

Text Analytics Solutions for the Control of Fake News: Materials and Methods

Emeka Ogbuju, Taiwo Abiodun, and Francisca Oladipo

Abstract— The increase in the rate of internet and social media use has given rise to a lot of fake news and misinformation available online. The internet and social media have made information and communications flow to be faster and easier. On the other hand, the internet and social media have also jeopardized the authenticity of the news that is being sent online, as it has given people the opportunity to intentionally spread fake news. This has caused a lot of social and national damage with destructive impacts. Hence, there is a need to apply data mining and text analytic techniques in the detection of fake news across news agencies that operate online. Literature has shown that the use of data mining and text analytic techniques can play important role in both the detection of fake news and the blockage of it. The leading data mining and text analytic techniques used in fake news detection are described in this paper by answering three (3) research questions from papers between 2017 to 2022 alongside recommendations for applications for newsagents. The result presents fourteen (14) techniques and twenty (20) state of the arts datasets for fake news research.

Keywords—Fake news, text analytic, data mining, machine learning algorithms, big data.

I. INTRODUCTION

Fake news are false stories that are spread to influence people's view [1]. There are lots of negative impact of fake news on individuals and the community at large as it can be spread to damage the reputation of an individual or even an organization. There are numerous online platforms where fake news can be spread, such platforms includes Twitter, Facebook, WhatsApp, Telegram, Instagram etc. Detecting these fake news has been a great challenge. Over the years, the spread of fake news online has been on the increase [2]. Although there has been a significant progress in the detection of fake news, but researchers are yet to establish a concrete solution to this problem. Many have explored the use of machine learning to detect fake news. Machine learning is an aspect of AI that can perform different actions based on what it has learnt [3]. There are lots of machine learning algorithms which can either be supervised, unsupervised or reinforcement machine learning algorithms [4]. These algorithms can be trained using dataset, and are applied to perform different task in different sectors. These algorithms are mostly used for prediction and detention purpose. Machine learning algorithms have been performing very well in the detection of fake news, as lot of researchers

have been using these algorithms to detect fake news. Also, some researchers have applied data mining techniques in the detection of fake news. Data mining is an aspect of machine learning which is concerned with the analysis of large amount of data in order to discern patterns and trends. Since internet users' produces big, noisy, incomplete and unstructured data, data mining techniques can be applied to mine these data before applying machine learning models. Additionally, deep learning techniques have been adopted by lot of researchers to detect fake news. Deep learning is an aspect of machine learning which uses multiple layers to extract features from input.

In this study, three (3) research questions will be answered. The paper will provide the importance of data mining techniques to detecting fake news online. Different machine learning algorithms that has been used by researchers to detect fake news and how well these machine learning algorithms have performed will also be discussed in this paper. Different fake news dataset that are being used for fake news detection will be discussed as well.

II. METHODOLOGY AND RESEARCH QUESTIONS

The method used in this paper is a systematic literature review. We reviewed papers from 2017 to 2022 in order to provide answers to the research questions. The inclusion and the exclusion criteria for this study are provided in the table below.

Table I: Inclusion and Exclusion criteria

Inclusion Criteria	Exclusion Criteria
Paper is written in English language.	Paper is not written in English language.
Paper is open source.	Paper is not open source.
Paper is relevant to machine learning.	Paper is not relevant to machine learning.
Paper is relevant to fake news detection.	Paper is not relevant to fake news detection.
Paper is relevant to data mining.	Paper is not relevant to data mining.
Papers from the year 2017 to 2022.	Papers are not from the year 2017 to 2022.

The papers reviewed in this study were considered based on the research work presented by these papers. Papers that discussed the application of data mining and machine learning techniques to fake news detection, and papers

where the state of art fake news dataset are discussed were considered as papers with good quality to be reviewed in this study. Papers from different online repositories were assessed. They were separated based on their titles and the contents of their abstracts; the papers that are relevant to the research questions were studied and reviewed.

Based on valid arguments, three research question were answered. The questions include:

RQ1: Why are data mining, machine learning and deep learning techniques needed for fake news detection?

RQ2: What are the data mining, machine learning, deep learning techniques that are being used for fake news detection?

RQ3: which dataset are being used for fake news detection?

The answers to these questions will be provided in the result and discussion part of this paper.

III. RESULT AND DISCUSSION

A. Answers to RQ1: Why are data mining, machine learning and deep learning techniques needed for fake news detection?

Fake news is a threat to our society, economy and democracies, and spreading fake news has an extremely negative impact on individuals and the society at large. For this reason, detecting fake news has become an emerging research that attracts lots of attentions [5]. Fake news is sent online intentionally to mislead internet users in believing false information, thus, it is difficult to detect it based on news content, as the content is diverse in styles, media platforms and topics, and may contain may also contain real news cited within fake news, therefore auxiliary information like social engagements and knowledge base can also be applied to improve the quality of fake news online. These auxiliary information produces big data, incomplete data, unstructured data and noisy data [5]. Thus data mining techniques which can be used to extract meaningful information from large amount of data is needed for decision making in fake news detection. Data mining is an aspect of machine learning, which is centered on exploration of data through unsupervised learning. With the use of data mining, machine learning or deep learning techniques, detecting fake news can be done automatically [6]. Machine learning and deep learning models can check context of a news and immediately detect if the news is real or fake. There are lots of machine learning techniques that could be used to detect fake news. Researchers have used different machine learning and data mining techniques in the detection of fake news, and the accuracy of these techniques depends on the way the models are being trained [7]. The accuracy is important because failure of a model to detect fake new can affect different people negatively. The state of art machine learning algorithms that are being used for fake news detection are reviewed in the next question.

B. Answers to RQ2: What are the data mining, machine learning, deep learning techniques that are being used for fake news detection?

Ordinarily, detecting fake news is a difficult task, but with the help of machine learning techniques, it can be detected easily. Researchers have used a lot of machine learning algorithms to detect if a news is fake or real. There are different machine learning techniques that can be used to detect fake news. [8] combined network metadata and sentiment analysis to classify real news and fake news. The Random Forest classifier algorithm was used to train the model. The FakeNewsCorpus and the GFRN dataset was used by the researchers. The authors used a scrapping tool to gather information that are related to the news and leverage four different sub-pipelines for feature extraction and feature engineering. The Random Forest classifier algorithm was able to achieve a F1 score above 88%. The author was able to develop a web interface that can take the web address of a news and display if the news is fake or not. In another study, different approaches to detecting fake news with the use of text mining techniques and text features was discussed by [9]. The researchers proposed an ensemble method that combine feature from set3 and word vector for the classification of fake news on FakeNewsNet dataset and McIntire dataset. The researchers concluded in their report that an ensemble method of stylometric features and text based word vector can predict fake news with an accuracy above 95.48%. likewise, [10] used six different machine learning algorithms to detect fake news based on text analytics. These algorithms were used on Liar dataset which is a benchmark dataset for fake news detection. The algorithms used includes; Random Forest, XGboost, KNN, Naïve Bayes, SVM and Decision Tree. From the result, the XGboost out perform all compared algorithm with an accuracy that is greater than 75%. The Random Forest and Support Vector Machined approximately gave 73% accuracy. In another work, [11] carried out a research on the detection of fake news using machine learning techniques. The researchers used four different open source fake news dataset from kaggle. The machine learning techniques used are Logistics Regression, SVM, KNN, Random Forest, Bagging Ensemble Classifiers, Voting Ensemble Classifier and Deep Learning algorithms like Linear SVM, Multilayer Perception, CNN and Bidirectional LSTM network. The performance of each model varies on each dataset. The Random Forest algorithm and Linear SVM achieved the maximum accuracy of 99% on the ISOT fake news dataset. The Multilayer Perception, Bagging Classifier, Boosting Classifier and Linear SVM were 98% accurate on Hereafter dataset. [12] used machine learning techniques to detect fake news in Korean text. The dataset used contain articles crawled from maeil businesses, Hankyoreh, Chosun Ilbo, Joongang Ilbo and Dong-A Ilbo. The dataset was divided into two parts, less than 40% of the dataset was named mission1 and more than 60% was named mission2. The algorithms used were trained separately on mission1 and mission2. The algorithms used are BCNN, LSTM+BCNN, Bi-LSTM+BCNN and BCNN with attentive pooling

similarity (ASP-BCNN). The models performed better on the larger dataset (mission2 dataset) with accuracies above 70%. The APS-BCNN model has the best performance with an accuracy of 72.6% on mission2, while the Bi-LSTM+BCNN has the least performance of 70.7%. A recent study with the aim of detecting fake news from online channels using machine learning algorithms was carried out by [13]. The study used Naïve Bayes, decision tree, KNN, Logistic Regression, and SVM classifiers to detect fake news in a dataset from Github's repository. The models Logistic Regression model outperformed all other compared models with an accuracy of 90.46%. The KNN obtained an accuracy of 89.98%, The SVM was 89.33% accurate and the decision tree was 73.33% accurate. Also, a machine learning based fake news detection model was developed by [14]. A logistic regression classifier was used to classify fake news. The open source Liar dataset was used to train and test the model. The logistic repressor classifier was 98% accurate with 98% precision. [15] developed a model that can detect fake news in Indonesia. The researchers used XGBoost to classify fake news and real news. The dataset used by the respecters contains real and fake news about Indonesia and the world at large from 2015 to 2020. The dataset contains 500 news which includes 250 fake and 250 real news written in Indonesia. The XGBoost algorithm was able to achieve 89% accuracy and 90% precise. Also, a model that can detect fake news was also developed by [16]. These researchers used four deep learning algorithms which are; CNN, unidirectional LSTM, Bi directional LSTM and Vanilla RNN. The researchers used two publicly available dataset containing fake and real news. The bi directional LSTM algorithm outperform other algorithms that was used on both datasets. [17] compared the performance of different machine learning algorithms in the detection of fake news. The algorithm compared are Naïve Bayes, Decision Tree, Random Forest, Stochastic Gradient Descent, KNN, XGBoost, and Logistic Regression. The fake news dataset used by the researchers was gotten from Kaggle and contains four attributes which includes; title, text, subject and date. The algorithms were evaluated based on accuracy, F-score and run time. The Logistic Regression outperforms other compared algorithms in terms of accuracy which was 99% accurate. In terms of F-score, Random Forest Performed best with an F-score of 99%. And Naïve Bayes has the best runtime with runtime of 0.11s. [18] conducted a survey on the detection of fake news on social media. The study reviewed existing data mining approaches to fake new detection, such approaches include future extraction and model construction. The researchers proposed the use of fake news benchmark dataset like BuzzFee, LIAR, BS Detector and GREDBANK datasets. There are other datamining techniques that have not been used by any researcher to detect fake news, but should perform well when used in the detection of fake news.

From the above reviewed papers, we provide a description of both machine learning and deep learning techniques that are commonly in use for the detection of fake news. Basically, we found out that both machine learning and deep

learning algorithms have been in use for the detection of fake news and that the most frequently used ones are the fourteen (14) described above as seen in resent literature.

Machine Learning techniques for fake news detection.

- 1) XGboost: This ensemble machine learning algorithm can be used to detect if a news is fake or not. It has been used by researchers like [10], [15] and [17] to detect fake news and it has performed well.
- 2) K-Nearest Neighbor: This supervised machine learning algorithm works by classifying data based on similarity to its nearest neighbor. The algorithm has performed well in the detection of fake news in [10] – [11], [13], [17] and others.
- 3) Naïve Bayes: This algorithm can also be used to classify if a news is fake or real. It has been used by a lot of researchers lie [10], [13], [17] and others and has proven to be a good algorithm in the detection of fake news.
- 4) Support Vector Machine (SVM): The SVM is a machine learning algorithm and it is mostly used for classification purpose. Many researchers like [10] – [13] and others have used it to classify fake news, and it has performed excellently.
- 5) Logistic Regression: This has been used by many researchers like [11], [13], [14] and [17] and, it has achieved high accuracy.
- 6) Decision Tree: This machine learning algorithm can be used in fake news detection. It works by breaking dataset into different smaller subsets. Researchers like [10], [13], and [17] have used it in the detection of fake news and it performed well.
- 7) Random Forests: This classifier algorithm is an ensemble of different Decision Tress and it has performed greatly in the detection of fake news. Researchers like [8], [10], [11], and [17] have used it in detection of fake news.

Deep Learning techniques for fake news detection.

- 1) Multilayer Perception: This is a deep learning algorithm that uses more than one hidden layer unlike the single layer perception. This algorithm has been used to detect fake news by researchers like [11].
- 2) Convolutional Neural Network (CNN): This is a deep learning model that can also be used to classify fake news; it has been used by researchers like [11] and [16]
- 3) Bi Convolutional Neural Network (BCNN): This deep learning algorithm is a CNN with two input and retrained word embedding. It uses 3g filters to extract feature maps from bodies and headlines. It has been used by [12] to detect fake news and it performed well.
- 4) Long Short-Term Memory (LSTM): This is another deep learning algorithm that can be used for fake news detection also known as the unidirectional LSTM. It has been used by different researchers like [11], [12] and [16], and it has performed very well.
- 5) Bidirectional Long Short-Term Memory (Bi-LSTM): unlike the unidirectional LSTM, the Bi-LSTM goes through the input sequence in two directions (right and left) at the same time. It has also been used by different

researchers like [12] and [16], and it has performed very well.

- 6) LSTM + BCNN: This is an ensemble of Long Short-Term Memory and Bi Convolutional Neural Network. These method has performed greatly in detecting fake news as it has been used by researchers like [12].
- 7) Bi-LSTM + BCNN: Bidirectional Long Short-Term Memory and Bi Convolutional Neural Network. It has also been used by [12], and it has performed well.

Data mining techniques for fake news detection.

- 1) Future extraction: A data mining techniques that can be used to extract useful attributes from social contents. The features could be source, headline, body text, images or videos. This method was proposed by [18].
- 2) Data cleaning and preparation: This is an important part of data mining process that deals with the formatting of data to make it useful. [8] and [10].
- 3) Data visualization: This is another important process in data mining. This process grant insight to users based on what the user can see. [8], [10], [11] and [17]

C. Answers to RQ3: Which dataset are being used for fake news detection?

Multiple Domain

A publicly available dataset known as LIAR dataset was published by [19]. This researcher collected 12.8k short statements that are manually labelled from POLITIFACT.COM. The statements collected contains detailed analysis reports and links to source documents for each cases. As at the time the LIAR dataset was published, it was the largest publicly available fake news dataset.

RealNews dataset is a fake news dataset that was introduced by [20]. It contains a large amount of news articles which were collected from Common Crawl. The authors used the dataset to classify fake news from real news.

Another fake news dataset is the FakeNewsNet which was introduced by [21]. This dataset was collected from Politifac and GossipCop websites. These websites contain social context, news content and dynamic information.

[22] introduced Fake news challenge stage1 (FNC-1) dataset. This fake news dataset contains 75,385 article pairs and labelled headline. The article pairs in the dataset are labeled as either unrelated, discuss, disagree or agree. And the headlines in the dataset are phrased as statement.

The Snopes dataset which was introduced by [23]. The dataset contains fact checking articles (multimodal tweets and FC article) from Snopes.com. The Politifact dataset was introduced by [23]. Just like the Snopes dataset, it also contains fact-checking articles (multimodal tweets and FC article) from politifact.com.

The Fakeddit is a multimodal fake news dataset that was introduced by [18]. The dataset consists of one million samples of fake news from multiple categories. These samples are labelled base on 2-way, 3-way and 6-way

classification categories.

“Some Like It Hoax” is another fake news detection dataset that was introduced by [24]. The dataset consists of 15,500 posts from Facebook and 909,236 users. The researchers used two classification algorithms on this dataset. Theses algorithms includes the Logistics Regression and the novel adaptation of Boolean crowd sourcing. They were able to obtain accuracy above 99%.

The NELA-GT-2018 is a multiple labelled dataset that was introduced by [25]. The fake news dataset consists of 714k articles that was collected between February and November 2018. The articles were collected from 194 media outlet including hyper-partisan, mainstream and conspiracy sources.

The NELA-GT-2019 dataset was introduced by [26]. This dataset is an updated version of NELA-GT-2018 dataset. It contains 1.12 million news articles, these articles are from 260 different sources and are collected between January to December 2019.

The NELA-GT-2020 dataset was introduced by [27]. This dataset is an updated version of NELA-GT-2019 dataset. The dataset contains 1.8 million news articles, these articles are from 519 different sources and are collected between January to December 2020.

UPFD (User Preference Aware Fake News Detection) dataset was introduced by [28]. This dataset has been integrated with deep learning library and Pytorch geometric. The dataset which includes real news and fake news from Twitter was curated for evaluating binary graph classification.

The UPFD-GOS (User Preference Aware Fake News Detection) dataset was introduced by [28]. This dataset simultaneously captured different signals from user preferences by graph modelling and joint content.

UPFD-POL (User Preference Aware Fake News Detection) dataset was introduced by [28]. Like the previous UPFD datasets, this dataset simultaneously captured different signals from user preferences by graph modelling and joint content.

The Weibo21 dataset was introduced by [29]. This fake news dataset is a benchmark for multiple domain fake news detection. The dataset contain of 4,640 genuine news and 4,488 fake news from nine domains.

Single Domain

The Covid19 fake news dataset was introduced by [30]. This dataset contains 10,700 social media articles and posts of real and fake news. The dataset was used to classify fake news from real news. The SVM, Decision Tree, Logistic Regression and the gradient boosting was used on the dataset. The SVM outperform other algorithms compared and achieved a F1 score of 93.32%.

MM-COVID was introduced by [31]. The covid19 related fake news dataset provides multilingual fake news and the relevant social context. The dataset consists of 3981 fake news and 7,192 real news contents in English, French, Italian, Hindi, Spanish and Portuguese.

Language Specific

AraCOVID19-MFH (Arabic Covid19 multiple labelled fake news and hate speech detection) dataset was introduced by [32]. The dataset which was annotated with ten different labels consist of 10,828 Arabic tweets.

BanFakeNews dataset was introduced by [33]. The dataset consists of 50,000 fake and real news in Bangla language. This dataset can be used to build automatic fake news detection systems in Bangla.

Finally, The Fake News Filipino dataset was introduced by [34]. The dataset was expertly curated for detection of fake news in Filipino. The dataset consists of 3,206 news articles including 1603 real articles and 1603 fake articles.

Twenty (20) state of the art dataset that are being used for fake news detection have been reviewed in this paper to provide answer to research question 3. The Table 2 below shows the state of the art dataset that are being used for fake news detection.

Table II: Fake News Detection Dataset.

S/N	Source	Dataset	Category
1	[19]	LIAR dataset	Multiple Domain
2	[20]	RealNews dataset	Multiple Domain
3	[21]	FakeNewsNet	Multiple Domain
4	[22]	Fake news challenge stage1 (FNC-1) dataset	Multiple Domain
5	[23]	The Snopes dataset	Multiple Domain
6	[23]	The Politifact dataset	Multiple Domain
7	[18]	The Fakeddit	Multiple Domain
8	[24]	Some Like it Hoax	Multiple Domain
9	[25]	NELA-GT-2018	Multiple Domain
10	[26]	NELA-GT-2019	Multiple Domain
11	[27]	The NELA-GT-2020	Multiple Domain
12	[28]	UPFD (User Preference aware Fake News Detection) dataset	Multiple Domain
13	[28]	The UPFD-GOS (User Preference aware Fake News Detection)	Multiple Domain
14	[28]	UPFD-POL (User Preference aware Fake News Detection)	Multiple Domain
15	[29]	The Weibo21 dataset	Multiple Domain

16	[30]	The Covid19 fake news dataset	Single Domain
17	[31]	MM-COVID	Single Domain
18	[32]	AraCOVID19-MFH	Language Specific
19	[33]	BanFakeNews dataset	Language Specific
20	[34]	Filipino dataset	Language Specific

IV. CONCLUSION

This work had shown that fake news can easily be detected using data mining techniques as lots of machine learning algorithms has been used by different researchers to detect fake news. Some of these algorithms has been reviewed in this paper and they have performed very well as seen in the works by [9] – [12] and [14]. The state of the art datasets that have been used for fake news detection were also reviewed in this work. These datasets are open source dataset which contains fake news and real news from different sources and in different languages. For local research works in the domain of fake news detection, we recommend the use of local datasets relevant to the community in focus in order to achieve an accurate and reliable result. Such local works may use the top five (5) performing machine learning algorithms, and then perform a comparative analysis to know the one that will achieve the best performance on the locally curated dataset.

REFERENCES

- [1] Maciej, S. (2018). FakeNewsCorpus: A dataset of millions of news articles scraped from a curated list of data sources. Retrieved from github:
- [2] Zhou, X., Zafarani, R., Shu, K., & Liu, H. (2019). Fake news: Fundamental theories, detection strategies and challenges. Proceedings of the 12th ACM International Conference on Web Search and Data Mining (pp. 836-837). Association for Computing Machinery, Inc.
- [3] Donepudi, P. K. (2019). Automation and Machine Learning in Transforming the Financial Industry. Asian Business Review, 129-138.
- [4] Alzubi, Jafar., Nayyar, Anand., & Kumar, Akshi (2018). Machine learning from theory to algorithms: an overview. Journal of Physics Conference Series.
- [5] Aniyath, A. (2019). A Survey on Fake News Detection by the Data Mining Perspective. International Journal of Information and Computer Science, 9-28.
- [6] Della, V. M., Tacchini, E., Moret, S., Ballarin, G., DiPierro, M., & De Alfaro, L. (2018). Automatic online fake news detection combining content and social signals. Proceedings of the 22st Conference of Open Innovations Association FRUCT (pp. 272–279). IEEE.
- [7] Kurasinski, L., & Mihalescu, C. (2020). Machine Learning explainability in text classification for Fake News detection. 19th IEEE International Conference on Machine Learning (pp. 775-781). IEEE.
- [8] Maniz, S. (2018). Detecting FAKE News with Sentiment Analysis and Network Metadata. Earlham Historical Journal.
- [9] Reddy, H., Raj, N., Manali, G., & Basava, A. (2020). Text-mining-based Fake News Detection Using Ensemble Methods. International Journal of Automation and Computing, 210-221.

- [10] Khanam, Z., Alwaseel, B. N., Sirafi, H., & Rashid, M. (2021). Fake News Detection Using Machine Learning Approaches. IOP Conference Series: Materials Science and Engineering. IOP.
- [11] Ifitikhar, A., Muhammad Yousaf, Sukail, Y., & Muhammad, O. A. (2020). Fake NEWS Detection Using Machine Learning Ensemble Methods. Complexity, 11 pages.
- [12] Dong-Ho, L., Yu-Ri, K., Hyeong-Jun, K., & Yu-Jun, Y. (2019). Fake News detection using Deep Learning. Journal of Information Processing Systems, 1119-1130.
- [13] Shalini, P., Sankeerthi, P., Subba, R. N., & Dinesh Acharya. (2022). Fake News Detection from Online media using Machine Learning Classifiers. 1st international Conference on Artificial Intelligence, Computational Electronics and Communication System (pp. 28-30). Manipal India: Journal of Physics: Conference Series.
- [14] Ali, H. H., & Heba, Y. A. (2022). Fake News Detection Based on the Machine Learning Model. Design Engineering, 1373-1378.
- [15] Haumahu, J. P., Silvester, D. H., & Yaddarabullah, Y. (2020). Fake news classification for Indonesian news using Extreme Gradient Boosting (XGBoost). The 5th Annual Applied Science and Engineering Conference. IOP Publishing.
- [16] Pritika, B., Preeti, S., & Raj, K. (2019). Fake News Detection using Bi-directional LSTM-Recurrent Neural Network. International Conference on recent trends in advanced computing 2019, ICRTAC 2019. India: Procedia Computer Science.
- [17] Zhibin, W., & Huatai, X. (2021). Performance comparison of different machine learning model in detecting fake news. Sweden: Open Access.
- [18] Kai, N., Sharon, L., & William, Y. W. (2020). Fakeddit. Retrieved from <https://fakeddit.netlify.app/>
- [19] William, Y. W. (2017). "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. arXiv.
- [20] Rowan, Z., Ari, H., Hannah, R., Yonatan, B., Ali, F., Franziska, R., & Yejin, C. (2020). Defending Against Neural Fake News. arXiv.
- [21] Kai, S., Deepak, M., Suhang, W., Dongwon, L., & Huan, L. (2018). FakeNewsNet: A data repository with news content, social context and dynamic information for studying fake news on social media. Journal of computer science.
- [22] Dean, P., & Delip, R. (2017, June 15). Fake News Challenge Stage 1 (FNC-1): Stance Detection. Retrieved from Fake News Challenge: <http://www.fakenewschallenge.org/>
- [23] Nguyen, V., & Kyumin, L. (2020). Where Are Facts? Searching for Fact-checked Information to Alleviate the Spread of Fake News. arXiv.
- [24] Eugenio, T., Gabriele, B., Marco, L. D., Stefano, M., & Luca, d. A. (2017). Some Like it Hoax: Automated Fake News Detection in Social Networks. arXiv.
- [25] Jeppe, N., Benjamin, D. H., & Sibel, A. (2019). NELA-GT 2018: A Large Multi-Labelled News Dataset for The Study of Misinformation in News Articles. arXiv.
- [26] Mauricio, G., Benjamin, D. H., & Sibel, A. (2019). NELA-GT-2019: A Large Multi-Labelled News Dataset for The Study of Misinformation in News Articles. arXiv.
- [27] Maurico, G., Benjamin D, H., & Sibel, A. (2020). NELA-GT-2020: A Large Multi-Labelled News Dataset for The Study of Misinformation in News Articles. arXiv.
- [28] Yingtong, D., Kai, S., Congying, X., Philip, S. Y., & Lichao, S. (2021). User Preference Aware Fake News Detection. arXiv.
- [29] Qiong, N., Juan, C., Yongchun, Z., Yanyan, W., & Jintao, L. (2022). MDFEND: Multi-domain Fake News Detection. arXiv.
- [30] Parth, P., Shivam, S., Srinivas, P., Vineeth, G., Gitanjali, K., Md, S. A., . . . Tanmoy, C. (2020). Fighting an Infodemic: Covid-19 Fake News Dataset. arXiv.
- [31] Yichuan, L., Bohan, J., Kia, S., & Huan, L. (2020). MM-COVID: A Multilingual and Multimodal Data Repository for Combating COVID-19 Disinformation. arXiv.
- [32] Mohamed, S. H., & Hassina, A. (2021). AraCOVID19-MFH: Arabic COVID-19 Multi-label Fake News and Hate Speech Detection Dataset. arXiv.
- [33] Zobaer, H., Ashraful, R., Saiful, I., & Sudipta, K. (2020). BanFakeNews: A Dataset for Detecting Fake News in Bangla. arXiv.
- [34] Jan, C. B., Julianne, A. T., & Charibeth, C. (2020). Localization of Fake News Detection via Multitask Transfer Learning. arXiv.

Taiwo Abiodun - Department of Computer Science, Faculty of Sciences, Federal University Lokoja, Lokoja, Nigeria (email:taiwo4real007@gmail.com)

Francisca Oladipo - Department of Computer Science, Faculty of Sciences, Federal University Lokoja, Lokoja, Nigeria (email:francisca.oladipo@fulokoja.edu.ng)

Emeka Ogbuju - Department of Computer Science, Faculty of Sciences, Federal University Lokoja, Lokoja, Nigeria (email:emeka.ogbuju@fulokoja.edu.ng)