Detecting Fraudulent Reviews in E-commerce Platforms

Bedadala Akhileswar Reddy, Gopireddy Someswara Reddy, Kollepara Lokesh, Meena Belwal

*Department of Computer Science and Engineering,  
Amrita School of Computing, Bengaluru,  
Amrita Vishwa Vidyapeetham, India.*

[*akhileswarreddy5016@gmail.com*](mailto:akhileswarreddy5016@gmail.com) *,* [*somu.gopireddy@gmail.com*](mailto:somu.gopireddy@gmail.com) *,* [*kolleparalokesh@gmail.com*](mailto:kolleparalokesh@gmail.com) *,* [*b\_meena@blr.amrita.edu*](mailto:b_meena@blr.amrita.edu)

***Abstract***— **Fake review detection in e-commerce reviews is a vital issue in preserving the trustworthiness of online platforms. The domain includes a wide range of products and services, each with its own difficulties in finding fraudulent or misleading reviews. The challenges are the complexity of the fake review techniques, the vast number of reviews that are created every day, and the need for the detection of the fake reviews to stop the spread of the misinformation. This work addresses these issues based on a multi-faceted approach that includes the use of advanced natural language processing techniques, machine learning algorithms, ensembled models trained on labeled datasets, and the analysis of user behavior to detect anomalies. Through the combination of these techniques, e-commerce platforms can successfully fight fake reviews, thus, protecting the consumers trust and keeping the online reviews credibility.**

Keywords- Fake reviews detection, supervised machine learning, feature engineering, Fake review observation.

# Introduction

During the age of ecommerce, internet-based product assessments are crucial in customer choices providing important basic information on different products/service quality, use value, and satisfaction rate [1]. Such evaluations have real customer experience descriptions that help you evaluate products or commercial services well before making purchasing decisions. Nonetheless, within this expanse relating to user-generated content, there has been an increase in counterfeit reviews that negatively affects trust as they occur everywhere making other platforms unreliable. Thus, the issue of detecting fake reviews presents itself as an important research field protecting the online business ecosystems’ credibility and enhancing consumer confidence.”

Main issue is the precise recognition and control of counterfeit reviews on internet-based product evaluating systems [2]. These wrong review practices involve a range of actions aimed at attracting consumers into making some judgements while buying items, right from enhancing ratings for no reason to making-up own testimonials as required for that purpose. There is a need to come up with sound methods that can make a distinction between authentic feedbacks and fraudulent reviews to lessen the negative effects of spurious comments on customer loyalty and commerce honesty [3].

Addressing this challenge requires robust methodologies that are strong enough to differentiate the real feedback from the fake reviews, thus, the fraudulent comments will not affect the trust of the consumers and the business [4]. The latest studies have been focused on the various detection methods, which are based on the linguistic, behavioral and contextual cues, to determine the fraudulent content. Researchers have been looking into sentiment analysis, linguistic style analysis, and semantics analysis to find the subtle markers of deceptive practices.

Recent research in this field has tried to find different ways of detecting false reviews using a mix of linguistic, behavioral and contextual markers [5]. Works have started to look at how well the techniques of sentiment analysis, linguistic style analysis or semantics can detect deceptive contents. In addition to this, machine learning models have seen tremendous growth such as deep learning models like Recurrent Neural Networks and Convolutional Neural Networks which can capture intricate patterns and relationships among words [ 6].

Furthermore, research has focused on the introduction of ensemble procedures and also the combination of various detection methods in order to promote the general scores of both accuracy and power. However, the detection of fraudulent behaviors continues to develop, whereas also knowledge suggests what is happening is similar to two magic dolls’ tricks, leaving it immensely difficult to seek out a reliable means of fake review detection as scammers continuously advance as they view the safest method to work without exactly getting caught.

The Key contributions of the proposed work includes:

1. The work highlights the utility of ensembled models like Voting Classifier on Naïve Bayes, Adaboost, SVM and comparing them with traditional Machine Learning Algorithms and Deep Learning Algorithms.
2. The work also uses a dataset consisting of Reviews from different kinds of products for which the models are performing well with around 90% accuracy, So, we can highlight that our proposed model performs well on different kinds of product reviews.

The article is organized into several subsequent parts as is explained below. There is a section 2 that attempts descriptions of the research. In section 3, the Methodology is described. Section 4 gives the analysis and results. This article finishes with section 5 conclusion and future works.

# Related works

The investigation by Ahmed M et.al [8] is a study of the role in decision making process of online reviews and how machine learning can detect fake review on the basis of their reviews. Moreover, it maintains that behavioral features, including caps count, punct count, and emojis present in the reviews, could enhance the accuracy of fake review detection. The KNN classifier reaches the highest f1 score and surpasses all the others in the experiment conducted in the Yelp dataset. The document discussed has focused on the influence of reviews on web-based data and how the non-textual observations like behavioral features affected the detection performance to create a fake review.

Another innovative approach by Z Khanam et.al [9] is about how a supervised machine learning model can be used in detecting fake news. It involves preprocessing, eliminating the noise using NLP, extracting features and applying different classification methods. Data is divided into training and testing sets and then model’s precision is measured. The main emphasis is put on the use of decision trees as a means for classifying things whereby their importance and relationships in data are determined. The document also mentions that another stuff that is worth exploring is part of speech taggers for different languages and that data mining techniques can be classified into supervised and unsupervised methods.

J. C. Rodrigues et al. [10] conducted another study concerning various machine learning techniques including SVMs, CNNs, LSTMs, GRNNs, BRNNs, and Bi-LSTMs in an attempt to detect fake reviews. These methods approach issues related to detecting fake reviews using techniques such as sentiment analysis, opinion spam detection and new algorithms or features. Additionally, researches on PU learning, calculating sentiment scores and behaviour indicators improves recognizing misleading opinions and reviews while preventing spam reviews.

The investigation by Muhammad Ahsan Saeed et al. [11] emphasizes various ways in which machine learning is done, using vulnerable systems like SVM, CNN, LSTM, GRNN, and BRNN on the analysis of fake reviews sent in by individuals. The aim of the methods is twofold: firstly, dealing with challenges in detecting fake appraisals such as PU learning, opinion spam detection, and sentiment analysis. They also assist in finding out if someone is truthful and preventing others from leaving fake reviews using techniques such as establishing behavioral signs which help reveal when reviews tell lies or applying algorithms targeted at spotting lies. Moreover, different tactics concentrate on ensuring accurate identification of fake views.

Nikhil Chandra Sai Ram et al. [12] proposed the algorithm of unprecedented sophistication to curb an information flow misled by linguistic and behavioural characteristics so that endusers are not subject to fake reviews online. It emphasizes the necessity for the feature extraction which is having an implemented python-based application able to identify anaemia with the Django infrastructure. Along with that, it highlights the use of the Naive Bayes classifier and K S Nearest Neighbours algorithm in classifying these applications. These classifiers are also compared as to their benefits as well as the field in which they can be applied.

Another approach by A. Zaheer et al. [13] proposed a methodology that discusses about the fake advertising has become a significant problem across diverse media platforms, notably smartwatches. These dishonest practices have far-reaching social and national implications especially on social media. This study aimed at using supervised machine learning models to differentiate between true and false ads which were not genuine ones. The obtained outcomes revealed that this technique could be applied effectively to differentiate valid advertisements from false ones, suggesting that other advertisement classification methods could use smartwatch ads having advanced characteristics as benchmarks in researches that are yet to come.

Poonam Tanwar et al. [14] proposed a methodology that focuses on the problem of the significance of consumer reviews in determining purchasing patterns, opinion spam problem, target of detection, classification of opinion spam and the influence of social networks on the spread of the spam data. Dr. Poonam Tanwar, and Ms. Priyanka Rai are well-experienced scholars in the field of opinion mining based on machine learning, NLP and classifiers who made their input via addressing the spam dynamic process of diffusion and the features that feed such a process, especially through social network Twitter.

Another study by Amol Gadewar et al. [15] provides a comprehensive literature survey, datasets, and techniques for further research in the detection of counterfeit evaluations at the point in time. Moreover, it reviews certain AI algorithms which include but not limited to Naive Bayes, Support Vector Machine, Maximum Entropy, K Nearest-Neighbor among other things as being helpful for this purpose exactly like it does with the issue of the impact that they have when users make decisions about purchasing products/services after reading reviews written by different individuals on various platforms and how these same businesses get affected.

The investigation by T. Nahian et al. [16] explores on detecting review spam is important due to the increase in fake reviews, which affect consumer decisions. With high precision and recall, the proposed ensemble model that combines active and supervised learning shows promising results. It outperforms recent methodologies by detecting review spam with 95.0% accuracy. The ability to create a hybrid dataset and use both fabricated and real-life reviews makes classifiers more accurate, versatile, and robust. This model also solves the problem of limited real-life labeled datasets by offering potential for trustworthy online reviews.

R. Agarwal et al. [17] proposed methodology evaluates the significance of detecting false reviews in e-commerce. It surveys various methods and algorithms employed for counterfeit assessment including supervised learning and deep learning. The analysis concentrates on the impacts of reviews on buying decisions, difficulties faced in identifying false ones and attributes related to deceitful reviews such as reciprocated ratings or reciprocal remarks. Moreover, it reveals that ensembled machine learning techniques as well as Sparse Additive Generative Models have been used which results into a detection rate ranging from 71.7% - 87.68% across different datasets.

Another approach by P. M. Kumar et al. [18] examines methods for identifying fake reviews through the use of machine learning algorithms like Naive Bayes, Decision Tree, and SVM. It covers their impact on Amazon as well as the movie industry and stresses how vital real customer feedback is. Many different techniques were tested on large datasets but it was found that Support Vector Machines had the highest accuracy rate when detecting fraudulent feedback. Researchers hope to be able to distinguish between what's genuine and what isn't in order to make sure products can be trusted- they suggest including aspects related to individual items for improved results.

Nicola Capuano et al. [19] proposed a methodology that discusses about automatic detection of fake news, including the vulnerability of NLP-based detection to adversarial attacks. It also delves into embedding and word tagging features, with a focus on algorithms and features with average accuracy higher than 80%. Additionally, the limitations of the systematic review are outlined, particularly regarding the quality assessment of articles and potential partial selection of works.

Another approach by C. Rekha et al. [20] discusses the need for identifying deceptive online critiques, underlining the grey area between authentic material and commercials that has been brought about by social media. It stresses the FTC's endeavor to fine dishonest schemes and the importance of spotting counterfeit appraisals in order to safeguard buyers as well as preserve business integrity. Furthermore, it introduces the application of Naive Bayes and K Nearest Neighbors algorithms in such areas as text classification or recommendation systems where their effectiveness can be demonstrated through different uses cases.

D. Radovanović et al. [21] proposed a methodology which is about the creation of a system to sort out unreal appraisals in order to put forward certified views on a product. It covers the down sides of Samsung Galaxy S series phone, and also elaborates on its advantageous side including quick Wifi6 connection, small design,[22] durable battery plus fantastic photo shooting that can be compared with Olympus mirrorless camera. Classifying counterfeit reviews by applying supervised machine learning and giving an outline of reports sentimentally are some of the methods used in this system.

The presented literature reviews bring to the fore a concern that is growing as to proliferation of fabricated reviews all over various online platforms. Researchers have applied numerous machine learning approaches ranging from supervised to unsupervised machine learning algorithms such as SVM, CNN, LSTM, decision trees, and ensemble methods for effective detection of deceitful information. Detection precision has been improved using key features including behavioral indicators, linguistic characteristics and sentiment analysis while a number of studies have shown how fake reviews affect customers’ choices necessitating more reliable detection techniques for platform credibility. Moreover, the detection of fake news and advertisements through machine learning algorithms is highlighted, showing how important it is to fight misinformation online.

# Methodology

The steps involved in the proposed methodology are represented in Figure 1. and detailed explanation is given below.

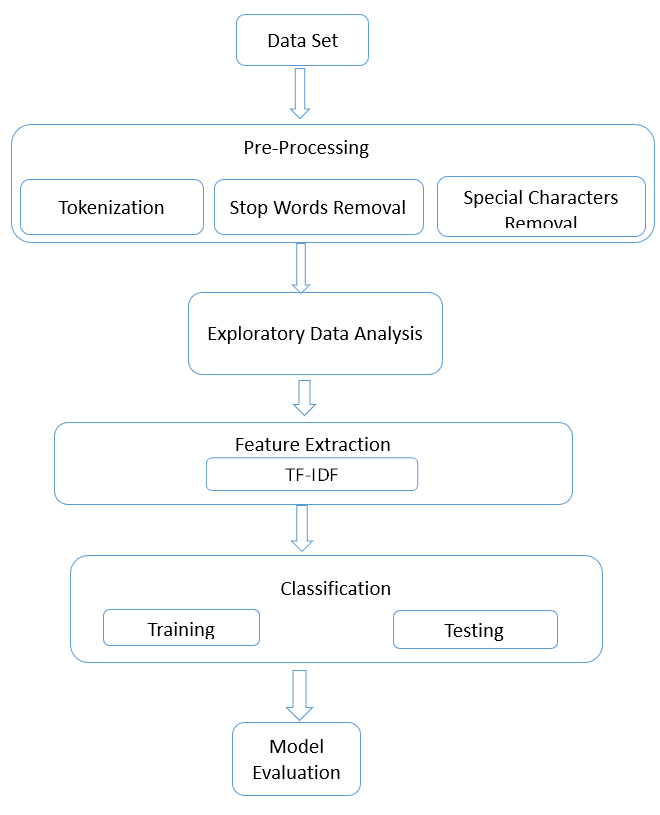


Fig 1: Flow diagram representing the proposed methodology

1. Dataset Description:

The dataset has both genuine and fake reviews mixed up in the various categories in it. Here is a description of the dataset:

Features:

* Category: This is the product’s group.
* Rating: This is the result given to the product.
* Label: This shows if the review is genuine (OR) or fake (CG).
* Text: The actual words of the review.

Categories Distribution:

* The set contains numerous categories like electronics, books, home and kitchen wares as well as other products from various industries.
* Analysing reviews’ allocation in different categories is necessary in order to identify features that are peculiar to each one.
* Rating: Users’ rating generally varies from one (least satisfaction) up to five (top satisfaction).
* Analysing the rating distribution in both fake and real reviews helps in understanding how fake reviewers behave.
* Both Original and Computer generated are equal in the dataset which is depicted in figure 2 which is good for training and prediction so that there won’t be any bias in the prediction.

Text Content:

* The reviews’ texts contain crucial details essential for analysis and model development.
* Text data preprocessing is very important in getting rid of useless information and pulling out important features for constructing a model.

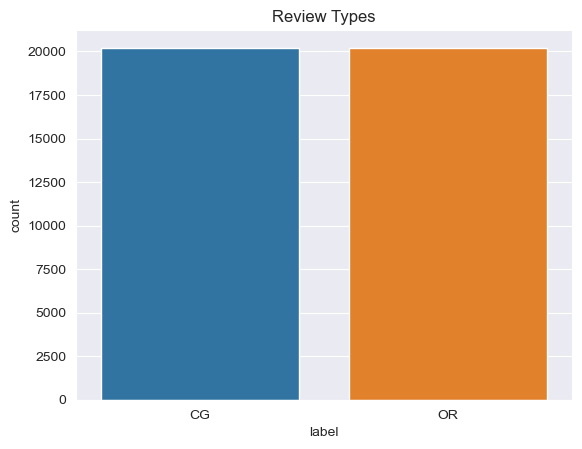


Fig 2: Bar Graph representing the count of target labels

1. Data Pre-processing:

Special Characters Removal:

* Exclude any special characters, punctuation marks, and other extra characters from the text received. Change text to lowercase so that text processing can be uniform

Tokenization:

* Divide the text data first into individual words or phrases so as to set it up for other operations. This involves splitting the text into tokens using white space and other indicators.

Stop Word Removal:

* Stop Words in English like am, is, are, which etc. which doesn’t contribute much to the differentiation of labels in the training process.

1. Exploratory Data Analysis (EDA):

* Distribution of categories: Studying how reviews are spread among various categories can help in understanding what forms a dataset.
* Distribution of ratings: Examine how r rates are distributed in fake and real comments.
* Text-length analysis: Research the number of words or characters contained in the reviews of both fake and genuine contents.
* Word-frequency analysis: Detect patterns by identifying common words or phrases in genuine and fake reviews.

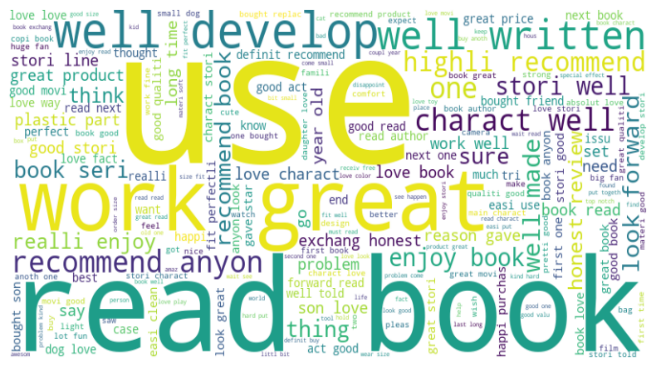


Fig 3: Word Cloud Representation of Computer Generated Reviews



Fig 4: Word Cloud Representation of Original Reviews

d. Feature Engineering:

* TF-IDF Vectorization: This would help us understand the level of significance that a given word has when we compare it to our whole collection of documents just by changing this text content into numbers based on Term Frequency-Inverse Document Frequency for each of them.

1. Model Selection and Training:

In the proposed work the following models evaluated for the following Machine Learning and Deep Learning models:

* Logistic Regression: A simple algorithm for classification that predicts the probability of having one out of two outcomes. Logistic regression is advisable as the initial model for recognising false alterations.
* Random Forest: Random forests, on the other hand, overcome this by using multiple decision trees, allowing them to model complex, non-linear relationships between the features. They are often used for classification tasks since their classification is not affected by overfitting.
* Gradient Boosting Machines (GBM): Most of the time, it provides better results than other types of machine learning algorithms because it minimizes the loss measure by gradually adding weak learners, giving more emphasis to the errors of the prior models.
* Support Vector Machines (SVM):

SVM tries to look for that most excellent hyperplane for increasing a breach in classes as well as minimizing misclassification.

They are proficient in carrying high-dimensional data besides being effective when binary classifications are concerned.

* Deep Learning Models:

When it comes to sequential information (RNNs) or text and/or image data sources (CNNs), fake reviews detection can be done using Recurrent Neural Networks (RNNs) or Convolutional Neural Networks (CNNs).

These models may discern intricate patterns and correlations within the amassed data; however, they necessitate massive computational capability as well as extensive amounts of reliable data for successful training.

f. Model Evaluation:

* The models should be evaluated by measuring metrics like accuracy, precision, recall and F1 score, performance of TF-IDF vectors against Word2Vec embeddings as a representation of model’s performance. The confusion matrix should be analysed to comprehend how good the model is at classifying fake versus real reviews based on two sets of data: TF-IDF vectors and Word2Vec embeddings.

# Results and Analysis

The performance of the classifiers such as Accuracy, Precision, Recall, F1-score are represented in the Table 1

`Table 1: Performance Metrics of Machine Learning Classifiers and Deep Learning models

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Algorithms** | **Accuracy** | **Precision** | **Recall** | **F1** |
| Naïve Bayes | 84 | 85 | 84 | 84 |
| **SVM** | **87** | **87** | **87** | **87** |
| Bagging naïve Bayes | 80 | 81 | 80 | 80 |
| AdaBoost | 77 | 77 | 77 | 77 |
| Random Forest | 85 | 85 | 85 | 84 |
| **Voting** | **86** | **86** | **86** | **86** |
| BILSTM | 88 | 88 | 88 | 88 |
| **CNN** | **89** | **89** | **89** | **89** |

As shown in table 1 CNN outperformed among the Deep Learning Models with 89% accuracy. Under machine learning SVM performed the best with an accuracy of 87%. *The Voting classifier, as proposed by this research, a combination of Naïve Bayes, SVM, Adaboost performed well with an accuracy of 86%, which is almost at par with SVM and CNN.*

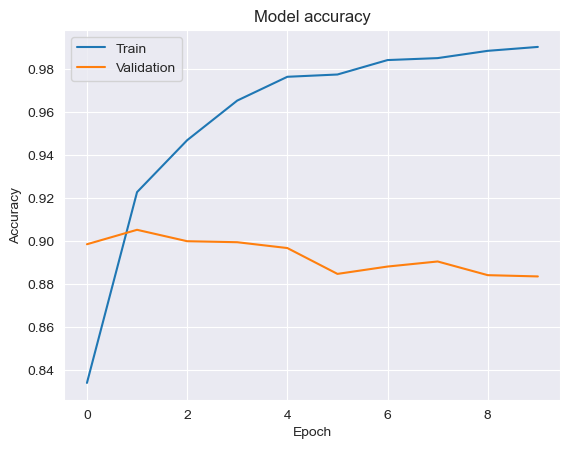


Fig 5: BI-LSTM Model Accuracy Graph

The Validation accuracy of BILSTM is presented in figure 5 which ranges from 90% to 88%. Similarly, the validation accuracy of CNN model is depicted in figure 6 ranging from 90% to 89%.

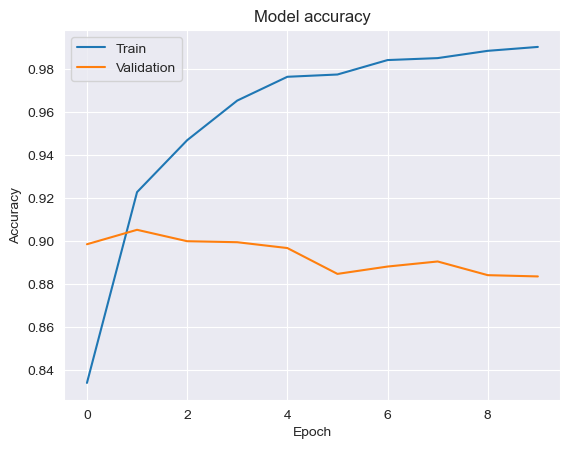


Fig 6: CNN Model Accuracy Graph

# Conclusions and Future Scope

Finally, our research demonstrated the applicability of a broad range of models in solving a wide range of Fake Review detection tasks. Both CNN and BiLSTM networks managed to obtain amazing results but using classic machine learning methods turned out to be equally good. This pretty much stresses that a combination of different research methods is a perfect approach to get the Fake Review Detection on the reviews of Ecommerce platforms.

Moving on, further research should explore hybrid models including a deep learning and classical machine learning algorithm. Therefore, the hybrid models have these potentials to better lead to the increase in the performance of Fake Review Detection by combining the qualities of each approach. Along with it, there could be a need to explore more advanced pre-processing approaches and utilize data of a great size to create the state-of-the-art Fake Review Detection algorithms which in turn would help us to know the good products in online platforms such as Amazon, Flipkart.

References

[1] S. R. R, M. R. L, D. S, A. D. P K and A. M. S, "Detection and Summarization of Honest Reviews Using Text Mining," 2022 8th International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 2022, pp. 01-05, doi: 10.1109/ICSSS54381.2022.9782167.

[2] S. R. R, M. R. L, D. S, A. D. P K and A. M. S, "Detection and Summarization of Honest Reviews Using Text Mining," 2022 8th International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 2022, pp. 01-05, doi: 10.1109/ICSSS54381.2022.9782167.

[3] A. Tripathi, T. Singh and R. R. Nair, "Optimal Pneumonia detection using Convolutional Neural Networks from X-ray Images," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2021, pp. 1-6, doi: 10.1109/ICCCNT51525.2021.9580140..

[4] K. S. Naveenkumar, R. Vijayakumar and K. P. Soman, "Amrita-CEN-SentiDB: Twitter Dataset for Sentimental Analysis and Application of Classical Machine Learning and Deep Learning," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, 2019, pp. 1522-1527, doi: 10.1109/ICCS45141.2019.9065337.

[5] N. B. Nair, T. Singh, A. Thakur and P. Duraisamy, "Deployment of Breast Cancer Hybrid Net using Deep Learning," 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2022, pp. 1-6, doi: 10.1109/ICCCNT54827.2022.9984513.

[6] N. K. E., K. M., P. P., A. R. and V. S., "Tomato Leaf Disease Detection using Convolutional Neural Network with Data Augmentation," 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2020, pp. 1125-1132, doi: 10.1109/ICCES48766.2020.9138030.

[7] Belwal, Meena, and T. K. Ramesh. "N-pir: a neighborhood-based pareto iterative refinement approach for high-level synthesis." Arabian Journal for Science and Engineering 48, no. 2 (2023): 2155-2171.

[8] Fake Reviews Detection using Supervised Machine Learning, Ahmed M. Elmogy and Usman Tariq and Ammar Mohammed and Atef Ibrahim, International Journal of Advanced Computer Science and Applications,2021,https://api.semanticscholar.org/CorpusID:231877755

[9] Z Khanam et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1099 012040,10.1088/1757-899X/1099/1/012040

[10] J. C. Rodrigues, J. T. Rodrigues, V. L. K. Gonsalves, A. U. Naik, P. Shetgaonkar and S. Aswale, "Machine & Deep Learning Techniques for Detection of Fake Reviews: A Survey," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 2020, pp. 1-8, doi: 10.1109/ic-ETITE47903.2020.063.

[11] Muhammad Ahsan Saeed1 , Farrukh Yousaf , Osama Bin Khalid , Mushhad Gilani , Qamar Nawaz , Isma Hamid3, International Journal of Advanced Trends in Computer Science and Engineering, f Advanced Trends in Computer Science and Engi

[12] Nikhil Chandra Sai Ram , Gowtham Vakati , Jagadesh Varma Nadimpalli , Yash Sah , Sai Karthik Datla, International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 ,Volume 10 Issue V May 2022

[13] A. Zaheer, S. Tahir, M. Humayun, M. F. Almufareh and N. Z. Jhanjhi, "A novel Machine learning technique for fake smart watches advertisement detection," 2022 14th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS), Karachi, Pakistan, 2022, pp. 1-5, doi: 10.1109/MACS56771.2022.10023151.

[14] Poonam Tanwar , Priyanka Rai, IAES International Journal of Artificial Intelligence (IJ-AI),Vol. 9, No. 4, December 2020, pp. 726~733,ISSN: 2252-8938, DOI: 10.11591/ijai.v9.i4.pp726-733

[15] Prof. Amol Gadewar, Pratima Jadhav, Pratiksha Kale, Dhanashree Kature, Kshitija Patil Department of Information Technology Engineering, Pune District Education Association's College of Engineering, Pune, India, International Journal of Scientific Research in Science, Engineering and Technology ,Online ISSN : 2394-4099

[16] M. N. I. Ahsan, T. Nahian, A. A. Kafi, M. I. Hossain and F. M. Shah, "An ensemble approach to detect review spam using hybrid machine learning technique," 2016 19th International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2016, pp. 388-394, doi: 10.1109/ICCITECHN.2016.7860229.

[17] R. Agarwal and D. K. Sharma, "Detecting Fake Reviews using Machine learning techniques: a survey," 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2022, pp. 1750-1756, doi: 10.1109/ICACITE53722.2022.9823633.

[18] P. M. Kumar, S. S. Harrsha, K. Abhiram, M. Kavitha and M. Kalyani, "Role of Machine Learning in Fake Review Detection," 2022 6th International Conference on Electronics, Communication and Aerospace Technology, Coimbatore, India, 2022, pp. 1026-1030, doi: 10.1109/ICECA55336.2022.10009174.

[19] Nicola Capuano, Giuseppe Fenza, Vincenzo Loia, Francesco David Nota, Content-Based Fake News Detection With Machine and Deep Learning: a Systematic Review, Neuro computing, Volume 530,2023,Pages 91-103,ISSN0925-2312, https://doi.org/10.1016/j.neucom.2023.02.005.

[20] C. Rekha, Mr. G. Lakshmikanth M. Tech Scholar , Associate Professor Department of Computer Science, Sri Ramachandra Engineering and Technology, Chennai, Tamil Nadu, India, International Journal of Scientific Research in Science and Technology ,Print ISSN: 2395-6011 | Online ISSN: 2395-602X

[21] D. Radovanović and B. Krstajić, "Review spam detection using machine learning," 2018 23rd International Scientific-Professional Conference on Information Technology (IT), Zabljak, Montenegro, 2018, pp. 1-4, doi: 10.1109/SPIT.2018.8350457

[22] Reddy, Bandi Rupendra, Daka Chandra Rup, Mathi Rohith, and Meena Belwal. "Indian sign language generation from live audio or text for tamil." In 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), vol. 1, pp. 1507-1513. IEEE, 2023.