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 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124134

Implementing A Passive Aggressive Classifier To Detect False Information

Alkesh S. Lajurkar[1], Abhijeet R. Rupune[2], Madhubala N. Lahabar[3], Akanksha S. Umak[4], Prof. A. U. Chaudhari[5]

B.E. Scholars, Department of Computer Science and Engineering, Prof. Ram Meghe Institute of Technology and Research Badnera, Maharashtra, India[1-4]

Assistant Professor, Department of Computer Science and Engineering, Prof. Ram Meghe Institute of Technology and Research Badnera, Maharashtra, India[5]

Abstract: Today is an information age, and people of all ages use smartphones and social media to convey information. This information is not always reliable; sometimes false information is spread over social media. The proliferation of fake news is a major problem all around the world. Fake news is infiltrating human minds via social media. After reading the news articles on Facebook, WhatsApp, Instagram, Twitter, and other social media platforms, Sometimes these publications convey the inaccurate or incorrect message, which is referred to as fake news. This fake news is being a very big scam in which many people are being trapped. To solve this big issue, it's important to develop a system called "Fake news detection system". This fake news detection system detects the fake news by the prefilled data by various articles, from various sources. The earliest available systems uses some common methodology like Data Collection, Feature Extraction, Model Building and Model Evalution but only variations are seen in the use of techniques to solve the problem. This paper presents a Passive Aggressive Classifier Model with TfidfVectorizer Text Classification to address fake news. As per result obtained by proposed model it gives better accuracy than previous model used by researchers.

Keywords: Fake News Detection, Data Collection, Feature Extraction, Passive Aggressive Classifier, TfidfVectorizer.

I. INTRODUCTION

Social media is an interactive platform that allows for the creation, exchange, and sharing of information, ideas, viewpoints, and emotions over the internet. Low cost, user-friendly, accessibility, and availability are the driving factors behind social media's rising popularity and quick expansion. However, because of the availability of the internet, the majority of people rely on online sources and social media for news in their daily lives. For many people in the world today, social media is one of the most accessible news sources. Therefore, it might be challenging to tell whether news is real or fake. Fake news and stories are being circulated on social media networks more frequently now because of a lack of understanding and information.[1,2]

Fake news denotes the wrong information which is populated through social media. The term "fake news" refers to reports which appear as authentic while actually being false. False stories tales about a variety of topics that are carefully crafted to appear authentic and attract traffic and social media shares on numerous platforms are considered fake news. With the increase in social networks, a greater number of people are creating and sharing information than ever before, many of them have no relevance to reality. Therefore, with the increase in popularity of social media, rate of circulating fake news is exponential [9].

As per 10th Report (2021) released by Reuters Institute on the state and trends of digital news media consumption worldwide the percentage of consuming news on social media was 63% and on online platforms (including social media) was 82%. It demonstrates that while being constantly overtook with information, we lack the tools, expertise, and skills necessary to check it. Fake news intentionally spread for a variety of purpose such as, it can impact their ability to distinguish what is legitimate or what is not legitimate [4].

To solve these problems, it's important to detect the fake news firstly and check whether the news is real or fake. This paper presents a Passive Aggressive Classifier Model with TfidfVectorizer Text Classification to address fake 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 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124134

II. LITERATURE REVIEW

Misinformation spreads quickly on social media. It has caused societal confusion in terms of economics and different perspectives. thus, addressing false information is necessary. system that has already been demonstrated and developed by several researchers are discussed in this section.

TABLE I: COMPARATIVE ANALYSIS OF FAKE NEWS DETECTION METHODOLOGY

Author Name, Year	Proposed Method	Description	Scope /Accuracy
Roshan Karwa, Sunil Gupta,2022 [1]	Hybrid Deep Neural Network Model, with Convolution Dynamic Semantic Structural Model (C-DSSM) and Deep Convolutional Neural Networks(DCNN) Models	The preliminary layers of the C-DSSM design reflect the feature extraction approach, whilst the D-CNN layers indicate the categorization process	According to experimental results, the proposed model obtained an accuracy of 92.60%
Sudhanshu Kumar, Thoudam Doren Sing, 2022 [2]	Naïve Bayes, logistic regression and Long Short-Term Memory (LSTM)	Different algorithm like Naïve bayes, logistic regression and Long Short-Term Memory (LSTM) are used to detect the fake news and compared the result.	LSTM algorithm gives the best accuracy of 92.36%.
Enjoy Maity, Ankush Tomar, Ruhi Peter,2022 [3]	Bidirectional Encoder Representations from Transformers (BERT) Model of Deep Neural Network	BERT model on its own is a complex and complicated model, it is based on deep neural network. If this model gets large data set, then it can give a good performance	The BERT model that have used in this work gives us an accuracy of 52%.
Reza Mansouri, Mahmood Naderan-Tahan, Mohammad Javad Rashti,2020 [4]	Hybrid of Convolutional neural network and semi-supervised Linear discriminant analysis (SLD-CNN)	Various features of text and image data are extracted using CNN. Then, linear discrimination analysis (LDA) is used to predict the classes of unclassified data.	The results of the evaluations show that the precision of this method is 95.6%, and the recall is 96.7%.
Vanya Tiwari, Ruth G. Lennon, Thomas Dowling, 2020 [5]	Logistic regression, decision tree, K-nearest Neighbor and random forest	Implementation of four classification algorithms with the different vectorizer gave the result differently.	The logistic regression algorithm when implemented after extracting feature with tf-idf vectorizer gave the highest accuracy of 71% while testing the model.
Ankit Kesarwani, Sudakar Singh Chauhan, Anil Ramachandran Nair,2020 [6]	K-Nearest Neighbor classifier.	The performance of K-Nearest Neighbor depends on the value of K	This approach achieved a maximum classification accuracy of 79%

DOI: 10.17148/IJARCCE.2023.124134

Anjali Jain, Avinash Shakya, Harsh Khatter , Amit Kumar Gupta,2019 [7]	Naïve Bayes classifier, Support Vector Machine (SVM).	It used the mix of naïve bayes classifier, support vector machines and semantic investigation.	The proposed model is working well and defining the correctness of results upto 93.6% of accuracy
Supanya Aphiwongsophon, Prabhas Chongstivatana, 2018 [8]	Naïve Bayes classifier, Neural Network,Support Vector Machine (SVM)	It uses three methods differently to measure the highest accuracy.	Naïve bayes achieve the accuracy of 96.08% and Neural Network and SVM achieve 99.90% accuracy
Mykhailo Granik, Volodymyr Mesyura,2017 [9]	Naive Bayes Classifier	a simple approach for fake news detection using naive Bayes classifier	It achieved classification accuracy of approximately 74% on the test set
Shlok Gilda,2017 [10]	Multiple classification algorithms like Support Vector Machines, Stochastic Gradient Descent, Gradient Boosting, Bounded Decision Trees, and Random Forests.	It generate PCFG and TFIDF feature for each article by using each classification models.	TF-IDF of bi-grams fed into a Stochastic Gradient Descent model identifies non-credible sources with an accuracy of 77.2%

III. METHODOLOGY

The primary goal of the research is to address the problem caused by the propagation of fake news. So, To detect the fake news, the Passive Aggressive Machine Learning algorithm with TFidfVectorizer for text classification is to be used.

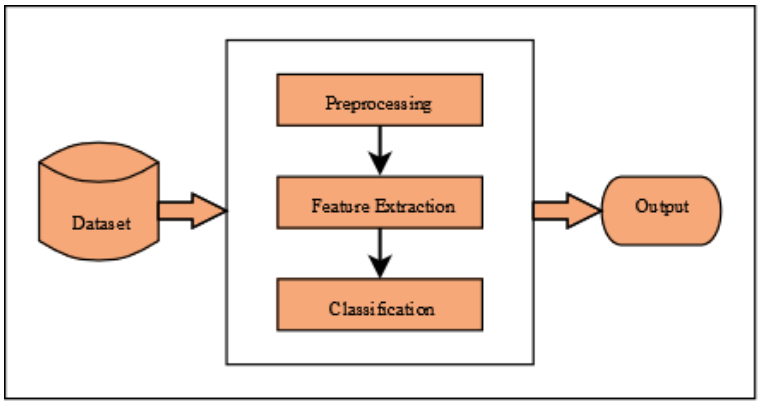

Figure 1: Basic Methodology of Model

A. Preprocessing

In this component, the collected data is cleaned, organized, and formatted in a way that makes it suitable for analysis. This may involve removing irrelevant information, such as ads or images, and converting the text to a machine-readable format. Preprocessing is a crucial step in machine learning that involves preparing the raw data to make it suitable for analysis by machine learning algorithms. Preprocessing includes a variety of techniques to clean, normalize, transform, and reduce the dimensionality of the data.

DOI: 10.17148/IJARCCE.2023.124134

B. Feature Extraction

This component involves identifying the relevant features or characteristics of the news articles that can be used to distinguish between real and fake news. The goal of feature extraction is to identify the most informative and relevant features that can be used to make accurate predictions. By applying feature extraction techniques, we can improve the performance of machine learning algorithms and make better predictions on new, unseen data.

C. Classsification

In this component, machine learning algorithms are used to analyze the features extracted from the news articles and to classify them as real or fake. A machine learning model is a mathematical representation of a system or process that is trained on a dataset to make predictions or decisions about new data. A machine learning model learns from the patterns and relationships in the training data to make predictions on new, unseen data.

IV. IMPLEMENTATION & ANALYSIS

A. Dataset Description

The dataset is simple. It contains the titles of the news, the body text, subject, date and a target field, which if the news is authentic, shows REAL and if inauthentic, shows FAKE. The dataset includes two kinds of articles: fake News and actual News. This dataset has been gathered from real-world sources; particular articles were obtained by crawling the news website Reuters.com. The fraudulent news pieces were gathered from a variety of sources. The false news stories were gathered from untrustworthy websites that had been reported by Politifact (a fact-checking organisation based in the United States) and Wikipedia. The dataset covers a variety of articles on various themes, with the majority of articles focusing on political and global news.

	title	text	subject	target
0	Remains of exiled Italian king return to Italy	MILAN (Reuters) - The remains of Victor Emmanu...	worldnews	true
1	'Might Have Been Faked By Liberals': Top Advi...	I mean, that headline, right? Did someone actu...	News	fake
2	North Korea warns of 'more gift packages' for ...	GENEVA (Reuters) - North Korea is ready to sen...	worldnews	true
3	Trump and spy chief differ on what was said in...	WASHINGTON (Reuters) - U.S. spy chief James Cl...	politicsNews	true
4	Greece moves asylum-seekers from Lesbos to mai...	PIRAEUS (Reuters) - Greek authorities on Thurs...	worldnews	true

Figure 2: Dataset

B. Whitespace Tokenization

After general preprocessing steps such as removal of unnecessary attributes, punctuation, stopwords,etc, checking null value; tokenization is performed. Whitespace tokenization is a simple and straightforward approach to tokenizing text data. It splits the input text into tokens based on whitespace characters such as spaces, tabs, and newlines. One advantage of using whitespace tokenization is that it is fast and easy to implement. Because it doesn't require complex pattern matching or regular expressions. The main objective of using this whitespace tokenization is that it preserves the original structure of the text also it can be useful in cases where the context of a word is important.

C. Feature Extraction with TF-IDF

The word count from text documents is very basic at the starting point. However simple word count is not sufficient for text processing because of the words like “the”, “an”, “your”, etc. are highly occurred in text documents. Their large word count is meaningless towards the analysis of the text. Tf-idf can be successfully used for stop-words filtering from the text document. The other way to solve this problem is word frequency. This method is called the TF-IDF stands for “*Term Frequency – Inverse Document Frequency”*. TF-IDF is a numerical statistic which measures the importance of the word in a document.

D. Classifier

Passive-aggressive algorithms are a family of great learning algorithms. They are similar to Perceptron because it does not require a reading scale. However, unlike Perceptron, they include parameter correction. Passive is used when the prediction is correct and there is no change in the model. But if there is any kind of change in the model, that is if the prediction is not correct then the aggressive part is called, which changes the model accordingly.

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 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124134

E. Model Building

The model is built through the train and test of the dataset, by ensuring that the training is done for 80% of the dataset and testing is done in the rest of the 20% of the dataset.

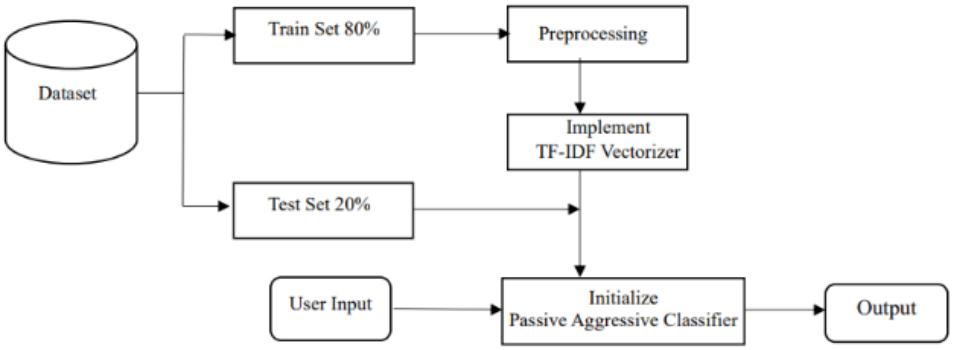

Figure 3: Model Architecture

F. Classification of Real and Fake News

In our study, predicted values are identified as Positive and Negative, whereas actual values are identified as True and False. Figure 4 shows the True Positive value to be 4723, the True Negative value to be 4215, the False Positive value to be 24, and the False Negative value to be 18. This confusion matrix is necessary for calculating performance metrics such as recall, precision, specificity, and accuracy.

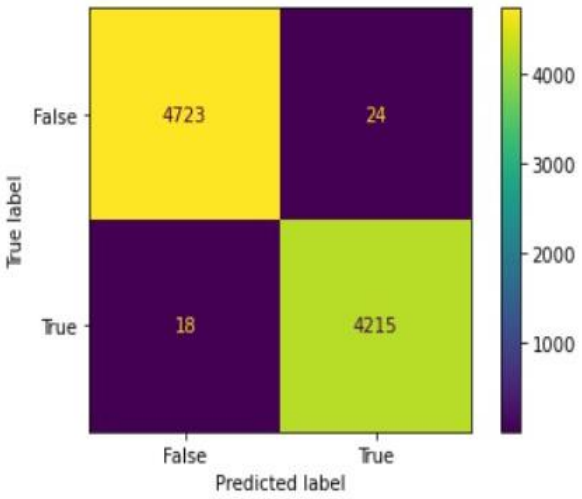

Figure 4: Classification of Real and Fake News

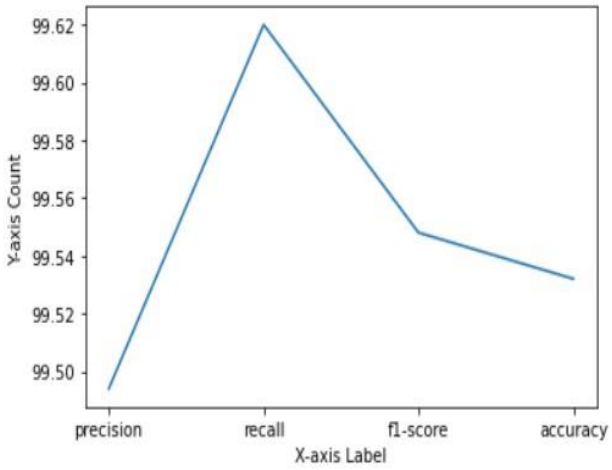

Figure 5: Performance Evaluation Metrics

The proposed technique's performance in detecting and classifying fake news is evaluated using the following evaluation parameters: accuracy, F1 Score, precision, and recall.

The performance evaluation measures for the proposed approach are shown in Figure 5. The results were 99.53%, 99.54%, 99.49%, 99.62% for accuracy, F1 Score, Precision, and Recall respectively. With the use of a TFidf Vectorizer and Passive Aggressive Classifier, our suggested system performs better in terms of accuracy, F1 score, precision, and recall.

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 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124134

G. Output Interface

The following output is displayed on the screen at the end, indicating if the news is true or false. When you access the web interface, Type or paste the news into the input area, after typing the news into the input box and press the 'Predict' button, the model will predict the outcome. If the news appears to be true, the result will be ‘News Headline is FAKE’. Otherwise, the message ‘News Headline is REAL’ will appear. That is how the UI detects fake or real news.

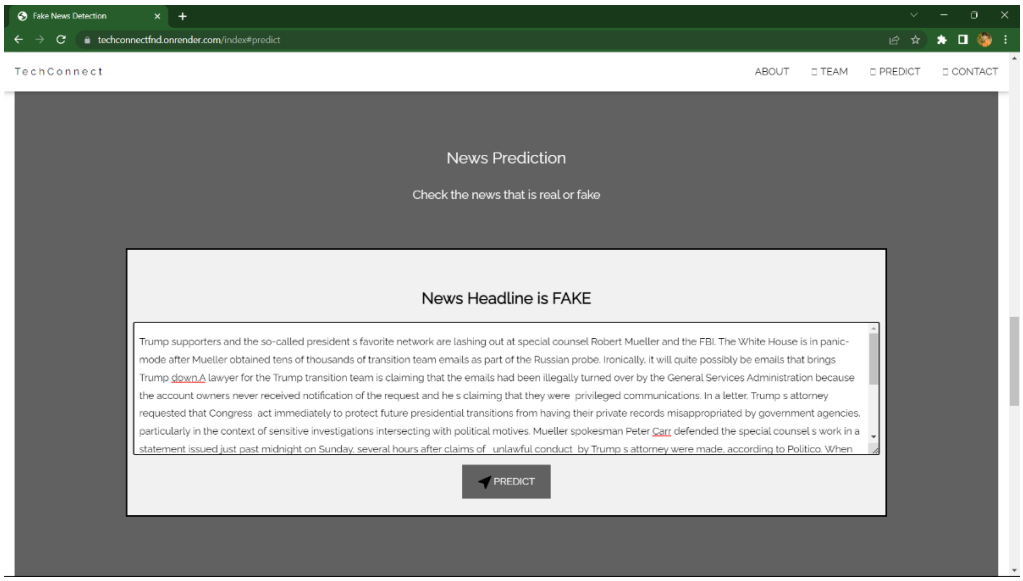

Figure 6. Result of Fake News Prediction

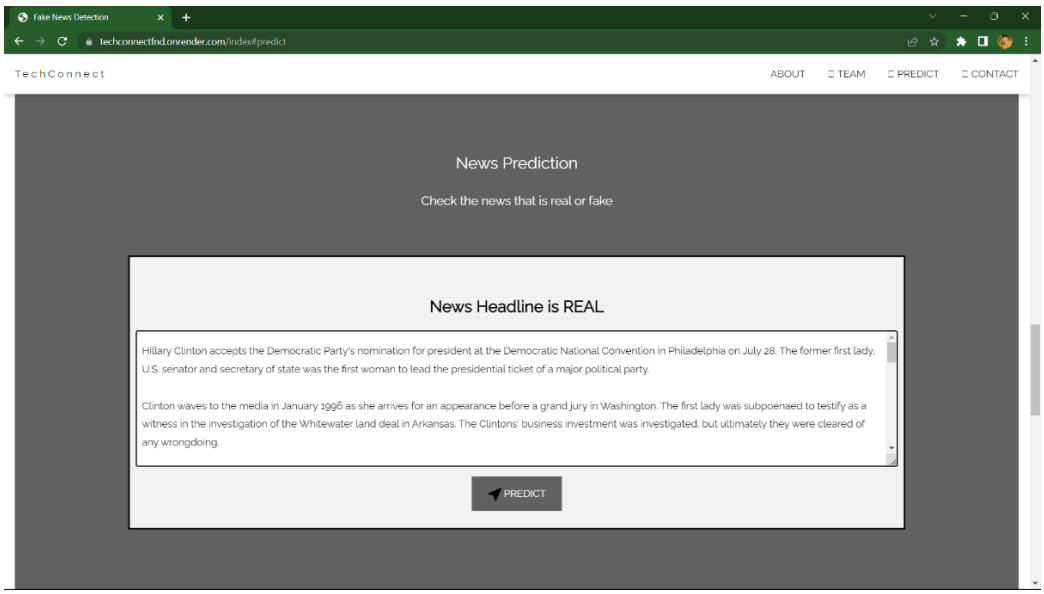

Figure 7: Result of Real News Prediction

V. CONCLUSION

Fake news is a term used to describe false information or propaganda presented as news. It is designed to deceive people into believing something that is not true. This can be done through social media platform and other. Regardless of the medium, the impact of fake news can be significant and far-reaching. In order to detect fake news, this paper presents a Passive Aggressive Classifier Model with TfidfVectorizer Text Classification to address fake news. The results show that our research performs better than other classifiers in terms of accuracy. The proposed outcomes of this study are significant, with a 99.53 percent accuracy rate when using the IOST Fake News Dataset, which is highly effective at detecting false news. Furthermore, in the future, the study can be improved with continued research for preserving the accuracy and trustworthiness of information in the digital age.

DOI: 10.17148/IJARCCE.2023.124134

ACKNOWLEDGMENT

We would like to extend our thanks **Dr. G. R. Bamnote,** Principal, PRMIT&R for their kind co-operation and encouragement.

We would like to express our deep and sincere thanks to **Dr. M. A. Pund**, Head of Department of Computer Science & Engineering for their tremendous support and valuable guidance.

We would like to express our deep and sincere thanks to **Prof. A. U. Chaudhari** for their guidance and constant supervision as well as for providing necessary information regarding the work & also for their support in completing the work.

We would like to express my very great appreciation **to Dr. R. R. Karwa** for his valuable and constructive suggestions during the planning and development of this work.

REFERENCES

[1] Roshan R. Karwa, Sunil R. Gupta, "Automated hybrid Deep Neural Network model for fake news identification and classification in social networks", Journal of Integrated Science and Technology, Volume 10, No 2, 2022.
[2] Sudhanshu Kumar, Thoudam DorenSingh, "Fake News detection on Hindi news dataset",Global Transition Proceedingd,2022,doi: https://doi.org/10.1016/j.gltp.2022.03.014
[3] Enjoy Maity, Ankush Tomar, Ruhi Peter, "Fake News Detection System: In Hindi Data Set Using BERT", International Research Journal of Modernization in Engineering Technology and Science, Volume 4, May-2022.
[4] Reza Mansouri, Mahmood Naderan-Tahan, Mohammad Javad Rashti, "A Semi-supervised Learning Method for Fake News Detection in Social Media", Iranian Conference on Electrical Engineering (ICEE), 2020, doi: 10.1109/ICEE50131.2020.9261053.
[5] Vanya Tiwari, Ruth G. Lennon, Thomas Dowling, "Not Everything You Read Is True! Fake News Detection using Machine learning Algorithms", Irish Signals and Systems Conference (ISSC),2020, doi: 10.1109/ISSC49989.2020.9180206.
[6] Ankit Kesarwani, Sudakar Singh Chauhan, Anil Ramachandran Nair, "Fake News Detection on Social Media using K-Nearest Neighbor Classifier", International Conference on Advances in Computing and Communication Engineering (ICACCE),2020, doi: 10.1109/ICACCE49060.2020.9154997.
[7] Anjali Jain, Avinash Shakya, Harsh Khatter , Amit Kumar Gupta, "A Smart System For Fake News Detection Using Machine Learning", International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT),2019, doi: 10.1109/ICICT46931.2019.8977659.
[8] Supanya Aphiwongsophon, Prabhas Chongstivatana, "Detecting Fake News with Machine Learning Method", International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON),2018, doi: 10.1109/ECTICon.2018.8620051
[9] Mykhailo Granik, Volodymyr Mesyura, "Fake News Detection Using Naive Bayes Classifier", Ukraine Conference on Electrical and Computer Engineering (UKRCON), 2017, doi: 10.1109/UKRCON.2017.8100379
[10] Shlok Gilda, "Evaluating Machine Learning Algorithms for Fake News Detection", Student Conference on Research and Development (SCOReD),2017, doi: 10.1109/SCORED.2017.8305411
[11] Monther Aldwairi, Ali Alwahedi, "Detecting Fake News in Social Media Networks", The 9th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2018), doi: 10.1016/j.procs.2018.10.171
[12] Ms. Aarti Chugh, Dr. Yojna Arora, Mr. Jaivardhan Singh, Mr. Shobhit, Mr. Ronak, "Media Manipulation Detection System Using Passive Aggressive", International Journal of Innovative Research in Computer Science & Technology (IJIRCST), 2021, doi: 10.21276/ijircst.2021.9.3.8, ISSN: 2347-5552, Volume-9, Issue-3, May 2021
[13] Rohit Kumar Kaliyar, "Fake News Detection Using A Deep Neural Network", 2018 4th International Conference on Computing Communication and Automation (ICCCA),2018,doi: 10.1109/CCAA.2018.8777343
[14] Sherry Girgis, Eslam Amer, Mahmoud Gadallah, "Deep Learning Algorithms for Detecting Fake News in Online Text", "2018 13th International Conference on Computer Engineering and Systems (ICCES)" 2018, doi: 10.1109/ICCES.2018.8639198
[15] Roshan R. Karwa and Sunil R. Gupta, "Artificial Intelligence Based Approach to Validate the Authenticity of News", 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), 2021, doi: 10.1109/ICAECT49130.2021.9392456

DOI: 10.17148/IJARCCE.2023.124134

[16] Mayur Bhogade, Bhushan Deore, Abhishek Sharma, Omkar Sonawane, Prof. Manisha Singh, "A Research Paper On Fake News Detection", International Journal Of Advance Scientific Research And Engineering Trends, 2021,doi: 10.51319/2456-0774.2021.6.0067, Volume 6, Issue 6, June 2021, ISSN (Online) 2456-0774

[17] Hadeer Ahmed, Issa Traore, Sherif Saad, "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques", International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, 2017, doi: 10.1007/978-3-319-69155-8_9

[18] Pedro Faustini, Thiago Covoes, "Fake News Detection Using One-Class Classification", 2019 8th Brazilian Conference on Intelligent Systems (BRACIS), 2019, doi: 10.1109/BRACIS.2019.00109

[19] Roshan R. Karwa, Sunil R. Gupta, "Detect Misinformation Using Two Stage Semantic Extractor Based On Neural Network Classification", Webology (ISSN: 1735-188X) Volume 18, Number 6, 2021

[20] Viera Maslej Kresnakova, Martin Sarnovsky, Peter Butka, "Deep learning methods for Fake News detection", IEEE Joint 19th International Symposium on Computational Intelligence and Informatics and 7th International Conference on Recent Achievements in Mechatronics, Automation, Computer Sciences and Robotics, 2019, doi: 10.1109/CINTI-MACRo49179.2019.9105317

[21] Suwarna Gothane, Vydehi Kavali, S. Dinesh Kumar, Ramya Jampani, "Fake News Detection and Classification using Natural Language Processing", International Journal for Research in Applied Science & Engineering Technology (IJRASET), 2021, doi: 10.22214/ijraset.2021.34700, Volume 9 Issue V, May 2021

[22] Jayashree M Kudari, Varsha V, Monica BG, Archana R, "Fake News Detection using Passive Aggressive and TF-IDF Vectorizer", International Research Journal of Engineering and Technology (IRJET), 2020, e-ISSN: 2395-0056 , p-ISSN: 2395-0072, Volume: 07 Issue: 09, Sep 2020

[23] Ms.CH.Uma Devi, R.Priyanka, P.S.Surendra, B.S.Priyanka, C.H.N.D.L.Nikhila, "Fake News Detection Using Machine Learning", Journal of Emerging Technologies and Innovative Research (JETIR), 2019, ISSN-2349-5162, Volume 6, Issue 4

[24] Kartik Rajesh, Aditya Kumar, Rajesh Kadu, "Fraudulent News Detection using Machine Learning Approaches", 2019 Global Conference for Advancement in Technology (GCAT), 2019, doi: 10.1109/GCAT47503.2019.8978436

[25] Sai Rama Krishna Indarapu; Jahnavi Komalla; Dheeraj Reddy Inugala; Gowtham Reddy Kota; Anjali Sanam, "Comparative analysis of machine learning algorithms to detect fake news", 2021 3rd International Conference on Signal Processing and Communication (ICPSC), 2021, doi: 10.1109/ICSPC51351.2021.9451690