



ISSN: 2321-2152

IJMECE

*International Journal of modern
electronics and communication engineering*

E-Mail

editor.ijmece@gmail.com

editor@ijmece.com

www.ijmece.com

GENERATING FAKE NEWS DETECTION USING A TWO STAGE EVOLUTIONARY APPROACH

Nagamallu Harshavardhan¹,

Chigurupally Yashwanth Goud²,

Yssr Prithvi Murthy³, Dr.T.Anil Kumar⁴

^{1,2,3} UG Student, Dept. of ECE, CMR Institute of Technology, Hyderabad

⁴ Professor, Dept. of ECE,

CMR Institute of Technology, Hyderabad

ABSTRACT

While fake news is morally reprehensible, irresponsible parties intentionally use it to achieve their goals by disseminating it to vulnerable and targeted groups. Machine learning techniques have been researched extensively to detect fake news. On the other hand, evolutionary-based algorithms are now gaining popularity in the research community. In this study, a two-stage evolutionary approach is proposed to generate and optimize a mathematical equation for fake news detection. In the first stage, tree-based Genetic Programming (GP) algorithm is used to generate mathematical expressions to detect correlations between the language-independent (Lang-IND) features, extracted from *Fake.my-COVID19* dataset, the newly curated fake news dataset in a mixed Malay - English language. The uniqueness of the proposed approach is that the mathematical expressions are formed by basic arithmetic operators or to include complex arithmetic operators such as addition, multiplication, subtraction, division, square, abs, log1p, sign, square root, and exponential together with Lang-IND features as the variables. Prior to second stage of the evolutionary

approach, a sensitivity analysis is applied to shorten the best equation while maintaining the F1-score performance. In the second stage, an Adaptive Differential Evolution (ADE), is used to fine-tune the mathematical model.

INTRODUCTION

The COVID-19 pandemic is the most critical health crisis in the whole world, affecting all aspects of life [1]. At this time, with the outbreak of the COVID-19 pandemic, social media have become widely used to obtain information about the pandemic. There is a massive amount of false information and fake news, which results in people being confused and raises the need for accurate and real information about the pandemic. Social media such as Facebook, Twitter, and Instagram have facilitated the interaction among people across the world. Thus, because of the huge amount of information and its instant spread, some of this information is real and some is fake. Fake news influences people and leads them in the wrong directions. Therefore, there is an urgent need to find an effective approach that can detect fake news about the COVID-19 pandemic.

Distinguishing fake news is not an easy task since it purposely aims to identify false information. Research on fake news detection has used different classification models. Fake news on social media can have significant negative social effects. As a result, the discovery of fake news on social platforms has attracted

researchers' interest. The existing techniques for fake news detection use both news and social media content as the sources for the learning process. Fake news plays a major role in misleading people and spreading false information.

Several research studies [2,3,4] used various classification algorithms to detect misinformation related to the COVID-19 pandemic. The study [2] used BERT embedding and a shallow neural network to classify COVID-19 tweets. Another study [3] used ten machine learning algorithms with seven feature extraction methods to classify fake news on COVID-19. Furthermore, the study [4] used four machine learning classifiers, decision trees, logistic regression, gradient boost, and support vector machine, to detect fake news on social media.

With the advent of deep learning, there has been a great development in the field of text classification, and thereby in fake news classification. The study [5] used convolutional neural networks (CNNs), long short-term memory (LSTM), and bidirectional encoder representations from transformers (BERT) to detect fake news on COVID-19. Furthermore, the study [6] developed an approach based on an ensemble of three transformer models (BERT, ALBERT, and XLNET) to detect fake news. The model was trained and evaluated in the context of the Constraint AI 2021 Fake News Detection dataset.

The motivation of this study was to use an evolutionary fake news detection technique to extract the most important features of fake news information. Because of the number of symmetrical features related to the COVID-19 pandemic information is large, the proposed approach aimed to reduce the number of symmetrical features and obtain a high classification accuracy simultaneously after implementing four evolutionary classifications, namely k-NN-BSSA, k-NN-BPSO, k-NN-BGA,

and k-NN, and three wrapper feature selection techniques (BGA, BPSO, and BSSA), with k-nearest neighbors (kNNs) as the main classifier to evaluate the output features. The fake news detection involved two primary sources of information: news websites and social media platforms. The experiments were conducted on a fake news website-based dataset called the Koirala dataset, then six datasets were constructed using different text tokenization forms (binary, TF, TF-IDF, and bag-of-words) and stemming techniques, namely Data 1, Data 2, Data 3, Data 4, Data 5, and Data 6.

LITERATURE SURVEY

Social networks have an important role in our daily lives as anyone can publish their ideas and spread information without verifying the authenticity of the content. Several studies involved the discovery of fake news in the context of social media. The study [7] proposed an approach to detect fake news sites to help users avoid such fake news. The study used some features of the news to detect fake information such as keywords and punctuation marks. Furthermore, the study [8] presented different approaches to distinguish between true and fake news, with high performance.

The research paper [9] developed an approach for classifying fake users and fake news on Twitter by implementing entity recognition and hashtag, emoji, and text sentiment analysis. The study by [10] proposed deep learning models based on a feed-forward neural network (FNN) and LSTM in conjunction with different word vector representations. The proposed model mainly focused on collecting information from the news article based on the content and title. The developed model obtained good results in terms of the evaluation measures used.

Additionally, another work [11] used Twitter's data to determine the most important features that affected the performance of the

machine learning methods when used for the classification of fake news. The study presented a set of new Twitter properties that could improve the classification of fake and real tweets. Furthermore, the work by [12] proposed a graph-based semi-supervised learning model to capture fake users on Twitter using certain valuable features.

Similarly, the researchers in [13] provided a text-based approach to fake news detection by using a two-layer classification. The first layer was used for detecting the fake topic, and the second layer was used for detecting the fake event. The research article [14] proposed a hybrid approach for classifying fake news on social platforms, which combined naive Bayes, support vector machines, and semantic analysis.

The study [15] introduced a two-phase technique for identifying fake news on social media. The first phase involved converting unstructured datasets into structured datasets. The second phase involved applying twenty-three supervised AI models on the BuzzFeed Political News dataset. The experimental results indicated that the J48 algorithm achieved the highest accuracy compared with the other AI models. Further, the authors in [16] described a deep learning model on a Kaggle fake news dataset. They performed text preprocessing by using word embedding (GloVe) to construct a vector space of words and created a linguistic relationship. For the classification, the authors proposed a new classification model based on two types of neural networks: convolutional and recurrent architectures. Another paper [17] captured the problem of malicious rumors that appear during breaking news. The study proposed a rumor-based propagation model called the HISBM model, which detected the propagation process of multiple rumors on online social networks to reduce the number of malicious information over a short period. Likewise, The spread of fake health-related news has a bad effect on people's

sentiments. The work introduced by [18] proposed a new approach to analyze Reddit, Facebook, and Twitter content. The results indicated that news focused on fake health information was often considered as bad messaging and that news based on the evidence of social influence was acceptable and respected. Furthermore, the study [19] proposed a theory-based model to early distinguish fake news. Based on the news content, the model detected fake news when it was published on news sites before spreading on social media platforms. The results were obtained by applying the model on two real-world datasets. In the same context, the research study [20] proposed a graph-based approach for unsupervised fake news detection over multiple datasets. The proposed approach employed different graph-based techniques, such as label spreading, a graph-based feature vector, and bi-clique identification. Particular to the AAI-2021 COVID-19 dataset, the study [21] used both machine and deep learning techniques including the SVM, CNN, BiLSTM, and CNN+BiLSTM techniques with the TF-IDF and Word2Vec embedding techniques. Additionally, the research work [22] applied sentiment analysis to identify misinformation in tweets on the COVID-19 Twitter discourse. They captured unreliable and misleading content based on fact-checking websites and investigated the narratives of misinformation tweets.

The aforementioned studies above did not introduce an evolutionary classification technique to show the most essential features of the COVID-19 fake news information. As a result, this study proposed an evolutionary-based detection approach to capture the most important features of fake news information. Our approach aimed to reduce the number of features and achieved a high accuracy. Four evolutionary classifications techniques (k-NN-BSSA, k-NN-BPSO, k-NN-BGA, and k-NN) were used. Furthermore, three wrapper feature selection techniques (BGA,

BPSO, and BSSA) were implemented to evaluate the output features.

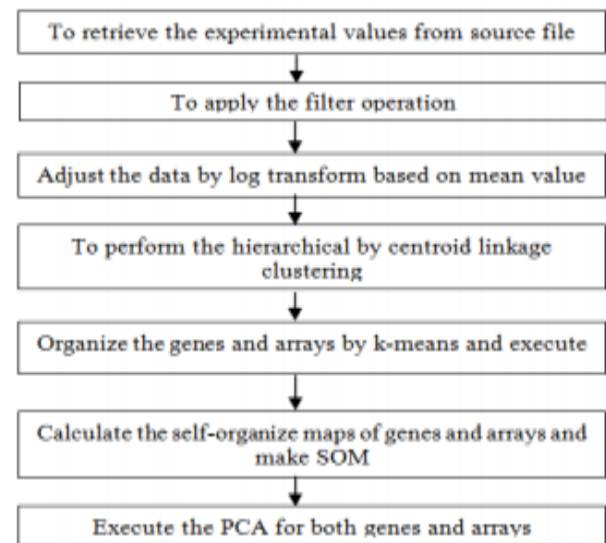
1.1.1 ProposedSystem:

In this paper author is describing concept to detect fake news from social media or document corpus using Natural Language Processing and attribution supervised learning estimator. News documents or articles will be uploaded to application and then by using Natural Language Processing to extract quotes, verbs and name entity recognition (extracting organizations or person names) from documents to compute score, verbs, quotes and name entity also called as attribution. Using supervised learning estimator we will calculate score between sum of verbs, sum of name entity and sum of quotes divided by total sentence length. If score greater than 0 then news will be consider as REAL and if less than 0 then new will be consider as FAKE.

1.1.2 Advantages of proposed system:

- It is desirable to use COX data for phylogenetic exploration.
- We use the data of COX experimental values.
- Security

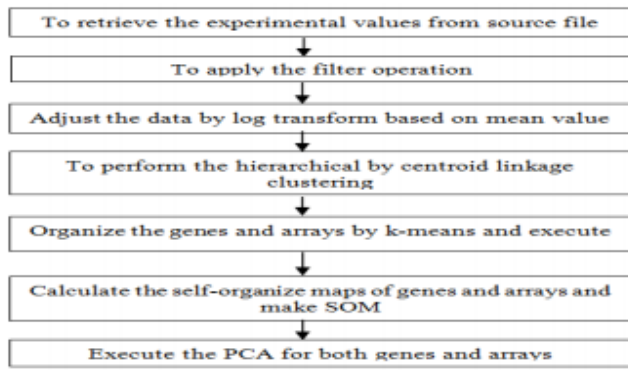
3.1 SYSTEMARCHITECTURE:



Algorithms used in this project :-

CLUSTERING ALGORITHM :

- Clustering is an unendorsed culture algorithm that finds the concealed arrangement in the unlabeled data. In this work, we used the filter the values, adjust the data values, then apply the hierarchical method, k-means algorithm, self-organizing maps (SOM), and finally apply the Principal Component Analysis (PCA) for avoid the unwanted values, adjust the data with help of log transform, for clustering genes and arrays with hierarchical clustering by centroid linkage,



TEST RESULTS

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, subassemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

CONCLUSION & FUTURE WORK

CONCLUSION:

This paper presented the results of a study that produced a limited fake news detection system. The work presented herein is novel in this topic domain in that it demonstrates the results of a full-spectrum research project that started with qualitative observations and resulted in a working quantitative model. The work presented in this paper is also promising, because it demonstrates a relatively effective level of machine learning classification for large fake news documents with only one extraction feature. Finally, additional research and work to identify and build additional fake news classification grammars is ongoing and

should yield a more refined classification scheme for both fake news and direct quotes.

Future Work

The work presented in this paper is also promising, because it demonstrates a relatively effective level of machine learning classification for large fake news documents with only one extraction feature. Finally, additional research and work to identify and build additional fake news classification grammars is ongoing and should yield a more refined classification scheme for both fake news and direct quotes.

REFERENCES

- [1] M. Balmas, "When Fake News Becomes Real: Combined Exposure to Multiple News Sources and Political Attitudes of Inefficacy, Alienation, and Cynicism," *Communic. Res.*, vol. 41, no. 3, pp. 430–454, 2014.
- [2] C. Silverman and J. Singer-Vine, "Most Americans Who See Fake News Believe It, New Survey Says," *BuzzFeed News*, 06-Dec-2016.
- [3] P. R. Brewer, D. G. Young, and M. Morreale, "The Impact of Real News about "Fake News": Intertextual Processes and Political Satire," *Int. J. Public Opin. Res.*, vol. 25, no. 3, 2013.
- [4] D. Berkowitz and D. A. Schwartz, "Miley, CNN and The Onion," *Journal. Pract.*, vol. 10, no. 1, pp. 1–17, Jan. 2016.
- [5] C. Kang, "Fake News Onslaught Targets Pizzeria as Nest of Child-Trafficking," *New York Times*, 21-Nov-2016.
- [6] C. Kang and A. Goldman, "In Washington Pizzeria Attack, Fake News Brought Real Guns," *New York Times*, 05-Dec-2016.
- [7] Radha Krishna Karne and Dr. T. K. Sreeja

(2022), A Novel Approach for Dynamic Stable Clustering in VANET Using Deep Learning (LSTM) Model. IJEER 10(4), 1092-1098. DOI: 10.37391/IJEER.100454.

[8] Reddy, Kallem Niranjana, and Pappu Venkata Yasoda Jayasree. "Low Power Strain and Dimension Aware SRAM Cell Design Using a New Tunnel FET and Domino Independent Logic." International Journal of Intelligent Engineering & Systems 11, no. 4 (2018).

[9] Reddy, K. Niranjana, and P. V. Y. Jayasree. "Design of a Dual Doping Less Double Gate Tfet and Its Material Optimization Analysis on a 6t Sram Cells."

[10] Reddy, K. Niranjana, and P. V. Y. Jayasree. "Low power process, voltage, and temperature (PVT) variations aware improved tunnel FET on 6T SRAM cells." Sustainable Computing: Informatics and Systems 21 (2019): 143-153.

[11] Reddy, K. Niranjana, and P. V. Y. Jayasree. "Survey on improvement of PVT aware variations in tunnel FET on SRAM cells." In 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), pp. 703-705. IEEE, 2017

[12] Karne, R. K. ., & Sreeja, T. K. . (2023). PMLC- Predictions of Mobility and Transmission in a Lane-Based Cluster VANET Validated on Machine Learning. International Journal on Recent and Innovation Trends in Computing and Communication, 11(5s), 477–483.
<https://doi.org/10.17762/ijritcc.v11i5s.7109>