A Comprehensive Scrutiny on Fake News Detection Techniques

Prof. Nilam Thakkar[1] Prof. Palak Parmar[2] Prof. Avni Patel[3]
[1,2,3]LDRP-ITR, KSV, Gandhinagar, India

Abstract— Social media has recently taken over as the main source for individuals to learn about what is occurring in the world. Every day, fake news appears on social media. Social media fake news has damaged a number of industries, including politics, the economy, and health. Furthermore, it has had a negative impact on society's stability. Although various research have provided valuable models for recognizing fake news in social networks using a variety of methodologies, there are still certain restrictions and difficulties. Data augmentation, feature extraction, and data fusion are some of the approaches explored in this review to improve detection accuracy. Moreover, it discusses the most prominent techniques used in detection models and their main advantages and disadvantages. This review aims to help other researchers improve fake news detection models. In this review, many techniques to increase detection accuracy were investigated, including data augmentation, feature extraction, and data fusion. Additionally, it addresses the most popular strategies employed in detection models as well as their primary benefits and drawbacks. This review is intended to assist other academics in developing false news detecting algorithms.

Key words: Social Media, Fake News Detection Techniques

I. INTRODUCTION

Many people throughout the world now primarily consume news through social media platforms. Social media networks provide the quick dissemination of posts to a large audience in a short amount of time without cost or validation constraints, in contrast to traditional media.[1] This aids in the spread of false information on social media sites.[2] Numerous people have problems telling the difference between true and false news, according to research. This problem is not dependent on education, gender, or a particular age group.[3] Fake news propagated 70% more quickly than legitimate news, according to researchers [4]. Recent studies have claimed that the propagation of fake news on social media has emerged as a pressing issue that requires immediate attention and intervention [5] because it is causing social unrest and economic instability.[6] The detection of fake news is a challenging task. As a result, this subject has received a lot of attention from the research community. It is regarded as one of the modern fields [7], and research is still expanding and evolving in this area. This requires more development and exploration of potential future research directions [1] to improve fake news detection techniques. However, there are various sectors where false news detection overlaps [8]. As a result, numerous researchers from various fields and disciplines have developed an interest in this subject [9]. According to earlier research, fake news models encounter various challenges because of the unique characteristics of this problem. These difficulties include the absence of standard datasets, their brevity, or their unbalanced distribution, which impairs the effectiveness of detection models. How to handle social media data, the features used, and the advancement of methods for extracting these aspects should also be highlighted. developing techniques for the merging of characteristics and decision-making. Despite the fact that there are several review articles that examine various areas of false news detection, the majority only recap the methods employed in detection models. Additionally, they classify characteristics according to kind or organize datasets according to labels or domains. This evaluation stands out because it addresses the difficult issues that continue to influence models for detecting false news. It also talks about how future proposals for enhancing the procedures could perhaps boost detection accuracy. The datasets, extracted characteristics, and data fusion are three crucial and important parts of fake news detection research that are examined in this review. The effectiveness of fake news detection methods is impacted by these factors. Studies that dealt with the three issues mentioned above that were published in English between 2017 and 2023 in the academic databases and digital libraries Web of Science, ACM Digital Library, Springer Link, IEEE Explore, and Science Direct were covered. The following are the review's significant contributions:

- Give a general overview of fake news, including its various forms, effects of its dissemination, and function in social media.
- Draw attention to the key components of the false news detection model and their most significant weaknesses.
- Examine the techniques employed in a number of disciplines that have shown encouraging results and potential future suggestions for improvement.

This review article's remaining sections are organized as follows: A description of fake news is given in Section II. Section III examines the primary components of fake news detection methods. The primary detection model strategies and their difficulties are examined. The popular methods employed in detection models are covered in Section IV. Future directions for enhancing detection models are suggested in Section V. This review is concluded in Section VI.

II. FAKE NEWS OVERVIEW

One of the biggest threats to journalism, freedom of speech, and independence is fake news. Studies have shown that fake news on social media, particularly political news, receives more retweets and shares than real news [10]. The contentious "Brexit" vote and the divided 2016 U.S. presidential election both contributed to the decline in public confidence in governments [11]. The spread of false news received the most attention during the pivotal months of the 2016 U.S. presidential election campaign.

The Oxford Dictionary selected the phrase "fake news" as the international word of the year in 2016 [12]. Fake news, which is associated with the volatile stock market and significant transactions, has the potential to harm this country's economy. For instance, false information claimed that US President Barack Obama was hurt in an explosion that erased shares worth $130 billion [13]. The capacity of fake news to win over the public's trust is influenced by psychological and sociological factors. They aid in the spread of false information. It has been demonstrated that when authenticity and fabrication are distinguished, people become less rational and more vulnerable. They also have to deal with phony news at the same time. According to research using 1,000 participants in more than 100 social psychology and communication experiments, humans are slightly better at recognizing falsehood than is possible, with a mean precision of 54% and common precision degrees ranging from 55% to 58% [14]. Developing public trust in the news is simpler when truth and objectivity are expected. People may also trust false information if they have frequently been exposed to it (validity satisfying or if the information matches their want bias, conforms to established principles, confirmation bias, or opinions (selective exposure [15]). Peer pressure can sometimes "control" people's opinions and actions (for example, the bandwagon effect[16]).

A. Fake News Types

The term "fake news" describes news stories that have been released but purposefully mislead readers [7] for nefarious reasons [17]. According to the literature, there are various categories of fake news, as shown in Fig. 1. These include clickbait, rumor, misinformation, deception, and hoax [7]. A rumor is an unverified or unsubstantiated statement that spreads quickly [18]. Misinformation is faulty information that is mistakenly spread, whereas disinformation is intentionally false information that is released to confuse others [7]. Disinformation is what happens when a user posts incorrect information with bad intentions [19]. Misinformation is spread because consumers are ignorant of a certain subject or industry [7]. A hoax, which aims to purposefully and maliciously mislead the reader, falls under the genre of fake news. This involves cheating customers and pocketing cash [20]. Clickbait is one of the types of bogus news that attracts readers in by piqueing their curiosity about the attention-grabbing news title, claim psychological studies. Additionally, it prompts users to click [21]. To increase traffic to websites with advertisements, clickbait directs users to fictitious websites[20].

III. COMPONENTS OF THE FAKE NEWS DETECTION MODEL

This section reviews the crucial components that make up the false news detection model, influence its functionality, and relate to one another in order to do this mission. A dataset and a model built on top of a supervised classifier make up the two primary parts of the detection model.

A. Dataset Used in Fake News Detection Models

Studies on the detection of fake news have shown that there is presently no benchmark dataset that provides resources for extracting all essential elements. On social media, fake news spreads in a variety of temporal patterns compared to actual news. Any model's success depends on the dataset, which is the most important factor in fake news identification [20]. The dataset's larger, more varied, feature-rich and low-noise dimensions all contribute to the model's improved performance and increased accuracy in spotting false information [22].

B. Features Used in Fake News Detection Models

Linguistic and visual elements from news material are used as features in news identification [23]. People like and tend to believe articles with interesting images along with the text, according to psychologists. An article reaches more users when it includes multimedia components like photographs rather than only text. Therefore, while visual content plays a crucial role in spotting fake news, textual content is still necessary for news verification [24].

IV. THE PROMINENT TECHNIQUES USED IN DETECTION MODELS

A. Dataset Augmentation Techniques

1) Generative Adversarial Network (GAN): Deep learning-based generative models are called GANs. Their architecture is made up of a generator model for producing new instances and a discriminator model for detecting whether the instances created by the generator model are real or fake. Adversarial networks are frequently utilized to produce images that match observed samples. The generator model creates new images that mimic the original image using features derived from training data. Whether the created image is fake or real is predicted by the discriminator model. In detail, a vanilla GAN is made up of two networks that cooperate during training: Generator and Discriminator. Generator: This network produces images with the same structure as the training set of images when a vector of random values is presented as input. Discriminator: This network attempts to classify observations as "real" or "fake" based on batches of images that include observations from the training set and images created by the generator. The generator output is directly connected to the discriminator input. The generator utilizes the discriminator classification as a signal by using a backpropagation process to update its weights[25].
2) LeakGAN: Tragically, there are two problems using GAN, while GAN is designed to produce continuous, real-valued data, it has difficulty directly generating sequences of discrete tokens, such as sentences. Because of this, GAN starts with random sampling before switching to a deterministic transform with model parameter control. For a partially generated sequence, it is challenging to reconcile the current performance with the anticipated score for the entire sequence in the future [26]. GAN can only deliver the score/loss for a complete sequence after it has been formed. LeakGAN, which

addresses the issue of producing lengthy text [87], LeakGAN is a revolutionary algorithmic framework that addresses the non-informative and sparsity issues associated with earlier GAN versions. A novel strategy known as LeakGAN builds on recent advances in hierarchical reinforcement learning. It gives the generator more data from the discriminator. There is now a hierarchical generator G, which consists of a high-level MANAGER module and a low-level WORKER module. An LSTM named MANAGER mediates disputes. It obtains the high-level feature representation for each step from generator D, such as the CNN feature map, and uses it to produce the guiding aim for the WORKER module at that timestamp. It is not intended for D to provide G access to such information because it preserves it and engages in an aggressive game. As a result, it is referred to as a D information leak. The ultimate action in the current state is then taken by the WORKER by fusing the LSTM output with the goal embedding. This is based on the goal embedding that the MANAGER defined. This is achieved by first encoding the words that are now being created with another LSTM. As a result, LeakGAN experienced substantial performance improvements when producing larger sentences.

B. Features Extraction Models

1) Bidirectional Encoder Representations from Transformers (BERT):

To develop the relationships between words in a phrase, BERT was developed. BERT employs a language representation strategy that only makes use of the transformer's encoder section in conjunction with semi-supervised learning. BERT, in particular, is based on a multi-layer bidirectional transformer encoder that effectively collects data from a token's left and right contexts at each layer at the same time [28]. A masked language model (MLM) and a sentence-next predictor using BERT will be used in an unsupervised prediction operation to carry out the pre-training. In MLM, word prediction comes after context knowledge [29]. To create a sequence of sentences, the BERT model frequently takes inputs of individual tokenized sentences. The BERT model considers context from both perspectives. The transformer examines each word in relation to every other word in the sentence rather than processing each word independently. The self-attention mechanism in BERT also facilitates the identification of sentence keywords. The pre-trained BERT model could be successfully adjusted for improved performance in a variety of Natural Language Processing (NLP) applications, demonstrating how versatile BERT models are. Words and their subwords are the foundation of BERT's tokenizer. As a result, if a word is missing from the lexicon, it will be divided into a number of smaller tokens that, when put together, will form the original term. The remaining new tokens, those linked with more uncommon words, are simply separated into smaller units to guarantee that OOV tokens do not appear and that all vocabulary units are consistently updated and appropriately taught during training. [30].

2) VGG-19 Model:

The VGG16-Visual Geometry Group CNN architecture is an improved version of the AlexNet architecture. The implementation of an enhanced convolution neural network involves increasing the network's depth to 16 or 19 trainable layers. In computer vision, VGG networks are still preferred for many challenging situations [31]. The ImageNet dataset is used to learn the 143 million parameters of the deep architecture. RGB photos of 224 224 pixels are sent to VGG. The 19 trainable weight layers that make up the VGG-19 [32] start with five stacks of convolutional layers and end with three fully connected layers (FC). At each point in the process, these convolutional stacking layers take image information, perform the operation, and then send the outcome to the subsequent layer. All convolution layers use a 3 by 3 filter size, and the number of filters rises by a factor of two. The non-linear activation function is followed by a max-pooling layer and a layer based on a Rectified Linear Unit (ReLU) activation function. After ReLU, max-pooling layers are applied using a 2 by 2 kernel filter with 2 strides (pixels) in between each stack of convolution layers.

V. FUTURE DIRECTION

We list some of the important problems in this area that require attention based on this review. Additionally, we think there is a lot more space for advancement in fake news identification methods. The following are these problems:

1) Detection models continue to face substantial problems, such as underfitting, overfitting, and poor classification, which impair their performance. This is due to the use of either small or unbalanced datasets.
2) Despite their very detrimental impact, image-based elements have not been utilized frequently in prior studies to identify bogus news.
3) Despite being widely used to create new samples, vanilla GANs still have some serious flaws, such as vanishing gradients, mode collapse, and inability to converge.
4) Pre-trained word embedding models can effectively extract features, but they are unable to fully utilize the semantic and structural aspects of the text.
5) A number of machine-learning techniques have been employed to identify false news. Lower detection accuracy, though, was offered.
6) Simply concatenating the features will not reveal the true significance of many modalities. While integrating pertinent information throughout the various ways, it is necessary to maintain the distinctive characteristics of each approach (text and image).

VI. CONCLUSION

An overview of fake news, its varieties, and its effects is provided in this review article. It was also highlighted how social media sites contribute to the propagation of false information. The most important elements influencing false news detection methods were also emphasized. The dataset, features, and supervised learning classifiers are a few of these variables. Reviewing

the most effective strategies and tactics indicated the crucial constraints that still need to be resolved. In a number of domains, these techniques produced intriguing results that can be incorporated into false news detection algorithms. Additionally, these methods were examined, and some of the difficulties they encountered were detailed. Future academics would be able to enhance them and increase the accuracy of fake news identification as a result.

REFERENCES

[1] A systematic mapping on automatic classification of fake news in social media, De Souza, J.V., 2020.
[2] Detecting fake news over online social media via domain reputations and content understanding. Xu, K., et al,2019.
[3] Identifying fake news and fake users on Twitter, Atodiresei, C.-S., A. Tănăselea, and A.J.P.C.S,2018.
[4] False information detection in online content and its role in decision making: a systematic literature review. Habib, A.,2019.
[5] Fake News Detection Model on Social Media by Leveraging Sentiment Analysis of News Content and Emotion Analysis of Users‟ Comments. Hamed, S.K., M.J. Ab Aziz, M.R.J.S,2023.
[6] Fake news detection on social networks with artificial intelligence tools: systematic literature review. in International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions, Goksu, M. and N. Cavus,2019.
[7] Deep learning for misinformation detection on online social networks: a survey and new perspectives, Islam, M.R,2020.
[8] Integrating Machine Learning Techniques in Semantic Fake News Detection, Braşoveanu, A.M. and R.J.N.P.L. Andonie,2020.
[9] Detecting fake news in social media networks. Aldwairi, M. and A.J.P.C.S. Alwahedi,2018.
[10] The spread of true and false news online. Vosoughi, S., D. Roy, and S. Aral,2018.
[11] How to Stamp Out Fake News, Pogue, D,2017.
[12] Post-truth‟named 2016 word of the year by Oxford Dictionaries, Wang, A.B.J.W.P.,2016.
[13] Can fake news „impact the stock market?”, Rapoza, K.J.F.N,2017.
[14] On deception and deception detection: Content analysis of computer-mediated stated beliefs. Proceedings of the American Society for Information, Rubin, V.L,2010.
[15] Cognitive dissonance or credibility? A comparison of two theoretical explanations for selective exposure to partisan news, Metzger, M.J., E.H. Hartsell, and A.J. Flanagin,2020.
[16] Bandwagon, snob, and Veblen effects in the theory of consumers' demand,2020.
[17] Detecting breaking news rumors of emerging topics in social media, Alkhodair, S.A,2020.
[18] Utilizing computational trust to identify rumor spreaders on Twitter, Rath, B.,2018.
[19] Magdy, Your stance is exposed! analysing possible factors for stance detection on social media, Aldayel, A. and W.J.P,2019.
[20] Automating fake news detection system using multi-level voting model, Kaur, S., P. Kumar, and P.J.S.C. Kumaraguru,2020.
[21] Fake news early detection: A theory-driven model, Zhou, X,2020.
[22] Social rumor detection based on multilayer transformer encoding blocks, Lin, L. and Z. Chen,2021.
[23] A comprehensive review on fake news detection with deep learning. Mridha, M.F,2021.
[24] Combating the menace: A survey on characterization and detection of fake news from a data science perspective, Ansar, W. and S.J.I.J.o.I.M.D.I. Goswami,2021.
[25] A review of medical image data augmentation techniques for deep learning applications, Chlap, P.,2021.
[26] Seqgan: Sequence generative adversarial nets with policy gradient. in Proceedings of the AAAI conference on artificial intelligence, Yu, L., 2017.
[27] Long text generation via adversarial training with leaked information. in Proceedings of the AAAI conference on artificial intelligence, Guo, J.,2018
[28] Automatic fake news detection model based on bidirectional encoder representations from transformers, Jwa, H,2019.
[29] Almeida. Deep learning models for representing out-of-vocabulary words. Lochter, J.V., R.M. Silva, and T.A,2020.
[30] Fine-Tuning BERT Models for Intent Recognition Using a Frequency Cut-Off Strategy for Domain-Specific Vocabulary Extension, Fernández-Martínez, F,2022.
[31] ImageFake: An Ensemble Convolution Models Driven Approach for Image Based Fake News Detection, Choudhary, A. and A. Arora,2021.
[32] Very deep convolutional networks for large-scale image recognition. Simonyan, K. and A.J.a.p.a. Zisserman,2014.