RESEARCH ARTICLE OPEN ACCESS

Fake News Detection Using Machine Learning and Deep Learning Classifiers

K.S.DHANRAJ
UG Student
Department Of
Computer Science & Engineering
E.G.S Pillay Engineering College
Nagapattinam,India
ksdhanraju@gmail.com

G.YOKESH
UG Student
Department Of
Computer Science & Engineering
E.G.S Pillay Engineering College
Nagapattinam,India
gyogesh200269@gmail.com

R.R.NAVEEN KUMAR
UG Student
Department Of
Computer Science & Engineering
E.G.S Pillay Engineering College
Nagapattinam,India
rrnaveenkumar07@gmail.com

Dr.M.RAJAKUMARAN, M.E., Ph.D.,
Assistant Professor
Department Of
Computer Science & Engineering
E.G.S Pillay Engineering College
Nagapattinam,India
rajakumaran@egspec.org

ABSTRACT

To read the news, most smartphone users prefer social media over the internet. The news is posted on news websites, and the source of the verification is cited. The problem is determining how to verify the news and publications shared on social media platforms such as Twitter, Facebook Pages, WhatsApp Groups, and other microblogs and social media platforms. It is damaging to society to hold on to rumors masquerading as news. The request for an end to speculations, particularly in developing countries such as India, with a focus on authenticated, accurate news reports. Fake news has spread to a larger audience than ever before in this digital era, owing primarily to the rise of social media and direct messaging platforms. Fake news detection requires significant research, but it also presents some challenges. Some difficulties may arise as a result of a limited number of resources, such as a dataset. In this paper, we survey a machine learning and deep learning technique for detecting fake news with Natural language processing steps that include text mining in the "Fake News Challenge" datasets and compare the algorithms such as Naives Bayes, Random Forest Multi-layer perceptron (MLP) model with CNN algorithms and used to determine the text accuracy value for the precision, recall, and f1 score.

INDEX TERMS: Fake News Detection, Natural Language Processing, Random Forest, Text Mining, Naives bayes classifier, Convolutional neural network

1. INTRODUCTION

Fake news detection is a difficult problem due to the nuances of language. Understanding the reasoning behind certain fake items implies inferring a lot of details about the various actors involved. We believe that the solution to this problem should be a hybrid one, combining machine learning, semantics and natural language processing. The purpose of this project is not to decide for the reader whether or not the document is fake, but rather to alert them that they need to use extra scrutiny for some documents. Fake news detection, unlike spam detection, has many nuances that aren't as easily detected by text analysis. The main objective is to classify the news data and predict the fake news using deep learning technique with Natural language processing. The news media evolved from newspapers, tabloids, and magazines to a digital form such as online news platforms, blogs, social media feeds, and other digital media formats.

Social media is a popular medium for the dissemination of real-time news all over the world. Easy and quick information proliferation is one of the reasons for its popularity. An extensive number of users with different age groups, gender, and societal beliefs are engaged in social media websites. Despite these favourable aspects, a significant disadvantage comes in the form of fake news, as people usually read and share information without caring about its genuineness. Therefore, it is imperative to research methods for the authentication of news.

Machine learning is a machine method of teaching and learning computer systems to do what humans do instinctively: learn by doing. Deep learning is an important technology behind self-driving cars, allowing them to recognise a stop sign or distinguish between a pedestrian and a lamppost. It is essential for voice control in customer devices such as phones, tablets, televisions, and hands-free speakers. Deep learning has received a lot of attention recently, and for valid reason. It is meeting expectations that were previously unthinkable. A desktop model can learn to perform tasks of classification directly from pictures, text, or sound in deep learning. Deep learning algorithms can achieve cutting-edge accuracy, sometimes outperforming humans. Deep learning is a method of instructing and training computers to perform what humans do naturally: learn by doing. Deep learning is a critical component of self-driving cars, allowing them to recognise a stop sign or distinguish between a pedestrian and a lamppost. Voice commands in client devices like phones, tablet devices, television sets, and hands-free speakers is critical. Deep learning has recently received a lot of attention, and for good reason. It is exceeding previously unthinkable expectations. In deep learning, a desktop model can learn to perform classification tasks directly from images, text, or sound. Algorithms for deep learning can achieve trying to cut accuracy, even outperforming humans in some cases. The fake news detection methods are shown in fig 1.

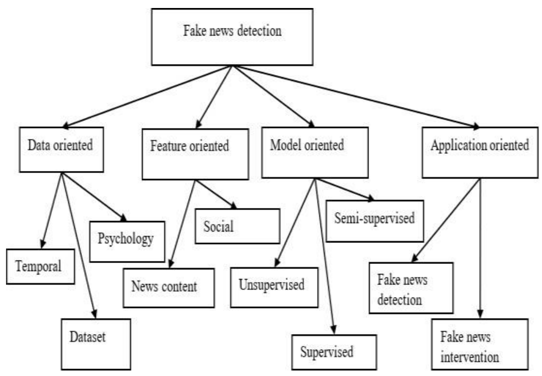

Fig 1: Fake News Detection Methods

2. RELATED WORKS

TodorMihaylov et al [1] could consistently use trolls, write fake posts and comments in public forums, thus making veracity one of the challenges in digital social networking. The practice of using opinion manipulation trolls has been reality since the rise of Internet and community forums. It has been shown that user opinions about products, companies and politics can be influenced by posts by other users in online forums and social networks. This makes it easy for companies and political parties to gain popularity by paying for “reputation management” to people or companies that write in discussion forums and social networks fake opinions from fake profiles. In Europe, the problem has emerged in the context of the crisis in Ukraine. There have been a number of publications in news media describing the behavior of organized trolls that try to manipulate other users’ opinion. Still, it is hard for forum administrators to block them as trolls try not to violate the forum rules. We have presented experiments in trying to distinguish trolls vs. non-trolls in news community forums. We have experimented with a large number of features, both scaled and non-scaled, and we have achieved very strong overall results using statistics such as number of comments, of positive and negative votes, of posting replies, activity over time, etc.

TodorMihaylov et al [2], shown that user opinions about products, companies and politics can be influenced by opinions posted by other online users in online forums and social networks. This makes it easy for companies and political parties to gain popularity by paying for “reputation management” to people that write in discussion forums and social networks fake opinions from fake profiles. Opinion manipulation campaigns are often launched using “personal management software” that allows a user to open multiple accounts and to appear like several different people. Over time, some forum users developed sensitivity about trolls, and started publicly exposing them. Yet, it is hard for forum administrators to block them as trolls try formally not to violate the forum rules. In our work, we examine two types of opinion manipulation trolls: paid trolls that have been revealed from leaked “reputation management contracts”and “mentioned trolls” that have been called such by several different people. Overall, we have seen that our classifier for telling apart comments by mentioned trolls vs. such by non-trolls performs almost equally well for paid trolls vs. non-trolls, where the non-troll comments are sampled from the same threads that the troll comments come from. Moreover, the most and the least important features ablated from all are also similar. This suggests that mentioned trolls are very similar to paid trolls (except for their reply rate, time and day of posting patterns). However, using just mentions might be a “witch hunt”: some users could have been accused of being “trolls” unfairly. One way to test this is to look not at comments, but at users and to see which users were called trolls by several different other users.

Peter Bourgonje et al [3] aim to contribute to a first step in battling fake news, often referred to as stance detection, where the challenge is to detect the stance of a claim with regard to another piece of content. Our experiments are based on the setup of the first Fake News Challenge (FNC). In FNC1, the claim comes in the form of a headline, and the other piece of content is an article body. This step may seem, and, in fact, is, a long way from automatically checking the veracity of a piece of content with regard to some kind of ground truth. But the problem lies exactly in the definition of the truth, and the fact that it is sensitive to bias. Additionally, and partly because of this, annotated corpora, allowing training and experimental evaluation, are hard to come by and also often (in the case of fact checker archives) not freely available. We argue that detecting whether a piece of content is related or not related to another piece of content (e. g., headline vs. article body) is an important first step, which would perhaps best be described as click bait detection (i. e., a headline not related to the actual article is more likely to be click bait). Following the FNC1 setup, the further classification of related pieces of content into more fine-grained classes provides valuable information once the "truth" (in the form of a collection of facts) has been established, so that particular pieces of content can be classified as "fake" or, rather, "false". Since this definitive, resolving collection of facts is usually hard to come by, the challenge of stance detection can be put to use combining the outcome with credibility or reputation scores of news outlets, where several high-credibility outlets disagreeing with a particular piece of content point towards a false claim. Stance detection can also prove relevant for detecting political bias: if authors on the same end of the political spectrum are more likely to agree with each other, the (political) preference of one author can be induced once the preference of the other author is known.

Sahil Chopra et al [4] propose a two-part solution to FNC-1. First, suggest a linear classifier to classify headline-article pairs as related or unrelated. Second, we suggest several neural network architectures built upon Recurrent Neural Network Models (RNNs) to classify related pairings as agree, disagree, or discuss. Overall, we scored a SF NC = 0.8658 which out performs the reported models on the FNC-1 Slack channel, which average .70 - .80. Moving forward, we hope to submit results on the test set for FNC-1 that will be released in June. Additionally, we are planning to perform greater qualitative analysis to determine potential strategies for correctly classifying disagree headline-article pairs and look into other potentially relevant network architectures. And additionally hope to try tuning our Bilateral Multi-Perspective Matching Model and look for more powerful GPUs on which to run all four layers of attention. In our FNC-1 models, we leveraged ideas proposed in Stance Detection with Bidirectional Conditional Encoding, where the authors used Bidirectional Recurrent Neural Networks (BiRNNs) to conditionally encode target phrases and tweets for the SemEval 2016 Stance Detection Challenge. Lastly, we implemented the Bilateral Multi-Perspective Matching Model (BiMpM) model and applied it to FNC-1. As discussed later in paper, the model takes word embeddings as inputs to a Bidirectional Siamese LSTM, applies four variants of attention on the output of the BiLSTM, feeds these attention-induced outputs through two separate BiLSTMs, concatenates the final hidden states, and uses a 2-Layer MLP for classification.

Konstantinovskiy et al [5] describes the iterative process we followed together with fact checkers to come with up an annotation schema that would effectively capture claims and non-claims. This annotation schema avoids factors that can be affected by personal biases, such as importance, in the manual annotation to produce an objective outcome. Following this annotation schema through a crowdsourcing methodology, we generated a dataset of 5,571 sentences labelled as claims or non-claims. Further, we set out to present the development of the first stage in the automated fact checking pipeline. It constitutes the first automated claim detection system developed by an independent fact checking charity, Full Fact, along with academic partners. And introduce the first annotation schema for claim detection, iteratively developed by experts at Full Fact, comprising 7 different labels. Then describe a crowdsourcing methodology that enabled us to collect a dataset with 5,571 sentences labelled according to this schema. And develop a claim detection system that leverages universal sentence representations, as opposed to previous work that was limited to word-level representations. Our experiments show that our claim detection system outperforms the state-of-the-art claim detection systems, Claim Buster and Claim Rank. With the annotation schema, crowdsourcing methodology and task definition, we set forth a benchmark methodology for further development of claim detection systems. Through leveraging the fact checkers at Full Fact, and through academia-industry collaboration, we have developed the first annotation schema for claim detection informed by experts. This has enabled us to create an annotated dataset made of sentences extracted from transcripts of political TV shows.

Jitendra Kumar Jaiswal and colleagues [6], implemented the system for built effective and improved prediction performance on the more cost-effective class variables, as well as more trustworthy data comprehension. Forest of chance has evolved into a highly effective and dependable algorithm capable of dealing with feature selection issues even when the number of variables is increased. Furthermore, it is especially useful when dealing with imputation, classification, and missing data analysis with regression. It can also effectively manage noisy data and outliers. In contrast to traditional supervised learning approaches, such as batch or online learning, which frequently require requesting class labels for each incoming instance, online active learning updates the classification model by querying only a subset of informative incoming instances. Throughout the online learning task, this approach aims to maximise classification performance while requiring the least amount of human labelling work. In this study, we offer a new family of online active learning algorithms called Passive-Aggressive Active (PAA) learning algorithms by modifying the Passive-Aggressive algorithms in online active learning situations. In contrast to standard Perceptron-based systems, which only use misclassified instances, the proposed PAA learning algorithms use both successfully classified cases with low prediction confidence and misclassified instances to update the classifier.

Janmenjoy Nayak and colleagues [9] conducted on the application of support vector machines to various data mining applications. Data mining is a promising and appealing study field due to its broad application areas and innovative nature. Tasks that are rudimentary. The Support Vector Machine (SVM) is critical because it provides strategies that are particularly well suited to obtaining results quickly and effectively while maintaining a high standard of quality. This paper examines the role of SVM in various data mining tasks such as classification, clustering, prediction, forecasting, and other applications. And specifically propose a number of PAA algorithm modifications to address three different types of online learning tasks: binary classification, multi-class classification, and cost-sensitive classification. Then provide theoretical error boundaries for the suggested algorithms and conduct extensive tests to evaluate their empirical performance on both conventional and large-scale datasets. The positive results validate the proposed algorithms' empirical effectiveness.

Aman Kataria and colleagues [10], categorise data, whether using neural networks or any biometrics application, such as handwriting classification or iris detection, machine learning techniques, such as the stockpile's most honest classifier or the Nearest Neighbor, may be used. A classifier that uses identification to achieve classification by using those as the query's closest neighbours to ascertain the query's class. K-NN Instances are classified based on how similar they are to instances in the training data. This paper provides a variety of output with varying algorithmic distances, which may help to understand how the classifier responds to the desired input. Furthermore, it demonstrates the computational difficulties in determining the closest neighbours and reducing the data dimension.

3. MACHINE LEARNING CLASSIFIERS

There have been numerous instances in the current fake news corpus where both monitored and unsupervised algorithms have been used to classify text. However, the majority of the literature focuses on specific datasets or domains, most notably the domain of politics. As a result, the trained algorithm performs best on a specific type of article's realm and does not produce optimum performance when revealed to publications from other domains. Because each domain's articles have a distinct text - based framework, it is difficult to develop a generic algorithm that performs well across all news domains. The main goal is to propose a solution to the problem of fake news detection using the machine learning ensemble method.

Due to linguistic nuances, detecting fake news is a difficult problem. Understanding the logic behind some of these fake items necessitates deducing a large amount of information about various actors implicated. We believe that a hybrid solution combining machine learning, semantics, and natural language processing should be used to address this issue. Fake news detection, with exception of spam detection, has several nuances that text analysis cannot detect. The main goal is to classify media data and predict false propaganda using deep learning and language generation. Newspapers, tabloids, and magazines gave way to online media systems, blogs, Twitter feeds, and other digital media formats.

RANDOM FOREST ALGORITHM

An algorithm for group learning is random forests. The algorithm's fundamental premise is that it is computationally inexpensive to construct a tiny decision-tree with few features. If we can construct multiple small, weak decision trees concurrently, we can then average or take the majority vote to join the trees to create a single, strong learner. It is frequently discovered in reality that random forests are the most up until now precise learning techniques.

Algorithm 1 provides an illustration of the pseudo code.

The procedure is as follows: we choose a bootstrap sample from S, where S (i) stands for the ith bootstrap, for each tree in the forest. Then, we discover a decision-tree. By modifying the decision-tree learning method. The algorithm is changed such that as follows: rather than looking at every potential feature-split at each node of the tree, pick a subset of the features f F at random, where F is the collection of features. After that, the node splits based on f's best feature rather than F's. In reality, f is very, significantly more modest than F. Choosing which feature to separate is frequently the most difficult an expensive computational feature of decision tree learning by reducing the range. We greatly accelerate the learning of the tree by the use of features.

Algorithm: Random Forest

Precondition: A training set S := (x1, y1), . . . ,(xn, yn), features F, and number of trees in forest B.

```
function Random Forest(S, F)
  H ← ∅
  for i ∈ 1, . . . , B do
    S (i) ← A bootstrap sample from S
    hi ← RandomizedTreeLearn(S (i) , F)
    H ← H U {hi}
  end for
  return H
end function
function RandomizedTreeLearn(S , F)
  At each node:
    f ← very small subset of F
    Split on best feature in f
  return The learned tree
end function
```

NAVIES BAYES ALGORITHM

The Nave Bayes algorithm is a supervised learning method for classification issues that is based on the Bayes theorem. It is mostly employed in text categorization with a large training set. The Naive Bayes Classifier is one of the most straightforward and efficient classification algorithms available today. It aids in the development of quick machine learning models capable of making accurate predictions. Being a probabilistic classifier, it makes predictions based on the likelihood that an object will occur. Spam filtration, Sentimental analysis, and article classification are a few examples of Naive Bayes algorithms that are frequently used.

Algorithm: Navies Bayes

Input:

Training dataset T,

F=(f1,f2,f3,……fn) // value of the predictor variable in the testing dataset.

Output:

A class of testing dataset.

Step:

1. Read the training dataset T;
2. Calculate the mean and standard deviation of the predictor variable in each class;
3. Repeat

   Calculate the probability of f_i using the gauss density equation in each class;

Until the probability of all predictor variables (f1,f 2 ,f 3 ,….,f n) has been calculated.

4. Calculate the likelihood for each class;
5. Get the greatest likelihood;

4. DEEP LEARNING CLASSIFIER

Technologies such as Deep learning and Natural Language Processing (NLP) tools offer great promise for researchers to build systems which could automatically detect fake news. However, detecting fake news is a challenging task to accomplish as it requires models to summarize the news and compare it to the actual news in order to classify it as fake. Moreover, the task of comparing proposed news with the original news itself is a daunting task as its highly subjective and opinionated. In this project, we can implement text mining algorithm to extract the key terms based on natural language processing and also include classification algorithms such as deep learning algorithm named as multi-layer perceptron algorithm

In the current fake news corpus, there have been multiple instances where both supervised and unsupervised learning algorithms are used to classify text. However, most of the literature focuses on specific datasets or domains, most prominently the politics domain. Therefore, the algorithm trained works best on a particular type of article's domain and does not achieve optimal results when exposed to articles from other domains. Since articles from different domains have a unique textual structure, it is difficult to train a generic algorithm that works best on all particular news domains. In this paper, we propose a solution to the fake news detection problem using the deep learning ensemble approach. Our study explores different textual properties that could be used to distinguish fake contents from real.

The datasets we used in this study are open source and freely available online. The data includes both fake and truthful news articles from multiple domains. The truthful news articles published contain true description of real world events, while the fake news websites contain claims that are not aligned with facts. The datasets we used in this study are open source and freely available online. The data includes both fake and truthful news articles from multiple domains. The truthful news articles published contain true description of real world events, while the fake news websites contain claims that are not aligned with facts.

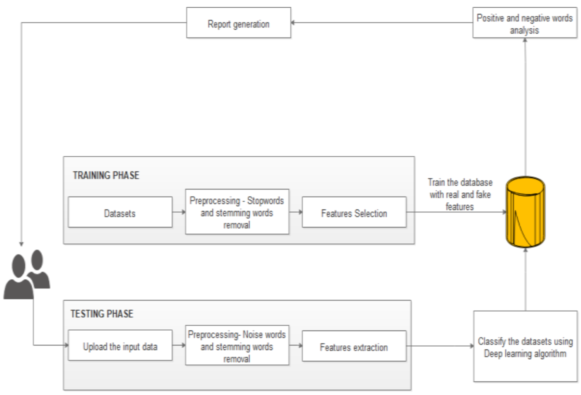

Fig 2: Proposed Architecture

5. EXPERIMENTAL RESULTS

In this study, we can analyze classifiers in fake news datasets that are collected from KAGGLE web sources. Table 1 shows the fake news datasets attributes

S.NO	Attributes	Specifications
1	id	unique identifier for a news item
2	title	the headline of a news story
3	text	the article's content
4	url	the article link
5	top_img	top image of the article
6	authors	writer of the news item
7	source	source of the article
8	publish_date	specifies the uploaded date
9	movies	movies types
10	images	image of the article
11	canonical_link	specifies the index link
12	meta_data	provides details regarding other data.
13	news_type	specifies the type of the news content

And evaluate the performance of the survey using following metrics

True positive (TP): the sensing system produces a positive diagnosis for the sample, and the text is present in the sample.

False positive (FP): the detection system produces a conclusive result for the sample despite the fact that the sample does not contain the text.

True negative (TN): the detection method generates a positive test result for the sample despite the fact that the sample does not contain the text.

False negative (FN): the detection method generates a positive test result for the sample despite the fact that the sample contains text.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

That FP is equal to zero. As FP increases, the precision value decreases while the denominator value increases, resulting in the opposite of what we want.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

A good classifier should have a recall of one (high). Only if the denominator and numerator are the same, as in TP = TP + FN, does recollect equal one, implying that FN is zero. As FN increases, the recall value decreases (which is undesirable) as well as the lowest common value increases.

As a result, the ideal precision and recall for a proficient classification model are one, implying that FP and FN are also zero. As a result, we need a statistic that takes precision and recall into account. The F1-score, a calculation that takes precision and recall into account:

$$\text{F1 Score} = 2 * \frac{\Pr ecision * \operatorname{Re} call}{\Pr ecision + \operatorname{Re} call}$$

ALGORITHM	F1 SCORE
Naives Bayes	60%
Random forest	70%
MLP algorithm	90%

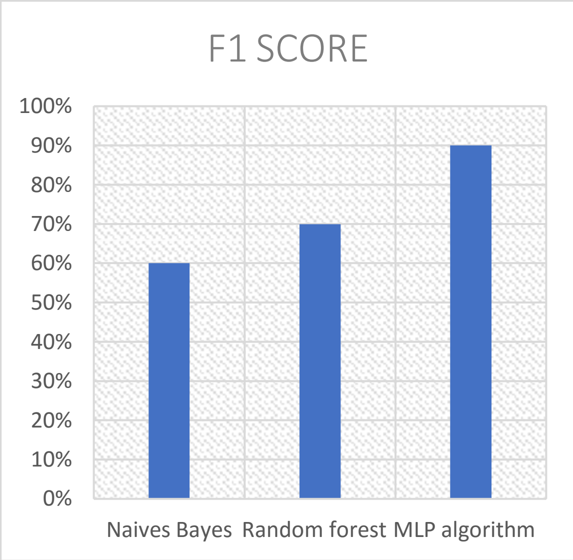

Fig 3: Performance Chart

From this performance chart, proposed system provides improved F1 score than the existing machine learning classifiers.

6. CONCLUSION

In this paper, we have studied the fake news article, creator and subject detection problem. Based on the news augmented heterogeneous social network, a set of explicit and latent features can be extracted from the textual information of news articles, creators and subjects respectively. Furthermore, based on the connections among news articles, creators and news subjects, a deep diffusive network model has been proposed for incorporate the network structure information into model learning. Deep learning model provides improved accuracy rate.

REFERENCES

[1] Mihaylov, Todor, Georgi Georgiev, and Preslav Nakov. "Finding opinion manipulation trolls in news community forums." Proceedings of the nineteenth conference on computational naturallanguage learning. 2015.

[2] Mihaylov, Todor, and Preslav Nakov."Hunting for troll comments in news community forums." arXiv preprint arXiv:1911.08113 (2019)

[3] Bourgonje, Peter, Julian Moreno Schneider, and Georg Rehm. "From clickbait to fake news detection: an approach based on detecting the stance of headlines to articles." Proceedings of the

2017 EMNLP workshop: natural language processing meets journalism. 2017.

[4] Chopra, Sahil, Saachi Jain, and John Merriman Sholar. "Towards automatic identification of fake news: Headline-article stance detection with LSTM attention models." Proc. Stanford CS224d Deep Learn. NLP Final Project. 2017.

[5] Konstantinovskiy, Lev, et al. "Toward automated factchecking: Developing an annotation schema and benchmark for consistent automated claim detection." Digital threats: research and practice 2.2 (2021): 1-16.

[6] Jaiswal, Jitendra Kumar, and Rita Samikannu. "Application of random forest algorithm on feature subset selection and classification and regression." 2017 world congress on computing and communication technologies (WCCCT). IEEE, 2017.

[7] Ul Haq, Muhammad Aziz, et al. "Indoor localization using improved multinomial naïve bayes technique." International afro- European conference for industrial advancement. Springer, Cham, 2016.

[8] Lu, Jing, Peilin Zhao, and Steven Hoi. "Online passive aggressive active learning and its applications." Asian Conference on Machine Learning. PMLR, 2015.

[9] Nayak, Janmenjoy, Bighnaraj Naik, and H. S. Behera. "A comprehensive survey on support vector machine in data mining tasks: applications & challenges." International

[10] Kataria, Aman, and M. D. Singh. "A review of data classification using k-nearest neighbour algorithm." International Journal of Emerging Technology and Advanced Engineering 3.6 (2013): 354-360.

[11] X. Zhou and R. Zafarani, ‘‘A survey of fake news: Fundamental theories, detection methods, and opportunities,’’ACM Comput. Surv., vol. 53, no. 5, pp. 1–40, Sep. 2020.

[12] A. Choudhary and A. Arora, ‘‘Linguistic feature based learning model for fake news detection and classification,’’ Expert Syst. Appl., vol. 169, May 2021, Art. no. 114171.

[13] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, ‘‘Fake news detection on social media: A data mining perspective,’’ ACM SIGKDD Explor. Newslett., vol. 19, no. 1, pp. 22–36, 2017.

[14] I. Singh, P. Deepak, and K. Anoop, ‘‘On the coherence of fake news articles,’’ in Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases. Cham, Switzerland:Springer, 2020, pp. 591–607.

[15] J. Pennebaker, R. Booth, R. Boyd, and M. Francis, Linguistic Inquiry and Word Count: LIWC2015. Austin, TX, USA: LIWC.Net, 2015.

[16] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, ‘‘Truth of varying shades: Analyzing language in fake news and political factchecking,’’ in Proc. Conf. Empirical Methods Natural Lang. Process., 2017, pp. 2931–2937.

[17] G. Pennycook and D. G. Rand, ‘‘Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking,’’ J. Personality, vol. 88, no. 2, pp. 185–200, Apr. 2020.

[18] G. Pennycook and D. G. Rand, ‘‘Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning,’’ Cognition, vol. 188, pp. 39–50, Jul. 2019.

[19] A. Ghanem, C. Asaad, H. Hafidi, Y. Moukafih, B. Guermah, N. Sbihi, M. Zakroum, M. Ghogho, M. Dairi, M. Cherqaoui, and K. Baina, ‘‘Realtime infoveillance of Moroccan social media users’ sentiments towards the COVID-19 pandemic and its management,’’ Int. J. Environ. Res. Public Health, vol. 18, no. 22, Nov. 2021, Art. no. 12172.

[20] F. Haouari, M. Hasanain, R. Suwaileh, and T. Elsayed, ‘‘ArCOV-19: The first Arabic COVID-19 Twitter dataset with propagation networks,’’ in Proc. 6th Arabic Natural Lang. Process. Workshop, 2021, pp.82–91.