A Comprehensive Survey on Machine Learning Techniques for Fake News Detection and Analysis

Saurabh Jaiswal[1], Mr. Vivek Rai [2]
[1]*M.Tech, Dept. of Computer Science & Engineering, B N College of engineering & Technology (AKTU), Lucknow, India*
[2]*Assistant Professor, Dept. of CSE,* B N College of *engineering & Technology, Lucknow, India*

***Abstract*— The proliferation of fake news poses significant challenges to society, undermining trust in media and impacting public opinion and behavior. As the digital landscape expands, the need for effective detection and analysis of fake news becomes increasingly critical. This comprehensive survey examines the state-of-the-art machine learning techniques employed in the identification and analysis of fake news. We systematically categorize various approaches, including supervised, unsupervised, and semi-supervised learning methods, and evaluate their performance in different contexts. Key features and data sets used in these studies are discussed, alongside an analysis of the strengths and limitations of each technique. Furthermore, we explore the challenges faced in the fake news detection landscape, such as the dynamic nature of fake news, data scarcity, and the need for real-time processing. Future research directions are proposed to address these challenges, emphasizing the integration of multi-modal data, advanced natural language processing techniques, and the potential of deep learning models. This survey aims to provide researchers and practitioners with a detailed understanding of current methodologies and inspire innovative solutions to enhance the effectiveness of fake news detection systems.**

Keywords: Fake News, media, public opinion, machine learning (ML), deep learning

1. INTRODUCTION

The digital revolution has transformed the way information is disseminated and consumed, enabling unprecedented access to news and opinions from around the globe [1]. However, this transformation has also facilitated the rapid spread of fake news—false or misleading information presented as news, often with the intent to deceive [2]. The pervasive nature of fake news can have profound consequences, including the erosion of public trust in media, manipulation of political outcomes, and social unrest. As such, developing effective methods to detect and mitigate fake news is a critical challenge for researchers, policymakers, and technology developers [3].

Machine learning has emerged as a promising tool in the fight against fake news, offering the ability to analyze vast amounts of data and identify patterns indicative of false information [4]. Machine learning algorithms can be trained to recognize the subtle linguistic, stylistic, and contextual features that differentiate fake news from legitimate news. Additionally, these techniques can leverage metadata and user behavior patterns to enhance detection accuracy[5].

This comprehensive survey aims to provide a detailed overview of the machine learning techniques employed in fake news detection and analysis. We begin by defining fake news and discussing its various forms and the contexts in which it occurs [6]. Following this, we categorize the primary machine learning approaches used in this domain, including supervised, unsupervised, and semi-supervised learning methods. We examine the key features and data sets that these models rely on and evaluate their performance across different scenarios [7].

Moreover, this survey addresses the inherent challenges in fake news detection, such as the evolving nature of fake news tactics, the scarcity of labeled data, and the necessity for real-time processing [8]. We also highlight the limitations of current approaches and suggest potential avenues for future research. These include the integration of multi-modal data sources, advancements in natural language processing (NLP), and the application of deep learning models [9].

By providing a thorough analysis of existing methodologies and their effectiveness, this survey seeks to inform and guide future research efforts in the field of fake news detection. Our goal is to foster the development of more robust and accurate detection systems, ultimately contributing to a more informed and resilient society [10]. The Facebook clients who posted the pictures guaranteed they showed a slaughter in progress in the Gashish area of Plateau State, Nigeria by Fulani Muslims who were murdering Christians from the locales Berom ethnic minority [11].

As an outcome, a slaughter occurred in Gashish that end of the week and somewhere close to 86 and 238 Berom individuals were executed, as per gauges made by the police and by neighborhood local area pioneers. Nonetheless, probably the most combustible pictures and recordings were absolutely immaterial to the brutality in Gashish [12]. The video showing a man's head was cut, was not occurred in Nigeria and it was recorded in Congo, in 2012.2 The earlier chips away at counterfeit news location have applied a few conventional AI techniques and neural organizations to distinguish counterfeit news. In any case, they have zeroed in on recognizing information on specific sorts, (for example, political) [19]. In like manner, they fostered their models and planned highlights for explicit datasets that match their subject of interest. All things considered, these methodologies would experience the ill effects of dataset predisposition and are probably going to perform ineffectively on information on another point [13]. A portion of the current investigations have likewise made correlations among various strategies for counterfeit news recognition. It has assembled a benchmark dataset specifically, Liar and tested some current models on that dataset. The examination result hints us how various models can perform on an organized dataset like Liar [14]. Be that as it may, the length of this dataset isn't adequate for neural organization investigation and a few models were found to experience the ill effects of overfitting. Gilda has investigated some conventional AI approaches [15]. Notwithstanding, many progressed AI models, e.g., neural organization based ones are not applied that have been demonstrated best in numerous content characterization issues. A significant limit of earlier relative examinations is that these are completed on a particular sort of dataset, it is hard to arrive at a decision about the exhibition of different models [16]. Additionally, these works have zeroed in on a predetermined number of highlights that have brought about the deficient investigation of expected attributes of phony news. In this examination, we will probably introduce a relative presentation investigation of existing strategies by carrying out every one on two of the accessible datasets and another pre-arranged by us consolidating information on circulated subjects. We likewise fuse various highlights from existing works and explore the exhibition of some effective content order strategies that are yet to be applied for counterfeit news recognition as far as we could possibly know [17].

There exists a huge assemblage of exploration on the subject of AI techniques for trickiness discovery, its vast majority has been zeroing in on ordering on the web audits and freely accessible online media posts. Especially since late 2016 during the American Presidential political race, the topic of deciding 'counterfeit news' has likewise been the subject of specific consideration inside the writing. Conroy, Rubin, and Chen [1] diagrams a few methodologies that appear to be encouraging towards the point of impeccably group the deceptive articles.

In this review paper section I contains the introduction, section II contains the literature review details, section III contains the details about methodologies, section and section IV provide conclusion of this paper.

2. RELATED WORK

Julio C. S. Reis, et. al., (2019) the genuine separating force of these highlights is at this point unclear: some are more broad, however others perform well just with explicit information. In this work, we lead a profoundly exploratory examination that created countless models from a huge and various arrangement of highlights. These models are unprejudiced as in their highlights are haphazardly browsed the pool of accessible highlights. While by far most of models are insufficient, we had the option to create various models that yield profoundly exact choices, hence successfully isolating phony news from genuine stories. In particular, we zeroed in our investigation on models that position a haphazardly picked counterfeit report higher than an arbitrarily picked reality with more than 0.85 likelihood. For these models we tracked down a solid connection among highlights and model expectations, showing that a few highlights are plainly custom fitted for distinguishing particular sorts of phony news, accordingly confirming that various mixes of highlights cover a particular area of the phony news space. At long last, we present a clarification of components adding to display choices, along these lines advancing community thinking by supplementing our capacity to assess computerized content and arrive at justified resolutions.

Adrian M.P. et. al. (2019) Fake news detection is a difficult problem due to the nuances of language. Understanding the reasoning behind certain fake items implies inferring a lot of details about the various actors involved. We believe that the solution to this problem should be a hybrid one, combining machine learning, semantics and natural language processing. We introduce a new semantic fake news detection method built around relational features like sentiment, entities or facts extracted directly from text. Our experiments show that by adding semantic features the accuracy of fake news classification improves significantly.

William Yang Wang (2018) [2] Automatic phony news identification is a difficult issue in misdirection discovery, and it has huge true political and social effects. Be that as it may, measurable ways to deal with battling counterfeit news has been drastically restricted by the absence of marked benchmark datasets. In this paper, we present LIAR: another, freely accessible dataset for counterfeit news recognition. We gathered a long term, 12.8K physically marked short explanations in different settings from POLITIFACT.COM, which gives nitty gritty examination report and connections to source records for each case. This dataset can be utilized for certainty checking research too. Prominently, this new dataset is a significant degree bigger than already biggest public phony news datasets of comparable sort. Observationally, we examine programmed counterfeit news recognition dependent on surface-level etymological examples. We have planned a novel, half breed convolutional neural organization to incorporate metadata with text. We show that this crossover approach can improve a book just profound learning model.

Costin BUSIOC et. al., (2020) [3] Fighting phony news is a troublesome and testing task. With an expanding sway on the social and world of politics, counterfeit news apply an unprecedently sensational effect on individuals' lives. Because of this marvel, drives tending to computerized counterfeit news discovery have acquired prominence, producing inescapable examination interest. Notwithstanding, most methodologies focusing on English and low-asset dialects experience issues when conceiving such arrangements. This examination centers around the advancement of such examinations, while featuring existing arrangements, difficulties, and perceptions shared by different exploration gatherings. Furthermore, given the restricted measure of computerized examinations performed on Romanian phony news, we review the materialness of the accessible methodologies in the Romanian setting, while at the same time recognizing future exploration ways.

Alim Al Ayub Ahmed (2020) [4] Web is one of the significant developments and countless people are its clients. These people utilize this for various purposes. There are diverse web-based media stages that are open to these clients. Any client can make a post or spread the word through these online stages. These stages don't confirm the clients or their posts. So a portion of the clients attempt to get out counterfeit word through these stages. These phony news can be a promulgation against an individual, society, association or ideological group. A person can't distinguish every one of these phony news. So there is a requirement for AI classifiers that can recognize these phony news naturally. Utilization of AI classifiers for distinguishing the phony news is depicted in this methodical writing survey.

Table 1: Previous Year Research Paper Comparison table based on Findings

Paper Title	Summary
1. "Fake News Detection on Social Media: A Data Mining Perspective"	Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). This paper presents an overview of fake news detection using data mining techniques, emphasizing the role of machine learning in identifying patterns in data that indicate fake news. It discusses various features such as linguistic and network information, and the use of supervised and unsupervised learning algorithms.
2. "Fake News Detection with Deep Learning: A Review"	Zhang, X., & Ghorbani, A. A. (2020). This review paper focuses on the application of deep learning techniques for fake news detection. It explores the effectiveness of convolutional neural networks (CNNs), recurrent neural networks (RNNs), and other deep learning models in distinguishing fake news from real news based on textual data.
3. "Machine Learning Approaches to Fake News Detection: A Survey"	Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). This survey provides an extensive review of various machine learning approaches applied to fake news detection. It includes traditional machine learning models such as support vector machines (SVMs) and decision trees, as well as ensemble methods and their comparative effectiveness.
4. "The Role of User Profile Information in Fake News Detection"	Ruchansky, N., Seo, S., & Liu, Y. (2017). This paper investigates the significance of user profile information in the detection of fake news. It highlights the integration of user characteristics with content-based features to improve the accuracy of machine learning models in fake news detection.
5. "Attention-based Recurrent Neural Networks for Rumor Veracity Classification"	Ma, J., Gao, W., Wei, Z., Lu, Y., & Wong, K.-F. (2016). This study explores the use of attention mechanisms in recurrent neural networks to classify the veracity of rumors on social media platforms. The paper demonstrates how attention-based models can enhance the understanding and detection of fake news.
6. "Exploiting Temporal Patterns for Fake News Detection on Social Media"	Yang, Y., Zheng, L., Zhang, J., Cui, Q., Li, Z., & Yu, P. S. (2018). The paper delves into the use of temporal patterns in social media posts for fake news detection. It discusses the application of temporal features in machine learning models to improve the identification of fake news by analyzing the evolution of news dissemination.
7. "A Multi-Source Approach to Fake News Detection in Social Networks"	Qi, P., Cao, J., Yang, T., Guo, J., & Li, J. (2019). This paper proposes a multi-source approach that combines data from different social networks to enhance fake news detection. It highlights the benefits of leveraging diverse sources of information and the challenges associated with integrating data from multiple platforms.
8. "Detection of Fake News on Social Media with Bi-Directional LSTM and Co-Attention Networks"	Karimi, A., & Tang, J. (2018). This research presents a model combining bi-directional long short-term memory (LSTM) networks with co-attention mechanisms to improve the detection of fake news. It demonstrates the model's effectiveness in capturing context and semantic meaning from text data.
9. "Using Contextual Information for Fake News Detection"	Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. (2018). The study focuses on incorporating contextual information, such as the surrounding text and metadata, to enhance fake news detection models. It evaluates the impact of context-aware features on the performance of various machine learning algorithms.
10. "A Benchmark Dataset for Fake News Detection"	Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2020). This paper introduces a comprehensive benchmark dataset specifically designed for fake news detection research. It discusses the dataset's creation, characteristics, and potential applications in evaluating and training machine learning models.

3. METHODOLOGY

- Decision Tree Algorithm

Decision Tree algorithm has a place with the group of managed learning calculations. In contrast to other administered learning calculations, the decision tree calculation can be utilized for tackling relapse and order issues as well. The objective of utilizing a Decision Tree is to make a preparation model that can use to anticipate the class or worth of the objective variable by taking in basic decision principles surmised from earlier data(training information). In Decision Trees, for anticipating a class name for a record we start from the foundation of the tree. We think about the upsides of the root trait with the record's characteristic. Based on examination, we follow the branch relating to that worth and leap to the following hub.

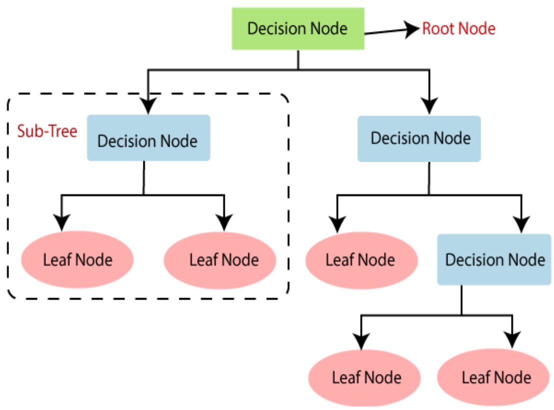


Figure 1: Structure of decision tree algorithm

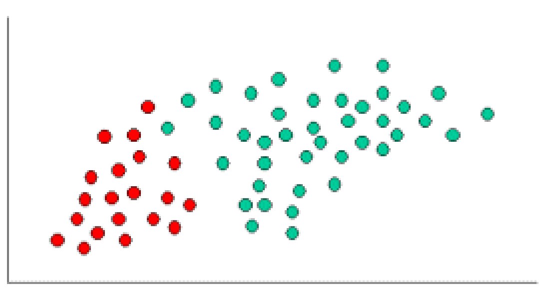
The complete process can be better understood using the below algorithm:

- Step-1: Begin the tree with the root node, says S, which contains the complete dataset.
- Step-2: Find the best attribute in the dataset using Attribute Selection Measure (ASM).
- Step-3: Divide the S into subsets that contains possible values for the best attributes.
- Step-4: Generate the decision tree node, which contains the best attribute.
- Step-5: Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

- NAÏVE BAYES

Naive Bayes Classifier Introductory Overview

The Naive Bayes Classifier procedure depends on the supposed Bayesian hypothesis and is especially fit when the dimensionality of the information sources is high. Notwithstanding its straightforwardness, Naive Bayes can frequently beat more complex arrangement strategies.



To show the idea of Naïve Bayes Classification, consider the model showed in the outline above. As demonstrated, the items can be delegated either GREEN or RED. Our assignment is to arrange new cases as they show up, i.e., choose to which class name they have a place, in view of the at present leaving objects.

Since there are twice as many GREEN items as RED, it is sensible to accept that another case (which hasn't been noticed at this point) is twice as liable to have enrollment GREEN instead of RED. In the Bayesian investigation, this conviction is known as the earlier likelihood. Earlier probabilities depend on past experience, for this situation the level of GREEN and RED items, and frequently used to anticipate results before they really occur.

Thus, we can write:

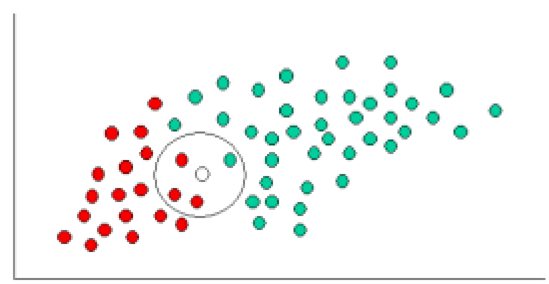
$$\text{Prior probability for GREEN} \propto \frac{\text{Number of GREEN objects}}{\text{Total number of objects}}$$

$$\text{Prior probability for RED} \propto \frac{\text{Number of RED objects}}{\text{Total number of objects}}$$

Since there is a total of 60 objects, 40 of which are GREEN and 20 RED, our prior probabilities for class membership are:

$$\text{Prior probability for GREEN} \propto \frac{40}{60}$$

$$\text{Prior probability for RED} \propto \frac{20}{60}$$



Having detailed our earlier likelihood, we are currently prepared to group another item (WHITE circle). Since the articles are very much bunched, it is sensible to expect to be that the more GREEN (or RED) objects nearby X, the more probable that the new cases have a place with that specific tone. To quantify this probability, we draw a circle around X which envelops a number (to be picked deduced) of focuses regardless of their group names. Then, at that point we ascertain the quantity of focuses in the circle having a place with each class name. From this we figure the probability:

$$\text{Likelihood of } X \text{ given GREEN} \propto \frac{\text{Number of GREEN in the vicinity of } X}{\text{Total number of GREEN cases}}$$

$$\text{Likelihood of } X \text{ given RED} \propto \frac{\text{Number of RED in the vicinity of } X}{\text{Total number of RED cases}}$$

From the illustration above, it is clear that Likelihood of X given GREEN is smaller than Likelihood of X given RED, since the circle encompasses 1 GREEN object and 3 RED ones. Thus:

$$\text{Probability of } X \text{ given GREEN} \propto \frac{1}{40}$$

$$\text{Probability of } X \text{ given RED} \propto \frac{3}{20}$$

Albeit the earlier probabilities show that X may have a place with GREEN (given that there are twice as many GREEN contrasted with RED) the probability demonstrates something else; that the class participation of X is RED (given that there are more RED articles nearby X than GREEN). In the Bayesian investigation, the last characterization is created by consolidating the two wellsprings of data, i.e., the earlier and the probability, to frame a back likelihood utilizing the alleged Bayes' standard (named after Rev. Thomas Bayes 1702-1761).

$$\text{Posterior probability of } X \text{ being GREEN} \propto$$

$$\text{Prior probability of GREEN} \times \text{Likelihood of } X \text{ given GREEN}$$

$$= \frac{4}{6} \times \frac{1}{40} = \frac{1}{60}$$

$$\text{Posterior probability of } X \text{ being RED} \propto$$

$$\text{Prior probability of RED} \times \text{Likelihood of } X \text{ given RED}$$

$$= \frac{2}{6} \times \frac{3}{20} = \frac{1}{20}$$

Finally, we classify X as RED since its class membership achieves the largest posterior probability.

Note. The above probabilities are not normalized. However, this does not affect the classification outcome since their normalizing constants are the same.

To index

Technical Notes

In the previous section, we provided an intuitive example for understanding classification using Naive Bayes. In this section are further details of the technical issues involved. Naive Bayes classifiers can handle an arbitrary number of independent variables whether continuous or categorical. Given a set of variables, X = {x1,x2,x...,xd}, we want to construct the posterior probability for the event Cj among a set of possible outcomes C = {c1,c2,c...,cd}. In a more familiar language, X is the predictors and C is the set of categorical levels present in the dependent variable. Using Bayes' rule:

$$p(C_j \mid x_1, x_2, \ldots, x_d) \propto p(x_1, x_2, \ldots, x_d \mid C_j) p(C_j)$$

where p(Cj | x1,x2,x...,xd) is the posterior probability of class membership, i.e., the probability that X belongs to Cj. Since Naive Bayes assumes that the conditional probabilities of the independent variables are statistically independent we can decompose the likelihood to a product of terms:

$$p(X \mid C_j) \propto \prod_{k=1}^{d} p(x_k \mid C_j)$$

and rewrite the posterior as:

$$p(C_j \mid X) \propto p(C_j) \prod_{k=1}^{d} p(x_k \mid C_j)$$

Utilizing Bayes' standard above, we name another case X with a class level Cj that accomplishes the most noteworthy back likelihood.

Albeit the suspicion that the indicator (free) factors are autonomous isn't generally exact, it improves on the grouping task drastically, since it permits the class contingent densities p(xk | Cj) to be determined independently for every factor, i.e., it diminishes a multidimensional errand to various one-dimensional ones. As a result, Naive Bayes diminishes a high-dimensional thickness assessment undertaking to a one-dimensional piece thickness assessment. Moreover, the suspicion doesn't appear to incredibly influence the back probabilities, particularly in areas close to choice limits, in this manner, leaving the arrangement task unaffected. Innocent Bayes can be demonstrated in a few distinct manners including ordinary, lognormal, gamma and Poisson thickness.

4. CONCLUSION

The fight against fake news is an ongoing battle, one that is becoming increasingly critical in our digitally connected world. Machine learning has demonstrated substantial potential in identifying and mitigating the impact of fake news, thanks to its ability to process and analyze large volumes of data to uncover patterns and signals indicative of false information. This comprehensive survey has highlighted the diverse array of machine learning techniques currently employed in fake news detection and analysis, showcasing their strengths, limitations, and areas of application.

Our examination revealed that while supervised learning methods have shown high accuracy in controlled environments, they often struggle with the dynamic and evolving nature of fake news. Unsupervised and semi-supervised methods offer promising alternatives, particularly in scenarios with limited labeled data, yet they too face challenges in achieving real-time detection and high precision. The integration of multi-modal data and advanced natural language processing techniques has been identified as a crucial step forward, providing richer context and enhancing the robustness of detection systems.

Furthermore, the survey underscored several critical challenges that persist in the field, including the scarcity of comprehensive and labeled datasets, the adaptability of fake news creators, and the need for scalable solutions that can operate in real-time. Addressing these challenges requires a multifaceted approach, combining advancements in algorithm development with collaborative efforts across academia, industry, and government.

In conclusion, while significant progress has been made in the application of machine learning to fake news detection, there remains much work to be done. Future research should focus on developing more sophisticated models that can adapt to the rapidly changing landscape of fake news, leveraging cross-disciplinary insights and technological innovations. By continuing to advance the state-of-the-art in fake news detection, we can contribute to a more informed public discourse and a more resilient information ecosystem. This survey serves as a foundation for such efforts, providing a detailed understanding of current methodologies and pointing towards the future directions in this vital area of study.

REFERENCE

[1] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P.A., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zhang, X.: Tensorow: A System for Large-Scale Machine Learning. CoRR abs/1605.08695 (2016), http://arxiv.org/abs/1605.08695

[2] Aghakhani, H., Machiry, A., Nilizadeh, S., Kruegel, C., Vigna, G.: Detecting Deceptive Reviews using Generative Adversarial Networks. CoRR abs/1805.10364 (2018), http://arxiv.org/abs/1805.10364

[3] Costin BUSIOC, Stefan RUSETI, Mihai DASCALU, " A Literature Review of NLP Approaches to Fake News Detection and Their Applicability to Romanian- Language News Analysis", 2020, *Romanian Ministry of Education and Research, CNCS - UEFISCDI, project number PN-III-P1-1.1-TE-2019-1794, within PNCDI III.*

[4] Alim Al Ayub Ahmed, Ayman Aljarbouh, Praveen Kumar Donepudi," Detecting Fake News using Machine Learning: A Systematic Literature Review",2020, IEEE conference.

[5] Razan Masood and Ahmet Aker," The Fake News Challenge: Stance Detection using Traditional Machine Learning Approaches", In Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (KMIS 2018), pages 128-135

[6] Sohan De Sarkar, Fan Yang," Attending Sentences to detect Satirical Fake News", Proceedings of the 27th International Conference on Computational Linguistics, pages 3371–3380 Santa Fe, New Mexico, USA, August 20-26, 2018.

[7] Abdullah-All-Tanvir, Ehesas Mia Mahir, Saima Akhter" Detecting Fake News using Machine Learning and Deep Learning Algorithms", 2019 7th International Conference on Smart Computing & Communications (ICSCC).

[8] Hadeer Ahmed, Issa Traore, and Sherif Saad. Detection of online fake news using n-gram analysis and machine learning techniques. In International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, pages 127–138. Springer, 2017.

[9] Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. Journal of Economic Perspectives, 31(2):211–36, 2017.
[10] Peter Bourgonje, Julian Moreno Schneider, and Georg Rehm. From clickbait to fake news detection: an approach based on detecting the stance of headlines to articles. In Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism, pages 84–89, 2017. 12
[11] Yimin Chen, Niall J Conroy, and Victoria L Rubin. Misleading online content: Recognizing clickbait as false news. In Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection, pages 15–19. ACM, 2015.
[12] Mathieu Cliche. The sarcasm detector, 2014.
[13] Niall J Conroy, Victoria L Rubin, and Yimin Chen. Automatic deception detection: Methods for finding fake news. In Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community, page 82. American Society for Information Science, 2015.
[14] Ethan Fast, Binbin Chen, and Michael S Bernstein. Empath: Understanding topic signals in large-scale text. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pages 4647–4657. ACM, 2016.
[15] Song Feng, Ritwik Banerjee, and Yejin Choi. Syntactic stylometry for deception detection. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2, pages 171–175. Association for Computational Linguistics, 2012.
[16] Johannes F¨urnkranz. A study using n-gram features for text categorization. Austrian Research Institute for Artifical Intelligence, 3(1998):1–10, 1998.
[17] Shlok Gilda. Evaluating machine learning algorithms for fake news detection. In Research and Development (SCOReD), 2017 IEEE 15th Student Conference on, pages 110–115. IEEE, 2017.
[18] Mykhailo Granik and Volodymyr Mesyura. Fake news detection using naive bayes classifier. In Electrical and Computer Engineering (UKRCON), 2017 IEEE First Ukraine Conference on, pages 900–903. IEEE, 2017.
[19] A´ ngel Herna´ndez-Castan˜eda and HiramCalvo. Deceptive text detection using continuous semantic spacemodels. Intelligent Data Analysis, 21(3):679–695, 2017.
[20] Johan Hovold. Naive bayes spam filtering using word-position-based attributes. In CEAS, pages 41–48, 2005.
[21] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759, 2016.
[22] Z Khanam1 , B N Alwasel1 , H Sirafi1 and M Rashid, "Fake News Detection Using Machine Learning Approaches", IOP Conf. Series: Materials Science and Engineering 1099 (2021) 012040 IOP Publishing doi:10.1088/1757-899X/1099/1/012040.
[23] Hadeer Ahmed, Issa Traore, Sherif Saad, "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques",International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments ISDDC 2017: Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments pp 127-138.
[24] Vo, N., Lee, K.: The Rise of Guardians: Fact-checking URL Recommendation to Combat Fake News. In: Collins-Thompson, K., Mei, Q., Davison, B.D., Liu, Y., Yilmaz, E. (eds.) The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08- 12, 2018. pp. 275{284. ACM (2018), http://doi.acm.org/10.1145/3209978.3210037
[25] Vosoughi, S., Roy, D., Aral, S.: The Spread of True and False News Online. Science 359(6380), 1146{1151 (2018)
[26] Wang, W.Y.: "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. CoRR abs/1705.00648 (2017), http://arxiv.org/abs/1705.00648