IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11615

Fake Media Detection Using NLP, CNN Algorithm And Blockchain

Prof. Pallavi N[1], P Lohith[2], Sanjay A[3], Yeshwanth P[4], Nitesh Kumar G[5]

[1]Assistant Professor, Dept of Computer Science and Engineering, Atria Institute of Technology, Bengaluru, India.

[2-3]Dept of Computer Science and Engineering, Atria Institute of Technology, Bengaluru, India.

[4,5]Dept. of Computer Science and Engineering, Atria Institute of Technology,Bengaluru, India.

Abstract: News and Media has become one of the most significant components of human existence, based entirely on the most recent generation and traits within the discipline of laptop generations. This area has become a famous platform for sharing information and statistics on a variety of issues, as well as daily reports, and is the most popular generation for transferring and generating data. There are various advantages to living in this environment. however there also are several fake statistics and data that mislead the reader and consumer with a view to acquire the numbers required. One of the system's primary trouble is the shortage of functional data and real statistics on social media data. To resolve this issue, we have got proposed an incorporated system with Various additives of the Convolutional Neural Network (CNN) and Natural Language Processing (NLP) to utilize machine learning analyzing strategies to discover faux statistics and better are anticipating faux money owed and posts. This approach is finished the use of the Reinforcement Learning method. To enhance the safety of this platform, the decentralized blockchain framework became implemented, which incorporates the definition of digital content material authority proofs. More specifically, the purpose of this system is to expand a dependable platform for awaiting and detecting social media networks.

Keywords: CNN (Convolutional neural network) Algorithm, NLP (Natural Language Processing)

I. INTRODUCTION

Assorted information has emerged as a fully noteworthy topic till now, which 300% usually utilized online. Struggling with fictitious datasets will become an insurmountable challenge in social networks, deep within the statistics and records consumption software layer, and a serious and difficult issue in record development in the diplomatic, economic, and political sectors. The fictitious data reveal an unneeded system as a must-have of communal resources. It also has a material of content depending on the available service. As a consequence , the improper records sharing has an impact on the Quality of Trust that should be used when disseminating information.

The size of the machine inspecting textual content category improves the security This is required in social media-based completely networking on a daily basis. Extracting sharing or ardor an opinion from the non-authoritarian league survey via social networking contains a lot of phoney money owing and records. dispensing the gateway solely over a satisfactory method In this set of affairs, the mutilating and unpleasant debts must be removed from the organization to make room for the information hub and to handle the mess and political concerns inside the community's boundaries. Propaganda, which is special for political goals, is another analogous place for records extraction. In order to stir and aggravate the fervor of consumers for disseminating false records, the mimicking information forging language might be quite sophisticated in words of predesignate. This is in charge of detecting fake information. Contents assessment based altogether on facts contained in the limits of shared deeds. With the growing amount of shady and unstructured statistics, as well as the growing number of consumers and information, there may be a need for an automated solution for extracting bogus traits. Based on current trends in system learning, deep learning, and synthetic intelligence, these terms appear to be restricted. One of the prerequisite course of action for record sharing is to verify the authorship of virtual materials.

As a case study, we gathered social media content from Facebook and Twitter, which are undoubtedly the most well-known record-sharing platforms, with millions of users uploading massive amounts of daily information and postings on a variety of themes. The goal of this study is to use blockchain, natural language processing, and tool analysis approaches to approve fraudulent clients and records. More specifically, the suggested device is a preventive strategy that combines gamification components with veiled techniques for the notion of bogus facts extraction, If the records are incorrect, the device limits the use of similar records in the future to minimise the number of phoney and incorrect records.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11615

II. RELATED WORK

A. [1] A Sensitive and Stylistic Approach to Detecting Falls media on online media platforms- They propose to employ both dimensionality reduction techniques, such as latent semantic analysis , and data compaction techniques, such as our proposed approaches, to original data. There are three different news classification approaches that are employed when the number of attributes was reduced by more than 80% - two using cascade or unique configurations of learning algorithms, and the other statistically measuring the difference between the news.

B. [2] In this paper A complete examination of the features has revealed that there are more. When it comes to detecting social media accounts that spread incorrect information on the Internet, this strategy is more accurate than an automated method. . We perform a statistical study in this work that reveals that these characteristics are more predictive than an automated technique. To obtain this conclusion, elements such as the users' personal and social information were taken from Twitter.

C. [3]The goal of this research is to provide a linguistic-based framework for detecting bogus news on social media. Its goal is to figure out what language markers may be utilised to spot false news. This methodology is used to analyse and forecast the many aspects of news material using quantitative and qualitative analysis.It also aims to figure out which linguistic characteristics may be used to detect fake news.They may then be used to filter and identify different aspects of news information. This research demonstrates that these indicators have a lot of potential.

D. [4] In smart cities, blockchain can help with waste management. In this article, we look at how blockchain technology may help smart cities manage garbage by providing traceability, immutability, transparency, and auditing in a decentralised, trustworthy, and safe way.

E. [5]A Review of Recent Advances in Fake News Detection Machine Learning Techniques Deep learning algorithms are widely used in today's culture because they can successfully solve text recognition problems such as false news and spam detection, since such models work on raw data effectively by analysing high-level characteristics on their own. Various classic machine learning approaches and deep learning strategies for detecting fraudulent and spam communications are covered in this study.

III. PROPOSED SYSTEM

A. Fake Media Detection Natural language processing

Natural language processing plays the major role in the process of data preparation as it carry out the functions like dividing the data into small clusters ,extracting the important and usefull feature that might help in the data analysis , and indexing of each words so they can be identified with their unique index.When extracting any of the features or information from the data, the specific data content must be cleaned and undesired material deleted. before passing into the next step where the usefull information from the data is extracted and it is converted into an computer understandable format like vectors and stores them in the database management systems ,The relevant data is extracted by sending queries to the data source, and the further data retrieval process is carried out with the help of this information that was stored throughout the feature extraction phase. and the similarity of the contents is measured with th help of feature extraction modules and sorted accordingly to form the list.

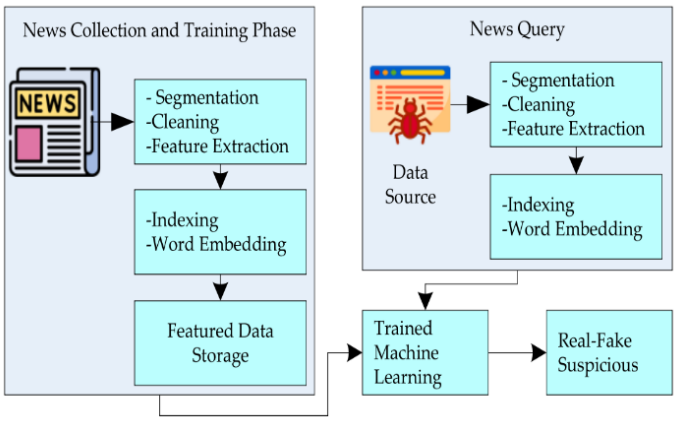

Figure 2: Block diagram for NLP

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11615

B. Fake Media Detection using CNN

To determine weather the content is true or fake ,The gathered data should be processed and trained using Convolutional Neural Network (CNN) algorithm in order to finish the training process as fast as possible. This involves 3 layers:

- Convolution layer: Here, the feature extraction will take place where only the useful features which are needed to the machine will be collected and unwanted features will be removed so that training period will be finished soon.
- Pooling layer: In this, the size of the data or image will be reduced and give us a compressed document with important features which is needed for the machine.
- Fully connected layer: Here, the above data which we get from the previous layer will be fed to fully connected layer in a vector form. Then these compressed features will be split and get trained using CNN and will produce us the final output.

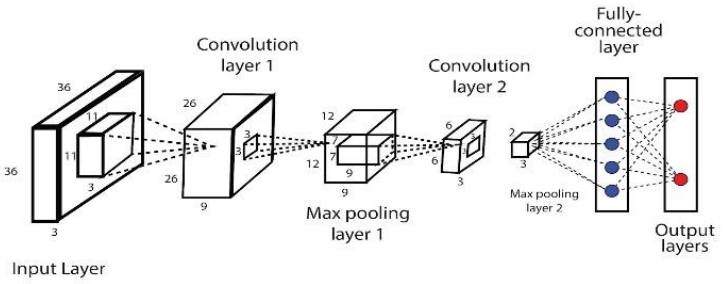

Figure 3: Block Diagram of CNN

IV. ARCHITECTURE

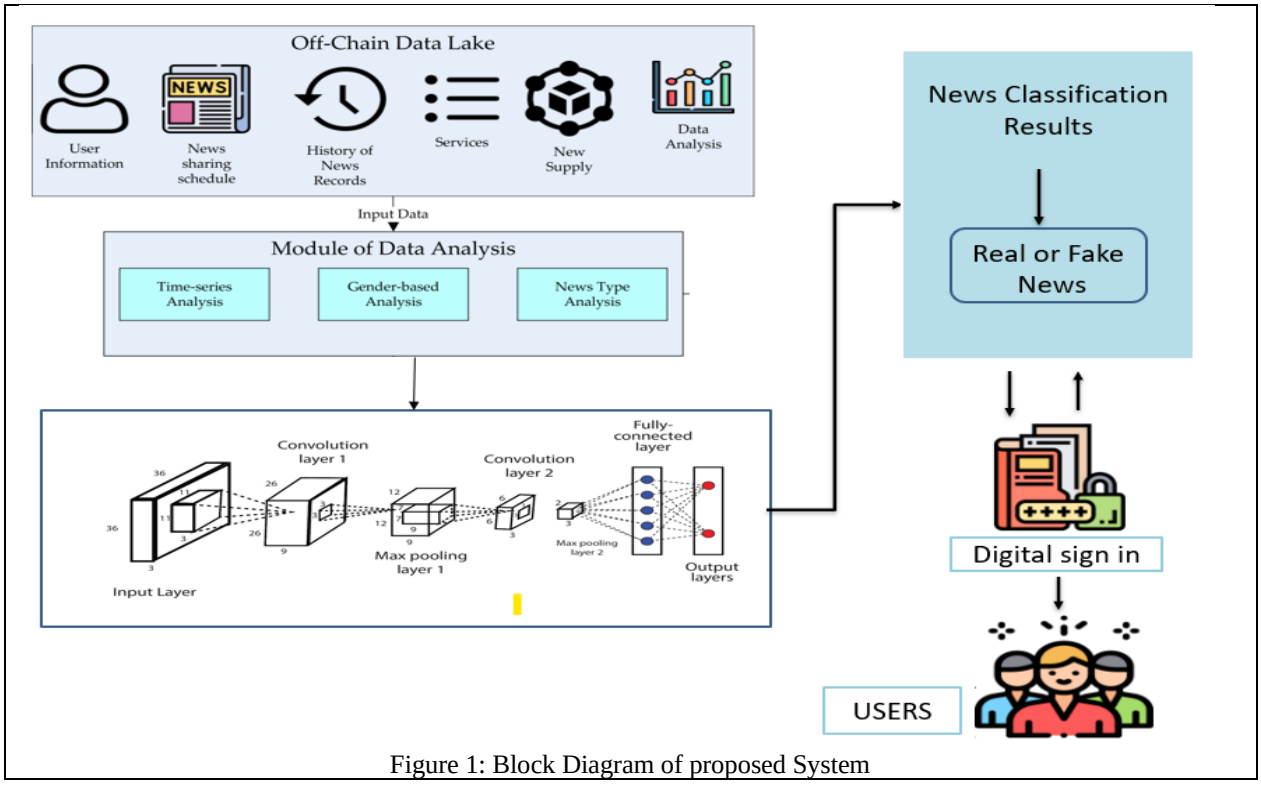

Figure 1: Block Diagram of proposed System

C. Fake Media Detection Using Blockchain

In today's generation, media organizations require specific information in order to operate or run their businesses; this information or data must then be verified and licensed by the respective organizations, and such reports and information are secured with the blockchain via e-sign in options or two-way authentication, ensuring that the information placed on the blockchain is secure, and cannot be altered by any of the attackers are fake news publishers, if any changes to be made on that applications it has to be done by the admin or owner of the organization who owns the

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11615

right of that information, so the data cannot be altered by the unknown users ,to identify weather the given media is true one has to go through the proper validation process.

Figure 4: Blockchain Authentication

V. DESIGN

A data flow diagram is a graphical representation of the data flows though the proposed system, which simulates the properties of the procedure It's commonly used as a first stage to generate a model of the process's features. It's typically used as a first step to give a high-level overview of the system without going into too much detail, with more depth added afterwards.. They'll even be used to analyze the processing of knowledge.
A data flow chart depicts the types of data that are input to and output from the system, as well as how the data will go through the system and where it will be kept. In contrast to a flowchart, it does not provide information about process time or whether processes will run in succession or in parallel.

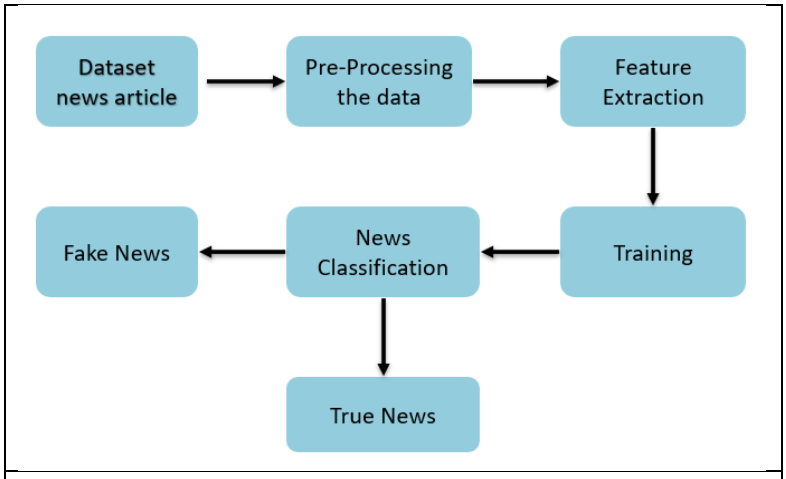

Figure 5: Block Diagram of Design Flow

VI. MEATHODLOGY

Data pre-processing- Data preprocessing is a phase in the data mining and data analysis process that turns raw data into a format that computers and machine can understand and evaluate**.**This involves 3 layers:

- First, the feature extraction will take place where only the useful features which are needed to the machine will be collected and unwanted features will be removed so that training period will be finished soon.
- Secondly, the size of the data or image will be reduced and give us a compressed document with important features which is needed for the machine.
- Next, the above data which we get from the previous layer will be fed to fully connected layer in a vector form. Then these compressed features will be split and get trained using CNN and will produce us the final output.

a. **Data gathering:** Gathering data is the most important step in solving any supervised machine learning problem. Your text classifier can only be as good as the dataset it is built from.

b. **Data Integration:** Data Integration is a data preprocessing technique that involves combining data from multiple heterogeneous data sources into a coherent data store and provide a unified view of the data.

c. **Data Cleaning:** The process of cleaning the involves the removal of extra and unwanted information ,and

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11615

wrong data by using tools to remove repeated data or duplicate data .

VII. SEQUENCE DIAGRAM

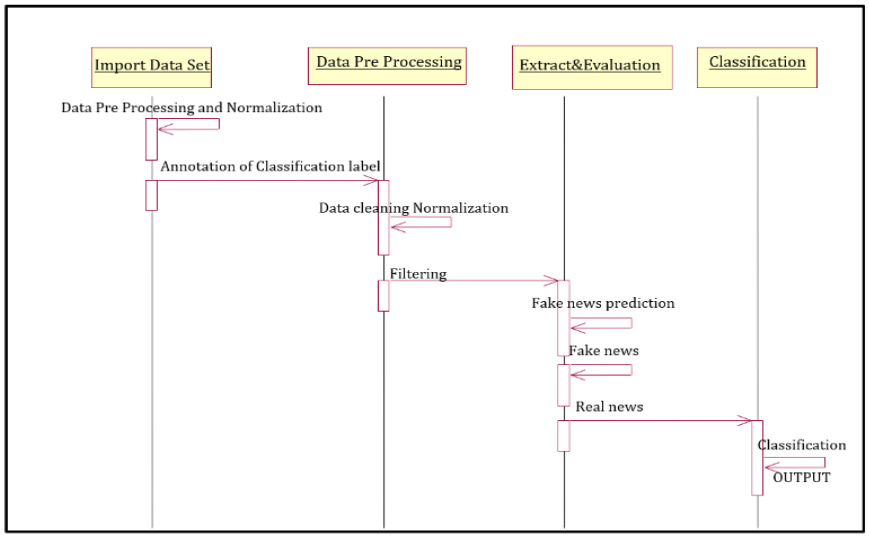

Figure 6: Block Diagram of CNN

VIII. OUTPUT FLOW DIAGRAM

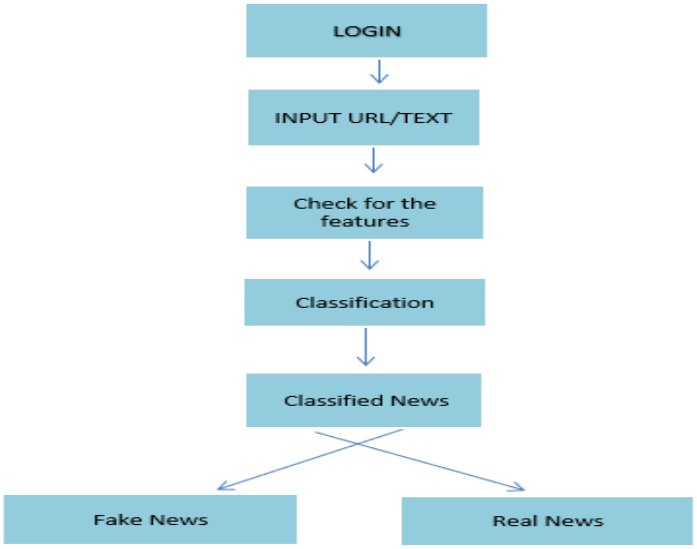

Figure 7: Block diagram of output flow

IX. RESULTS

The result and analysis section demonstrate the results and outputs obtained from the experiment upon the implementation of fake media detection system

Snapshot 1: User interface for the fake media detection system

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11615

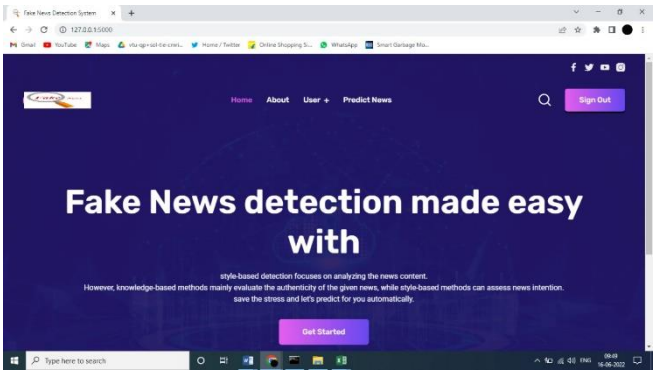

Figure 8: UI for Fake media detection

Snapshot 2: This is the registration terminal where users can register themselves and login to the fake media detection system

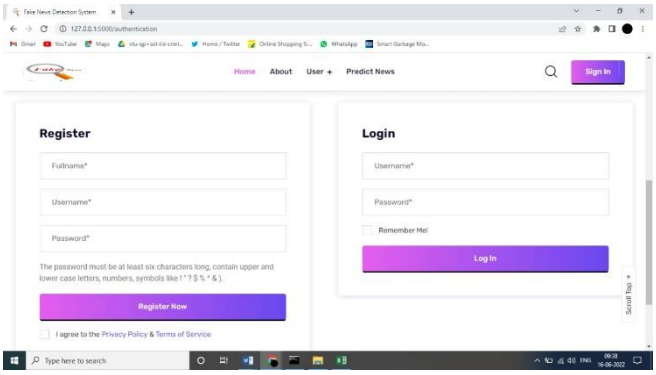

Figure 9: Registration and Login page

Snapshot 3 and 4: If the input news is true then the system displays the news from the particular news and display the result as news is Real and if the news is not original then the system displays the result as the news is Fake.

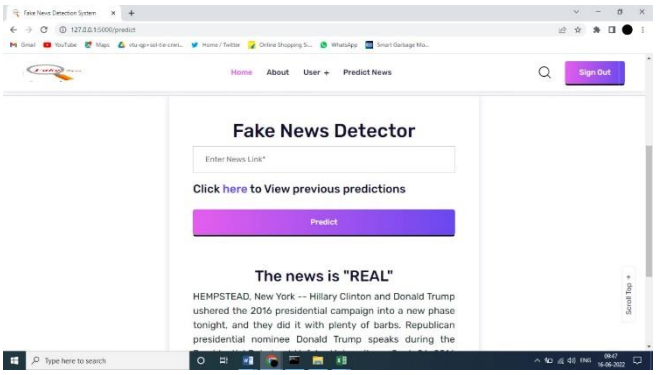

Figure 10: Output for the True News

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11615

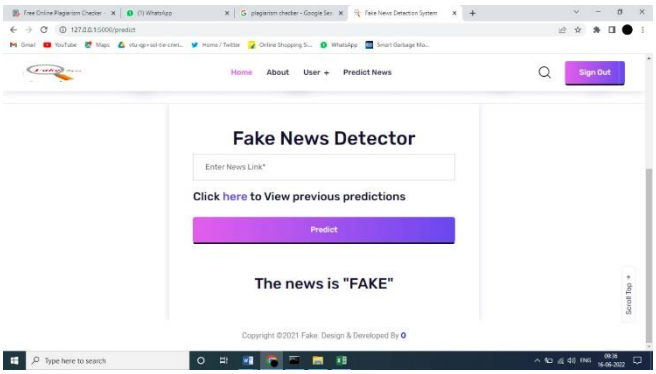

Figure 11: Output for the Fake News

X. CONCLUSION

Fake News and media are one among the most popular issues in the society with the ongoing generations,. In this article, we have got supplied the combinationture of blockchain and system studying strategies to deliver answers and fashion a agree with-primarily based totally structure in the direction of shared information online. we have used the combination convolutional neural network for image analysis algorithm and for the news which is in the form of text is analysed using thrnatural language procesing toolkit , a study-based approach to form a efficient decision-making architecture and connecting it with a blockchain technology that is well-suited to the Proof-of-Authority protocol,Social media performs a key position in the course of this process. The shared statistics platform incorporates faux information, and its a useful venture to enhance and inspect the Proof-of-Authority protocol and consumer validation

XI. ACKNOWLEDGEMENT

We would like to offer our thanks to the project guide **Prof. Pallavi N,** Assistant Professor: Dept. of CSE, AIT thank you for guiding us through such great support and research. She gave the paper the insights and expertise needed to make it look good. Her advice, professional insight, and encouragement proved to be invaluable as a guide. And we would also like to thank our project coordinator **Dr. Devi Kannan,** Associate Professor: Dept. of CSE, AIT Thank you for guiding us through this process. Thanks to **Dr. Aishwarya P**, Hod - Dept. of computer science and engineering, Atria institute of Technology, Bengaluru, India.

XII. REFERENCES

[1] Xishuang Dong, ‘‘Deep Two-path Semi-supervised Learning for Fake News Detection(2019),’’ in Cornell
[2] Nikolaos Augoustis,Katia Lida Kermanidis, ‘‘Fake News Detection Regarding the Hong Kong Events from Tweets(2020),’’ in Springer,
[3] J. A. Vijay, H. A. Basha, and J. A. Nehru, ‘‘A dynamic approach for detecting the fake news using random forest classifier and NLP,’’ in Computational Methods and Data Engineering. Springer, 2021.
[4] M. Mahyoob, J. Algaraady, and M. Alrahaili, ‘‘Linguistic-based detection of fake news in social media,’’ Int. J. English Linguistics.
[5] Koirala, ‘‘COVID-19 fake news classification using deep learning,’’ Tech. Rep., 2020.
[6] F. Gasparetti, Sansonetti, G. D’Aniello, and A. Micarelli, ‘‘Unreliable users detection in social media: Deep learning techniques for automatic detection,’’ IEEE Access,
[7] H. Rojas, and Gill ‘‘Chatting in a mobile chamber: Effects of instant messenger use on tolerance toward political misinformation among south Koreans,’’ Asian J. Commun.
[8] Y. Jin, J. Shen, and C. Li, ‘‘Application of combined model of stepwise regression analysis and artificial neural network in data calibration of miniature air quality detector,’’ Sci. Rep.
[9] G. Augasta, A. R. Merryton ‘‘A survey on recent advances in machine learning techniques for fake news detection,’’ Test Eng. Manag.