

A Framework for Predicting and Analyzing Fake News Using Machine Learning

Vaishali Singh

Assistant Professor, Department of Computer Science & Engineering, Ambalika Institute of Management & Technology, Lucknow, Uttar Pradesh, India

ABSTRACT

In today's world, social media is the most effective way to express him. And this is the finest area to provide information about yourself, your society, your faith, and your customs. It is involved in the rapid exchange of information, in which news from all fields is available. Social media has a major impact on our lives and society nowadays. And, in today's world, social media is the most effective way to express him. Furthermore, social media has evolved into a platform for sharing current events. People in the other location are informed about what is going on in the other location. People also learn about the culture of other places as a result of this. However, some nefarious elements utilize social media to promote false information, which has an impact on both our lives and society. And if Fake News isn't dealt with quickly enough, it spreads like a forest fire. And this fake news hurts some people's sentiments, and it has also been known to trigger riots in society. It is vital in today's world to have some instruments that can verify any news, whether it is factual or not. And I'd like to accomplish the same thing with this algorithm.

Keywords : Machine Learning, Support Vector Machine, Naive Bayes Algorithm, Fake News, Prediction.

Article Info

Volume 9, Issue 2

Page Number : 204-209

Publication Issue :

March-April-2022

Article History

Accepted : 01 April 2022

Published: 09 April 2022

I. INTRODUCTION

Online media is a unique form of communication. This creates a virtual world [1] that can be accessed via the Internet. The online media industry is a massive organization that keeps the entire world connected. It's everything but a reliable mode of communication. It is involved in high-speed data [2] trading, in which information on each field is available. Web-based media has a massive impact on our lives and society today. In addition, web-based media is the ideal vehicle for communicating your viewpoints at this

moment. Furthermore, web-based media has become a channel for sharing what is going on in the world around you. Which folks at the other location are aware of what is going on in the other location [3]. Alongside this, individuals additionally get information about the way of life of other spot. However, some devilish components utilize online media to spread their off-base words and it influences both our life and society. Also, in the event that this Fake News isn't dealt with at the ideal opportunity, it's anything but a woods fire [4]. Furthermore, this phony news offends of certain individuals and some of the

time this phony news additionally causes riots in the general public. In the present time it is important that we have a few apparatuses that can check any news whether it is genuine information or phony news. Also, I need to do exactly the same thing through this calculation of mine.

II. SIGNIFICANCE OF THE STUDY

In today's time social media is a good tool to express your views. But some people are also misusing this tool in today's time. Some wrong people are spreading their personal hatred to the people through Fake News. And this fake news hurts the sentiments of some innocent people, society or people of any particular religion [5]. That's why in today's time it is necessary that we make some such tool which can catch any fake news in social media before it spreads. In today's time computer technology has progressed so much that through which we can analyze fake news well. And by analyzing we can easily identify any fake news.

III. REVIEW OF RELATED STUDIES

Ethar Qawasmeh et. al. (2019) [3] In this exploration paper, the creator has utilized an enormous dataset. Essentially in this paper the dataset has been cleaned first by taking the dataset and afterward by utilizing vectorisation procedure. Yet, in this paper it is seen that there is no discussion about the wellspring of information. Which is important to decide any news that on the off chance that we don't have a clue what is the wellspring of the news, how might we realize that the news is phony. In this paper, the author has given more attention to the subject of news, writing on political news, sports news, etc. But to predict any fake news it is necessary that we pay attention to every fact. In this paper, the author has used the system algorithm and the accuracy is 85.3%.

William Yang Wang (2018) [4] Automatic phony news identification is a difficult issue in misdirection

discovery, and it has huge true political and social effects. Be that as it may, measurable ways to deal with battling counterfeit news has been drastically restricted by the absence of marked benchmark datasets. In this paper, we present LIAR: another, freely accessible dataset for counterfeit news recognition. We gathered a long term, 12.8K physically marked short explanations in different settings from site, which gives nitty gritty examination report and connections to source records for each case. This dataset can be utilized for certainty checking research too. Prominently, this new dataset is a significant degree bigger than already biggest public phony news datasets of comparable sort. Observationally, we examine programmed counterfeit news recognition dependent on surface-level etymological examples. We have planned a novel, half breed convolutional neural organization to incorporate metadata with text. We show that this crossover approach can improve a book just profound learning model.

Costin BUSIOC et. al., (2020) [5] Fighting phony news is a troublesome and testing task. With an expanding sway on the social and world of politics, counterfeit news apply an unprecedentedly sensational effect on individuals' lives. Because of this marvel, drives tending to computerized counterfeit news discovery have acquired prominence, producing inescapable examination interest. Notwithstanding, most methodologies focusing on English and low-asset dialects experience issues when conceiving such arrangements. This examination centers around the advancement of such examinations, while featuring existing arrangements, difficulties, and perceptions shared by different exploration gatherings. Furthermore, given the restricted measure of computerized examinations performed on Romanian phony news, we review the materialness of the accessible methodologies in the Romanian setting, while at the same time recognizing future exploration ways. Z Khanam, et. al., (2021) [6] in this research

paper, the author has used a very large dataset. Basically in this paper the dataset has been cleaned first by taking the dataset and then by using vector is ratio technique. In this paper, Researcher has used Python's scikit learn library, Torch Library. But in this paper it is seen that there is no talk about the source of news. Which is necessary to determine any news that if we do not know what the source of the news is, how will we know that the news is fake.

IV. OBJECTIVES OF THE STUDY

We want to design a tool through which we can analyze any news and we can know what kind of news it is [7]. Through this tool, we can test any fake news at the time of posting and find out which area a news is related to. And the purpose of this study is that we can know whether any news is fake or not and whether that news is harmful to society or any person.

V. HYPOTHESES OF THE STUDY

Today social media has become a medium through which even a common man can express his views. But in today's time there is no such tool that can detect any fake news at the time of posting. Fake news hurts the sentiments of a common man, society or any section of the people. Even if that post is removed later, the news still spreads. That's why a tool is needed that can recognize the news at the time of the post.

VI. METHODOLOGY

First we will enter the Fake News, and then we will create a data set by twiping the data collected from that news and scraping the web data. In the next step we will clean the dataset and remove the grammarless data from the dataset. Then we will create a meaning full dataset and after that we will do feature selection of each data by doing feature selection (source, date, time). Next we will do a testing dataset preparation and create a test dataset. In this step we create a model in

which, considering the date of news, location of news and timing of news. In the next step, we analyze the entered news from the model created by taking the location, timing and source. If we get matching news related to the features of the news entered from Faithful Sources then we create its score. If the score is between 1 to 3 then it will be fake news, and if the score is between 3 to 20 then it may be true news and if the news score is more than 20 then the news is true [8].

VII. RESULT ANALYSIS AND FINDINGS

First, he will enter the news he wants to check, then the related news is fetched through tweepy and web scraping [9]. Then a dataset of truncated data is obtained, and then cleaning [10] of the fetch dataset is done through Imp [11]. The cleaned dataset is then classified using a support vector machine [12]. In classification we have taken trained dataset 80% and test data set 20%. Here True News is represented by 1 and False News by 0 in the trained dataset. Then using Naive Bayes to predict whether the news is true or false.

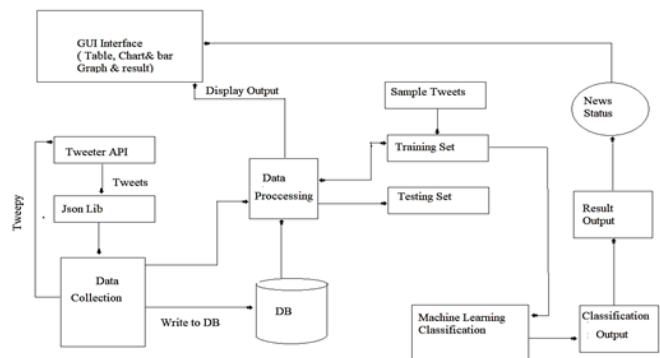


Figure 1: Process Diagram Of fake News Prediction

VIII. COMPARATIVE ANALYSIS BETWEEN ALGORITHM'S ON THE BASIS OF RELIGION AND POLITICS DATASET

In this section, we've collected political and religious data and written different algorithms on top of it, such as Decision Tree and Logistic Regression, and

compared our algorithms against each other. Table 1 compares the accuracy of the following algorithms using 2500 political and religious news stories [14].

Algorithm	Decision Tree	Logistic Regression	Naïve Bayes	Naïve Bayes+SVM
Political	1768	897	1922	2039
Religions	1698	873	1823	2343

It can be seen that the outcome is pretty accurate when we utilize our naïve bias and support vector machine [15]. Among the other algorithms, our technique outperforms them all because we clean the data first with slp and then [16]. As a result, our method has an accuracy of 89 % [17]. He primarily uses social media to disseminate fake news focused on politics and religion.

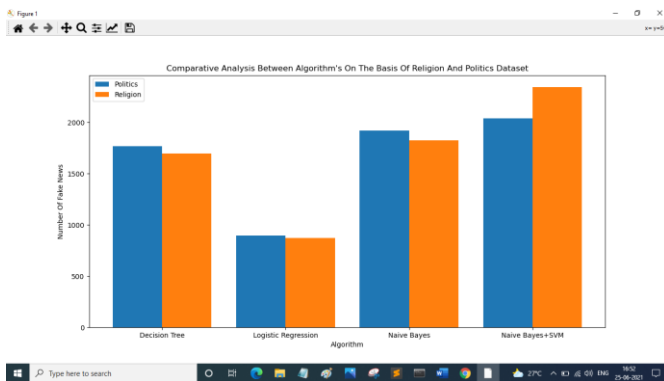


Figure 2 . The Comparative Analysis Between Algorithm's On The Basis Of Religion And Politics Dataset

IX. CONCLUSION

Our investigation reveals that everything has two sides, good and evil, and it is up to us to decide which side we will take. The same thing happens with social

media platforms, and some people today are misusing them even more. By spreading fake news, some bad people are causing significant harm to society. This study of mine could be very useful in combating fake news. Any bogus news can be stopped before it reaches the general public if this feature is deployed on the social media platform. People can also use it to double-check whatever news they've seen on any social media site, allowing them to make the best decision possible. We can also prevent anything bad from happening. I achieved the main goal of this method, and the accuracy of my algorithm was 89 percent, which satisfied me to a large extent.

X. REFERENCES

- [1]. L. Atzori, A. Iera and G. Morabito, "The internet of things: A survey", *Comput. Netw.*, vol. 54, no. 15, pp. 2787-2805, 2010
- [2]. Yusuf Perwej, "An Experiential Study of the Big Data", *International Transaction of Electrical and Computer Engineers System (ITECES), USA, Volume 4, No. 1, PP. 14-25, 2017*, DOI: 10.12691/iteces-4-1-3
- [3]. Ethar Qawasmeh, Mais Tawalbeh, Malak Abdullah, "Automatic Identification of Fake News Using Deep Learning", 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), 978-1-7281-2946-4/19/\$31.00 ©2019 IEEE
- [4]. William Yang Wang, "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection", arXiv:1705.00648v1 cs.CL1 May 2017.
- [5]. Costin BUSIOC, Stefan RUSSETI, Mihai DASCALU, "A Literature Review of NLP Approaches to Fake News Detection and Their Applicability to Romanian- Language News Analysis", 2020, Romanian Ministry of Education and Research, CNCS - UEFISCDI, project number PN-III-P1-1.1-TE-2019-1794, within PNCDI III.

- [6]. Alim Al Ayub Ahmed, Ayman Aljarboub, Praveen Kumar Donepudi, "Detecting Fake News using Machine Learning: A Systematic Literature Review", 2020, IEEE conference.
- [7]. Razan Masood and Ahmet Aker, "The Fake News Challenge: Stance Detection using Traditional Machine Learning Approaches", In Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (KMIS 2018), pages 128-135
- [8]. Z Khanam et.al. , " Fake News Detection Using Machine Learning Approaches", IOP Conference Series: Materials Science and Engineering, Volume 1099, International Conference on Applied Scientific Computational Intelligence using Data Science (ASCI 2020) 22nd-23rd December 2020, Jaipur, India
- [9]. Sohan De Sarkar, Fan Yang, " Attending Sentences to detect Satirical Fake News", Proceedings of the 27th International Conference on Computational Linguistics, pages 3371-3380 Santa Fe, New Mexico, USA, August 20-26, 2018.
- [10]. Abdullah-All-Tanvir, Ehasas Mia Mahir, Saima Akhter " Detecting Fake News using Machine Learning and Deep Learning Algorithms", 2019 7th International Conference on Smart Computing & Communications (ICSCC).
- [11]. Hadeer Ahmed, Issa Traore, and Sherif Saad. Detection of online fake news using n-gram analysis and machine learning techniques. In International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, pages 127-138. Springer, 2017.
- [12]. Yusuf Perwej, "The Hadoop Security in Big Data: A Technological Viewpoint and Analysis", International Journal of Scientific Research in Computer Science and Engineering (IJSRCSE) , E-ISSN: 2320-7639, Volume 7, Issue 3, Pages 1-14, 2019, DOI: 10.26438/ijsrcse/v7i3.1014
- [13]. Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211-36, 2017.
- [14]. Yusuf Perwej, Nikhat Akhtar, Firoj Parwej, "The Kingdom of Saudi Arabia Vehicle License Plate Recognition using Learning Vector Quantization Artificial Neural Network", *International Journal of Computer Applications (IJCA)*, USA, Volume 98, No.11, Pages 32 - 38, 2014, DOI: 10.5120/17230-7556
- [15]. Peter Bourgonje, Julian Moreno Schneider, and Georg Rehm. From clickbait to fake news detection: an approach based on detecting the stance of headlines to articles. In Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism, pages 84-89, 2017. 12
- [16]. Yimin Chen, Niall J Conroy, and Victoria L Rubin. Misleading online content: Recognizing clickbait as false news. In Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection, pages 15-19. ACM, 2015.
- [17]. Yusuf Perwej, "An Evaluation of Deep Learning Miniature Concerning in Soft Computing", *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)*, ISSN (Online): 2278-1021, Volume 4, Issue 2, Pages 10 - 16, 2015 DOI: 10.17148/IJARCCE.2015.4203
- [18]. Niall J Conroy, Victoria L Rubin, and Yimin Chen. Automatic deception detection: Methods for finding fake news. In Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community, page 82. American Society for Information Science, 2015.
- [19]. Ethan Fast, Binbin Chen, and Michael S Bernstein. Empath: Understanding topic signals in large-scale text. In Proceedings of the 2016 CHI Conference on Human Factors in

Computing Systems, pages 4647-4657. ACM, 2016.

Cite this article as :

Vaishali Singh, "A Framework for Predicting and Analyzing Fake News Using Machine Learning", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 9 Issue 2, pp. 204-209, March-April 2022. Available at doi : <https://doi.org/10.32628/IJSRSET229233>
Journal URL : <https://ijsrset.com/IJSRSET229233>