Sentiment Analysis for Fake News Detection in Online Media Networks: A survey, fusion techniques and quality metrics

Mahmoud Ibrahim

Department of information systems, Zagazig University
Email: mahibrahim@zu.edu.eg

Abstract

The development of Online media sites in recent years has led to the spread of content sharing like commercial advertisements, political news, celebrity news, and so on. Various social media applications, such as Facebook, Instagram, and Twitter, have been impacted by fake news. Due to the easier access and rapid expansion of data through online media platforms, distinguishing between fake and real data has become difficult. The massive volume of news transmitted over online media portals makes manual verification impractical, which has prompted the development and deployment of automated methods for detecting fake news. Given the clear dangers of misleading and deception, fake news study has seen an increase in efforts that employ machine learning approaches, and sentiment analysis. In this study, we review the many implementations of sentiment analysis and machine learning methodologies in the fake news detection, as well as the most pressing difficulties and future research prospects.

Keywords: Fake News; Sentiment Analysis; Social Media; Fusion Technique

1. Introduction

Nowadays, online media portals (Facebook, Twitter, Snapchat, YouTube, and so on) are now the dominant source of information for people all around the world, particularly in developing nations. Furthermore, these online media portals are perfect ways for people can publish their feelings, stories, and concerns, as well as give greater benefit for getting quick response and feedback on various international problems, although some factually inaccurate thoughts may well be placed forward on aim. Furthermore, the low cost of sending and receiving data at online networks enables information to circulate more rapidly and substantially than before. Users need to pick the news, but they have to wait for a while, in this situation, people will turn to social media networks for obtaining breaking news in a matter of seconds. Because greatest of our survives are chiefly related online with public platforms, an increasing number of people are seeking out and consuming news from online media portals rather than conventional news organizations such as newspapers or television. Fake news and misleading information on online media can be less expensive, allow for discussion with friends or booklovers, and take less time than old-style organizations.

On the other hand, these online media networks must be trustworthy, and users have faith in them; the primary concern that undermines public's trust is trying to deal with disinformation. These social media systems must prevent fake news in their portals, even if it means losing users. Avoiding false propaganda in social media is not an easy problem, so we will follow some researchers who have done research at fake news detection on online media networks.

several misinformation or rumors may be produced and expanded throughout the web, going to lead other readers to trust and continue to spread them, resulting in a sequence of unintended lies. False news can lead to illusory beliefs and opinions, widespread panic, and other serious repercussions. To prevent such occurrences, particularly during political activities like elections, scholars have studied the flow of information and propagation on social medias in recent years, concentrating on disciplines such as

Doi: https://doi.org/10.54216/NIF.010205
Received: January 18, 2023 Accepted: March 15, 2023

opinion mining, user relationships, sentiment analysis, hateful spread, and so on, and taking an interest in evaluating methodologies for technical training to notice fake news, concentrating on the features of distinguishable approaches and methodologies, cognitive design ideas for identification, and so on. Fake news is definite as "any form of false, inaccurate, or misleading information intended to cause public harm or profit." It should be noted that fake news is a global issue that affects people worldwide. The widespread propagation of fake news can have major consequences for both people and communities. Fake news has the possible to upset the balance of authenticity in the news ecosystem. Furthermore, fake information intentionally leads viewers to receive prejudiced or erroneous opinions. Propagandists frequently employ false information to demonstrate political statements or fluency.

Sentiment Analysis (SA) is the field of Natural Language Processing (NLP) responsible for the building and development of approaches, methodologies, and methods to find out whether a text covers emotional or cognitive information and, if so, whether certain information is demonstrated positively, neutrally, or negatively, as well as how strong or weakly. Sentiment Analysis is also recognized as Opinion Mining (OM) because a large portion of the subjective topics represented by people on online media is about opinions (on review websites, discussion boards, online forums, chatrooms, and so on). The word of sentiment is significant in fake news. When there is information that people on online media find arousing but over which they have less control, they need to comment on articles. Readers, on the other hand, feel the need to share an article once they feel further in control [1]. Dickerson et al. [2] demonstrated that incorporating numerous sentiment indicators was enough for unique real accounts from virtual bot accounts. To keep increasing the propagation of news, headlines might catch the reader's curiosity and sentimentally relate them. It is not by chance that the propagation of fake information is repeatedly related to the existence of clickbait, in which content creators deliberately use the setups of sentimental polarization "Positive or Negative" and arousal "Strong or Weak" to mislead users [3], considering that a significant part of the false news viewers does not comprehend far beyond stories [4]. As a result, SA offers useful information on news article content to identify whether it is reliable or should be assumed fake news. In the case of detecting fake reviews, SA was thought to be a helpful tool not so much for detecting fake texts as it was for detecting fake negative reviewers, who overproduced negative feelings words when particularly in comparison to factual reviews due to overstatements of the sentiment they were attempting to demonstrate.

Human fact-checking is one solution to the issue of false news, which has seemed a major worry for both business and research. The essential nature of false news on online media, on the either side, tends to make detecting online false information much more difficult. Because of its inefficiency, expert fact-checking may be of little assistance. Furthermore, human fact-checking is time-consuming and expensive. The goal of fully automated fake news detection processes is significant for two reasons: initially, to minimize energy and time in detecting false data [5] and classify data along a veracity continuum with a related way of measuring of certainty, take into consideration that veracity is affected by deliberate deceits [6]. Consequently, fake news detection can be described as the procedure of determining whether a specific news piece through any issue from any scope is intentional or unintentional misrepresentative [7]. To automate this process, we must employ Machine Learning (ML) and Deep Neural Networks (DNN) models. Fake news can take the form of text, an image, a video, or audio. As a result, there is a need to conduct fake news analysis on various kinds of data through ML or DNN. ML and DNN have had a lot of success in various research fields. Through using these techniques, several more studies try to detect misleading information. The propagation of fake profiles aids in the distribution of spam emails, fake news, reports, and unlawful money demands. There is also cyberbullying, which can be harmful to people. All of these problems must be identified and investigated [8].

1.1 Related Review Studies

Various review studies have recently been established for the discovery, categorization, identification, and prevention of fake news. The concentrate of each review study differs in terms of its objective data, which can be text-based, vision-based, voice-based, or a combination of the three. Another survey looks at the various fake news dimensions that are associated with a certain language. As a result, the much more thorough and latest associated review studies are explored and reviewed in the following, and the key dimensions that differentiate this study from the other review studies are highlighted.

Doi: https://doi.org/10.54216/NIF.010205
Received: January 18, 2023 Accepted: March 15, 2023

Shu et al. [9] published a review of related research on social media misleading information detection through popular data mining methodologies to detect false information like feature extraction and model construction. They also introduced the categorization misinformation relying on psychology and behavioral ideas, current data mining techniques, assessment measures, and typical datasets in traditional organizations and social media platforms. They believed that SA should be used to determine post-based characteristics because users reveal their feelings or viewpoints about false information via online media news, like sceptical opinions or spectacular reactions. In terms of SA, they still assumed sentiment to be one of the features which may be obtained from content for false news identification., owing to the fact that differing sentiments between many posts' spreaders can sometimes reflect a high chance of fake news.

Bondielli and Marcelloni [10] discussed the characteristics deemed in fake news and rumor detection methodologies, obtainable a study of the different approaches often used to accomplish certain tasks, and illustrated how the gathering of related data for accomplishing them is troublesome. They believed that SA approaches could be utilizedd to extract one of the major significant semantic characteristics of false article stories.

Sharma et al. [11] discussed approaches for detecting fake news and that highlight computational tools for tackling these types of work, transcribed a wide selection data sets related to detect fake news, and discussed a range of difficulties and key challenges. They discovered that SA was a helpful context for detecting fake news to positive predictive words generally tend to be overstated in affirmative false reviews especially in comparison to their truthful equivalents, whereas online media responses to fake news seemed negative sentiment.

Da Silva et al. [12] discussed ML methods for fake news detection, discovering that the best models implicated neural networks constituted of traditional classification approaches that typically depend on lexicon-based approach of text input as key predictor features. SA was frequently utilized as a textual feature in the shape of tokens from sentiment lexicon approach or as the outcome of SA program relying on ML.

Zubiaga et al. [13] offered a review study on the detection and evaluation of rumours in online social networks by keeping in mind two kinds of rumours: i) long-standing rumors, ii) rumor type that spreads swiftly, such as important news, which is referred to as "emergent rumours". They discussed rumor techniques and tried to address four aspects of classification system of rumor. They examined the attempts and accomplishments done so far in the advancement of rumor classifying approaches.

Sunidhi and Kumar [14] provided a thorough comprehension of fake news definitions and foundations and its different kinds in the news spectrum, its features, few existing datasets and basic identifiers, and compiled a list of current approaches for detecting and identifying fake news. Lillie and Middelboe [15] discussed numerous ML techniques have investigated the activity of classing the opinion of a crowd towards a rumour, with positive results in classifying stance and very valuable outcomes.

Zhang and Ghorbani [16] constructed a review study of online misinformation characterization throughout aspects of misinformation providers and objective audience, news information, and social context. They also investigated existing datasets for identifying fake news. They described each news item as comprising of physical and non-physical data, in which physical data are the bearers and shape of the news and non-physical data are the news creators' viewpoints, feelings, behaviours, and emotions. As little more than a result, they concluded that SA is a helpful technique for illustrating the feelings, behaviours, and viewpoints transmitted by online media networks, and that sentiment aspects are important parameters for identifying suspect accounts.

Meel and Vishwakarma [17] constructed a high-level overview of cutting-edge technology solutions, methods, benchmark data, and empirical outlines for online social content and misinformation analysis. They also introduced a taxonomy for categorising misinformation and debated the influence on community of fake, enthusiasm for publicizing misleading information, user discernment, and cutting-edge approaches for rumor diffusion fact-checking and categorization, ML, and DL methodologies. They deemed SA to be among the most important origin of information for detecting misinformation.

Zhou and Zafarani [18] investigated and surveyed approaches for detecting misleading information from four different aspects "knowledge, style, propagation, and source". They saw sentiment as a significant

Doi: https://doi.org/10.54216/NIF.010205
Received: January 18, 2023 Accepted: March 15, 2023

semantic-stage characteristic of textual data, which argued that the development of effective and understandable misinformation detection techniques necessitates the collaboration of specialists in computer, social, political science, and news reporting. Antonakaki et al. [19] constructed an overview study on recent research topic areas in Twitter, deciding that SA was among of the four major aspects of research including Twitter and that the pervasiveness of fake news through it is among the serious risks to online media networks. However, they examined both concepts separately, without drawing any conclusions about the utility of SA in fake news detection.

Guo et al. [20] conducted a systematic examination of the multidisciplinary concept of social dishonesty and categories of online social deception attacks, as well as their distinguishing characteristics in comparison to other social outbreaks and cybercrimes They also examined various defense techniques for the inhibition, recognition, and reaction mitigation of online social deception attacks, as well as the relevant benefits and drawbacks, as well as the legitimate and moral concerns associated with that field.

Medeiros and Braga [21] provided a systematic literature review providing a review study of this research field, and an analysis of high-quality studies on fake news detection, which were then classified based on their type of contribution and model According to the results of this survey, Twitter and Weibo are the most common broadcasting platforms used by the collected articles, with Long Short-Term Memory (LSTM) providing the finest consequences. Hoy and Koulouri [22] presented a systematic literature review study to find what existing techniques for detecting misleading information exist and their effectiveness. They proposed combining manual and automated searches across a wide textual search to collect as several studies as possible associated to the research questions, as well as a set of inclusion criteria and quality assessment of the papers collected. Data extraction was performed to obtain various ML approaches, datasets, feature extraction techniques, and performance measures, as well as narrative analysis and discussion of the work. Providel and Mendoza [23] introduced a comprehensive literature review that conveys the research work described to discuss False information propagation over online media, with a central objective on the Spanish language. They also collaborate to recognize pending responsibilities for this community and difficulties that necessitate collaboration among predominant researchers on the subject. Oliveira et al. [24] emphasize the review of fake news detection in the framework of NLP by categorising the conventional procedures for recognizing fake news, provoking the primary data, and employing features to describe misinformation. They also examined the chief vectorization methods and tools for converting natural language information into methodically feasible data.

Rohera et al. [25] conducted a thorough assessment of current false news detecting methods. They also trained various ML models on the self-aggregated fake news dataset. Stitini et al. [26] offered a multiclass, semi-supervised framework for enhancing the precision of trust-aware recommendation systems.

1.2 Motivations and Contributions

Online media sites are incredibly fast data generation and broadcasting portals, with millions, if not billions, of people chatting on a variety of online portals and channels per second, creating massive amounts of data. However, unlike conventional news places (like newspapers and news portals), the integrity of data shared via online media is called into doubt due to the emancipation of free expression. Recently, there was a significant growth in the persons existing on online media seeking for or posting news and knowledge. The public information of online media has a significant impact on users' preferred decisions. Fake news can have serious consequences, whether intentionally or unintentionally associated to the financial problems or psychological health. It is widely used for a variety of goals like political parties propagating misleading information to own an edge in elections which make unfairness in election processes. As a result, developing ways to tackle the problem of false news transmission became critical. So, this work reviews and analyses the use of sentiment analysis and various ML approaches used to address the issues caused by fake news by seeing the current state of info pollution in numerous platforms for data publishing and generation, smart analysis, and fact-checking schemes. This work concentrates on fake news detection in online media platforms with many study accomplishments from sentiment analysis and ML perspectives, deliberating the architecture, enhancements, challenges, and chances for trustworthy social networks to provide a thorough and systematic interpretation.

The below are the study's significant contributions.

Doi: https://doi.org/10.54216/NIF.010205
Received: January 18, 2023 Accepted: March 15, 2023

- Address of the key role of the Sentiment Analysis, Text Analytics, and NLP approaches in the social media networks and fake news detection.
- This work identified, classified, and described the false info thoughts in social networks to grasp the fundamental notions of false info and its necessary elements. We also talk about the numerous sources of fake news, various types of data at online media, and fake news detection stages.
- This review suggests a complete overview of the current literature to explain the role of ML algorithms and DL models to detect fake news, rumors, and other forms of online false information.
- Finally, we discuss the research tasks and future work for the state-of-the-art in identifying and detecting fake news on social media networks.

1.3 Paper Organization

In this study we will look at the many ways that have been utilized to include sentiment and ML into the detection of fake news. We also cover Sentiment Analysis in Section 2 as a significant tool that has been employed well in the broad field of text analytics and NLP. Section 3 begins with a definition of fake news and what it means, as well as the implications of its spread in today's society. Section 4 is devoted to systems that employ Sentiment Analysis to detect false news, including both those in which Sentiment

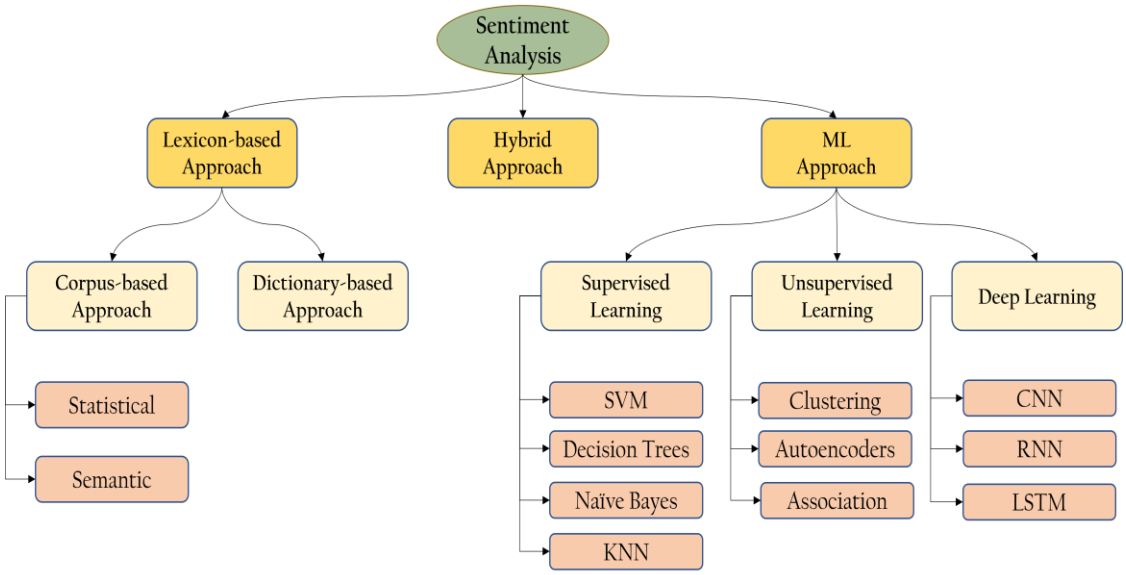

Figure 2: Most of the methodologies used in sentiment analysis disciplines are categorized.

Analysis is the foundation of the system and those in which the findings of a Sentiment Analysis system are used as a feature in ML approaches. Then in section 5, we discuss the most common ML techniques and studies used in detecting fake news. Section 6 continues with a discussion, difficulties, and future directions. Finally, we end with the conclusion in section 7. FIGURE 1 presents the structure of the paper.

2. Sentiment Analysis as a Critical Dimension in NLP

Subjective statements are usually defined in NLP and text analytics as pieces of natural language expressions that expose viewpoints, emotions, perspectives, and stances on a specified area of interest. Automatically trying to analyze such expressions and comprehension the sentiment presented there is known as SA, so it is significant not just from a research standpoint but also from an industrial perspective. It is thus able to process vast volumes of data to monitor societal attitudes toward public issues, activities, or products. Sentiment Analysis (opinion mining) is an aspect of NLP that involves the

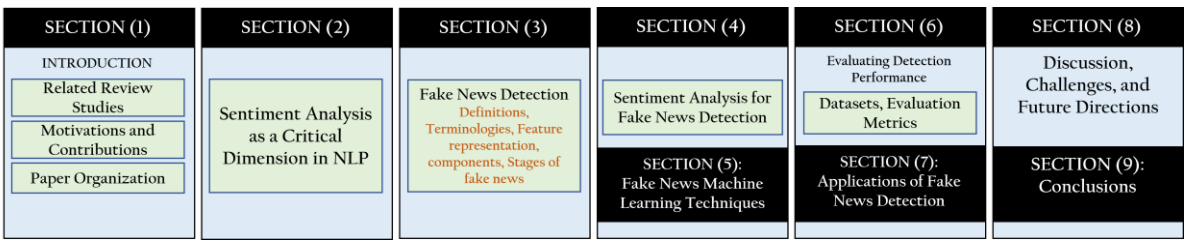

Figure 1: Structure of the Survey

extraction of users' expressed feelings and emotions from a particular text. This research domain has grown quickly given the rise of data produced by online users, that impacts a range of areas like medical research, industry, politics, and much more. Along with this wealth of data, several approaches were generated to automatically discover and detect stated viewpoints. Figure 2 depicts a hierarchical representation of some cutting-edge techniques to sentiment analysis.

Sentiment is typically a factor with values such as "Positive", "Negative", and "Neutral", or even more certain values such as "Happy", and "Angry". Each factor has a several range of values, enabling for numerous opinion projects in one unique word. This allows a single token can have both "positive" and "negative" connotations. Furthermore, relying on the sentiment values, we can produce extra meta-features. These are referred to as "Subjectivity" and "Polarity", which the ratio of "Positive" and "Negative" posts to "Neutral" posts is used to calculate "Subjectivity". "Polarity" is refer as the ratio of "Positive" to "Negative" posts. Which SA can evaluate a specific user's "Positiveness" as "Personality Trait" or monitor a community's attitude toward a particular issue.

The regular sentiment analysis methodology involves the pre-processing and lexical features extraction from posts. Tokenization, acronym expansion, and cleanup of stop-words as well as other factors without linguistic value, such as URLs and mentions, are examples of preprocessing stage that can have a massive impact on approach performance. This step is a hot topic in NLP and is popularly known to as "Text Normalization", which the tweets and posts include several "Out Of Vocabulary" words, and approaches involve obtaining data from lexicon-based approach for acronym extension, using spell checking models for token correction, using machine translation for phrase normalization, and using Word Vectors and embeddings to measure similarity [27]. Sarawagi et al. [28] presented lexicon-based tool, which the approach involved automatically producing semantic-based lexicons (dictionary) that navigated tokens to relativism ranks (score). The sum of token scores could then be used to measure the total sentiment of a given set of inputs. Vilares et al. [29] introduced lexical rule-based system "SentiStrength" that concentrates on short texts and simply compatible to variety of languages. Cambria et al. [30] created semantic graphs that link notions to obtain their semantics for the intriguing issue of text polarity detection.

Subjective analysis can be handled through ML approaches, a direction that has become increasingly prominent since the widespread deployment of DL in NLP. The ML approach applicable to SA which mostly related to supervised classification in general and text classification approaches in particular. To classify the reviews, a variety of ML approaches were used. Text classification has seen great results with ML approaches, and the N-gram models are some of the other well-known ML methods in NLP.

SA tools, schemes, and approaches have already been deployed effectively in the case of textual analysis; in solutions like the analysis of demand and supply review sites [29]; online media articles [31], Facebook posts [32], Instagram [33], and other platforms; detecting social spam for preventing normal people from being unjustly overwhelmed with undesired or false content through the use of social network [34]; political [35], community [36], and industrial analysis [37]; Cybersecurity [38]; and healthcare [39].

3. Fake News Detection

The concept "Fake News" refers to misinformation propagate through traditional media portals, particularly online media, and web portals [10]. Also, "Fake News" has been known widely as "news piece that is purposely and verifiably untrue", as well as data shared as a content story that is completely false and intended to deceive readers into believing it is correct. The principle of fake news dates back to the 15th century. There are numerous sources of fake news, including radio, newspapers, television broadcasting, and diverse media sites. Previously, human-to-human communication was the major important origin of propagating misinformation. These days, online media bots take a critical part in the spread of false information. This concept takes into account recent advancements in fake news detection, particularly following the 2016 U.S. presidential election [40]. Fake news is already viewed as among the most severe risks to the public, democratic, and media. At the America's presidential elections in 2016, several more occurrences were noted that propagate false news via reputable online portals. Millions of emotions, reviews, and shares produced by fake news articles on the internet were produced by election stories published by major news portals [41]. The propagation of fake news has cast doubt on the credibility of published news content on online media portals. It should also be acknowledged that

Doi: https://doi.org/10.54216/NIF.010205
Received: January 18, 2023 Accepted: March 15, 2023

online media portals contribute a key part in the spread of misleading information across users all over the world.

TABLE 2
KEY COMPONENTS OF FAKE NEWS

Term	Description
The Originators	False information on the internet may be distributed by both actual people and non-humans. It's possible for innocent authors and viewers to accidentally spread inaccurate information, as well as malicious individuals who deliberately spread misinformation. The most prevalent non-human originators are social bots and cyborgs.
Victims	Fake news on the internet mostly targets people named "Target Victims." They might be social media users or users of other online media platforms. Depending on the purpose of the news, its audience may include teens, voters, family members, the elderly, and so on.
News Content	The information that makes up the bulk of a news story is considered to as the news's content. It consists of tangibly present elements such as the author, headline, message body, and picture or video, as well as intangibly present elements such as the themes, sentiments, and objectives of the communication.
Social Context	social context analysis involves two forms of news dissemination on the Internet: In the first case, we looked at how the news gets spread between internet users, and in the second, we looked at how it is aired over time.

TABLE 3
TYPES OF DATA IN NEWS

Term	Description
Text	Text linguistics is a subfield of linguistics that focuses on the use of written language in the context of communication. There is more to it than simply a sentence and some tokens; it also includes the tonality, grammatical structures, and semantics that are necessary for analysis of conversation.
Multimedia	Media amalgamation is exactly what it sounds like. Graphics, audio and video are all part of the package. That's a great way to get the audience's attention right away.
Hyperlinks or Embedded Content	Using hypertext connections, writers may connect to a wide range of information sources and gain the trust of their readers by stating the hypothesis of their news articles. Because of the popularity of online media such as social networking sites like Facebook, Twitter, YouTube, and Soundcloud, writers are increasingly using screenshots of relevant posts from these sites in their work.
Audio	utilizing audio as a medium for reporting the news is possible despite the fact that it is a component of the category known as multimedia. This component includes radio services, broadcast networks, and podcasts; the dissemination of information via this medium reaches a greater number of individuals.
Bot	like posting or reposting other users messages. Social bots are frequently used to generate fake reviews automatically.
Information Manipulation	In order to influence public opinion or distort the facts, it is common practice to spread false information discreetly, such as by "planting a rumour." Disinformation distributed by malicious people is generally used for financial or political benefit by malicious users.
Deceptive Online Comments or Fake Reviews	False comments, thoughts, and reviews are often posted on social media by dishonest publishers to mislead consumers. An opinion-based misleading information is often regarded as a fake review.
Conspiracy Theories	includes attempts to explain certain phenomena that are factually unverified (or unverifiable). This type of text, content, belief, or discourse typically takes the form of narratives alleging secret plans by powerful entities, to harm or destroy a segment of the population.

The study of miss information and false facts encompasses a variety of notions that commonly intersect. We distinguish such notions by offering community-accepted definitions as in TABLE 1:

Misinformation is widely distributed on online media sites and can income the form of tweets, comments, posts, blogs, photos, chat, stories, or feisty news. It is most often referred to as information pollution and manifests itself in a variety of preparations that are not contradictory but exhibit some heterogeneity that expresses them below an unmistakable community. To better grasp the scale and diversification of misleading information on the internet, several major elements for categorizing fake news should be understood, which TABLE 2 represent that, and TABLE 3 represent the numerous kinds of data that make up media articles.

Doi: https://doi.org/10.54216/NIF.010205
Received: January 18, 2023 Accepted: March 15, 2023

TABLE 5
STAGE OF FAKE NEWS DETECTION APPROACHES

Features Type	Term	Description
News Content	**Knowledge-based**	Knowledge-based techniques rely on third-party sources to substantiate statements made in news reports. Assigning a truth value in the context of an argument is the goal of fact-checking. Expert-oriented, crowdsourcing-oriented, and computational-oriented methods of fact-checking are all now in use.
	Style-based	False news is created by many people who are not journalists, and this is why style-based techniques to identifying fake news aim to detect tricksters in news content writing style. Deception-oriented methodology and objectivity-oriented methodologies are two of the most common style-based approaches.
Social Context	Stance-based	Stance-based type is similar to other style-based approaches, but its primary focus is on the development of claims inside a news story. Real news articles are structured so that readers have all the facts they need to develop their own opinions about what happened. By design, stance-based publications provide little substantiation for their bold assertions (fake arguments).
	Propagation-based	the ability to establish a propagation process via either a homogeneous or a heterogeneous reputation network; also known as structure-based characteristics that identify false news by comparing the flow of information through info cascades utilising the social network's propagation structure.

The techniques used to detect fake news are mainly concerned with news content, social context, and structure. Different kinds of feature representations can be constructed to obtain discriminatory features and characteristics of fake news as described in TABLE 4. Also, TABLE 5 represent the stages of fake news detection approaches.

So, researchers looking for automatic false news categorization and intelligent approaches with diverse accuracies. So, the process of the classification of false news is as follows: Deciding available datasets or collecting news stories from online portals, where FIGURE 3 depicts a graphical representation of the collection of data from the online media portals. When the data has been collected, it requires some pre-processing techniques for data such as removing stop_words, stem, and tokenization, which involves eliminating noisy, incorrect entries, and anomalies to organize and improve dataset integrity, which helps to improve approach predictive performance. then feature extraction step, which neglects unneeded arguments and extraneous features out of the data in order to enhance performance and improve the prediction model's efficacy. At last, the ML or DL approach is applied as a last layer, determining whether the text is FALSE or TRUE.

TABLE 4
FEATURE REPRESENTATION OF FAKE NEWS

Features Type	Term	Description
News Content	**Linguistic-based**	create linguistic features that can be derived from textual material at several levels of document structure and are often used to identify false news, such as certain writing styles, sensational sentiments, and headlines. characters, token levels, and syntactic features (phrases, and documents levels) Attempting to sum together the many distinguishing features of disinformation.
	Visual-based	uses visual analysis extensively since information and features are extracted from images. consisting of a combination of still images and moving ones.
Social Context	User-based	user-based attributes are gathered from their profiles to acquire both global and local adjustments for user profile characteristics throughout propagation routes to detect false news, and these features are targeted towards a particular audience for fake profiles specific on their age ranges, gender, religion, and so on.
	Post-based	For the identification of probable false news based on wide public interactions, social networking sites were the primary focus of the investigation It is important that posts contain the following qualities: attitude (the users' views on current events), subject, and authority (degree of reliability). Whether it's a photo or video, a tweet, a meme, or anything else, a post is a kind of content.
	Network-based	Groups of friends on Facebook and groups of mutually linked persons on LinkedIn are all examples of this idea being applied to groups of people. There are three ways to construct network-based features: diffusion networks, interactions, and propagation networks.

Doi: https://doi.org/10.54216/NIF.010205
Received: January 18, 2023 Accepted: March 15, 2023

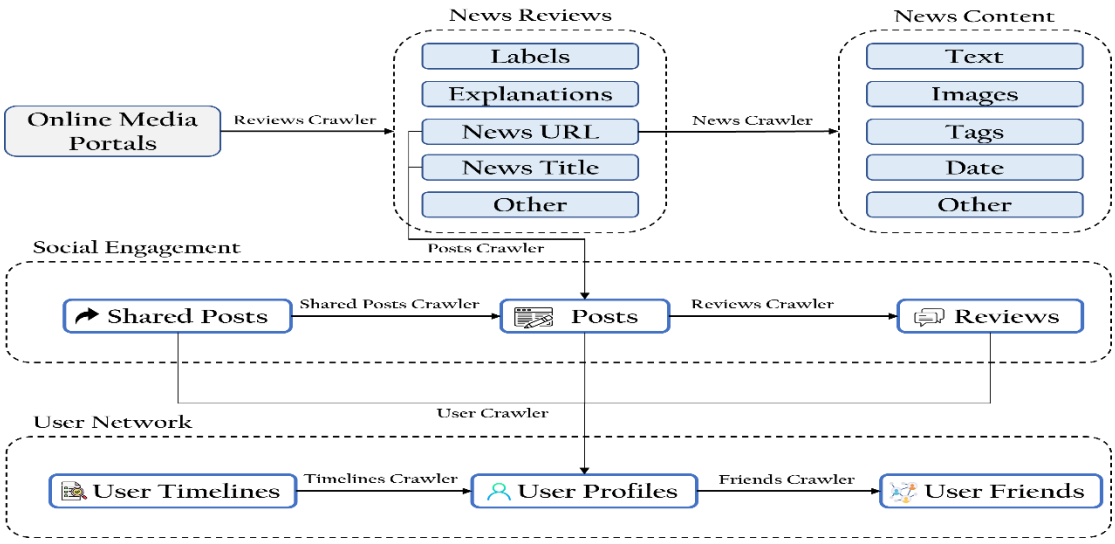

Figure 3: Procedure of the collection pipeline

4. Sentiment Analysis for Fake News Detection

The problem will be identified for this literature review by establishing SA that can be utilised to develop an approach that can recognise whether information is true or false, particularly with this current inaccuracy. Because SA simply uses the idea of computers to analyse Natural Language in order to discover and retrieve features about the creators' emotional mood, i.e., as to if content is False or True. The study includes text mining as it corresponds to SA and variety of techniques to identifying intention of the user in writing meanings for false news identification. It is the process of trying to obtain emotions and feelings from text or publisher stances. Because publishers of misleading information are more concerned with impressing the readership and propagating the information quickly, the sentiment in real and misrepresented information can vary. As a result, misleading information usually includes either extreme emotions that could quickly reflect with the community, or controversial text intended to elicit strong feeling in receivers. As a result, sentiment analysis can be used to identify misinformation in both content and user reviews. Table 6 present most studies in fake news detection using SA. Guo et al. [42]

TABLE 6
MOST STUDIES IN FAKE NEWS DETECTION SYSTEMS USING SA

Reference	Media Platform & Dataset	Language	SA Method	Detection Approach
AlRubaian et al. [43]	Twitter	Arabic	Lexicon-based	NB
Popat et al. [124] [125]	Snopes and Wikipedia	English	Lexicon-based	LR, CRF
Horne and Adah [4]	Buzzfeed and Political News (Facebook)	English	Lexicon-based	SVM
Rashkin et al. [126]	Fact checking	English	Lexicon-based	LSTM
Dey et al. [44]	Twitter	English	Lexicon-based	KNN
Bhutani et al. [45]	LIAR (Politifact, Facebook, Twitter)	English	NB	RF
Cui et al. [127]	PolitiFact and GossipCop	English	Rule-based	DNN
Vicario et al. [128]	Facebook	Italian	Dandelion API	LR, DT, and KNN
Reis et al. [129]	Buzzface	English	Rule-based	NB, SVM, KNN, RF, XGBoost
Anoop et al. [130]	HWB	English	Lexicon-based	NB, SVM, KNN, RF, AdaBoost, CNN, LSTM
Zhang et al. [46]	Weibo-20	Chinese	Lexicon-based	BiGRU, BERT, NileTMRG

presented a feelings Misinformation Identification system for learning content- and statement mappings for producers and viewers, respectfully, in order to utilise both information and social feelings for misleading identification. AlRubaian et al. [43] utilized sentiment analysis to detect unrealistic Arabic scripts in twitter in way to stop the spread of misleading and deceptive information, because they believed that sentiment provides a measurement of user behavior, resulting in high accuracy in credibility analyses, which sentiment was graded at 30% for all of the features that were deemed. Dey et al. [44] used many NLP techniques (portion tags, named-entity recognition, and SA) on 200 twitter posts about the 2016 U.S. National Election. They discovered that real twitter posts tended to be positive or neutral sentiment, whereas false posts tended to be negative. Bhutani et al. [45] built their false information detection on SA, assuming that the sentiment expressed while ability to write a news piece would be a critical deciding element in classifying the news posts as false or true. They utilized a NB model to find text emotion and afterwards used it as a key focus of Multinomial NB and RF models for false information detection, with the last one producing the better outcomes.

5. Fake News Machine Learning Techniques

There are a variety of techniques employing diverse ML and DL methodologies for misinformation classification. Firstly, Supervised learning is commonly used to solve regression and classification issues, in which the model is trained and learned from labeled data with each input data matched to a fixed outcome. It only works well if a prespecified dataset is utilized to train and construct the approach. By leveraging input features, regression challenges can be fixed through forecasting factual or continuous values. In conversely, classification separates each data input relying on its labels [47]. There are diverse supervised learning approaches for fake news classification which we will discuss like SVM, CNN, NB, DT, RF, and LR.

The SVM model is preferred for complicated tasks such as classifying fake news. SVM is a discriminative classifier is known to be the great text classification methodology. The approach presented in most of work was assessed on the available datasets and the SVM approach gave greater outcomes than other classifiers like in [48] and [49]. Prasetijo [50] examined SVM approach effectiveness to detecting false information relying on textual analytics, where SVM is a powerful technique approach for binary classification. Deokate [51] presented an SVM approach to detecting misinformation on online media portals, particularly Twitter. It accomplishes reliable text preprocessing on twitter posts by transforming slang in the article to standard forms. Furthermore, regular expressions returned the original word to words with redundant letters. The n-grams tool was then used to separate the posts. Their feature extraction step was included structural, user, and content features about the article. They also looked at the publisher's profile to determine whether the article was false or factual. They tested their introduced method on the BuzzFeed dataset and got the best results.

Among the other supervised learning architectures, Convolutional Neural Network (CNN) CNN is the major commonly utilized. For many years, CNN has been crucial in identifying misinformation. Although CNN is primarily deployed in Computer Vision applications like image processing and object detection, they also accomplish well in several NLP applications. The convolutional method enable the neural network to generate local features over each token of the adjacent token and then merge them through max function to generate a fixed sized word-level embedding [52], where one-dimensional CNN (Conv1D) is commonly used in text classification or NLP and Conv1D is concerned with one-dimensional arrays that indicate word vectors. Yang et al. [52] conducted a new TI-CNN technique to detecting misinformation that combined textual and pictorial information with apparent and latent features. They used a dataset from Kaggle that was concentrated on data about the United States election. The hybrid deep learning models achieve better detection results like [53] that mix convolutional and recurrent neural networks to identifying false information. Deligiannis et al. [54] a new technique of Graph Convolutional Neural Network (GCNN) introduced to tackle the user geolocation issue for classifying fake news through associated events and event producers.

Decision Trees (DT) employ a hierarchical chart to split a dataset into smaller groups, which is a helpful strategy for tasks such as classification. Branching is determined by the outcomes of tests performed on each attribute specified in the DT's nodes. The leaf node has a class label once all attributes have been computed. The root-to-leaf distance is used as a criterion for categorization [55]. Hakak et al. [56] introduced a model that extracts significant characteristics from fake news datasets and then classifies the collected features using an ensemble model built of three standard ML techniques. It is well-known

Doi: https://doi.org/10.54216/NIF.010205
Received: January 18, 2023 Accepted: March 15, 2023

that the Random Forest (RF) is a classifier that is trained by creating a large number of decision trees. The bagging strategy, which allows the model to be trained again with the same dataset but with randomly selected features, is made possible by the RF during training. RF is a mixture of decision trees. A random subset of the training dataset will be generated for each tree. DT models break up the whole set of data into discrete nodes, each of which has its own random set of variables.

When estimating a categorical value, a classifier like Logistic Regression (LR) might be utilized. For example, it may tell you if a prediction is correct or not [57]. Tacchini et al. [58] used the Facebook platform to create a logistic regression classifier to distinguish between "Hoaxes" and "Non-Hoaxes." Using publicly accessible Facebook articles from July 2016 to December, we honed our skills using just that data. When developing a model for the false news classifier, Ogdol et al. [59] presented an LR technique that includes sentiment neutrality, page rank, and the content length to structure error ratio as independent variables for each data set.

Simply said, Naïve Bayes (NB) classifies data in an efficient and effective manner. Probabilistic approaches are used in text categorization using the Bayes theorem. When it comes to input and output data, they are concerned with the probability distribution and forecasting of the response variable values. The NB classifier has the capacity to operate with less training data in order to get the desired classification parameters. The three most prevalent naive bayes techniques are Gaussian, Multinomial, and Bernoulli. Gaussian NB is used when the features have a continuous value and are expected to have a Gaussian distribution. To solve document classification issues, multinomial NB is usually used. This includes guessing which category a particular document belongs to, such as politics or industry. The number of tokens in a text may be used as a classifier in this scenario. The only difference between Bernoulli NB and multinomial naive Bayes is that features seem to be Boolean variables describing inputs. It has been proved that the use of the NB model on Facebook postings may be used as a simple way to identify false news. According to Yuslee and Abdullah [60], NB with n-gram improves the accuracy of TF-IDF and Count Vectorizer when used as a classifier in a false news detection model.

To use the k-Nearest Neighbors (k-NN) technique, you must first choose a parameter (k) that determines how many samples from your closest neighbours will be used. Analyzing samples that haven't yet been categorised and all of the samples that have previously been categorised is done by measuring the distance or similarity between them [24]. Casillo et al. [61] news profiling is enabled by an approach that uses the syntactic and semantic characteristics of news to classify true and false news using the Syntactic Analyzer, Sentiment Analyzer, and Topic Analyzer. Then K-NN classifier applied on the processed data. Mladenova et al. [62] examines the detection of fake news and click-bait headlines from Bulgarian Facebook Pages through using KNN classifier and 4 diverse distance measures are evaluated and analysed. Weakly Supervised Learning is a form of machine learning in which noisy, limited, or imprecise sources are used to provide supervision signals for labelling large-scale training data in a supervised learning setting [63]. These datasets might be expensive or impossible to acquire manually and our solution alleviates that burden. A better approach would be to employ low-cost, imprecise labels to build a strong prediction model rather of a more costly, more robust one. There is a possibility of introducing both false positives and false negatives as a result of the dissemination of mixed news by unreliable sources (fake news propagation by trusted sources, e.g., by accident) [64][65][66].

Autoencoders and cluster analysis methods like K-means, K-medoids, and fuzzy C-means models are examples of unsupervised ML approaches. DNNs may also be used for unsupervised learning. A few of unsupervised efforts to identify false news have been attempted [67]. False news may take many forms, and creating an unsupervised technique is necessary because of the wide variety of forms it might take. A model built primarily on the news's content and based on text analysis may not be applicable to all fields. Li et al. [68] presented an autoencoder-based methodology for solving the issue of unsupervised fake news detection. Gangireddy et al. [69] focus on the issue of unsupervised false news detection without labeled data on online media portals, where the approach draws on graph-based methodologies like graph-based features vector learning and label propagation and identifying bicliques. [70] introduced a simple unsupervised method for identifying Twitter profiles that spread propaganda, which cluster twitter posts at each timeframe to obtain a set of publishers who published similar information using K-means.

Semi-supervised learning attempts to characterize some data given a set of labelled and unlabeled examples [71]. In general, semi-supervised approaches may be broken down into the following classes:

Doi: https://doi.org/10.54216/NIF.010205
Received: January 18, 2023 Accepted: March 15, 2023

Both the standard support vector machine (SVM) method and the graph-based method, which treats known tags as vertices and edges and unlabeled data as arcs, are viable options. Hence, the semi-supervised approach utilises both tagged and pseudo-labeled data to train a genetic algorithm, like a fuzzy SVM or DNN. Dong et al. [72] presented deep two-path semi-supervised learning approach, where first path is for supervised and the second is for unsupervised learning, and the two paths are jointly deployed with CNN on twitter datasets to improve detection efficiency. Benamira et al. [73] presented a semi-supervised false news identification approach using graph neural networks, which concentrate on content-based techniques for identifying false news for reducing the issue to a binary classification (article is fake or real). Also, there are other approaches on semi-supervised learning like [74][75].

Reinforcement learning has previously proven to be incredibly effective in boosting the performance of NLP approaches. So, Li et al. [76] introduced an inverse reinforcement learning technique for

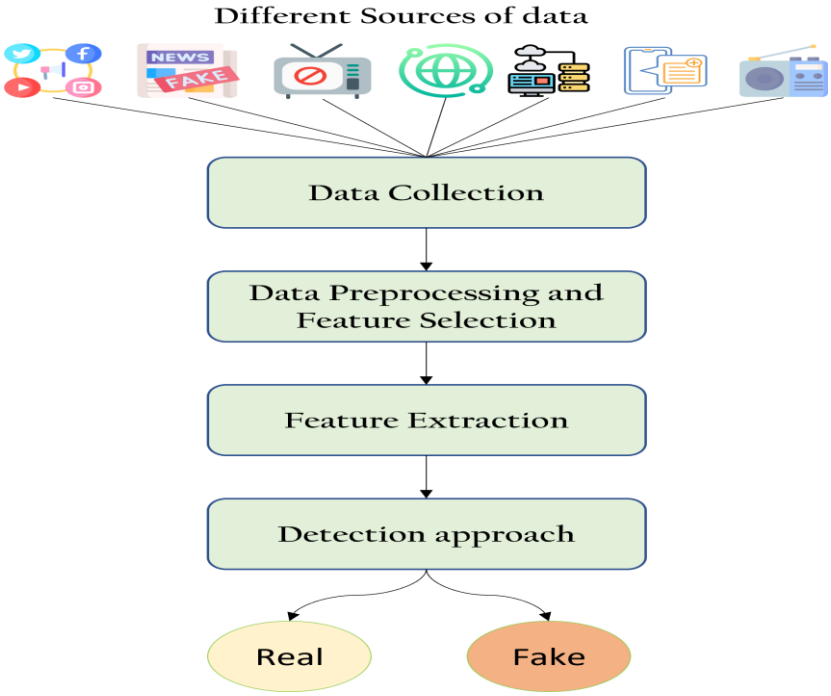

Figure 4: Procedure of fake news detection phases

developing paraphrases. Fedus et al. [77] employ RL to fine-tune the LSTM-based generative adversarial network's parameters for text production. Also, Cheng et al. propose used RL to tune a classifier's parameters for trying to remove different kinds of biases [78]. Mosallanezhad et al. [79] developed a domain adaptation approach with Reinforcement learning agent that adapts the representations of the article depending on feedback from both the adversarial domain model and the misinformation detection element. Meirom et al. [80] demonstrates how merging RL with GNNs gives an effective method for regulating diffusive processes on graphs. Figure 4 depicts the graphical representation of the procedure of fake news detection phases.

6. Evaluating Detection Performance

In this part, we'll look at how to evaluate the approaches for detecting fake news. our attention is drawn to the current existing datasets and evaluation measures.

A. Datasets

The lack of readily existing data is cited as the biggest obstacle academics face in creating new ways for detecting false news, according to the study. Developing a good supervised learning model necessitates having a dataset that can be trusted. It is possible to get online news from a variety of sources, including news agency homepages and search engines. However, manually verifying the truthfulness of news is a difficult operation that normally requires annotators with subject knowledge who do meticulous

Doi: https://doi.org/10.54216/NIF.010205
Received: January 18, 2023 Accepted: March 15, 2023

examination of assertions and supplementary evidence, context, and reporting from authoritative sources, such as reputable news organizations. In general, the following methods may be used to obtain news data with annotations: Detectors of industry and crowdsourced workers, as well as seasoned journalists and fact-checking sites. Fake news identification is a difficult issue since there are no agreed-upon benchmark datasets. In order to evaluate an approach's performance for the sake of getting the work done, some kind of dataset is necessary. To the same extent, validating its findings requires collecting data on both genuine and fabricated news. Listed below are publicly existing datasets.

- **CREDBANK:** Mitra and Gilbert [81] presented his dataset is comprised of streaming twitter posts collected through (October 2014 - February 2015), which contains over than 60 million twitter posts covering 1049 actual facts, and the veracity of the twitter posts is assessed by 30 annotators, and annotated with credibility scores. Their labels about five classes (Certainly Inaccurate; Probably Inaccurate; Uncertain (Doubtful); Probably Accurate; Certainly Accurate).
- **PHEME:** This dataset present Twitter posts were gathered during the 2014 Ferguson unrest in the United States Which the samples about 330 rumours conversations (159 are true, 68 are false and 103 remained unverified), where 297 in English and 33 in Germany with three classes (true, false, or unverified), and the results of the annotation task were presented, analyzed, and discussed in Zubiaga et al. [82].
- **Emergent:** Ferreira and Vlachos [83] introduced an Emergent dataset that serves as a legitimate dataset for a wide range of NLP tasks in the regard of fact-checking. Its size about 300 claims, and 2,595 associated article headlines (47.7 % true, 15.2% false, 37.1 unverified) with three classes (true, false or unverified). Which it could be developed for stance classification.
- **LIAR:** Wang et al. [84] proposes and publishes this data for detecting online fake news. It includes 12,800 manually labelled short claims from PolitiFact.com in diverse situations. Each data sample is labelled with one of six scores: true, mostly true, half true, barely true, false, or pants-faire. It can also be utilized for stance classification, argument mining, topic modelling, rumour detection, and political natural language processing research.
- **BuzzfeedNews:** it is compiled a list of news stories from Buzzfeed's post articles on fake election events on Facebook through looking for real and fake stories with the greatest engagement on Facebook via multiple methodologies during the nine months preceding the 2016 US Presidential Election, divided into three three-month segments. It contains other related data like URL of the news article, shared data, number of shares, reactions and comments. It is about 2,283 news samples from Facebook with four classes (mostly true, not factual content, mixture of true and false, and mostly false) [4].
    - **BuzzFace:** Santina and Williams [85] presented the dataset that was compiled by BuzzFeed from major sources such as ABC News Politics, Addicting Info, CNN Politics, Eagle Rising, Freedom Daily, Occupy Democrats, Politico, Right Wing News, and the 2016 Presidential Election and published on Facebook in September. Sample size is around 2,263, with the news coming from 9 different Facebook news sites (73.18% mainly genuine) (mostly true, mostly false, mixture of true and false, and no factual content).
    - **FAKENEWSNET:** Shu et al. [86] introduced two thorough datasets with a wide range of features in news content, social context, and spatiotemporal data, which use fact-checking websites for obtaining news stories for false news and real news like PolitiFact (political news) and GossipCop (entertainment news). Size of this dataset of about 422 news (211 fake news and 211 real news), and Labels are two classes (fake and real). It also includes important features such as publisher information, news content, and social engagements information for each news sample.
- **FEVER:** Thorne et al. [87] presented a freely released dataset called FEVER for extracting and verifying facts from textual sources by modifying Wikipedia sentences and then giving proof for or against such claims in Wikipedia articles. The size of this dataset about 185,445 claims extracted from Wikipedia with Three labels (supported, refuted, and notenoughinfo).
- **FCV-2018:** Papadopoulou et al. [88] presented a systematic process combining text search and near-duplicate video retrieval was used to create the dataset, which was then manually annotated using a set of journalism-inspired guidelines. Following the creation of the dataset, machine learning was used to perform automatic verification over a set of well-established features. And many Languages like English, Russian, Spanish, Arabic, German, Catalan, Japanese, and Portuguese. Data size: 380 videos and 77258 tweets with two labels.

Doi: https://doi.org/10.54216/NIF.010205
Received: January 18, 2023 Accepted: March 15, 2023

TABLE 7
BENCHMARK DATASETS

No.	Dataset	YEAR	Source & Platform	Extraction Time	Language	Scope Coverage	Size	Labels	Purpose	Type of disinformation	Content Type
1	CREDBANK [81]	2015	Twitter	October 2014 to February 2015	English	Society	60 million	Five-Grade	Veracity Classification	Rumors	Text
2	PHEME [82]	2016	Twitter	August 2014	English and German	Society, politics	330	Three-Grade	Rumor detection	Rumor	Text
3	Emergent [83]	2016	Twitter snopes.com	*	English	Society, technology	300	Three-Grade	Rumor detection	Rumors	Text
4	BuzzfeedNews [4]	2017	Facebook	2016 to 2017	English	Politics	2,283	Four-Grade	Fake detection	Fake news articles	Text
5	LIAR [84]	2017	POLITIFACT Facebook Twitter	2007 to 2016	English	Politics	12,800	Six-Grade	Fake detection	Fake news articles	Text
6	FEVER [87]	2018	Wikipedia	June 2017	English	Society	185,445	Three-Grade	Fact-Checking	Fake news articles	Text
7	Buzzface [85]	2018	Facebook	September 2016	English	Politics, Society	2,263	Four-Grade	Veracity classification	Fake news articles	Text
8	FAKENEWSNET [86]	2018	Twitter	*	English	Society, politics	422	Two-Grade	Fake detection	Fake news articles	Text, image
9	MisInfoText [89]	2019	Snopes Facebook	2016	English	Society	1,692	Four-Grade	Fact Checking	Fake news articles	Text
10	NELA-GT-2018 [91]	2019	Mainstream	February 2018 to November 2018	English	Politics	713,000	Two-Grade	Fake detection	Fake news articles	Text
11	FCV-2018 [88]	2019	Twitter Facebook YouTube	April 2017 To July 2017	Many Languages	Society	380 videos and 77258 tweets	Two-Grade	Fake detection	Fake news articles	Text, Video
12	NELA-GT-2020 [90]	2021	Mainstream	January 2020 to December 2020	English	Politics, Society, Covid19	1.8 million	Two-Grade	Fake detection	Fake news articles	Text

- **MisInfoText:** Torabi and Taboada [89] focusing on datasets including articles that have been individually vetted by consultants for veracity. This dataset was scraped, cleaned up, and organized individual articles harvested from fact checking sites, along with their labels (true, false, or similar labels) for text classification activities. Data samples about 1,692 news articles (1,380 from BuzzFeed dataset and 312 from Snopes dataset), and their four labels for BuzzFeed and five labels for Snopes ([fully] true, mostly true, mixture of true and false, mostly false, and [fully] false).
- **NELA-GT-2020:** this dataset [90] is a large-scale dataset of English news articles with source-level reliability labels, which improves on its precedents, NELA-GT-2019 and NELA-GT-2018 [91], in several ways. For starters, it employs a more robust scraper that is less prone to failures and sporadic data outages. NELA-GT-2020 includes roughly 1.8 million media stories gathered from 519 sources among 1st January 2020 and Dec. 31st, 2020 and these sources obtained from a variety of mainstream and alternative news sources, which it is the tweets embedded in the media stories provide an additional layer of information to the data.

The below table 7 outline the common datasets.

B. Evaluation metrics

In order to estimate a model's effectiveness, accuracy, sensitivity, and specificity are often used metric values. In circumstances when there is an unequal distribution of classes in the dataset, utilizing these matrices to assess the model's efficiency is not acceptable [92]. In such cases, the model may provide high accuracy since it is biassed towards the dominant class. As a result, in such an unbalanced domain, the confusion matrix is a valuable tool for understanding the model. To illustrate, consider a model that can tell whether or not a particular news story is a phony. True Positive (TP): The set of false positives forecasted by the approach that were really false positives. True Negative (TN): instances in which the model properly forecasted that something unfavorable was true and that it was, in fact, a true article. False Positive (FP): the set of erroneous negative predictions made by the approach that turned out to be fake. False Negative (FN): the set of positive occurrences that the approach mistook for fakes but were in fact true.

In the machine learning field, metrics are widely used to assess the effectiveness of a classifier from several aspects. Specifically, accuracy identify the similarity among forecasted false data and factual false data.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \qquad (1)$$

Doi: https://doi.org/10.54216/NIF.010205
Received: January 18, 2023 Accepted: March 15, 2023

Precision is defined as a percentage of false news stories that the model accurately identifies.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Using Recall, we can determine the sensitivity of the prediction of false news by looking at how many cases are found in comparison to how many are expected.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

False positives and false negatives are also taken into account when calculating the F1 score, which may provide insight into the overall prognosis for detecting false news.

$$F1 - score = \frac{2 \times Precison \times Recall}{Precison + Recall} \quad (4)$$

Classification approach performance at different threshold values may be evaluated using this measure. Both the Area Under the Curve (AUC) and the Receiver Operating Characteristic Curve (ROC) are measures of the degree to which two groups may be distinguished. The best the model predicts articles, the greater the AUC (as fake or real). False Positive Rate (FPR) is represented on the X-axis, and True Positive Rate (TPR) on the Y-axis. Separability and AUC = 1 are achieved when the curves of positive and negative classes do not overlap. According to the AUC value of 0.6, the model is able to distinguish between positive and negative classes with a 60% probability.

$$TPR = \frac{|TP|}{|TP| + |FN|} \quad (5)$$

$$FPR = \frac{|FP|}{|FP| + |TN|} \quad (6)$$

$$AUC = \frac{\sum(n_0 + n_1 + 1 - r_i) - n_0(n_0 + 1)/2}{n_0 n_1} \quad (7)$$

There are $n_0(n_1)$ false (actual) news articles and r_i is the rank of the i_{th} piece. False news classification is a good example of how AUC may be used in situations when the distribution of ground truth fake news and factual news is imbalanced, since it is statistically more consistent and discriminating than accuracy.

7. Applications of Fake News Detection

More and more people are signing up for social media accounts on a daily basis, including Facebook, Twitter and Instagram. However, it has been proven that majority of these accounts are false and may affect both the network and its users on social media. There is a belief that phoney accounts constitute the foundation for the spread of false information. As a result, it is critical to distinguish between actual and false accounts. The characteristics of fraudulent accounts have been studied extensively. In this part, we'll talk about contributions to the identification of bogus accounts. It is possible to classify the false account characteristics into two broad categories: textual and account related [93]. A username, profile image, the number of followers, the number of likes, and the location may all be account characteristics. Sender, mentions, hashtags, links, and the number of answers are all examples of textual features [93]. Spearman's rank-order correlation is used by Khaled et al. [94] to identify fake accounts and obtain excellent accuracy. Fake accounts may be detected using a new approach called relaxed functional dependencies (RFD) [95]. Profile similarity communication matching to detect duplicate accounts is another account-based feature technique [96]. Fraudulent project data is used by Rahman et al. [97] to identify fake Twitter accounts using a username feature-based approach. Many classifiers were used in this strategy, but the most accurate one was random forest (RF). The inclusion of an unrealistically limited number of accounts in the data meant that the outcomes would be unpredictable in real time. The emotion characteristic of RF classifiers may be used to identify bogus accounts. The fake account is characterised by words like "hatred" and "ugly" [98]. Swe and Myo [99] presented a textual feature-

Doi: https://doi.org/10.54216/NIF.010205
Received: January 18, 2023 Accepted: March 15, 2023

based methodology for detecting bogus accounts based on a blacklist that was implemented using a keyword topic. Time and money are saved since there are no network-based features required in this method. Khan et al. [100] used the Hyperlink-Induced Topic Search (HITS) technique to differentiate spammers from bloggers. A seventh attribute was employed to identify a hacked profile. In order to identify fraudulent or no accounts, this method produces a user history. For example, some research uses a mix of visuals and content-based characteristics and several classifiers to identify phoney Twitter accounts. The findings indicated that the RF classifier performed better than the others. Using the firefly algorithm, Aswani et al. [101] devised a way to identify twitter spammers. Using an RF classifier, eighteen characteristics retrieved from spam messages and false accounts may be used to detect spam and fraud.

It is a robot that is designed to carry out a certain duty without any human intervention. In social media, bots may be employed to run a platform or to publish current articles or news stories. Sybil bots, spam bots, social bots, and cyborg bots may all be created by bad bots. Fraudulent accounts may be generated by Sybil's bots, which can then be used to disseminate malware or fake data over the network. In order to protect against Sybil assaults, a complex authentication method must be implemented. Bots capable of sending spam messages, links, or any other garbage data are known as spambots. If you send a large amount of false data, you risk clogging up the network. There are a variety of ways that social bots might obtain sensitive information from consumers, including by posing as legitimate websites. Is any form of robot that can be operated by a person. These bots are being prevented in several research. Graph techniques, ML methods, crowdsourcing methods, and anomaly approaches [102] are all subcategories of these investigations. As a representation of a social network, a graph is all that is required. Many contributions are geared on detecting dangerous bots via the use of graph-based approaches. Methods from Cornelissen et al. [103] combined ML with network metrics. The Twitter dataset yielded findings with an unsatisfactory level of accuracy. Post-to-post and user-to-user models were used to identify political bots in another investigation [104]. There is also a Bootcamp to detect the campaigns of bots, which uses graph methods for topographical modelling to do clustering, by Abu-El-Rub and Mueen [105].

As social media platforms have grown in popularity, so has the prevalence of cyberbullying. The scope and speed with which cyberbullying is gaining traction puts others at risk. Removing cyberbullying information by hand may be seen as a waste of time and effort by others. As a result, models for automatically detecting bullying are an absolute need. In this area, you'll find the most research on how to spot online bullying. Unbalanced datasets were used by Rosa et al. [106] to evaluate the performance of Fuzzy Fingerprints (FFP) with LR, NB, and SVM for detecting cyberbullying. The results reveal that SVM is outperformed. Use class term-occurrence information to build a non-sparse, discriminative model for documents, according to Escalante et al. [107] Sub-profiles, such as sexual predators, or hostile text, are examples of this kind of profile. Sub-profile-based representations are better than profile-based representations in terms of accuracy. On Twitter streaming API, Cheng et al. [78] compared SVN, KNN, RF, and LR on congested data. Additionally, a PI-Bully model was developed to identify the peculiarities of individual users in order to better anticipate instances of cyberbullying. Table 8 summarization of most studies in applications of fake news detection.

Doi: https://doi.org/10.54216/NIF.010205
Received: January 18, 2023 Accepted: March 15, 2023

Fake accounts, malicious bots, network congestion, and other concerns related to social media platform security may all be the result of security flaws in social media platforms. Attacks on social media that potentially harm users. Users and mobile internet technologies are connected via a social network, which is defined by web 2.0. The current issue of trustworthiness and secrecy in social media necessitated the implementation of security control systems [108]. For example, Ma and Yan [109] use ubiquitous social networking to manage undesired information in terms of security. LSTM on-location analysis on the NN platform was used in another DeepScan to discover rogue accounts [110]. For the first time, a strategy based on spatial-temporal aspects was presented by Zhou et al. [111]. The findings reveal a high rate of detection and a low number of false positives. Many classifiers are used in the work by Campos and colleagues to detect harmful bots [112]. Distinct wavelength transformation is used in this method to identify writing patterns. In Zhang et al. [113]'s work, COLOR+ is introduced to identify bogus accounts based on the contact between users and their neighbours. Mobile devices may take use of this quick response strategy. Even more recently, Liu & colleagues presented a technique that analyses tweets to identify spam and retweet user behaviour [114]. In addition to network security concerns, users' knowledge and conduct also play a role in ensuring security and privacy. Any links or spam communications sent to a user should raise red flags. Security notifications are also ignored by many users. Also, do not divulge any private or sensitive information, whether it comes from a medical or political source, that has not been approved by you. Most of the victims of theft or cyberbullying had a different perspective on the security and privacy of social media [115].

8. Discussion, Challenges, and Future Research Directions

A significant impact on society has been the proliferation of incorrect information disseminated through websites and online media portals. Several scholars have worked to automate the early detection and identification of bogus news using AI techniques. Fake news detection is a time-consuming and challenging process. The detection method demands multiple steps to classify a given batch of news items. Depending on the kind of data and the language used, the preprocessing of the acquired news items will vary. Text is the most common kind of data in news reports. By putting the articles into a different form, feature vectors with enough information for accurate categorization and machine maintenance must be extracted. We address some of the outstanding concerns and obstacles encountered during the detection of false news using AI approaches.

- Deep learning research has shown a great deal of interest in making outcomes more interpretable. However, in the case of fact-checking platforms and social media, it is possible to get interpretability by mining social response such as the stance taken in tweets and postings and by mining expert investigations.
- According to Zhou et al. [116], An assault on NLP for recognizing false news is prone to three sorts of attacks: factual distortion, subject-object interchange, and cause misunderstanding. Overestimation or alteration of certain tokens are examples of distortions. When linguistic characteristics like letters and time are tampered with, an incorrect meaning might be conveyed. The

TABLE 8
SUMMARIZATION OF MOST STUDIES IN APPLICATIONS OF FAKE NEWS

Reference	Application Type	Media Platform	Language	Methods
Khaled et al. [94]	Fake Accounts, and Bot Detection	Twitter	English	SVM-NN
Wani et al. [98]	Fake Accounts	Facebook	English	SVM, NB, and RF
Cornelissen et al. [103]	Bot Detection	Twitter	English	Clustering and Graph-based
Kheir et al. [120]	Bot Detection	Twitter	English	Deep Forest
Bozyigit et al. [121]	Cyberbullying detection	Twitter	Turkish	SVM, LR, KNN, NB, AdaBoost, and RF
Kumari et al. [122]	Cyberbullying detection	Twitter	English	Pre-trained VGG and CNN
Haider et al. [123]	Cyberbullying detection	Facebook, and Twitter	Arabic	NB, and SVM

Received: January 18, 2023 Accepted: March 15, 2023

purpose of this back-and-forth between subject and object is to confuse the viewer as to who is doing the acting and who is being harmed by it. The intruder attempts to show the audience just the portions of the tale that the intruder wants them to see, which results in the attack of cause confusion, which entails creating causal linkages between unrelated events that don't exist.

- To help people better understand and interpret time-sensitive data, an online false news monitoring system should include visualization as an important component. Analyzing data via visual representations is a powerful tool that may provide a wide range of perspectives on the information at hand, facilitate human comprehension, and reveal temporal-based patterns and behaviors of data [117][118][119].
- Current research for the extraction of textual characteristics is concentrated on embedding techniques like word embedding and deep learning approaches, which can succeed better depicting for the features [17]. Visual features obtained from pictures may also be used to distinguish between real and fabricated news stories. In the mining of visual data for the identification of false news, deep learning offers a study opportunity [11].

9. Conclusions

Fake news was nothing new, but with the advent of online media platforms, it saw extraordinary development during the 2016 US presidential election. This cleared the path for scholars and other parties to work together to develop a long-term solution. According to the findings of this work, the growth of fake news on online media portals has frequently made users hesitant to engage in legitimate news and information exchange for fear of receiving inaccurate and misleading information. We demonstrated the topic of fake news detection from the perspective of how sentiment analysis and machine learning technologies are utilized to address the issue. Finally, we discussed the most pressing challenges, and future research direction in our view.

References

[1] M. Guerini and J. Staiano, "Deep feelings: A massive cross-lingual study on the relation between emotions and virality," 2015, doi: 10.1145/2740908.2743058.

[2] J. P. Dickerson, V. Kagan, and V. S. Subrahmanian, "Using sentiment to detect bots on Twitter: Are humans more opinionated than bots?," 2014, doi: 10.1109/ASONAM.2014.6921650.

[3] Y. Chen, N. J. Conroy, and V. L. Rubin, "Misleading online content: Recognizing clickbait as 'false news,'" 2015, doi: 10.1145/2823465.2823467.

[4] B. Horne and S. Adali, "This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news," in *Proceedings of the International AAAI Conference on Web and Social Media*, Mar. 2017, vol. 11, no. 1.

[5] R. Oshikawa, J. Qian, and W. Y. Wang, "A survey on natural language processing for fake news detection," 2020.

[6] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," *Proc. Assoc. Inf. Sci. Technol.*, 2015, doi: 10.1002/pra2.2015.145052010082.

[7] M. K. Elhadad, K. Fun Li, and F. Gebali, "Fake News Detection on Social Media: A Systematic Survey," 2019, doi: 10.1109/PACRIM47961.2019.8985062.

[8] R. K. Kaliyar, A. Goswami, and P. Narang, "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach," *Multimed. Tools Appl.*, 2021, doi: 10.1007/s11042-020-10183-2.

[9] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake News Detection on Social Media," *ACM SIGKDD Explor. Newsl.*, 2017, doi: 10.1145/3137597.3137600.

[10] A. Bondielli and F. Marcelloni, "A survey on fake news and rumour detection techniques," *Inf. Sci. (Ny).*, 2019, doi: 10.1016/j.ins.2019.05.035.

Doi: https://doi.org/10.54216/NIF.010205
Received: January 18, 2023 Accepted: March 15, 2023

[11] K. Sharma, F. Qian, H. Jiang, N. Ruchansky, M. Zhang, and Y. Liu, "Combating fake news: A survey on identification and mitigation techniques," *ACM Transactions on Intelligent Systems and Technology*. 2019, doi: 10.1145/3305260.

[12] F. C. D. da Silva, R. V. da Costa Alves, and A. C. B. Garcia, "Can machines learn to detect fake news? A survey focused on social media," 2019, doi: 10.24251/hicss.2019.332.

[13] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter, "Detection and resolution of rumours in social media: A survey," *ACM Computing Surveys*. 2018, doi: 10.1145/3161603.

[14] S. Sharma and D. K. Sharma, "Fake News Detection: A long way to go," 2019, doi: 10.1109/ISCON47742.2019.9036221.

[15] A. E. Lillie and E. R. Middelboe, "Fake news detection using stance classification: A survey," *arXiv Prepr. arXiv1907.00181*, Dec. 2019.

[16] X. Zhang and A. A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion," *Inf. Process. Manag.*, 2020, doi: 10.1016/j.ipm.2019.03.004.

[17] P. Meel and D. K. Vishwakarma, "Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities," *Expert Systems with Applications*. 2020, doi: 10.1016/j.eswa.2019.112986.

[18] X. Zhou and R. Zafarani, "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities," *ACM Comput. Surv.*, 2020, doi: 10.1145/3395046.

[19] D. Antonakaki, P. Fragopoulou, and S. Ioannidis, "A survey of Twitter research: Data model, graph structure, sentiment analysis and attacks," *Expert Syst. Appl.*, 2021, doi: 10.1016/j.eswa.2020.114006.

[20] Z. Guo, J. H. Cho, I. R. Chen, S. Sengupta, M. Hong, and T. Mitra, "Online Social Deception and Its Countermeasures: A Survey," *IEEE Access*. 2021, doi: 10.1109/ACCESS.2020.3047337.

[21] F. D. C. Medeiros and R. B. Braga, "Fake news detection in social media: A systematic review," 2020, doi: 10.1145/3411564.3411648.

[22] N. Hoy and T. Koulouri, "A Systematic Review on the Detection of Fake News Articles," *arXiv Prepr. arXiv2110.11240*, Oct. 2021.

[23] E. Providel and M. Mendoza, "Misleading information in Spanish: a survey," *Social Network Analysis and Mining*. 2021, doi: 10.1007/s13278-021-00746-y.

[24] N. R. de Oliveira, P. S. Pisa, M. A. Lopez, D. S. V. de Medeiros, and D. M. F. Mattos, "Identifying fake news on social networks based on natural language processing: Trends and challenges," *Information (Switzerland)*. 2021, doi: 10.3390/info12010038.

[25] D. Rohera *et al.*, "A Taxonomy of Fake News Classification Techniques: Survey and Implementation Aspects," *IEEE Access*, vol. 10, pp. 30367–30394, 2022.

[26] O. Stitini, S. Kaloun, and O. Bencharef, "Towards the Detection of Fake News on Social Networks Contributing to the Improvement of Trust and Transparency in Recommendation Systems: Trends and Challenges," *Information*, vol. 13, no. 3, p. 128, 2022.

[27] A. Fang, C. Macdonald, I. Ounis, and P. Habel, "Using word embedding to evaluate the coherence of topics from twitter data," 2016, doi: 10.1145/2911451.2914729.

[28] S. Sarawagi *et al.*, "Thumbs up or thumbs down? Semantic Orientation applied to Unsupervised Classification of Reviews," *Expert Syst. Appl.*, 2014.

[29] D. Vilares, C. Gómez-Rodríguez, and M. A. Alonso, "Universal, unsupervised (rule-based), uncovered sentiment analysis," *Knowledge-Based Syst.*, 2017, doi: 10.1016/j.knosys.2016.11.014.

[30] E. Cambria, Y. Li, F. Z. Xing, S. Poria, and K. Kwok, "SenticNet 6: Ensemble Application of

Doi: https://doi.org/10.54216/NIF.010205
Received: January 18, 2023 Accepted: March 15, 2023

Symbolic and Subsymbolic AI for Sentiment Analysis," 2020, doi: 10.1145/3340531.3412003.

[31] S. Hamidian and M. T. Diab, "Rumor identification and belief investigation on Twitter," 2016, doi: 10.18653/v1/w16-0403.

[32] M. Meire, M. Ballings, and D. Van den Poel, "The added value of auxiliary data in sentiment analysis of Facebook posts," *Decis. Support Syst.*, 2016, doi: 10.1016/j.dss.2016.06.013.

[33] M. AbdelFattah, D. Galal, N. Hassan, D. S. Elzanfaly, and G. Tallent, "A sentiment analysis tool for determining the promotional success of fashion images on instagram," *Int. J. Interact. Mob. Technol.*, 2017, doi: 10.3991/ijim.v11i2.6563.

[34] X. Hu, J. Tang, H. Gao, and H. Liu, "Social Spammer Detection with Sentiment Information," 2014, doi: 10.1109/ICDM.2014.141.

[35] M. A. Alonso and D. Vilares, "A review on political analysis and social media," *Procesamiento de Lenguaje Natural*. 2016.

[36] J. S. Jang, B. I. Lee, C. H. Choi, J. H. Kim, D. M. Seo, and W. S. Cho, "Understanding pending issue of society and sentiment analysis using social media," 2016, doi: 10.1109/ICUFN.2016.7536944.

[37] P. D. Azar and A. W. Lo, "The wisdom of twitter crowds: Predicting stock market reactions to FOMC meetings via twitter feeds," *Journal of Portfolio Management*. 2016, doi: 10.3905/jpm.2016.42.5.123.

[38] S. Sharma and A. Jain, "Role of sentiment analysis in social media security and analytics," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 2020, doi: 10.1002/widm.1366.

[39] A. H. Alamoodi *et al.*, "Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review," *Expert Systems with Applications*. 2021, doi: 10.1016/j.eswa.2020.114155.

[40] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *Journal of Economic Perspectives*. 2017, doi: 10.1257/jep.31.2.211.

[41] C. Silverman, "This analysis shows how viral fake election news stories outperformed real news on Facebook," *BuzzFeed news*, vol. 16, 2016.

[42] C. Guo, J. Cao, X. Zhang, K. Shu, and H. Liu, "Dean: Learning dual emotion for fake news detection on social media," *arXiv e-prints*, p. arXiv--1903, 2019.

[43] M. AlRubaian, M. Al-Qurishi, M. Al-Rakhami, S. M. M. Rahman, and A. Alamri, "A multistage credibility analysis model for microblogs," 2015, doi: 10.1145/2808797.2810065.

[44] A. Dey, R. Z. Rafi, S. Hasan Parash, S. K. Arko, and A. Chakrabarty, "Fake news pattern recognition using linguistic analysis," 2019, doi: 10.1109/ICIEV.2018.8641018.

[45] B. Bhutani, N. Rastogi, P. Sehgal, and A. Purwar, "Fake News Detection Using Sentiment Analysis," 2019, doi: 10.1109/IC3.2019.8844880.

[46] X. Zhang, J. Cao, X. Li, Q. Sheng, L. Zhong, and K. Shu, "Mining dual emotion for fake news detection," 2021, doi: 10.1145/3442381.3450004.

[47] X. Yu, Y. Chu, F. Jiang, Y. Guo, and D. Gong, "SVMs Classification Based Two-side Cross Domain Collaborative Filtering by inferring intrinsic user and item features," *Knowledge-Based Syst.*, 2018, doi: 10.1016/j.knosys.2017.11.010.

[48] V. Agarwal, H. P. Sultana, S. Malhotra, and A. Sarkar, "Analysis of Classifiers for Fake News Detection," 2019, doi: 10.1016/j.procs.2020.01.035.

[49] H. Ahmed, I. Traore, and S. Saad, "Detecting opinion spams and fake news using text classification," *Secur. Priv.*, 2018, doi: 10.1002/spy2.9.

Doi: https://doi.org/10.54216/NIF.010205
Received: January 18, 2023 Accepted: March 15, 2023

[50] A. B. Prasetijo, R. R. Isnanto, D. Eridani, Y. A. A. Soetrisno, M. Arfan, and A. Sofwan, "Hoax detection system on Indonesian news sites based on text classification using SVM and SGD," 2017, doi: 10.1109/ICITACEE.2017.8257673.

[51] S. B. Deokate, "Fake News Detection using Support Vector Machine learning Algorithm," *Int. J. Res. Appl. Sci. Eng. Technol.*, 2019, doi: 10.22214/ijraset.2019.7067.

[52] Y. Yang *et al.*, "Combating," *2017 IEEE 15th Student Conf. Res. Dev.*, 2018.

[53] J. A. Nasir, O. S. Khan, and I. Varlamis, "Fake news detection: A hybrid CNN-RNN based deep learning approach," *Int. J. Inf. Manag. Data Insights*, 2021, doi: 10.1016/j.jjimei.2020.100007.

[54] N. Deligiannis, T. Do Huu, D. M. Nguyen, and X. Luo, "Deep Learning for Geolocating Social Media Users and Detecting Fake News," *NATO Work.*, 2018.

[55] Z. Khanam, B. N. Alwasel, H. Sirafi, and M. Rashid, "Fake News Detection Using Machine Learning Approaches," *IOP Conf. Ser. Mater. Sci. Eng.*, 2021, doi: 10.1088/1757-899x/1099/1/012040.

[56] 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," *Futur. Gener. Comput. Syst.*, 2021, doi: 10.1016/j.future.2020.11.022.

[57] S. Kaur, P. Kumar, and P. Kumaraguru, "Automating fake news detection system using multi-level voting model," *Soft Comput.*, 2020, doi: 10.1007/s00500-019-04436-y.

[58] E. Tacchini, G. Ballarin, M. L. Della Vedova, S. Moret, and L. de Alfaro, "Some like it Hoax: Automated fake news detection in social networks," 2017.

[59] J. M. G. Ogdol, B.-L. T. Samar, and C. Cataroja, "Binary Logistic Regression based Classifier for Fake News," *Anal. Standar Pelayanan Minimal Pada Instal. Rawat Jalan di RSUD Kota Semarang*, 2018.

[60] N. S. Yuslee and N. A. S. Abdullah, "Fake News Detection using Naive Bayes," 2021, doi: 10.1109/ICSET53708.2021.9612540.

[61] M. Casillo, F. Colace, B. B. Gupta, D. Santaniello, and C. Valentino, "Fake News Detection Using LDA Topic Modelling and K-Nearest Neighbor Classifier," 2021, doi: 10.1007/978-3-030-91434-9_29.

[62] T. Mladenova and I. Valova, "Analysis of the KNN Classifier Distance Metrics for Bulgarian Fake News Detection," 2021, doi: 10.1109/HORA52670.2021.9461333.

[63] Z. H. Zhou, "A brief introduction to weakly supervised learning," *National Science Review*. 2018, doi: 10.1093/nsr/nwx106.

[64] Z. Kou, D. Yue Zhang, L. Shang, and D. Wang, "ExFaux: A Weakly Supervised Approach to Explainable Fauxtography Detection," 2020, doi: 10.1109/BigData50022.2020.9378019.

[65] S. Helmstetter and H. Paulheim, "Weakly supervised learning for fake news detection on Twitter," 2018, doi: 10.1109/ASONAM.2018.8508520.

[66] K. Shu *et al.*, "Leveraging multi-source weak social supervision for early detection of fake news," *arXiv Prepr. arXiv2004.01732*, 2020.

[67] S. Yang, K. Shu, S. Wang, R. Gu, F. Wu, and H. Liu, "Unsupervised fake news detection on social media: A generative approach," 2019, doi: 10.1609/aaai.v33i01.33015644.

[68] D. Li, H. Guo, Z. Wang, and Z. Zheng, "Unsupervised Fake News Detection Based on Autoencoder," *IEEE Access*, 2021, doi: 10.1109/ACCESS.2021.3058809.

[69] S. C. R. Gangireddy, P. Deepak, C. Long, and T. Chakraborty, "Unsupervised fake news detection: A graph-based approach," 2020, doi: 10.1145/3372923.3404783.

Doi: https://doi.org/10.54216/NIF.010205
Received: January 18, 2023 Accepted: March 15, 2023

[70] M. Orlov and M. Litvak, "Using behavior and text analysis to detect propagandists and misinformers on twitter," 2019, doi: 10.1007/978-3-030-11680-4_8.

[71] W. S. Paka, R. Bansal, A. Kaushik, S. Sengupta, and T. Chakraborty, "Cross-SEAN: A cross-stitch semi-supervised neural attention model for COVID-19 fake news detection," *Appl. Soft Comput.*, 2021, doi: 10.1016/j.asoc.2021.107393.

[72] X. Dong, U. Victor, S. Chowdhury, and L. Qian, "Deep two-path semi-supervised learning for fake news detection," *arXiv Prepr. arXiv1906.05659*, 2019.

[73] A. Benamira, B. Devillers, E. Lesot, A. K. Ray, M. Saadi, and F. D. Malliaros, "Semi-supervised learning and graph neural networks for fake news detection," 2019, doi: 10.1145/3341161.3342958.

[74] X. Li, P. Lu, L. Hu, X. G. Wang, and L. Lu, "A novel self-learning semi-supervised deep learning network to detect fake news on social media," *Multimed. Tools Appl.*, 2021, doi: 10.1007/s11042-021-11065-x.

[75] M. C. de Souza, B. M. Nogueira, R. G. Rossi, R. M. Marcacini, B. N. dos Santos, and S. O. Rezende, "A network-based positive and unlabeled learning approach for fake news detection," *Mach. Learn.*, 2021, doi: 10.1007/s10994-021-06111-6.

[76] Z. Li, X. Jiang, L. Shang, and H. Li, "Paraphrase generation with deep reinforcement learning," 2018, doi: 10.18653/v1/d18-1421.

[77] W. Fedus, I. Goodfellow, and A. M. Dai, "MaskGaN: Better text generation via filling in the," 2018.

[78] L. Cheng, A. Mosallanezhad, Y. N. Silva, D. L. Hall, and H. Liu, "Mitigating bias in session-based cyberbullying detection: A non-compromising approach," 2021, doi: 10.18653/v1/2021.acl-long.168.

[79] A. Mosallanezhad, M. Karami, K. Shu, M. V Mancenido, and H. Liu, "Domain Adaptive Fake News Detection via Reinforcement Learning," in *Proceedings of the ACM Web Conference 2022*, 2022, pp. 3632–3640.

[80] E. Meirom, H. Maron, S. Mannor, and G. Chechik, "Controlling graph dynamics with reinforcement learning and graph neural networks," in *International Conference on Machine Learning*, 2021, pp. 7565–7577.

[81] T. Mitra and E. Gilbert, "CREDBANK: A large-scale social media corpus with associated credibility annotations," 2015.

[82] A. Zubiaga, M. Liakata, R. Procter, K. Bontcheva, and P. Tolmie, "Towards detecting rumours in social media," 2015.

[83] W. Ferreira and A. Vlachos, "Emergent: A novel data-set for stance classification," 2016, doi: 10.18653/v1/n16-1138.

[84] W. Y. Wang, "'Liar, liar pants on fire': A new benchmark dataset for fake news detection," 2017, doi: 10.18653/v1/P17-2067.

[85] G. C. Santia and J. R. Williams, "Buzzface: A news veracity dataset with facebook user commentary and egos," Jun. 2018.

[86] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media," *Big Data*, 2020, doi: 10.1089/big.2020.0062.

[87] J. Thorne, A. Vlachos, C. Christodoulopoulos, and A. Mittal, "FEVER: A large-scale dataset for fact extraction and verification," 2018, doi: 10.18653/v1/n18-1074.

[88] O. Papadopoulou, M. Zampoglou, S. Papadopoulos, and I. Kompatsiaris, "A corpus of debunked

Doi: https://doi.org/10.54216/NIF.010205
Received: January 18, 2023 Accepted: March 15, 2023

and verified user-generated videos," *Online Inf. Rev.*, 2019, doi: 10.1108/OIR-03-2018-0101.

[89] F. Torabi Asr and M. Taboada, "Big Data and quality data for fake news and misinformation detection," *Big Data Soc.*, 2019, doi: 10.1177/2053951719843310.

[90] M. Gruppi, B. D. Horne, and S. Adal\i, "NELA-GT-2021: A Large Multi-Labelled News Dataset for The Study of Misinformation in News Articles," *arXiv Prepr. arXiv2203.05659*, 2022.

[91] J. Nørregaard, B. D. Horne, and S. Adal, "Nela-gt-2018: A large multi-labelled news dataset for the study of misinformation in news articles," in *Proceedings of the international AAAI conference on web and social media*, Apr. 2019, vol. 13, pp. 630–638.

[92] F. B. Gereme and W. Zhu, "Early detection of fake news "before it flies high," 2019, doi: 10.1145/3358528.3358567.

[93] P. K. Roy and S. Chahar, "Fake Profile Detection on Social Networking Websites: A Comprehensive Review," *IEEE Trans. Artif. Intell.*, 2021, doi: 10.1109/tai.2021.3064901.

[94] S. Khaled, N. El-Tazi, and H. M. O. Mokhtar, "Detecting Fake Accounts on Social Media," 2019, doi: 10.1109/BigData.2018.8621913.

[95] L. Caruccio, V. Deufemia, F. Naumann, and G. Polese, "Discovering Relaxed Functional Dependencies Based on Multi-Attribute Dominance," *IEEE Trans. Knowl. Data Eng.*, 2021, doi: 10.1109/TKDE.2020.2967722.

[96] S. Revathi and M. Suriakala, "Profile Similarity Communication Matching Approaches for Detection of Duplicate Profiles in Online Social Network," 2018, doi: 10.1109/CSITSS.2018.8768751.

[97] M. D. Rahman, A. M. Likhon, A. S. Rahman, and M. H. Choudhury, "Detection of fake identities on twitter using supervised machine learning," Brac University, 2019.

[98] M. A. Wani, N. Agarwal, S. Jabin, and S. Z. Hussain, "Analyzing Real and Fake users in Facebook Network based on Emotions," 2019, doi: 10.1109/COMSNETS.2019.8711124.

[99] M. M. Swe and N. Nyein Myo, "Fake Accounts Detection on Twitter Using Blacklist," 2018, doi: 10.1109/ICIS.2018.8466499.

[100] M. U. S. Khan, M. Ali, A. Abbas, S. U. Khan, and A. Y. Zomaya, "Segregating Spammers and Unsolicited Bloggers from Genuine Experts on Twitter," *IEEE Trans. Dependable Secur. Comput.*, 2018, doi: 10.1109/TDSC.2016.2616879.

[101] R. Aswani, A. K. Kar, and P. Vigneswara Ilavarasan, "Detection of Spammers in Twitter marketing: A Hybrid Approach Using Social Media Analytics and Bio Inspired Computing," *Inf. Syst. Front.*, 2018, doi: 10.1007/s10796-017-9805-8.

[102] M. Orabi, D. Mouheb, Z. Al Aghbari, and I. Kamel, "Detection of Bots in Social Media: A Systematic Review," *Inf. Process. Manag.*, 2020, doi: 10.1016/j.ipm.2020.102250.

[103] L. A. Cornelissen, R. J. Barnett, P. Schoonwinkel, B. D. Eichstadt, and H. B. Magodla, "A network topology approach to bot classification," 2018, doi: 10.1145/3278681.3278692.

[104] S. Hurtado, P. Ray, and R. Marculescu, "Bot detection in reddit political discussion," 2019, doi: 10.1145/3313294.3313386.

[105] N. Abu-El-Rub and A. Mueen, "BotCamp: Bot-driven interactions in social campaigns," 2019, doi: 10.1145/3308558.3313420.

[106] H. Rosa, P. Calado, R. Ribeiro, J. P. Carvalho, B. Martins, and L. Coheur, "Using fuzzy fingerprints for cyberbullying detection in social networks," 2018, doi: 10.1109/Fuzz-Ieee.2018.8491557.

[107] H. J. Escalante, E. Villatoro-Tello, S. E. Garza, A. P. López-Monroy, M. Montes-y-Gómez, and L. Villaseñor-Pineda, "Early detection of deception and aggressiveness using profile-based

Doi: https://doi.org/10.54216/NIF.010205
Received: January 18, 2023 Accepted: March 15, 2023

representations," *Expert Syst. Appl.*, 2017, doi: 10.1016/j.eswa.2017.07.040.

[108] S. R. Sahoo and B. B. Gupta, "Classification of various attacks and their defence mechanism in online social networks: a survey," *Enterprise Information Systems*. 2019, doi: 10.1080/17517575.2019.1605542.

[109] S. Ma and Z. Yan, "PSNController: An unwanted content control system in pervasive social networking based on trust management," *ACM Trans. Multimed. Comput. Commun. Appl.*, 2015, doi: 10.1145/2808206.

[110] Q. Gong *et al.*, "DeepScan: Exploiting Deep Learning for Malicious Account Detection in Location-Based Social Networks," *IEEE Commun. Mag.*, 2018, doi: 10.1109/MCOM.2018.1700575.

[111] Y. Zhou *et al.*, "Analyzing and Detecting Money-Laundering Accounts in Online Social Networks," *IEEE Netw.*, 2018, doi: 10.1109/MNET.2017.1700213.

[112] S. Barbon, G. F. C. Campos, G. M. Tavares, R. A. Igawa, M. L. Proença, and R. C. Guido, "Detection of human, legitimate bot, and malicious bot in online social networks based on wavelets," *ACM Trans. Multimed. Comput. Commun. Appl.*, 2018, doi: 10.1145/3183506.

[113] J. Zhang, Q. Li, X. Wang, B. Feng, and D. Guo, "Towards fast and lightweight spam account detection in mobile social networks through fog computing," *Peer-to-Peer Netw. Appl.*, 2018, doi: 10.1007/s12083-017-0559-3.

[114] B. Liu *et al.*, "Analysis of and defense against crowd-retweeting based spam in social networks," *World Wide Web*, 2019, doi: 10.1007/s11280-018-0613-y.

[115] Z. Zhang and B. B. Gupta, "Social media security and trustworthiness: Overview and new direction," *Futur. Gener. Comput. Syst.*, 2018, doi: 10.1016/j.future.2016.10.007.

[116] Z. Zhou, H. Guan, M. M. Bhat, and J. Hsu, "Fake news detection via NLP is vulnerable to adversarial attacks," 2019, doi: 10.5220/0007566307940800.

[117] P. Cogan, M. Andrews, M. Bradonjic, W. S. Kennedy, A. Sala, and G. Tucci, "Reconstruction and analysis of Twitter conversation graphs," 2012, doi: 10.1145/2392622.2392626.

[118] R. Nishi *et al.*, "Reply trees in Twitter: data analysis and branching process models," *Soc. Netw. Anal. Min.*, 2016, doi: 10.1007/s13278-016-0334-0.

[119] V. Gómez, H. J. Kappen, N. Litvak, and A. Kaltenbrunner, "A likelihood-based framework for the analysis of discussion threads," *World Wide Web*, 2013, doi: 10.1007/s11280-012-0162-8.

[120] K. E. Daouadi, R. Z. Rebaï, and I. Amous, "Bot detection on online social networks using deep forest," 2019, doi: 10.1007/978-3-030-19810-7_30.

[121] A. Bozyiğit, S. Utku, and E. Nasibov, "Cyberbullying detection: Utilizing social media features," *Expert Syst. Appl.*, 2021, doi: 10.1016/j.eswa.2021.115001.

[122] K. Kumari and J. P. Singh, "Identification of cyberbullying on multi-modal social media posts using genetic algorithm," *Trans. Emerg. Telecommun. Technol.*, 2021, doi: 10.1002/ett.3907.

[123] B. Haidar, M. Chamoun, and A. Serhrouchni, "Multilingual cyberbullying detection system: Detecting cyberbullying in Arabic content," 2017, doi: 10.1109/CSNET.2017.8242005.

[124] K. Popat, S. Mukherjee, J. Strötgen, and G. Weikum, "Credibility assessment of textual claims on the web," 2016, doi: 10.1145/2983323.2983661.

[125] K. Popat, S. Mukherjee, J. Strötgen, and G. Weikum, "Where the truth lies: Explaining the credibility of emerging claims on the web and social media," 2017, doi: 10.1145/3041021.3055133.

[126] 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," 2017, doi: 10.18653/v1/d17-1317.

Doi: https://doi.org/10.54216/NIF.010205
Received: January 18, 2023 Accepted: March 15, 2023

[127] L. Cui, S. Wang, and D. Lee, "Same: Sentiment-aware multi-modal embedding for detecting fake news," 2019, doi: 10.1145/3341161.3342894.

[128] M. Del Vicario, W. Quattrociocchi, A. Scala, and F. Zollo, "Polarization and fake news: Early warning of potential misinformation targets," *ACM Trans. Web*, 2019, doi: 10.1145/3316809.

[129] J. C. S. Reis, A. Correia, F. Murai, A. Veloso, F. Benevenuto, and E. Cambria, "Supervised Learning for Fake News Detection," *IEEE Intell. Syst.*, 2019, doi: 10.1109/MIS.2019.2899143.

[130] K. Anoop, P. Deepak, and L. V. Lajish, "Emotion cognizance improves health fake news identification," 2020, doi: 10.1145/3410566.3410595.

Doi: https://doi.org/10.54216/NIF.010205
Received: January 18, 2023 Accepted: March 15, 2023