International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 07 | July 2024 www.irjet.net p-ISSN: 2395-0072

Detecting Deception: A Comprehensive Review of Machine Learning Approaches to Fake News Detection

Osama Zaheer[1], Yash Bhagia [2], Khan Shah Ahmad Shakir Abu Asim[3], Mohd Azhan Umar Kamil[4]

[1,2,3,4]Zakir Hussain College of Engineering and Technology, Aligarh Muslim University, Aligarh (202001) -U.P., INDIA

Abstract - *It was challenging to distinguish between false and true information due to social media networks' easy access and rapid proliferation of content. The rapid spread of fake news has been facilitated by the simplicity of information dissemination on social media. Fake news has consequently grown to be a major issue with important repercussions for national security, public confidence, and media credibility. Thus, creating trustworthy and precise methods to identify fake information and stop its spread has become a big scientific challenge. To identify fake news, various machine-learning techniques have been proposed. However, the majority of those concentrated on a certain category of news (like political news), which raises the issue of dataset bias in the models employed. We want to provide a thorough analysis of the many methods for identifying false news using machine learning, which is essential in the classification of information. We also compared the performance of a variety of advanced pre-trained language models for fake news identification in addition to standard and deep learning models.*

Key Words: Natural Language Processing, Machine Learning, Deep Learning, Neural Networks, Convolution Neural Network

1.INTRODUCTION

The widespread use of social media platforms has greatly impacted the dissemination of information, including news. The ability to access the most recent news has been made simpler for consumers by different online news platforms including social media, blogs, and other digital media formats. Approximately 68% of Americans as of August 2018 got their news from social media, up from 62% in 2016 and 49% in 2012 [1]. However, these platforms have also been used to spread fake news for monetary gain or to manipulate mindsets. Fake news has become a major issue in today's society, significantly affecting how people think and make decisions, leading to erroneous beliefs, mistrust, and confusion. The main challenge in the existing content-based analysis is the difficulty in the detection of fake news, as it may encompass political, cinematic, and other issues. Therefore, common sense needs to be there for the detection of news datasets. The propagation of false information is a diverse topic on media platforms like Facebook, Instagram, and others. Fake news poses one of the greatest threats to freedom of speech, the media, and democracy. It can even erode the public's trust in governments. One notable instance of the effects of propagating fake news is the 2016 US Presidential Election. Gaining public's trust and disseminating false information depend heavily on social and psychological aspects. Humans, for instance, have been found to become gullible and illogical in their ability to distinguish between truth and untruth when they are repeatedly exposed to erroneous information [1]. In this field of study, the application of machine learning algorithms, such as deep learning algorithms, graph-based techniques, multi-modal approaches, and explainable AI methods, have been concentrated on the detection of fake news. These developments in the sector greatly encourage future efforts to effectively tackle fake news. In general, fake news is detected by confirming videos and images as well as the identity of the sender by checking the account's age on social media. Whether an account has blue account verification, validating the friends and followers as well as content that is frequently shared, mainly validates the URL. In order to solve this issue, machine learning algorithms have been developed for the automatic detection of fake news. These algorithms examine the trends and characteristics of both accurate and false news. On the basis of this information, they predict the validity of fresh publications. These algorithms can be trained using a large dataset to learn the traits and patterns of both real and fake news, recognizing the fake news with good precision and recall [2].

The complexity of the algorithms used for fake news detection varies; some rely on simple concepts, while others use complex deep learning models. One of the most popular techniques for examining news articles' text is natural language processing (NLP). The text's structure and substance, as well as the words used, their usage patterns, and their placement, can all be examined by NLP algorithms. Machine learning algorithms can also look at news sources, how they spread on social media, and other contextual information. Support vector machine (SVM), linear support vector machine (LSVM), K-nearest neighbor (KNN), logistic regression (LR), decision tree (DT), and Naive Bayse models combining SVM and LR are examples of machine learning (ML) models [3].

Another popular approach is to use network analysis to examine how news is shared on social media. This can

provide helpful information on how the news is disseminated and its origins, which can be used to identify fake news. Fact-checking websites and databases can also be included into machine learning models to further boost their accuracy.

2. Methodology

2.1. Research Goal

The purpose of this research is to review existing studies and their conclusions, as well as to summarize research efforts in the identification of fake news using machine learning.
RQ1: What exactly is social media fake news, and what kind of fake news are there?
RQ2: How reliable are deep learning versus conventional machine learning models at spotting fake news?
RQ3: What methods are there to detect fake news on social media?
RQ4: Which model works best with a limited amount of training data?

2.2. Search Criteria

We conducted our research across a seven-year span, however, we made assured that around two-thirds of the research publications we analyzed were published in or after 2020. Furthermore, because we focused on reviewing the existing state of the art as well as the difficulties and future directions, we established a list of keywords to search for the aforementioned scientific resources. The following terms are included in the set of keywords: fake news, detection techniques, social media, and literature review.

2.3. Inclusion and Exclusion Criteria

A basic list of articles was found. To choose the acceptable study papers from the initial list of studies we obtained, we followed a set of inclusion/exclusion criteria as shown in Figure 1. As a result, the inappropriate papers were excluded in compliance with the SLR requirements. The chosen paper had to be written in English, have undergone peer review, and concerns have been raised about the use of machine learning to detect fake news. On the other hand, the exclusion criteria were set based on studies that did not address machine learning-based fake news identification. Additional exclusion criteria include reusing previously published works, publishing unreviewed articles, and publishing works that are not in English.

3.Related Work

With its capacity to sway public opinion and decision-making, false news dominance in today's society has given rise to serious concerns. Researchers have been looking into various ways to identify fake news, including machine learning algorithms, to address this problem. A review of the literature provides a summary of the work that has been done so far on using machine learning algorithms to identify fake news. The use of natural language processing (NLP) methods for fake news identification has been the subject of numerous studies. NLP algorithms examine the content of news stories to find patterns and traits that can differentiate between genuine and false news. The use of network analysis to investigate the dissemination of news on social media is another strategy that has been examined. According to research, fake news frequently comes from unreliable sources and tends to spread more quickly and broadly than legitimate news. Researchers have created algorithms that can identify fake news based on its patterns of dissemination by studying the network structure of the distribution of news.

Asha J. et al. proposes a procedure for detecting fake news using unsupervised and deep learning techniques due to the difficulty of detecting fake news with limited publicly available resources. The proposed method has an accuracy rate of 58% with the unsupervised SVM algorithm and 96.4% with the Hybrid CNN-RNN deep learning algorithm [4].

Uma Sharma et al. discuss the growing problem of fake news and its impact on people's opinions and attitudes towards digital technology. To address this issue, the authors propose a system for fake news detection that uses NLP and machine learning techniques to determine whether news articles are true or false. The system's performance is evaluated using various performance measures, and the best model is determined to be Logistic Regression with an accuracy of 75% [5].

Iftikhar Ahmad et al. describe a study on the categorization of false news stories using machine learning algorithms and ensemble approaches. The study evaluates the accuracy, precision, recall, and F1-score of several algorithms on four datasets. The results show that the best-performing algorithm is the decision trees with an accuracy of 94%, while Wang-Bi-LSTM, with an accuracy of 62% [6,] is the worst-performing algorithm.

Berahmand K. et al. propose a link prediction technique based on neighbors and mutual influence nodes. Their study takes a similar approach to rate publishers' influence and credibility in disseminating fake news based on data from their social media followers [7]

According to Qawasmeh et al., identifying fake news is one of the tough issues when using standard content-based analysis approaches. Due to the exceptional capability of feature extraction, neural network models have recently outpaced conventional machine learning techniques in

terms of performance. The research on spotting fake news in the news and during time-sensitive situations is still lacking. We have therefore looked into the automatic detection of fake news on internet communication channels in this study. Moreover, we suggest adopting cutting-edge machine learning algorithms to automatically identify bogus news. The suggested bidirectional LSTM concatenated model that performs with 85.3% accuracy on the FNC-1 dataset [8].

Sequential and recurrent layers are both used in Sedik et al.'s deep learning technique. The CNN-LSTM model, which is a recurrent model, utilizes stacked CNN along with LSTM and Dense layers. Similarly, the GRU model, another type of recurrent model, employs a basic GRU along with a Dense layer. On the other hand, sequential models employ concatenated CNN, stacked CNN, or both. Based on experimental results using binary labeled Kaggle and Fake News Challenge datasets, the C-CNN and CNN-LSTM models demonstrate superior performance. Specifically, the C-CNN model achieves an accuracy of 99.90% on the Kaggle dataset, while the CNN-LSTM model attains an accuracy of 96% on the Fake News Challenge dataset.[9].

Kaliyar, R.K. conducted an analysis of benchmarked data that included several attributes such as ID for a unique value, title for a specific content heading, and author for the name of the creator. The primary attribute used to differentiate labels was the text, and this dataset was utilized to train and test another dataset. The process began with word embedding, followed by a 1D convolution of three convolutions. Convolutional neural networks aim to extract valuable information from the trained model by utilizing pooling and dense layers. To evaluate the performance of the model, accuracy, F1 score, recall, and precision are considered as parameter assessment metrics. The proposed approach was able to achieve an accuracy of 98.32% [10].

Himank Gupta et al. proposed a framework based on a different machine learning approach to overcome several issues such as low accuracy, time lag, and lengthy processing time. They collected data from 400,000 tweets from the HSpam14 dataset, out of which 250,000 were non-spam and 150,000 were spam tweets. The authors also extracted Top-30 terms from the Bag-of-terms model, which yielded the highest information gain, and identified several lightweight features. Their proposed solution outperformed the previous solution by approximately 18% and achieved an accuracy of 91.65% [11].

Using Twitter data, Monti, F. experimented with geometric deep learning to determine fake news. The results of the studies suggest that social network structure and propagation play a key role in achieving the maximum level of fake news detection accuracy. The second step involves detecting fake news as soon as it starts to spread. The third step involves testing the model once more to keep the level of accuracy, so it is founded on propagation. Rather than content for the purpose of detecting fake news. For non-Euclidean deep learning techniques that achieve an accuracy of 97% in terms of ROC, the phrase" geometric deep learning" is frequently used as a catch-all word [12].

To detect fake news on Twitter, a deep neural network architecture that can handle multiple input methods has been proposed. This architecture includes word embeddings of news headlines and bodies, linguistic attributes, and network account features like user profiles. The input can be combined at different network layers. A significant contribution of this approach is the creation of a new Twitter dataset containing both real and fake news related to the Hong Kong protests [13].

Additionally, it is suggested to utilize a linguistic model [14] to identify the content features of news material, including its syntactic, grammatical, sentimental, and readability features. This model can then be used as a neural-based sequential learning model to detect fake news. Similarly, Hakak et al. proposed an ensemble classification model based on linguistic variables for fake news detection. They used 26 linguistic features from the text and fed them into a Decision Tree, Random Forest, and Extra Tree Classifier ensemble model.[15].

Inclusion Criteria	Exclusion Criteria
Paper written in English.	Papers in a different language than English.
Peer review of the paper	Peer review is not done.
Clearly describes fake news using ML	Does not focus on fake news detection
Have been published in last 7 years (2017-23)	Have been published before 2017.
Published in a journal or in the proceedings of a conference	Editorial comments, government documents, book reviews and blog posts.

Figure 1: Inclusion and Exclusion Criteria.

4. Models Studied

•Machine learning has proven to be a valuable tool in detecting fake news, using two primary methods: supervised learning and unsupervised learning. Unsupervised learning makes getting training data easier by using unlabeled data to extract features, but it might not be as efficient as supervised learning. Labeled data are

necessary for supervised learning, but gathering enough of them can be time and money-consuming. Deep learning has become a prominent tool for detecting fake news as a result of recent advancements in natural language processing. A variety of algorithms are employed to detect fake news.[16].

• To analyze the text of the news stories, NLP techniques are applied. Techniques like sentiment analysis, topic modeling, and semantic analysis can be used in this. These methods can aid in spotting the use of emotive language, content discrepancies, and factual mistakes in reports. NLP, a cutting-edge field of machine learning, enables computers to learn, analyze, manipulate, and even synthesize human languages. In this process, various techniques such as pre-processing, word embedding, and feature extraction are used.[16].

• The traditional machine learning models, also referred to as shallow models, encompass various algorithms including supervised and unsupervised learning. Unsupervised learning involves k-means, while supervised learning comprises techniques such as evolution tree analysis, support vector machine (SVM), hybrid SVM, Bernoulli's naive Bayes [17], LDA [18], and voting classifier [19]. These algorithms have been utilized in multiple studies, including Zhang et al. (2019) [20], Jang et al. (July 2018) [21], Kauffmann et al. (October 2020) [22] to detect fake news.

• A variety of deep learning models have been utilized in research, such as unsupervised learning algorithms like RNN (Shu et al. 2019; Agarwal et al. 2020) [23,24]), LSTM (Umer et al. 2020) [25,], and GRU (Vereshchuk et al. 2020) [26], as well as AC-BiLSTM (Trueman et al. 2021) [27]; and Ensemble models (Ozbay and Alatas February 2020[28]. Additionally, CNN (Kaliyar et al. June 2020[29]; models have also been implemented in studies to achieve effective results.

• Deep learning algorithms have the ability to learn features directly from source texts and images without the need for manual feature engineering. This process can be executed end-to-end. Compared to shallow models, deep learning techniques excel in interpretability, learning capacity, feature representation, number of parameters, and execution time, especially for large datasets.[16].

• According to Devlin et al. [33], BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained model designed to learn contextual word representations of unlabeled text. The BERT-Base variant was studied due to the significant time and memory requirements of the BERT-Large model. BERT-Base is one of two BERT variants originally introduced.[32].

5. Bibliometric Analysis

We performed a bibliometric study to determine the most significant papers and researchers in the area of fake news detection using machine learning. The following search terms were used to retrieve all publications pertaining to fake news detection using machine learning.

("fake news" OR "machine learning" OR "deep learning" OR "natural language processing" OR "social media").
Our research was restricted to journal articles and survey papers across a ten-year span, however, most of the research papers reviewed were published in or after 2020.

We then used the subsequent inclusion criteria:
• The publication must be written in English.
• The article must discuss fake news detection using machine learning.
• The article or research paper must have undergone peer review.
The most influential papers and researchers were then found by looking at the citation information of these publications.

6. Result

In this part, we answer four research questions:
• RQ1: What exactly is social media fake news, and what kind of fake news are there?

According to Allcott and Gentzkow [30], the first definition of fake news is news pieces that are purposefully and demonstrably inaccurate and may lead to reader confusion. Other definitions were later offered in the literature, but they all concur that fake news is untrue (i.e., non-factual) in its authenticity. They dispute, nevertheless, on whether certain related ideas, such as rumors, satire, conspiracy theories, hoaxes, and incorrect information, should be included or excluded from the definition.

It is possible to list and identify the terms that are most frequently used in the literature to refer to, describe, and determine fake news. In Figure 2, we classify on the basis of various terminologies based on two essential traits that they all share: the purpose behind the news's publication and its validity. To precisely distinguish false news, terms such as disinformation, misinformation, and inaccurate information are used. The authenticity factor of a statement refers to its factual component and whether it can be verified as true or false. On the other hand, the intent aspect refers to the motivation behind the statement and whether it aims to deceive or harm.[31].

• RQ2: How reliable are deep learning versus conventional machine learning models at spotting fake news?
Generally, deep learning models perform better than traditional models in terms of overall accuracy,

particularly on large datasets like the Combined Corpus. However, deep learning models are more susceptible to overfitting on smaller datasets. Despite this, Naive Bayes (with n-gram), a traditional model, shows promising potential in detecting fake news, with a performance that is nearly on par with deep learning models and achieving an accuracy of 93% on the combined corpus.

Upon further investigation, it was found that as the amount of training data increases, the performance of the Bi-LSTM deep learning model improves at a faster rate compared to Naive Bayes. However, at some point, the performance of Naive Bayes reaches a saturation point and further improvements are minimal even with more training data. As a result, it can be concluded that deep learning models have the potential to outperform Naive Bayes when trained with sufficient examples [32].

• RQ3: What methods are there to detect fake news on social media?
Using the methodologies, we divide the detection methods into three categories as shown in Figure 3.

Human-based methods: This group generally consists of the employment of fact-checking and crowdsourcing methods, which are based on human understanding to examine and confirm the accuracy of news information.

Artificial intelligence-based methods: The group of artificial intelligence-based methods is commonly used in academic publications to detect fake news. It category includes a combination of traditional machine learning (ML), deep learning methods such as convolutional neural networks (CNN) and recurrent neural networks (RNN), and natural language processing (NLP) techniques.

Blockchain-based techniques: This category contains ways that employ blockchain technology to identify and counteract fake news on social media by establishing the legitimacy of the news's sources and their ability to be tracked back.[31].

• RQ4: Which model works best with a limited amount of training data?

We looked into how various models performed with a small amount of training data. Notably, even with little datasets, pre-trained BERT-based models showed good performance. Particularly on smaller datasets such as fake or real news, these algorithms are useful in identifying fake news. Furthermore, we analysed models from all three categories: Bi-LSTM (a deep learning model) and Naive Bayes with n-gram (a conventional model). Our findings validate the effectiveness of artificial intelligence-based methods in this context [32].

7.Conclusion

Detecting fake news can be challenging for humans as it requires expertise and in-depth knowledge to recognize anomalies and classify news reports. Manual validation of a single article is difficult, which makes the classification of fake news using machine learning models and ensemble techniques necessary. It is essential to be aware that not all news on social media is true and to exercise critical thinking. This can help the public make informed decisions and protect them from those who spread misinformation.

This study conducted a comprehensive analysis of false news detection models, building upon previous research. Various machine learning and deep learning approaches

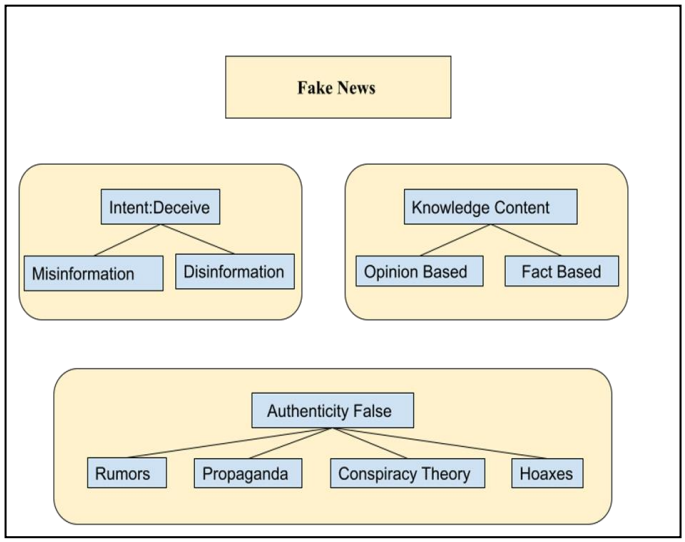

Figure 2: Classification of Fake News Types.

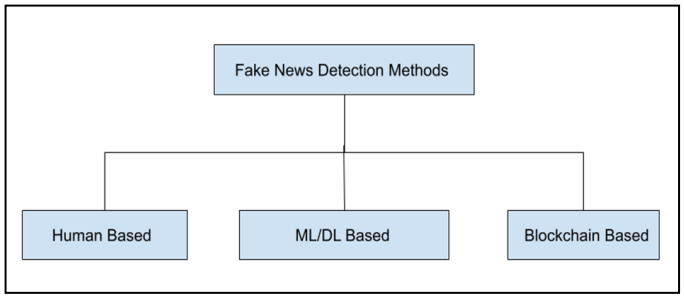

Figure 3: Fake News Detection Methods.

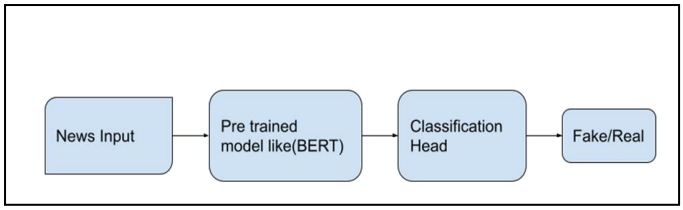

Figure 4: Fine-tuning of BERT Model.

Were explored, along with datasets, simulation environments, techniques, and their respective strengths and limitations. Performance measures were also presented for evaluating model effectiveness, as well as identifying gaps and challenges in developing new false news detection algorithms. The findings of this survey can serve as a guide for future researchers to develop innovative and effective false news detection programs. By improving the ability to identify false news, researchers can gain a better understanding of current issues, potential solutions, and future directions.

7.REFERENCES

[1] Jarrahi, A., Safari, L. Evaluating the effectiveness of publishers' features in fake news detection on social media. *Multimedia Tools and Applications, 82(3)*, 2913–2939.

[2] de Beer D, Matthee M.(2020) Approaches to Identify Fake News: A Systematic Literature Review. *Integrated Science in Digital Age*.

[3] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu(2017). Fake news detection on social media. *ACM SIGKDD Explorations Newsletter,* 19(1), 22–36.

[4] Asha. J1, Meenakowshalya (2021). Fake News Detection Using Unsupervised and Deep Learning Algorithms.

[5] Sharma, U., Saran, S., & Patil, S. M. (2021). Fake News Detection using Machine Learning Algorithms. *International Journal of Engineering Research & Technology (IJERT) NATSU – 2020 (Volume 09 – Issue 03).*

[6] Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M. O. (2020). Fake News Detection Using Machine Learning Ensemble Methods.

[7] Berahmand K et al (2021) A preference random walk algorithm for link prediction through mutual influence nodes in complex networks. *Journal of King Saud University - Computer and Information Sciences.*

[8] Qawasmeh E., Tawalbeh M. Abdullah M. (2019). October. Automatic identification of fake news using deep learning. 2019 *Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS).*

[9] Sedik, A., Abohany, A. A., Sallam, K. M., Munasinghe, K., & Medhat, T. (2022). Deep fake news detection system based on concatenated and recurrent modalities. *Expert Systems with Applications*, 208, 117953.

[10] Kaliyar, R. K., Goswami, A., Narang, P., & Sinha, S. (2020). FNDNet – A deep convolutional neural network for fake news detection. *Cognitive Systems Research*, 61, 32-44.

[11] Gupta, H., Jamal, M. S., Madisetty, S., & Desarkar, M. S. (2018). A framework for real-time spam detection in Twitter. *2018 10th International Conference on Communication Systems & Networks (COMSNETS)*, Bengaluru, 380-383.

[12] Monti F., Frasca F., Eynard D., Mannion D. Bronstein M.M. (2019). Fake news detection on social media using geometric deep learning. *Social and Information Networks.*

[13] Mouratidis D, Nikiforos MN, Kermanidis KL (2021) Deep learning for fake news detection in a pairwise textual input Schema. *Computation* 9(2),20.

[14] Choudhary, A., & Arora, A. (2021). Linguistic feature-based learning model for fake news detection and classification. *Expert Systems with Applications*, 169, 114171.

[15] Hakak, S., Alazab, M., Khan, S., Gadekallu, T. R., Maddikunta, P. K. R., & Khan, W. Z. (2021). An ensemble machine learning approach through effective feature extraction to classify fake news. *Future Generation Computer Systems,* 117, 47-58.

[16] Shah, N., & Ganatra, A. (2022). A systematic literature review and existing challenges toward fake news detection models. *Social Network Analysis and Mining,* 12, 168.

[17] Singh, M., Bhatt, M. W., Bedi, H. S., & Mishra, U. (2020). Performance of Bernoulli's naive bayes classifier in the detection of fake news.

[18] Reddy, H., Raj, N., Gala, M., & Basava, A. (2020). Text-mining-based fake news detection using ensemble methods. *International Journal of Automation and Computing*, 17, 210-221.

[19] Verma, P. K., Agrawal, P., Amorim, I., & Prodan, R. (2021). WELFake: word embedding over linguistic features for fake news detection. I*EEE Transactions on Computational Social Systems,* 8(4), 881-893.

[20] Zhang, C., Gupta, A., Kauten, C., Deokar, A. V., & Qin, X. (2019). Detecting fake news for reducing misinformation risks using analytics approaches. *European Journal of Operational Research,* 279(3), 1036-1052.

[21] Jang, S. M., Geng, T., Li, J. Y., Xia, R., Huang, C. T., Kim, H., & Tang, J. (2018). A computational approach for examining the roots and spreading patterns of fake news: evolution tree analysis. *Computers in Human Behavior*, 84, 103-113.

[22] Kauffmann E, Pearl J, Gil D, Fernandez A, Sellers R, Mora H (2020) A framework for big data analytics in commercial social networks: a case study on sentiment analysis and fake review detection for marketing decision-making. *Industrial Marketing Management,* 90, 523-537.

[23] Shu, K., Magudeswaran, D., & Liu, H. (2019). FakeNewsTracker: a tool for fake news collection, detection, and visualization.

[24] Agarwal, A., Mittal, M., Pathak, A., & Goyal, L. M. (2020). Fake news detection using a blend of neural networks: an application of deep learning.

[25] Umer, M., Imtiaz, Z., Ullah, S., Mehmood, A., Choi, G. S., & On, B. W. (2020). Fake news stance detection using deep learning architecture (CNN-LSTM). IEEE Access, 8, 156695-156706.

[26] Vereshchaka, A., Cosimini, S., & Dong, W. (2020). Analyzing and distinguishing fake and real news to mitigate the problem of disinformation.

[27] Trueman, T. E., Kumar, A., Narayanasamy, P., & Vidya, J. (2021). Attention-based C-BiLSTM for fake news detection. *Applied Soft Computing, 110.*

[28] Jiang, T., Li, J. P., Haq, A. U., Saboor, A., & Ali, A. (2021). A novel stacking approach for accurate detection of fake news.

[29] Kaliyar, R. K., Goswami, A., Narang, P., & Sinha, S. (2020). FNDNet—a deep convolutional neural network for fake news detection. *Cognitive Systems Research*, 61, 32-44.

[30] Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives.*

[31] Aimeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining,* 13, 30.

[32] Khan, J. Y., Khondaker, M. T. I., Afroz, S., Uddin, G., & Iqbal, A. (2021). A benchmark study of machine learning models for online fake news detection.