

An Analysis and Identification of Fake News using Machine Learning Techniques

Ashish

Yogananda School of AI, Computers and
Data Sciences
Shoolini University
Solon, India
vashistashu64@gmail.com

Sonia

Yogananda School of AI, Computers and
Data Sciences
Shoolini University
Solon, India
soniacsit@yahoo.com

Monika Arora

IT/Operations
Amity Business School
Amity University Haryana
marora@ggn.amity.edu

Hemraj

Yogananda School of AI, Computers and
Data Sciences
Shoolini University
Solon, India
rajputvishal846@gmail.com

Anurag Rana

Yogananda School of AI, Computers and
Data Sciences
Shoolini University
Solon, India
anuragrana@shooliniuniversity.com

Gaurav Gupta

Yogananda School of AI, Computers
and Data Sciences
Shoolini University
Solon, India
solan.gaurav@gmail.com

Abstract- Fake news is one of the major issues in today's world because a piece of false information can ruin someone's life easily. So, to identify these types of crimes, researchers introduced a fake news detection system through machine learning. Fake news identification is becoming more and more popular and widely used. Many businesses are investing in the sector, either for their needs or to offer it as a service to others. Machine learning (ML) and deep learning (DL) are two methods used for determining whether the news results to be authentic or not. Numerous methodologies exist for discerning false news through the utilization of both Machine Learning and Deep Learning methodologies. Assessing the need of the time, through this paper, an identification of fake news and analysis has been done using machine learning techniques. After a detailed review, it has been discovered that numerous Machine Learning and Deep Learning algorithms are applied. The most often used Machine Learning approach is SVM (Support Vector Machine), and the most widely used Deep Learning technique is LSTM (Long Short-Term Memory).

Keywords- News Data Analysis, Machine Learning, Deep Learning, Long Short Term Memory, Artificial Intelligence.

I. INTRODUCTION

In this digital world, the biggest concern is fake news, which can spread at lightning speed and can impact many millions of lives [1]. False information, often misleading or inaccurate, masquerades as genuine news when it is, in fact, fabricated or deceptive [2]. Today internet gives us many opportunities, so the young generation doesn't hesitate to get their news from the Internet and leave the newspapers and genuine sources behind [3]. Perhaps it is very common to receive news from the Internet, and the number of people receiving news through the Internet has increased, to 62% of America in 2016, however in 2012 it was 49% [4]. Nowadays, the Dominance of social media is very heavy in every

individual's life [5]. So for information sharing and communication social media has become a crucial tool [6]. Because of this fake news escalates through social media [5]. Fake News (FN) is the forged news spread on social media and other platforms. It seeks to deliver incorrect information to the user. It can take the shape of photographs, movies, or sounds. Researchers from all over the world are interested in how to determine whether a news story, blog post, or other piece of content is real or fraudulent Over the past few years, the automated detection of misleading, misinformation termed as fake news on social media has emerged as a highly anticipated area of study [8]. Keep in mind that the spread of fake news often occurs when publishers are focused on maximizing their profits and consumers are inclined to confirm their existing beliefs [9]. False information may originate from an individual or a collective effort. The misinformation is mainly created for personal, political, and economic profit. Nowadays fake news has become an important issue. As we get every type of information via social media, misinformation is also spreading at a very fast rate. There are many different types of tweets, such as those about a government, globally popular themes, mental abuse, urban legends, and tragic events [10]. To control misinformation, Fake news detection has become a very important research topic. Take the example of Facebook, there are 2.91 billion users active on Facebook, so spreading a piece of fake news is very easy and fast. Presently, primary sources of news include social media platforms and online news portals, and other online media, with breaking and exciting news being shared quickly [11]. In Bengaluru, on 30 March 2018, the cofounder of 'Postcard News' Mahesh Vikram Hegde was arrested for spreading fake news. The person Hegde has 7, 78, 000 followers on the micro-blogging social media platform. They aim to create a communist clash and spread hate for the government. The person who was arrested uploaded a post that Muslims attacked a Jain monk. However,

the case is related to the monk “Munisriupa dyaya Mayank” and his associates being hit by a two-wheeler. After the investigation, the Police said that this was a piece of fake news aimed at creating a communist clash between two regions and giving a bad impression to the Government that nobody is safe in the state. This news is seen by thousands of people and shared on other social media platforms. After that cyber-crime police registered a case against the case owner and arrested him. Distributing such fabricated content often serves as a tactic for shaping political strategies, diverging from conventional discourse aimed at reaching consensus based on factual information [12].

Everyone who is posting misinformation and has a motive behind it is responsible for fake news. These days fake news is spread by celebrities, Politicians, and Social media influencers who are paid for it [13]. But only they are not responsible for spreading fake news, all who share these posts can be responsible for the spread of illegitimate news. To stop the word on the street the fake news detection system is created [14]. WhatsApp: - According to previous research WhatsApp is the main source of misinformation. If you find plain text without any source that is related to any kind of news, then the chances are it can be fake. Especially when someone is targeted in it. For confirmation go to Google and search the news and see the source of the news. The influence of fake news extends to people's interpretation and reaction to genuine news. For instance, certain instances of false information are crafted specifically to sow seeds of doubt and confusion, complicating the distinction between truth and falsehood for individuals [15]. Photoshop: - Photoshop is also an important tool for spreading fake news on social media. Those who know about Photoshop take the original picture and after that make the changes according to them. The image contains bogus logos and signatures; to determine whether it is real or fake, do a reverse Google image search. You just must download the image and upload it on Reverse Google Image Search and then you can see the original content if the image is Photoshop. Following the 2016 US Presidential elections, Instances of fake news and the spread of misinformation dominated headlines. Allegedly, Russia orchestrated the creation of millions of counterfeit accounts and social bots to amplify misleading information during the elections [16].

In [17], Mykhailo Granik et al. Illustrating the application of Bayes' Theorem exemplifies a strategic approach. Machine Learning encompasses a range of algorithms tailored for fake news detection, discerning patterns within data and making predictions accordingly. The field of machine learning incorporates three primary learning paradigms: semi-supervised learning, unsupervised learning, and supervised learning. Notable algorithms include Decision Trees (DT), Linear Regression (LR), and Logistic Regression (LR)[18], Random Forest (RF), etc. are examples of supervised learning. Hierarchical clustering is an example of unsupervised learning. A hybrid convolution NN model developed by Wang [19] surpasses other conventional ML techniques. A thorough analysis of linguistic aspects was carried out by Hannah Rashkin et al [20] network featuring three hierarchical levels of attention: one for words, another for phrases, and a third for

the headline of a news story by Singhania et al [21]. The characteristics stability index model was created by Ruchansky et al [22] And is based on user behavior. It captures content, an article's response, and source attributes.

II. ROLE OF MACHINE LEARNING TECHNIQUES IN FAKE NEWS DETECTION

Samuel in 1959 coined the term machine learning [23]. Machine learning is the ability of machines to learn without being programmed and it is said by Arthur Samuel. In machine learning Data gets handled more efficiently through machine learning techniques [24]. Machine learning can assist us in achieving more accurate outcomes[25]. We can classify Machine learning into two mainly different groups that is supervised and un-supervised machine learning [26]. Again supervised learning is divided into two different groups; classification and regression [27]. Classification have many more algorithms like Linear Classifier, Logistic Regression, Decision Tree [28], Support Vector Machine, Random forest [29]. During data processing, every decision tree generated by the Random Forest (RF) algorithm issues a singular prediction, which is subsequently compared against predictions generated by other trees. A final prediction is then made based on the output given by the majority of individual decisions. This technique can be utilized to pinpoint patient characteristics that were considered significant during the model's construction for predicting a particular outcome. SVMs employ a method called the "kernel trick" to proficiently perform both linear and non-linear classifications. This technique implicitly transforms inputs into high-dimensional feature spaces, enabling versatile classification [24]. The main purpose of this review is to know about the deep learning and machine learning techniques and implementations that were used in Fake news detection from 2019 to 2022. In this paper, fake news detection areas are addressed which focuses on the fake and actual news covered in the gathered publications. Furthermore, the detection of fake news leverages advance and innovative algorithms. Furthermore, this paper focuses on the accuracy found in the papers and the dataset used in the collected papers. Eminent authors have contributed by providing several applications of Machine Learning techniques [30-34].

III. PREVIOUS WORKS DONE ON FAKE NEWS DETECTION USING ML TECHNIQUES

This paper explores the classification of news articles sourced from social media, employing binary classification. Utilizing Artificial Intelligence (AI), Natural Language Processing (NLP), and Machine Learning (ML), the study aims to categorize news as either false or genuine.[7] The research focuses on identifying fake news tweets on Twitter through the application of various machine learning algorithms.[10] Five ML techniques, namely SVM, LR, Naive Bayes (NB), and Recurrent Neural Network (RNN), are compared to determine the most accurate model.[11] The paper introduces multiple ML algorithms for fake news identification and conducts a benchmark study across three datasets to assess their effectiveness. Additionally, advanced Deep Learning (DL) techniques are employed alongside supervised and

unsupervised ML methods. Overall, the primary objective of this paper is to enhance the accuracy of fake news detection.[12] This study presents a model that is more accurate at identifying fake news. Random Forest, Decision Tree, and Extra Tree Classifier are the three ML models used. On the ISOT dataset, it offers accuracy for testing and training that is 99.8% and 44.15 percent, respectively. The classification of news articles using a machine learning ensemble approach is demonstrated in this paper [13]. We leverage linguistic features to differentiate between fake news and actual news. On four separate datasets, it applies machine learning techniques. This article demonstrates that it is difficult for humans to identify bogus news one by one. Machine learning algorithms are employed in the classification of fake news, including techniques such as Naive Bayes (NB), Passive

Aggressive Classifiers, and Deep Neural Networks [35]. This study uses eight distinct datasets to determine its correctness. This research article demonstrates how fake news spreads and the motivations behind it. Using AI, NLP, and ML binary classification is performed on internet news items to differentiate between bogus and true news [15]. The purpose of this paper is to give the reader the ability to categorize news as true or fake. This manuscript discusses the definition of false news as well as its significance [16]. It discusses the effects of fake news in many industries as well as several algorithms for spotting it on social media. This study compares three distinct module text classifiers, stance detection software, and fact-checking methods currently in use to identify false information.

TABLE I. SUMMARY OF LITERATURE REVIEW

Paper	Dataset	Year	Description	Difference
R[16]	Fake News	2022	What constitutes false news, the importance of addressing fake news, its widespread impact across multiple domains, and the array of techniques available for detecting bogus news on social media platforms?	It covers Text classifiers, stance detection software, and fact-checking methods from three separate modules.
R[15]	Fake News	2021	This encompasses the propagation of false information and methodologies employed to differentiate between fake and authentic news.	It delves into binary classification methodologies facilitated by machine learning algorithms.
R[21]	Fake News	2020	It focuses on the identification of fake news and also focuses analysis and results of each model.	In this paper, Passive Aggressive classifier, NB, and Deep neural networks are covered, which are the part of machine learning algorithms.
R[13]	ISOT	2020	This paper explores textual properties that identify fake content. It performs the model on 4 different datasets.	It covers fake news detection and its accuracy using Multilayer Perceptron, KNN, and Linear Regression.
R[12]	LIAR	2020	This paper focuses on achieving better accuracy	This section addresses three prominent machine learning models: Decision Tree, Random Forest, and Extra Tree Classifier.
R[11]	Fake News	2019	Within this article, we undertake a benchmark analysis to evaluate the effectiveness of different approaches across three diverse datasets.	It covers neural networks and deep learning models.
R[5]	Twitter	2019	This paper discusses how fake news is spreading on Twitter. It proposed a model for recognizing forged news.	It covers the comparison between ML algorithms like SVM, Logistic Regression, and Naïve Bayes method.

TABLE II. ANALYSIS AND RESULTS DERIVED THROUGH THE SURVEY CONDUCTED IN THIS PAPER

Paper	Dataset	ML technique	Accuracy
R[16]	Fake News	Text classifiers, stance detection applications & fact-checking	Not Defined
R[15]	Fake News	Naïve Bayes	0.60
		Random forest	0.59
		Confusion Matrix	0.65
R[21]	Fake News	Naïve Bayes	85.6%
		Passive Aggressive	89.2%
		DNN	89.8%
R[8]	ISOT	Logistic Regression	0.97
		Linear SVM	0.98
		Multilayer Perceptron	0.98
		KNN	0.88
R[12]	LIAR	Decision Tree	99.93%
		Random Forest	99.93%
		Extra Tree Classifier	99.93%
R[11]	Fake News	SVM	0.56
		Linear Regression	0.56
		Decision Tree	0.56
		CNN	0.58
		LSTM	0.54
R[5]	Twitter	LSTM	76%
		Recurrent Neural Network	73%
		Logistic Regression	69.47%
		SVM	89.34%

IV. RESULTS AND DISCUSSION

This review examined the effectiveness of various machine learning techniques in identifying fake and true news. Our analysis highlights the strengths of models in detecting features like unusual language patterns and emotional appeals. However, limitations exist, particularly concerning the model's ability to handle deepfakes, synthetic media etc.

A. Fake and Real News Detection Areas

We find that some topics have been studied and addressed from the gathered publications. These include the definition of fake news, its consequences, and how to recognize it. The term "fake news" refers to misinformation spread through websites and social media outlets like Twitter and Facebook. The last ten years have seen a quick surge in the spread of fake news, which was most dramatically seen during the 2016 US elections [36]. Mykhailo Granik et al. demonstrate a simple method for Nave Bayes classifier fake news identification in their study [17].

B. Algorithm for DL and ML.

ML and DL provide a variety of ways to detect fake and true news. As shown in the Table I, different ML and DL techniques are utilized in the collected papers, including Decision Tree, LR, SVM, and KNN[37].

C. Estimation and Accuracy

Table II summarizes ML and DL algorithms used by researchers in the detection of fraudulent and true news. SVM, Logistic Regression, and Nave Bayes were employed in the majority of the studies.

V. CONCLUSION AND FUTURE SCOPE

Identifying and categorizing fake news within social media networks presents a challenge due to its evolving attributes and characteristics. Nonetheless, the inherent capabilities of Machine Learning (ML) and Deep Learning (DL) facilitate the computation of these features, attributes, and characteristics. Through this study, an analysis reveals that leading researchers in this domain employ techniques such as the deep Boltzmann machine DL method, deep neural networks, Convolutional Neural Networks (CNN), and Deep Autoencoder models. Recognizing this necessity and upon thorough review, it is strongly recommended that future research endeavors explore the integration of metaheuristic models for text-based fake news classification.

REFERENCES

- [1] M. Choudhary, S. Jha, D. Saxena, and A. K. Singh, "A review of fake news detection methods using machine learning," in 2021 2nd International Conference for Emerging Technology (INCET), 2021: IEEE, pp. 1-5.
- [2] S. Raza and C. Ding, "Fake news detection based on news content and social contexts: a transformer-based approach," International Journal of Data Science and Analytics, vol. 13, no. 4, pp. 335-362, 2022.
- [3] L. Waikhom and R. S. Goswami, "Fake news detection using machine learning," in Proceedings of International Conference on Advancements in Computing & Management (ICACM), 2019.
- [4] A. Abdulrahman and M. Baykara, "Fake news detection using machine learning and deep learning algorithms," 2020 International Conference on Advanced Science and Engineering (ICOASE), pp. 18-23, 2020.
- [5] J. Shaikh and R. Patil, "Fake news detection using machine learning," in 2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC), 2020: IEEE, pp. 1-5.
- [6] S. Gupta, N. Yadav, S. S. Reddy, and S. Kundu, "FakEDAMR: Fake News Detection using Abstract Meaning Representation," Authorea Preprints, 2023.
- [7] S. I. Manzoor and J. Singla, "Fake news detection using machine learning approaches: A systematic review," in 2019 3rd international conference on trends in electronics and informatics (ICOEI), 2019: IEEE, pp. 230-234.
- [8] M. K. Singh, J. Ahmed, M. A. Alam, K. K. Raghuvanshi, and S. Kumar, "A comprehensive review on automatic detection of fake news on social media," Multimedia Tools and Applications, pp. 1-34, 2023.
- [9] S. K. Uppada, K. Manasa, B. Vidhathri, R. Harini, and B. Sivaselvan, "Novel approaches to fake news and fake account detection in OSNs: user social engagement and visual content centric model," Social Network Analysis and Mining, vol. 12, no. 1, p. 52, 2022.
- [10] E. M. Mahir, S. Akhter, and M. R. Huq, "Detecting fake news using machine learning and deep learning algorithms," in 2019 7th international conference on smart computing & communications (ICSCC), 2019: IEEE, pp. 1-5.
- [11] J. Y. Khan, M. T. I. Khondaker, S. Afroz, G. Uddin, and A. Iqbal, "A benchmark study of machine learning models for online fake news detection," Machine Learning with Applications, vol. 4, p. 100032, 2021.
- [12] S. Hakak, M. Alazab, S. Khan, T. R. Gadekallu, P. K. R. Maddikunta, and W. Z. Khan, "An ensemble machine learning approach through effective feature extraction to classify fake news," Future Generation Computer Systems, vol. 117, pp. 47-58, 2021.
- [13] I. Ahmad, M. Yousaf, S. Yousaf, and M. O. Ahmad, "Fake news detection using machine learning ensemble methods," Complexity, vol. 2020, pp. 1-11, 2020.
- [14] A. Jain, A. Shakya, H. Khatter, and A. K. Gupta, "A smart system for fake news detection using machine learning," in 2019 International conference on issues and challenges in intelligent computing techniques (ICICT), 2019, vol. 1: IEEE, pp. 1-4.
- [15] U. Sharma, S. Saran, and S. M. Patil, "Fake news detection using machine learning algorithms," International Journal of Creative Research Thoughts (IJCRT), vol. 8, no. 6, pp. 509-518, 2020.
- [16] S. Ahmed, K. Hinkelmann, and F. Corradini, "Combining machine learning with knowledge engineering to detect fake news in social networks-a survey," arXiv preprint arXiv:2201.08032, 2022.
- [17] A. Alsharif, Sonia, H. Nassour and J. Sharma, "Exploring the Efficiency of Text-Similarity Measures in Automated Resume Screening for Recruitment," 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2023, pp. 36-42.
- [18] R. Sharma and H. K. Meena, "Utilizing graph Fourier transform for automatic Alzheimer's disease detection from EEG signals," International Journal of Information Technology, pp. 1-7, 2024.
- [19] W. Y. Wang, "liar, liar pants on fire": A new benchmark dataset for fake news detection," arXiv preprint arXiv:1705.00648, 2017.
- [20] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, "Truth of varying shades: Analyzing language in fake news and political fact-checking," in Proceedings of the 2017 conference on empirical methods in natural language processing, 2017, pp. 2931-2937.
- [21] S. Singhania, N. Fernandez, and S. Rao, "3han: A deep neural network for fake news detection," in Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14-18, 2017, Proceedings, Part II 24, 2017: Springer, pp. 572-581.
- [22] N. Ruchansky, S. Seo, and Y. Liu, "Csi: A hybrid deep model for fake news detection," in Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 2017, pp. 797-806.

- [23] J. Marwah, K. Kanwar, Sonia, P. Vaidya and G. Gupta, "An Analysis on Animal Welfare, Production and Cost Through Statistical and Machine learning Techniques," 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2023, pp. 1853-1857.
- [24] B. Mahesh, "Machine learning algorithms-a review," International Journal of Science and Research (IJSR), [Internet], vol. 9, no. 1, pp. 381-386, 2020.
- [25] Z. Khanam, B. Alwasel, H. Sirafi, and M. Rashid, "Fake news detection using machine learning approaches," in IOP conference series: materials science and engineering, 2021, vol. 1099, no. 1: IOP Publishing, p. 012040.
- [26] S. Hasan, T. Siddiqui, M. Mustaqeem, and N. A. Khan, "A novel framework for highly contagious diseases deaths prediction using machine learning techniques," International Journal of Information Technology, pp. 1-8, 2023.
- [27] A. A. Soofi and A. Awan, "Classification techniques in machine learning: applications and issues," Journal of Basic & Applied Sciences, vol. 13, no. 1, pp. 459-465, 2017.
- [28] A. Zabian and A. Z. Ibrahim, "Karnauph classifier for predicting breast cancer based on morphological features," International Journal of Information Technology, vol. 16, no. 1, pp. 353-359, 2024.
- [29] S. Deepak, A. Ojha, K. Acharjya, R. Mishra, A. Gantra et al., "A novel and proposed triad machine learning-based framework for the prognosis of Huntington's disease," International Journal of Information Technology, pp. 1-8, 2024.
- [30] S. R. Zahra and T. J. Saleem, "Demystifying Cyber terrorism: Causes, Costs & its impact on individuals' psyche and public confidence," in 2023 International Conference on Recent Advances in Electrical, Electronics & Digital Healthcare Technologies (REEDCON), 2023: IEEE, pp. 482-486.
- [31] T. Kanwar, G. Gupta, and R. Kumar, "A Review About Privacy and Security within WBAN Utilizing Cryptology," in 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2023: IEEE, pp. 1838-1843.
- [32] A. Alsharif, Sonia, M. Arora, and K. Aggarwal, "Predicting Time-Series Data Using Linear and Deep Learning Models—An Experimental Study," in Data, Engineering and Applications: Select Proceedings of IDEA 2021: Springer, 2022, pp. 505-516.
- [33] R. Kaur and C. Iorio, "StudentCoin Price Prediction and Relation to Blockchain Education," in 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), 2023: IEEE, pp. 28-35.
- [34] D. D. Solomon, Sonia, K. Kumar, K. Kanwar, S. Iyer et al., "Extensive Review on the Role of Machine Learning for Multifactorial Genetic Disorders Prediction. Arch Computat Methods" Archives of Computational Methods in Engineering Vol. 31, pp. 623–640, 2024. <https://doi.org/10.1007/s11831-023-09996-9>
- [35] V. Lakshmi and C. Kumari, "Detection of Fake News using Machine Learning Models," International Journal of Computer Applications, vol. 183, pp. 22-27, 2022.
- [36] A. Holan, "The media's definition of fake news vs. Donald Trump's," First Amend. L. Rev., vol. 16, p. 121, 2017.
- [37] P. N. Singh and R. P. Mahapatra, "A novel deep learning approach for arrhythmia prediction on ECG classification using recurrent CNN with GWO," International Journal of Information Technology, vol. 16, no. 1, pp. 577-585, 2024.