and machine learning play a crucial role in the fight against fake reviews. They enable the analysis of review patterns, language intricacies, and other relevant features to flag suspicious content accurately. Moreover, considering the volume of reviews generated daily, automation is key to efficiently process this vast amount of data. By developing robust detection mechanisms, we can maintain the authenticity and credibility of online reviews, ensuring that consumers can continue to rely on them as a valuable resource in their decision-making process. A recent study has shown that 80% of customers tend not to buy products with high negative reviews (Tang and Cao; 2020). Hence, detecting fake reviews which mislead or deceive the customers becomes a crucial area for research. Preserving the integrity of online reviews is vital for both consumers and businesses. It's a shared responsibility to create an environment where genuine feedback prevails, enabling a transparent and honest dialogue between consumers and the businesses they engage with. Through continuous advancements in technology and a collective commitment to combat fake reviews, we can strive for a digital marketplace that operates on trust, honesty, and genuine consumer experiences.

A. The Merits of Deep Learning in Fake Review Detection

Deep learning has revolutionized various domains by unlocking the potential to understand and interpret complex data. This transformative power also extends to fake review detection. What sets deep learning apart is its exceptional capability to automatically extract intricate features from raw, unstructured data. In the context of fake reviews, this translates to the ability to discern nuanced linguistic, semantic, and contextual cues that are often characteristic of deceptive content. Deep learning models such as BERT, BiLSTM are designed with multiple interconnected layers, can extract high-level features, enabling a more nuanced understanding compared to traditional, shallow models. The adaptability and scalability of deep learning models add to their appeal. They can continually improve their performance by learning from new data, rendering them robust against evolving deceptive techniques.

©IJRASET: All Rights are Reserved

B. How Deep Learning Revolutionizes Fake Review Detection

Deep learning has revolutionized fake review detection by leveraging advanced neural networks to uncover intricate patterns in textual data. Unlike traditional methods, deep learning models autonomously learn features from raw data, making them highly effective in discerning nuanced characteristics of fake reviews.

1) *Feature Learning:* Deep learning models automatically learn relevant features from the input text, eliminating the need for manual feature engineering. This enables them to capture subtle linguistic cues, sentiment nuances, and contextual information.
2) *Hierarchical Representation:* Models like BERT, long short-term memory (LSTM) networks construct hierarchical representations of text, understanding relationships between words, phrases, and sentences. This helps in grasping the overall context and identifying deceptive patterns.
3) *Model Complexity:* Deep learning models are inherently complex, allowing them to handle intricate linguistic structures and variations in review texts. The deep layers of these models enable them to learn increasingly abstract and nuanced features, enhancing their ability to differentiate between genuine and fake reviews.
4) *Adaptability:* Deep learning models can adapt to evolving tactics used by spammers to generate fake reviews. As spammers modify their approaches, deep learning models can be retrained with new data, ensuring continued accuracy and effectiveness.

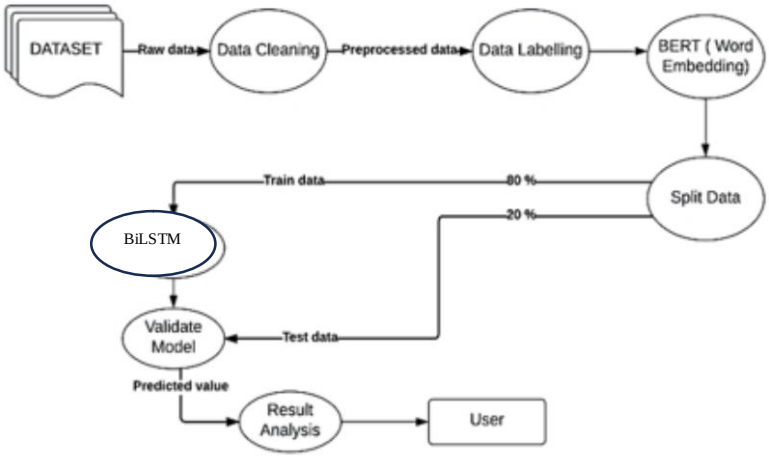


Fig.1 Fake Review Detection System Module Diagram

II. RELATED WORK

1) *Reviews Detection Using NLP Model and Neural Network Model [1] Abhijeet A Rathore* This study focuses on the detection of reviews using both NLP (Natural Language Processing) and Neural Network models. The study aims to address the growing concern of identifying authentic reviews amidst the proliferation of fake ones. NLP methods are employed to preprocess and analyze textual data, while Neural Network models are utilized for predictive analysis and classification tasks. The study involves the preprocessing of review texts, including tokenization, stemming, and removal of stopwords, to extract relevant features for analysis. Additionally, advanced NLP techniques such as sentiment analysis may be employed to understand the overall tone and sentiment expressed in reviews. On the other hand, Neural Network models are leveraged for their ability to learn complex patterns and relationships within the data. These models are trained on labeled datasets to classify reviews as authentic or fake based on learned features and characteristics. The integration of both NLP and Neural Network models allows for a robust and accurate approach to review detection. By effectively distinguishing between genuine and fraudulent reviews, the research aims to enhance consumer trust in online platforms and mitigate the impact of fake reviews on businesses.
2) *Examining Review Inconsistency for Fake Review Detection [2] Guohou Shan* This study addresses the issue of inconsistency in online consumer reviews (OCRs), which can lead to uncertainty among consumers. By examining various aspects such as rating-sentiment, content, and language, the research proposes hypotheses regarding their impact on detecting fake OCRs. Utilizing 22 features to operationalize review inconsistency, machine learning models are employed to test these hypotheses. Results confirm the presence of inconsistency and highlight its significant positive effect on the detection of fake OCRs. These findings are crucial for enhancing consumer decision-making processes and bolstering the trustworthiness of OCRs.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

3) *Content-Aware Trust Propagation Toward Online Review Spam Detection [3] HAO XUE, QIAOZHI WANG, BO LUO [3]* This study addresses the issue of review spamming on online platforms by proposing a detection scheme based on analyzing the deviation between individual reviews' aspect-specific opinions and the aggregated opinions on those aspects. By modeling the impact of users' opinions deviating from the majority on their trustworthiness, the scheme integrates a deviation-based penalty into a trust propagation framework across three layers: users, reviews, and review targets. This framework iteratively computes trust scores, which serve as effective indicators of spammers by reflecting users' overall deviation from aggregated aspect-specific opinions. Experimental results conducted on a dataset from Yelp.com demonstrate the effectiveness of the proposed detection scheme, showcasing its ability to measure users' trustworthiness based on the opinions expressed in reviews.
4) *Fake Reviews Detection using Supervised Machine Learning [4] Ahmed M. Elmogy* As E-commerce systems continue to evolve, online reviews have emerged as crucial determinants of reputation and decision-making for consumers. However, the proliferation of fake or deceptive reviews aimed at manipulating perceptions and attracting customers has underscored the need for robust detection mechanisms. This paper proposes a machine learning approach to identify fake reviews, integrating features extracted from both review content and reviewer behaviors. By conducting experiments on a real Yelp dataset of restaurant reviews, various classifiers including KNN, Naive Bayes, SVM, Logistic Regression, and Random Forest are evaluated. Additionally, different language models such as bi-gram and tri-gram are considered. The results demonstrate that incorporating features extracted from reviewer behaviors significantly improves the performance of the classifiers, with KNN achieving the highest f-score of 82.40%. This represents a notable improvement of 3.80% in f-score compared to models without behavioral features. This research highlights the effectiveness of considering reviewer behaviors alongside review content in identifying fake reviews, thus enhancing the reliability of online review systems and promoting consumer trust in E-commerce platforms.
5) *An Ensemble Model for Fake Online Review Detection Based on Data Resampling, Feature Pruning, and Parameter Optimization [5] JIANRONG YAO, YUAN ZHENG* The detection of fake online reviews has emerged as a prominent research area due to the proliferation of deceptive practices. However, existing studies often overlook the challenges posed by imbalanced data and feature pruning. To bridge this gap, our study introduces an ensemble model tailored for fake review detection. This model comprises four key steps aimed at optimizing the base classifiers. Firstly, we propose a novel approach to address data imbalance by combining resampling and grid search techniques. Secondly, an ablation study is conducted to eliminate irrelevant features through feature pruning. Thirdly, we employ the grid search algorithm to optimize parameters for each base classifier. Lastly, we integrate the optimized base classifiers using majority voting and stacking strategies to form the ensemble model. Notably, the data resampling method is also extended to the meta-classifier in the stacking ensemble. Our study represents a significant advancement by combining diverse methods and algorithms into a unified model. Experimental results demonstrate that our proposed ensemble model outperforms existing techniques, offering a novel solution to the challenges of data imbalance and feature pruning in the domain of fake review detection.

III.LITERATURE SURVEY

Table 1. Findings of survey papers

Sl. No.	Author	Title	Journal	Findings
1	Abhijeet A Rathore, Gayatri L Bhadane, Ankita D Jadhav, Kishor H Dhale, Jayashree D Muley (2023)	Fake Reviews Detection Using NLP Model and Neural Network Model	International Journal of Engineering Research & Technology (IJERT)	• The research specifies and reviews various sentiment analysis techniques, including sentiment dictionaries, rule-based systems, traditional machine learning methods, and deep learning approaches. This finding underscores the importance of employing diverse methodologies to accurately analyze sentiment in online consumer reviews (OCRs). • The study explores essential feature extraction techniques such as the bag-of-words model, TF-IDF, word embedding, attention mechanism, and Transformer. This highlights the significance of employing advanced feature extraction methods to capture nuanced information from OCRs for

				sentiment analysis tasks. • The research reveals that star ratings significantly influence hotel bookings, review helpfulness, and enhance the detection of fake OCRs. This finding emphasizes the pivotal role of star ratings as a key indicator for both consumers and platforms in assessing the quality and credibility of online reviews. • The study underscores the critical importance of sentiment analysis methods in understanding consumer attitudes expressed in OCRs.
2	Yuxin Jin, Kui Cheng and Xinjie Wang, Lecai Cai.(2023)	A Review of Text Sentiment Analysis Methods and Applications	Frontiers in Business, Economics and Management	• It specifies and reviews diverse sentiment analysis methods, spanning sentiment dictionaries, rule-based, traditional, and deep learning approaches. • explores key feature extraction methods, such as the bag-of-words model, TF-IDF, word embedding, Word2Vec, attention mechanism, and Transformer
3	Guohou Shan, Lina Zhou, Dongsong Zhang(2021)	Examining Review Inconsistency for Fake Review Detection	Elsevier on computers and security	• Star ratings influence hotel bookings, review helpfulness, and improve fake OCR detection. • Sentiment analysis methods (lexicon, machine learning, statistics, and rule-based) are crucial for understanding consumer attitudes in OCRs. • Linguistic features predict review helpfulness and contribute to fake OCR detection, with machine learning methods like Naïve Bayes and SVM dominating this task.
4	Hina Tufail, M. Usman Ashraf, Khalid Alsubhi, Hani Moaiteq AljahdalI (2022)	The Effect of Fake Reviews on e-Commerce During and After Covid-19 Pandemic: SKL-Based Fake Reviews Detection	IEEE Access	• Proposed SKL-based algorithm for detecting fake e-commerce reviews using SVM, KNN, and Linear Regression. • Used NLTK for punctuation removal and feature selection, including length count, bigram type, relationship words, sentiment count, and noun-verb count. • SKL-based solution considered relationship words and performed sentiment analysis, achieving 95% accuracy on Yelp and 89.03% on TripAdvisor. • Achieved 95% and 89.03% classification accuracy on Yelp and TripAdvisor datasets using 80-20 training-testing split and 10-fold cross-validation.
5	Hao: Xue, Qiaozhi Wang, Bo Luo, Hyunjin Seo, Fengjun Li (2019)	Content-Aware Trust Propagation Toward Online Review Spam Detection	ACM Journal of Data and Information Quality	• In the paper, it specifies Spammers manipulate ratings over time to evade detection. • The study proposes using opinion deviation for trust, considering both rating and sentiments. • A trust framework assesses user reliability based on opinions and aggregates. • Findings emphasize the challenges and solutions in online review deception.
6	Ahmed M. Elmogy, Usman Tariq, Atef Ibrahim (2022)	Fake Reviews Detection using Supervised Machine Learning	International Journal of Advanced Computer Science and Applications	• Developed a fake reviews detection model on Yelp data (4,709 real, 1,144 fake). • Extracted reviewer behavior features, evaluated with five classifiers; KNN(K=7) excelled with 83.73% f1-score. • Incorporating behavioral features increased f1-score by 3.80%, emphasizing their role in enhancing detection accuracy.
7	Jianrong Yao, Yuan Zheng, Hui Jiang (2022)	An Ensemble Model for Fake Online Review Detection Based on Data Resampling, Feature Pruning, and Parameter Optimization	IEEE Access	• Ensemble model achieves higher F1-scores after optimization, especially with stacking strategy. • Stacking RF, Lightgbm, and Catboost results in the most effective ensemble model. • Optimization of base classifiers significantly enhances ensemble model performance. • The proposed model compares favorably with other fake review detection approaches.

Table 2. Comparison table of various machine learning algorithms

Authors	Dataset Used and Period	Techniques	Advantages	Limitation
Ahmed M. Elmogy , Usman Tariq , Atef Ibrahim	Yelp dataset	KNN, Naive Bayes (NB), SVM, Logistic Regression and Random forest.	genuine and fraudulent reviews with high accuracy.	time-consuming, labor-intensive, and expensive
Guohou Shan Lina Zhou	Yelp and OpSpam datasets.	decision tree, linear regression, and support vector regression algorithms	enhances the accuracy and effectiveness of identifying fraudulent reviews.	lack of a systematic and empirical investigation
HINA TUFAIL , M. USMAN ASHRAF	Yelp dataset	Vector Machine, K-Nearest Neighbor, and logistic regression (SKL)	Its scalability and efficiency in processing large volumes of data.	decreased detection accuracy over time.
HAO XUE, QIAOZHI WANG, BO LUO	Yelp dataset	SVM classifier	fine-grained detection capability enhances the accuracy and reliability of spam detection	unfairly penalized as potential spammers due to their deviation from the aggregated opinions.
Ahmed M. Elmogy1 , Usman Tariq	Yelp dataset	Random Forest, Logistic regression	the enhanced accuracy and effectiveness in detecting deceptive content.	training and testing complex machine learning models

IV. CONCLUSIONS

In conclusion, our research endeavors to combat the pervasive issue of fake online reviews by proposing a robust ensemble model that harnesses the capabilities of advanced deep-learning algorithms. With the prevalence of fake reviews casting doubt on the credibility of online platforms, it has become imperative to distinguish between genuine and deceptive content. Our ensemble model, which incorporates BERT, and LSTMs, offers a comprehensive approach to identifying and filtering out fake reviews. Using BERT (Bidirectional Encoder Representations from Transformers) for feature extraction and BiLSTM (Bidirectional Long Short-Term Memory) for model creation provides a powerful approach for fake review detection. BERT uses contextual embeddings which process input sequences bidirectionally. This bidirectional processing allows BERT to capture the contextual relationships between words in both directions, including information from both preceding and following words when generating the representation for each word. By processing the input sequence in both directions, BiLSTM can capture information from both past and future elements in the sequence, allowing it to model complex sequential dependencies more effectively The combination of BERT for contextualized embeddings and BiLSTM for sequential analysis allows the model to learn intricate patterns in language, enhancing its ability to accurately detect fake reviews that may exhibit complex linguistic manipulations. After the Bidirectional Long Short-Term Memory (BiLSTM) layer in a fake review detection project, the output typically goes through a prediction step where the model makes predictions regarding the authenticity or sentiment of the reviews. Moreover, our strategy involves training and validating the model using diverse datasets sourced from reputable platforms, each representing unique linguistic styles and expressions. Through the amalgamation of powerful algorithms and diverse datasets, we strive to significantly improve the credibility and reliability of our fake review detection system. Ultimately, our overarching goal is to restore consumer trust in online platforms and empower individuals to make more informed decisions in the vast landscape of online consumerism.

V. ACKNOWLEDGMENT

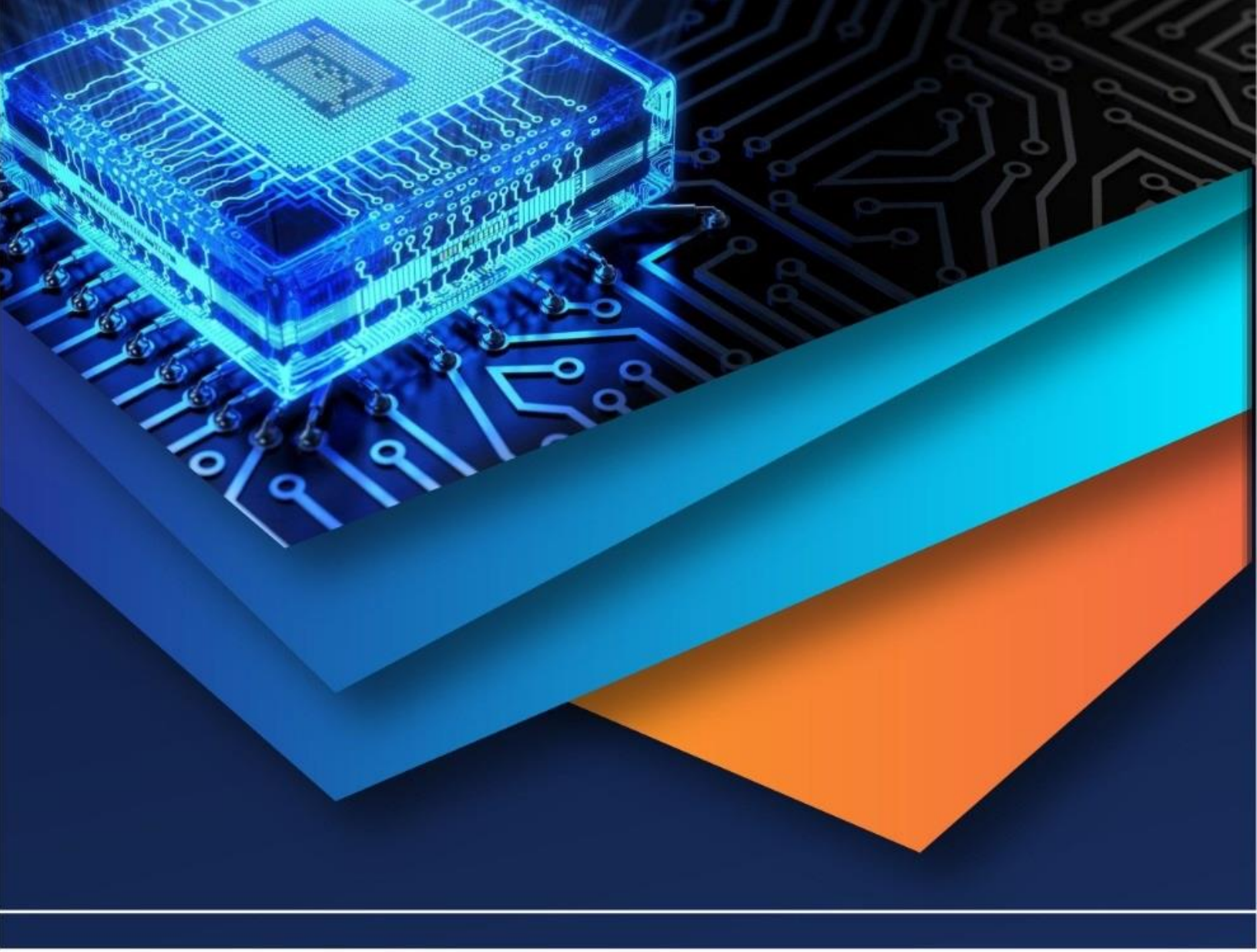
We would like to express our sincere gratitude to all those who contributed to the successful completion of this project. Our heartfelt thanks go out to our guide, Prof Dr. P. Boobalan, Information Technology Department, PTU, for their invaluable guidance and support throughout the entire research process. Their expertise and constructive criticism were instrumental in shaping this project and helping us to stay on track.

We would also like to extend our appreciation to the Puducherry Technological University (PTU) for providing us with the necessary resources to carry out this research. Their support was essential in enabling us to complete this project.
Finally, we would like to thank all the participants who volunteered their time and provided us with the data necessary to conduct this research. Without their support, this study was not possible. Once again, we express our gratitude to all those who have contributed to this project and hope that our findings will be of use to the wider research community.

REFERENCES

[1] Abhijeet A Rathore, Gayatri L Bhadane, Ankita D Jadhav, Kishor H Dhale, Jayashree D Muley(2023)“Fake Reviews Detection Using NLP Model and Neural Network Model” in International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 12 Issue 05, May-2023
[2] Jin, Kui Cheng and Xinjie Wang, Lecai Cai.(2023)"A Review of Text Sentiment Analysis Methods and Applications” In Frontiers in Business, Economics and Management ISSN: 2766-824X | Vol. 10, No. 1, 2023.
[3] Guohou Shan, Lina Zhou, Dongsong Zhang(2021) “Examining Review Inconsistency for Fake Review Detection”; in Elsevier on computers and security, vor :https://www.sciencedirect.com/science/article/pii/S0167923621000233.
[4] Hina Tufail, M. Usman Ashraf, Khalid Alsubhi, Hani Moaiteq AljahdalI (2022); “The Effect of Fake Reviews on e-Commerce During and After Covid-19 Pandemic: SKL-Based Fake Reviews Detection” in IEEE Access, vol. 10, pp. 25555 – 25564, doi:10.1109/ACCESS.2022.3152806.
[5] Hao: Xue, Qiaozhi Wang, Bo Luo, Hyunjin Seo, Fengjun Li (2019); “Content-Aware Trust Propagation Toward Online Review Spam Detection” in ACM Journal of Data and Information Quality, J. Data and Information Quality 11, 3, Article 11, 31 pages, pp. 1936-1955, https://doi.org/10.1145/3305258.
[6] Ahmed M. Elmogy, Usman Tariq, Atef Ibrahim (2022); "Fake Reviews Detection using Supervised Machine Learning" in (IJACSA) International Journal of Advanced Computer Science and Applications, vol. 12, No. 1, 2021, pp. 601 – 60
[7] Jianrong Yao, Yuan Zheng, Hui Jiang (2022); "An Ensemble Model for Fake Online Review Detection Based on Data Resampling, Feature Pruning, and Parameter Optimization” In IEES Access, vol. 9, pp. 16914 – 16927, doi: 10.1109/ACCESS.2021.30511
[8] Abrar Qadir Mir,Furqan Yaqub Khan, Mohammad Ahsan Chishti(2023)" Online fake review detection using supervised machine learning and bert model” In Frontiers in Business, Economics and Management.
[9] 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.
[10] E. I. Elmurngi and A.Gherbi, “Unfair Reviews Detection on Amazon Reviews using Sentiment Analysis with Supervised Learning Techniques,” Journal of Computer Science, vol. 14, no. 5, pp. 714–726, June 2018.
[11] Monica, C., Nagarathna, N. Detection of Fake Tweets Using Sentiment Analysis. SN COMPUT. SCI. 1, 89 (2020).
[12] M. Ott, Y. Choi, C. Cardie, and J.T. Hancock. 2011. 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.
[13] M. Ott, C. Cardie, and J.T. Hancock. 2013. Negative Deceptive Opinion Spam. In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.
[14] Mohawesh, Rami & Xu, Shuxiang & Tran, Son & Ollington, Robert & Springer, Matthew & Jararweh, Yaser & Maqsood, Sumbal. (2021). Fake Reviews Detection: A Survey. IEEE Access. 10.1109/ACCESS.2021.3075573.
[15] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. Advances in neural information processing systems, 30.
[16] M. A. Friedl and C. E. Brodley, “Decision tree classification of land cover from remotely sensed data,” Remote sensing of environment, vol. 61, no. 3, pp. 399–409, 1997.
[17] "Natural Language Processing." Natural Language Processing RSS. N.p., n.d. Web. 25 Mar. 2017

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



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)