Machine Learning-Based Effective Detection Scheme of Fake News

Khalid Abood Kamel[a], Jumana Waleed[b]

[a]*Department of Computer Science, College of Science, University of Diyala, Iraq. Email: scicompms23@uodiyala.edu.iq*

[b]*Department of Computer Science, College of Science, University of Diyala, Iraq. Email: jumanawaleed@uodiyala.edu.iq*

A R T I C L E   I N F O

A B S T R A C T

Today, extremely large amounts of false news are consistently uploaded by malevolent people with fraudulent intentions, endangering democracy, justice, and public confidence while having highly harmful social impacts on both individuals and society. Due to the inherent uncontrollable posting processes of social media sites (such as Facebook, Twitter, and Snapchat), this is particularly pertinent to them. The need for earlier false news identification has considerably fueled efforts in academia and business to create more accurate fake news detection techniques. Unfortunately, there isn't much information available regarding how news spreads. There are benefits and drawbacks to relying on social media to follow the news. Social media platforms do make it possible for information to flow swiftly among users. However, these websites could be used to disseminate "fake news," which is low-quality content that contains errors. The widespread dissemination of false information has an extremely damaging effect on both people and society. As a result, the detection of fake news posted on numerous social media platforms has recently emerged as a highly regarded field of study. In this paper, an effective detection scheme of fake news based on the commonly utilized techniques of machine learning has been presented. This proposed scheme involves diverse phases; Dataset preprocessing phase, extracting the phase of the feature, and Naive Bayes (NB) and K-nearest Neighbor are used in the categorization step (KNN). The obtained results in this presented scheme exhibit that the utilization of the NB classification technique exceeds the K-NN classification technique with an accuracy of 94% using the same dataset.

MSC..

1. Introduction

Because it is so accessible, a lot of people read the news on various social media sites. Social media platforms are cited as being superior to conventional media. Not all news that is reported is accurate, though. Data manipulation and information alteration occur for a variety of economic, social, and political reasons. Thus, news that is based on these modified statistics is neither totally truthful nor entirely untrue. Misinformation is then spread on social media, which further complicates society's problems. There are many different kinds and manifestations of this false information, sometimes known as "fake news." There are many distinct sorts of fake news, such as rumors, false political reporting, satire, and phony adverts. In comparison to actual news, fake news is now more widely

---

∗Corresponding author: Khalid Abood Kamel

Email addresses: kh.abood0801@gmail.com

Communicated by 'sub etitor'

circulated, causing many researchers to concentrate on developing reliable automated algorithms to recognize bogus news It is challenging to identify bogus news, in fact. The fake news detection model looks at both actual and fake content that has already been assessed in order to identify material that is being purposely misrepresented. This prompts us to point out one of the key pillars as being the ubiquitous availability of high-quality training data. A methodology for detecting false news can be thought of as doing simple binary classification or micro-classification in a challenging setting. Several data sets for fake news were uploaded after 2017. Using these several datasets, which are well-known and widely accessible datasets (ISOT, Kaggle, and LIAR datasets), the researchers aimed to enhance the performance of the published models [1].

Social media platforms provide a convenient channel for users to create, access and share various information. As the use and convenience of social media increases, more and more people are searching for it and receiving timely news information online. For example, the Pew Re-Research Center reported that approximately 68% of adults in the United States obtained news from social media in 2018, while only 49% reported seeing news on social media in 2021. Fake news has harmful societal effects. First, it significantly weakens the public's trust in governments and the press. For example, the reach of fake news during the 2016 US presidential election campaign for the top 20 fake news stories was, ironically, greater than the top 20 true stories discussed. Second, fake news may change the way people respond to the legitimacy of news [2]. A study showed that people's trust in the media had a significant impact and significantly deteriorated across different age groups and political parties. Third, "online" fake news can lead to "offline" community events. Therefore, It has become crucial to be able to stop fake news from spreading on social media and increase trust across the board in the news industry. However, identifying bogus news on social media poses particular difficulties. First, because fake news is created with the intent to deceive readers, it is not easy to spot fake news based on its tent. Second, social media data is extensive and mostly user-generated multimedia, sometimes anonymous and noisy. To address these challenges, recent research developments are collecting social users' posts in news articles to help deduce fake articles, giving some promising early results [3]. Generally, any machine learning-based detection scheme involves diverse phases [4], as shown in Fig. 1.
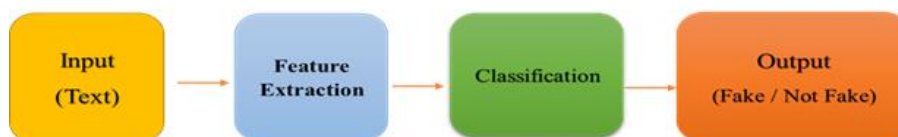

Fig. 1 - The general structure of the fake news detection-based algorithms.

2. Related Works

Fake news (also known as hoaxes, rumors, etc.) and the spread of misinformation have dominated the news cycle since the US presidential election (2016). Social media and online social networking sites, for example, Facebook and Twitter, have faced scrutiny due to their inability to curb the spread of false news [5]. There are many motives for activating and spreading fake news for political gain, damage to business reputation, earning clicks to increase ad revenue, and attention-seeking [6]. As a realistic instance, recently, Starbucks has been a victim of counterfeit advertisements for providing coffee for free to immigrants. Starbucks worked fast for denying this declaration on social media via replying to individuals, however, this claim was spread at lightning speed on the sites of social media. The severity of the issue and the requirement for developing improved techniques to cope with this challenge have been highlighted by online social networks. Additionally, Facebook has recently declared a series of attempts concerning these issues [7]. Over the last years, distinct research have been performed on the detection of fake news using the techniques of machine learning.

In 2019, According to the article, Shubham Bauskar et al.[8] used To create a special machine-learning model for recognizing "fake news," based on Natural Language Processing (NLP) methods, we consider both content-based and social components of news. On a typical dataset, the proposed model performed admirably, has a 90.62 percent average accuracy and a 90.33 percent F1 score. In this paper, Ray Oshikawa et al. [9] addressed the problem of detecting fake news as well as associated topics. The issue formulations, datasets, and NLP solutions established for this purpose, as well as their strengths and drawbacks, were thoroughly examined in this paper. More fine-grained, detailed, fair, and practical detection models were highlighted as potential research avenues based on the findings.

This research also highlights the contrast between fake news detection and other related problems, as well as the importance of NLP solutions in false news detection. The data was separated into three categories: assertions, whole articles, and social networking data.

In 2020 Hua Yuan et al. [10] suggested the "domain-adversarial and graph-attention neural network" (DAGA-NN). Its fundamental benefit is that in a text environment with numerous events and domains, it just requires a small quantity of partial domain sample data to train the model. Extensive testing was done on two Twitter and Weibo multimedia datasets XINYI ZHOU et al. [11] carried out this investigation in order to detect bogus news. The technique examines news material on several levels, such as vocabulary, grammar, semantics, and discourse. They transmit breaking news on all levels using well-established social and forensic psychology ideas. Then, to identify fake news, a supervised machine-learning architecture is deployed. Their study explores future fake news trends, improves the interpretability of false news feature engineering, and investigates the links between fake news, deception/disinformation, and click baits as multidisciplinary research. Experiments on two real-world datasets show that the proposed method exceeds existing best practices and can detect false news early, even with insufficient content information. Iftikhar Ahmad et al. [12] Several algorithms have been proposed in the field of machine learning, and their results have varied in terms of analyzing and detecting fake news. Z Khanam et al. [13] utilized a Python library named "scikit-learn" to tokenize the text data and then extracted features. After that, the methods of feature selection were implemented. Finally, the conventional models were utilized of machine learning to categorize fake news as false or true.

In 2021 Arianna D'Ulizia et al.[14] compared twenty-seven typical datasets for false news identification and provided insights into their properties. A collection of eleven features obtained from the datasets under consideration, as well as a set of comparison and production criteria, are included in a characterization of fake news detection datasets. The findings of the investigation will be useful to many academics in directing the selection or construction of acceptable datasets for testing their fake news detection algorithms, given the continued interest in the problem. Amila Silva et al. [15] presents an early false news detection scheme that gives various relevance levels for the cascades and nodes in the networks of propagation and rebuilds the knowledge of whole networks depending on their partial networks. This presented scheme was implemented using two public datasets (GossipCop.4 and PolitiFact3), and the obtained outcomes show that this scheme outperforms state-of-the-art performance.

3. The Proposed Detection Scheme

The proposed machine learning-based detection scheme of fake news involves diverse phases; Dataset preprocessing phase (normalization and tokenization), extracting the features' phase Frequency-Inverse Document Frequency (TF-IDF), and Naive Bayes (NB) and K-nearest Neighbor are used in the categorization step (KNN). The diagram in Fig. 2 demonstrates the general structure of the proposed scheme.
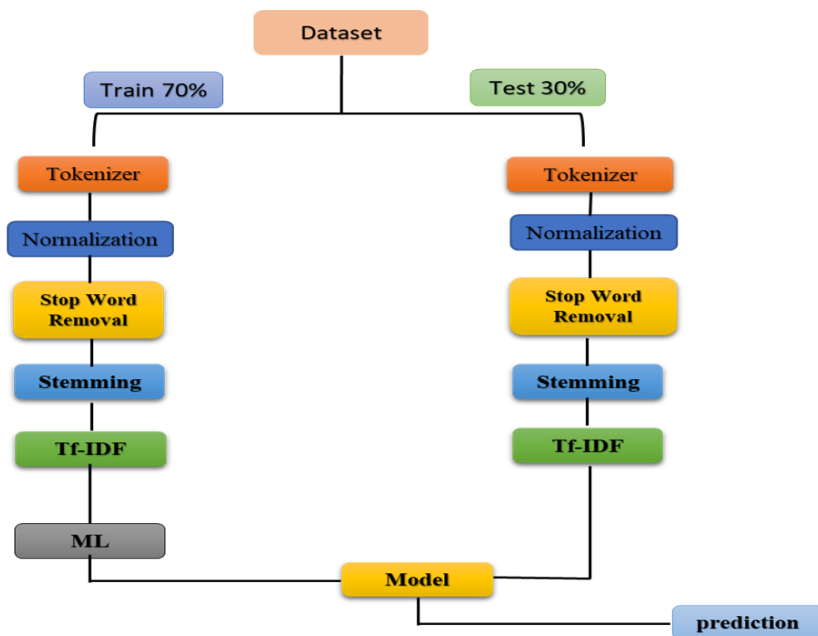
Fig. 2 - Diagram of fake news categorization.

3.1. Pre-processing

Social media data is largely unstructured; the majority of it consists of casual communications with mistakes in grammar, slang, and other areas. It is essential to establish methods for resource utilization so that judgments may be made with knowledge due to the hunt for enhanced performance and dependability. Prior to being used for predictive modeling, the data must be cleaned to produce better insights. For this reason, The News training data underwent simple pre-processing. This procedure comprised: Data on cleaning When reading data, it can be received in a manner structured or unstructured. When the format represents structured data, then it holds an obviously established pattern, and when it represents unstructured data, then it lacks the right framework. We have semi-structured data and it is approximately better than the unstructured format. In order to emphasize on the attributes that the utilized machine learning models to recognize, the input text should be clean. Pre-processing (data cleaning) includes several processes:

- Tokenization: It is the process of separating a stream of text into tokens, which can be words, phrases, symbols, or other significant pieces of information. The goal of tokenization is to examine each word in a sentence. For additional processing, like parsing or text mining, the list of tokens serves as the input. Tokenization is helpful in linguistics (where it is a type of text segmentation), computer science (where it is a component of lexical analysis), and other fields as well. Textual data only consists of a beginning block of characters. The words in the data set are needed for every phase in information retrieval. Tokenizing documents is therefore necessary in order for a parser to function. It might seem unnecessary because the text is already stored in machine-readable formats. However, several issues still remain, such as the omission of punctuation. Additionally, characters like brackets and hyphens must be processed. The tokenizer can also account for document consistency. Tokenization is mostly used to find significant terms. There may be a discrepancy due to varying time and number formats. Another issue is the need to convert acronyms and abbreviations into a uniform format. separating the split text into units like phrases or words (tokenization). It provides previously unstructured text structure. for instance: Plata o Plomo-> "Plata," "o," "Plomo.".

- Normalization: This stage requires numerous preprocessing operations that must be carried out on each word; these operations are detailed in the following:

1. Changing Uppercase to Lowercase: The case folding operation is the key concern at this point in the data pretreatment process. Case-Folding is a method for converting all uppercase characters in a word to lowercase, allowing comparisons to be made regardless of case. All news text, including that found in the corpus and that to be classified as phony or real, is changed to lowercase. This process is necessary since the news corpus's entire collection of text needs to be represented in a uniform manner. The matching algorithm will not be able to match two words that have the same meaning, are used in the same context, but have distinct representational formats, and will instead treat them as separate terms. For instance, "Automobile" and "automobile" will be treated as two different terms because of how they are represented, which is incorrect. As a result, altering "Automobile" to a lower case "automobile" will improve the matching process. Case-folding operations have a significant disadvantage because many proper nouns are derivations of common nouns and are only distinguishable through cases. For instance, even if "General Motors" and "General Motors" refer to the same business, the first term needs to be treated differently than the second. But after using the case-folding process, both words will have the same structure and be treated in a comparable way. Case folding actions cause the proper noun's information to be lost.

2. Removing Special Characters: Before using the classification technique, special characters should be deleted because they are meaningless during content-based matching. Although some context-based information may be lost when special characters like "$" are removed, content-based information is not lost. For instance, suppose news item A states, "The cost of a book is $100." Now imagine that preprocessing has eliminated the special character "$" from this sentence. The currency information will be lost, and the preprocessed phrase "cost of an item is 100" may now be referring to dollars, rupees, or any other currency.

3. Date Format Normalization: The most important factor enhancing the efficiency of the classification model may be dates that are mentioned in the news. With the help of this information, we can establish whether the news is an original work of fiction or a derivative. In the news, the date format could be dd/mm/yy, mm/dd/yyyy, or dd/mm/yyyy. All of these formats ought to be combined into a single standard so that, in the event that two dates are reported in the news and both pertain to the same day, they may be compared using content-based matching. Consider the case where the dates 10.31.2018 and 31.10.18 were listed in both News A and News B. Both of these dates will match during content-based categorization if these date formats are not normalized to a standard form that will be used throughout the dataset, i.e., convert 10.31.2018 (mm.dd.yyyy) to 31.10.18 (dd.mm.yy).

- Remove stop words: Many words that appear repeatedly in texts are virtually worthless because they are only employed to connect words in a phrase. Stop words don't add to the context or content of written works, as is often accepted. Their appearance in text mining creates a barrier to comprehension of the documents' content because of how frequently they occur. Common words like "and," "are," "this," etc. are often employed as stop words. They are useless for classifying documents. So, they need to be taken out. The creation of such a list of stop words is challenging and inconsistent among textual sources. Additionally, by doing this, text data is reduced and system speed is enhanced. These terms appear in every text document but are not necessary for text mining applications. Stop words are frequent words that almost always exist in texts. They don't tell us much about our data, so we remove them. e.g.: silver or lead is fine for me-> silver, lead, fine.

- Stemming: It is the process of combining a word's several spellings into its basic form, the stem. For instance, the phrases "presentation," "presented," and "presenting" might all be distilled into the one word "present." Using the phrase "presenting" in a search signals interest in documents that also contain the words "presentation" and "presented," according to a widely used text processing method for information retrieval (IR). A word can be reduced to its stem form by stemming. Treating terms that are related similarly often makes sense. It suffices to delete, like "ing," "ly," "s," etc. via a straightforward rule-based methodology. Although the number of words decreases, the actual words are frequently overlooked. for instance, Entitling, Entitled -> Entitle Be aware that certain search engines consider words with the same stem to be synonyms.

3.2. Feature Extraction

A word's relative frequency in a document vs its frequency across all papers is determined by the TF-IDF. The TF-IDF displays the relative weight of a term in the text and across the entire corpus. The word "term frequency,"

sometimes known as "TF," refers to how frequently a term appears in a document. A term may appear more frequently in a long text than in a short one due to the difference in document sizes. Thus, the length of the document often divides Term frequency.

$$TF(t, d) = \frac{\text{Number of times t occurs in document 'd'}}{\text{Total word count of document 'd'}} \qquad (1)$$

IDF stands for Inverse Document Frequency, which states that a word is not very useful if it appears in all of the documents. The words "a," "an," "the," "on," and "of," for example, are used frequently in documents yet have little meaning. IDF places a lower value on these phrases and a higher value on uncommon ones. IDF has a uniqueness factor that increases with its value.

$$IDF(t, d) = \log_e \left(\frac{\text{Total number of documents}}{\text{Number of documents with term t in it}} \right) \quad (2)$$

The relative count of each word in the sentences is recorded in the document matrix after TF-IDF is applied to the body text.

$$TFIDF(t, d) = TF(t, d) \times IDF(t) \qquad (3)$$

3.3. Classification Using Machine Learning

The Bayes theorem-based NB classification strategy, which makes the assumption that each characteristic in a class exists independently of each other, is used initially in this step. It offers a method for figuring out the posterior probability. The estimation of one feature distribution is totally disconnected from the estimation of others, which is the beauty of the naive Bayes technique. The effectiveness of Bayesian classifiers has been demonstrated in numerous real-world applications, such as text categorization, medical diagnosis, and system performance management. They assign the most likely class to a given example that is defined by its feature vector:

$$P(c \backslash x) = \frac{P(x \backslash c) P(c)}{P(x)} \qquad (4)$$

Where $P(c|x)$ denotes the posterior probability of the class given the predictor, $P(x|c)$ the likelihood of the predictor given the class, $P(c)$ the prior probability of the class, and $P(x)$ the prior probability of the predictor.

While the KNN classification technique is considered a lazy learner due to it does not have a training phase and performs well if all the data have the same scale. The simplicity of the KNN concept has made it a favorite tool for classification in different applications. As an example, for classifying "Si" which is a sample; Firstly, the strategy investigates its KNN in the space of features using the defined distance and the vectors of the features. Then this strategy performs "voting" for the neighbors concerning their labels. The sample of object can be categorized into a group with the highest number of same-label neighbors. There are multiple distance metrics utilized in the KNN technique; the finest metrics are the Euclidean and the Manhattan distance:

$$D(X_u, X_n) = \sqrt{\sum mr} = 1(ar(X_u) - ar(X_n))^2 \qquad (5)$$

Where X indicates the input and is known by an m-dimensional vector of features (a1, a2, a3…an), and the query vector Xu, Xn is the index vector in training data, where are indicates the number value of the feature of the vector X.

4. Evaluation and Assessment

Data for false news can be acquired from a variety of sources, including news agency websites and various social media platforms including Twitter, Facebook, and Instagram. Interest, manually separating the various types of news is difficult. So it is necessary to have an expert annotator who can evaluate the arguments, supporting data, and historical context from reliable sources. In general, there are numerous ways to gather news data, including professional, fact-checking websites and crowdsourcing employees. There are only so many benchmark datasets

that can be used to identify fake news issues. It makes use of the dataset (fake news articles.CSV file) gathered from kaggel.com. This dataset contains roughly 45,000 records that were gathered from numerous online publications, and its characteristics include (text, author, title, and label). After the missing rows were removed, there were 24,919 records in total, 21,417 for legitimate news and 23,502 for false news [16]. Only one attribute (text) is used to detect fake news.

In this paper, several measures of evaluation are used (F1-measure (F), Accuracy (A), Recall (R), and Precision (P)). The evaluation measures are computed by utilizing a confusion matrix in which the TP (true positives) indicates the news that is predicted as fake and is fake news, TN (true negatives) indicates the news that is predicted as not fake and is not fake, FP (false positives) indicates the news that is predicted as fake, however, it is non-fake, and FN (false negatives) indicates the news that is predicted as non-fake, however, it is fake. These utilized measures are specified as follows;

The measure of Accuracy (A) indicates the ratio of rightly categorized instances of not fake and fake news over entire rightly and non-rightly categorized instances, as in the following Equation;

$$A = \frac{TN+TP}{TN+TP+FN+FP} \qquad (6)$$

P indicates the ratio of news rightly specified as fake over entire fake (positive) news instances, as in the following Equation;

$$P = \frac{TP}{TP+FP} \qquad (7)$$

R indicates the sensitivity of the model, means how well the model can identify the positive instances (fake) of news, as in the following Equation;

$$R = \frac{TP}{TP+FN} \qquad (8)$$

F is calculated using the weighted average of the measures (R and P), as in the following Equation;

$$F = 2 * \frac{P \times R}{P+R} \qquad (9)$$

Table 1 presents the results of applying a data set to two machine learning techniques (NB and KNN) using the evaluation measures.

Table 1 - The obtained Precision, Recall, F1-score.

Proposed Scheme	Precision	Recall	F1-score	Accuracy
NB	0.94	0.94	0.94	0.94
K-NN	0.87	0.68	0.71	0.68

The attained results illustrated that NB is the suitable performing technique of machine learning via realizing the highest accuracy of 94%.

Furthermore, Table 2 presents the results of the previous works, while Fig. 3 illustrates the bar chart of the previous works performance, and the attained results illustrate that adopted NB classifier has better performance comparing with other related techniques of machine learning.

Model	Naïve Bayes	KNN

		Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score	Accuracy
Author	Shubham Bauskar [8]	0.92	0.90	0.87	0.90	-	-	-	-
	Ray Oshikawa [9]	0.88	0.88	0.88	0.88	0.78	0.78	0.78	0.78
	Z. Khanam [13]	-	-	-	0.68	-	-	-	0.70
	Iftikhar Ahmad [12]	-	-	-	-	0.91	0.87	0.89	0.68

Table 2 - The obtained Precision, Recall, F1-score from previous research.

The resul... matrix, as illustrated in ...
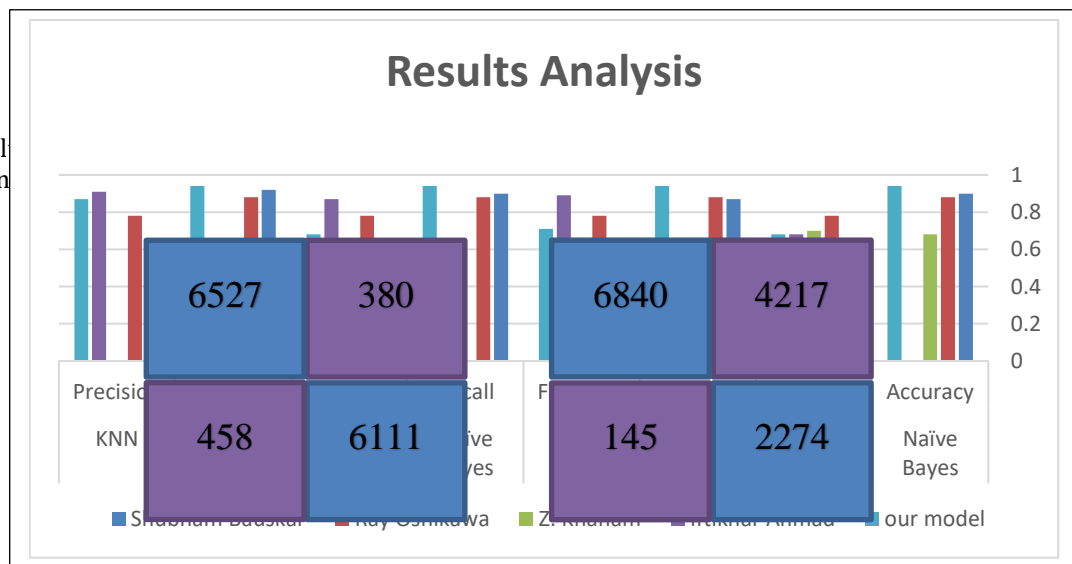
Fig. 4 - Confusion matrix, (a) NB, and (b) K-NN.

5. Conclusion

The process of manually classifying news requires in-depth knowledge of dominance and recognition of anomalies in the text. In this paper, we discuss the problem of classifying fake news articles using machine learning techniques. The utilized dataset is collected from the World Wide Web and contains news articles from various fields to cover most of the news rather than categorizing political news specifically. The aim of the research is to identify patterns in the text to distinguish fake articles from real news. Various text features were extracted from the articles (trained and parameterized learning models) to obtain optimum accuracy. Two commonly utilized techniques of machine learning were used (NB and K-NN) in the presented scheme, and the NB classification technique achieved an accuracy higher than K-NN classification technique.

References

[1] A. Albahr and M. Albahar, "An empirical comparison of fake news detection using different machine learning algorithms," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 9, (2020), pp. 146-152

[2] H. Allcott and M. Gentzkow, "Nber Working Paper Series Social Media and Fake News in the 2016 Election," J. Econ. Perspect., vol. 31, no. 2, (2017), pp. 211-236.

[3] K. Shu, L. Cui, S. Wang, D. Lee, and H. Liu, "Defend: Explainable fake news detection," Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., (2019), pp. 395-405.

[4] A. J. Abdullah, T. M. Hasan, and J. Waleed, "An Expanded Vision of Breast Cancer Diagnosis Approaches Based on Machine Learning Techniques," 2019 International Engineering Conference (IEC), (2019), pp. 177-181.

[5] F. Tchakounté, A. Faissal, M. Atemkeng, and A. Ntyam, "A reliable weighting scheme for the aggregation of crowd intelligence to detect fake news," Inf., vol. 11, no. 6, (2020), pp. 1-25.

[6] X. Zhou, A. Jain, V. V. Phoha, and R. Zafarani, "Fake News Early Detection," Digit. Threat. Res. Pract., vol. 1, no. 2, (2020), pp. 1-25.

[7] W. Han and V. Mehta, "Fake news detection in social networks using machine learning and deep learning: Performance evaluation," Proc. - IEEE Int. Conf. Ind. Internet Cloud, ICII 2019, (2019), pp. 375-380.

[8] S. Bauskar, V. Badole, P. Jain, and M. Chawla, "Natural Language Processing based Hybrid Model for Detecting Fake News Using Content-Based Features and Social Features," Int. J. Inf. Eng. Electron. Bus., vol. 11, no. 4, (2019), pp. 1–10.

[9] K. D. Patel, "Fake News Detection on Natural Language Processing A Survey," Int. J. Comput. Sci. Eng., vol. 7, no. 9, (2019), pp. 115-121.

[10] H. Yuan, J. Zheng, Q. Ye, Y. Qian, and Y. Zhang, "Improving fake news detection with domain-adversarial and graph-attention neural network," Decis. Support Syst., vol. 151, no. 2020, (2021), pp. 113633.

[11] X. Zhou and R. Zafarani, "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities," ACM Computing Surveys, vol. 53, no. 5, (2020), pp.1-40.

[12] Iftikhar Ahmad, Muhammad Yousaf, Suhail Yousaf, and M. O. Ahmad, "Fake News Detection Using Machine Learning Ensemble Methods", Hindawi, vol. 2020, (2020).

[13] Z. Khanam, B. N. Alwasel, H. Sirafi, and M. Rashid, "Fake News Detection Using Machine Learning Approaches," IOP Conference Series: Materials Science and Engineering, vol. 1099, no. 1, (2021), pp. 1-13.

[14] A. D'Ulizia, M. C. Caschera, F. Ferri, and P. Grifoni, "Fake news detection: A survey of evaluation datasets," PeerJ Computer Science, vol. 7, (2021), pp. 1–34.

[15] W. H. Bangyal et al., "Detection of Fake News Text Classification on COVID-19 Using Deep Learning Approaches," Computational Intelligence for Health Care, vol. 2021, (2021), pp. 1–14.

[16] P. JAIN, "Fake News Detection." kaggle datasets download -d jainpooja/fake-news-detection.