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.2022.11682

FAKE NEWS DETECTION USING MACHINE LEARNING

Bharathi C[1], Bhavana B K[2], Anusha S T[3], Aishwarya B N[4]

[1]Associate Professor, Department of Electronics and Communication Engineering, Global Academy of Technology, Bangalore

[2, 3, 4]Student, Department of Electronics and Communication Engineering, Global Academy of Technology, Bangalore

Abstract: Fake News is one of the major problems in the current situation. Fake News has capacity to change pinions, facts and can act as the strongest weapon in today's society. Our project uses NLP techniques for detecting the 'fake news', that is, misleading news stories which come from the non-reputable sources. The data science community has responded by taking actions against the problem. It is impossible to determine a news as real or fake accurately. So, the proposed project uses the datasets that are trained using Count Vectorizer method for the detection of fake news and its accuracy will be tested using machine learning algorithms.

Keywords: Misleading news, NLP techniques, Count Vectorizer, Train-Test split, Machine learning algorithms

I. INTRODUCTION

As an increasing quantity of our lives is spent interacting online through social media structures, an increasing number of humans generally tend consume news from social media in place of conventional information businesses. The reasons for this variation in consumption behaviours are inherent inside the nature of those social media structures. it's regularly extra well timed and less high priced to consume news on social media as compared with conventional journalism, like newspapers or tv. And it's less complicated to similarly share, discuss the information with friends or other readers on social media. It has been additionally determined that social media now outperforms tv and has become the major information source. However, as it's less expensive to deliver news online and a long way quicker and simpler to propagate through social media, large volumes of fake information, i.e., those information articles with deliberately fake records, are produced online for spread of purposes, like economic and political gain.

The considerable spread of fake information may have a giant bad impact on people and society. First, fake information can shatter the authenticity equilibrium of the news atmosphere for instance: it is obtrusive that the maximum famous fake information turned into even greater outspread on Facebook than the most standard actual mainstream information at some stage in the U.S. 2016 presidential election. 2nd, fake news deliberately persuades consumers to without a doubt accept biased or false ideals. Fake news is usually manipulated through propagandists to deliver political messages or have an impact on for instance, a few reports indicates that Russia has created fake accounts and social bots to unfold fake tales. Third, fake information adjustments the manner people interpret for instance, some fake news became just created to cause human's distrust and cause them to be confused. It's essential that we are building up methods to mechanically come across fake news broadcast on social media.

This bring about generating of the news articles that isn't absolutely genuine or maybe completely false. There even exist many websites that produce fake information nearly exclusively. They intentionally submit hoaxes, half of-truths, propaganda and disinformation affirming to be real news – regularly the usage of social media to drive web visitors and amplify their effect. The maximum desires of fake news websites are to influence the majority opinion on certain subjects (generally political). Samples of such web sites could also be in Ukraine, US, Germany, China and much of other nations. hence, fake information perhaps a global trouble also as a global challenge. Many scientists consider that fake information trouble may also be addressed through means of system gaining knowledge of AI.

We aim to provide the user with the potential to categorise the news as fake or real and check the authenticity of the internet site publishing the news, classify the information as fake or actual and also check the authenticity of the internet site publishing the information. And to create a system or model which can use the facts of beyond information reviews and expect the probabilities of an information file being fake or no longer. Numerous researchers have attempted fixing this venture in many ways to check which approach works and get desired outcomes.

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.2022.11682

II. LITERATURE SUMMARY

	Paper Details	Methodology	Results
1.	Shankar M. Patil et. al.: They aimed to provide the user with the ability to classify the news as fake or real-International Journal of Engineering Research & Technology (IJERT)-2021	To perform binary classification with the help of concepts Artificial Intelligence, Natural Language Processing and Machine Learning	They obtained the accuracy of approximate 75%.
2.	Krishna AN et.al.: They aimed at identification of fake news by proposing a system that can reliably classify fake news -2020	Machine Learning algorithms such as Naive Bayes, Passive Aggressive Classifier and Deep Neural Networks have been used on eight different datasets acquired from various sources.	The accuracy obtained is 86%.
3.	Rhethiqe Chalippatt et.al.: They proposed a model that classifies unreliable news into real and fake news after computing a score and will be able to distinguish between real and fake news -2019.	Employed SVM, KNN, Decision tree and Random Forest to build four models and compared them.	Training data consists of 66.6% of the data and testing data consists of 33.3% of the data.
4.	Mykhailo Granik et.al.: They proposed a model which collected news from three large Facebook pages each from the right and from the left, as well as three large mainstream political news pages (Politico, CNN, ABC News)	They have proposed a simple approach for fake news detection using naïve Bayes classifier. This approach was implemented as a software system and tested against a data set of Facebook news posts.	They achieved classification accuracy of approximately 74%.
5.	Madalina Erascu et.al. : They proposed a machine learning techniques, in particular supervised learning, for fake news detection	More precisely, they used a dataset of fake and real news to train a machine learning model using Scikit-learn library in Python.	Training data consists of 66.6% of the data and testing data consists of 33.3% of the data
6.	Shaikh Ansar Ahesanuddin et.al.: Some challenges can be due to a smaller number of resources like an available dataset and published literature. They proposed this in their paper, a fake news detection using machine learning techniques.	They compare three different machine learning classification techniques. Not only that, but We will be working with three different models that are Logistic Regression, Decision Tree Classifier and Random Forest Classification.	The approximate accuracy obtained by Decision Tree Classifier 99.6%.
7.	Rupali Patil :Their work helps us to detect the accuracy of the fake news using different classification techniques.	In this model, they have used classification techniques like Support Vector Machine(SVM), Naïve Bayes, Passive Aggressive Classifier.	Output of model using feature extraction techniques as Term Frequency-Inverted Document Frequency (TF-IDF) and Support Vector Machine (SVM) as classifier, has accuracy of 95.05%.
8.	Julio C. S. Reis et.al : They present a new set of features and measure the prediction performance of current approaches and features for automatic detection of fake news	It includes k-Nearest Neighbors (KNN), Naive Bayes (NB), Random Forests (RF), Support Vector Machine with RBF kernel (SVM), and XGBoost (XGB)	They obtain the accuracy of XGB-81%.

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.2022.11682

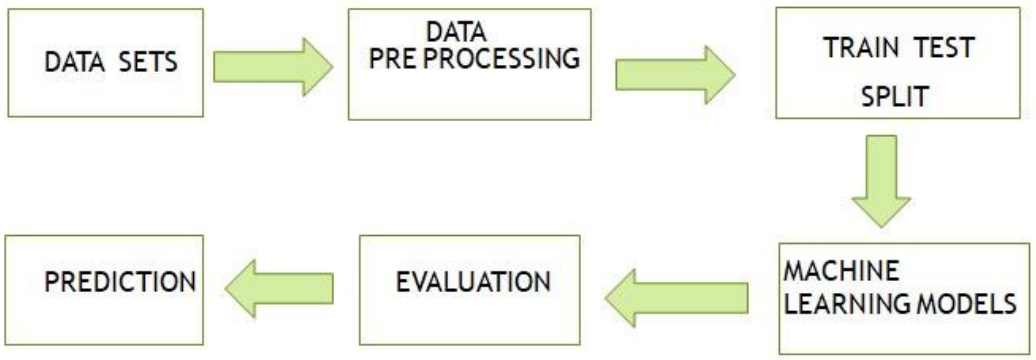
III. METHODOLOGY

We can find many datasets for fake news detection on Kaggle or many other sites. We downloaded these datasets from Kaggle. There are two datasets one for fake news and one for true news. In true news, there is 21417 news, and in fake news, there is 23481 news. Both datasets have a label column in which 0 for fake news and 1 for true news. We are combined both datasets using panda's built-in function. Then from data set few data are removed for manual testing which is used for predicting the accuracy of the system. Data set consists of subject, title, text, date so to predict it as fake or true title and date are not required hence, we drop those two columns.

We can't use text data directly because it has some unusable words and special symbols and many more things. If we used it directly without cleaning, then it is very hard for the ML algorithm to detect patterns in that text and sometimes it will also generate an error. So that we must always first clean text data.

Then we must pre-process our data which includes stemming stop words removal and punctuation removal, then we do train test split training data consists of 80% of the data set and testing data consists of 20% of the data set then they are converted into vectorized form using the TF IDF vectorizer and then they are fed to the machine learning models.

In our project we have used decision tree classifier, logistic regression, random forest and gradient boosting algorithms all these machine learning models are supervised learning methods that is it maps the input data to output data. after this process we evaluate the models using the testing dataset and we train our model using training datasets and then predicts its accuracy accordingly and if the label is zero it predicts as fake news and if the label is one then it predicts as true news.



Block Diagram

IV. CONCLUSION

Fake news research has in no way been more crucial than it is now. The concept of fake news detection in social media is particularly new and there is many research going on to accurately detect the fake news. So, in our paper we have proposed a method which may be used by other researchers to discover which combination of methods should be used to correctly predict it as fake or true. It also helps the people to take accurate decision and not believe in fake news and stop advertisers take the advantages of Fake News and stops the people from sharing links of money-making apps etc., Incidents like riots or videos that incite communal disharmony are trends that easily get millions of views within minutes such spread of false videos can also be stopped.

V. REFERENCES

[1].Shanker M Patil et. al , NTASU-2020 conference proceedings International journal of Engineering research and technology.
[2].Rhethiqe Chalippatt et. al. 2019,International conference on vision towards emerging trends in communication and networking.(ViTECoN)
[3]. Krishna A N, Member IEEE. (2020) ,Fake news detection using machine learning.
[4].Parikh, S. B., & Atrey, P. K. (2018, April). Media-Rich Fake News Detection: A Survey. In 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR).

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.2022.11682

[5]. Conroy, N. J., Rubin, V. L., & Chen, Y. (2015, November). 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 . American Society for Information Science.

[6].Ultimate guide to deal with Text Data (using Python) – for Data Scientists and Engineers by Shubham Jain, February 27, 2018.

[7].Understanding the random forest by Anirudh Palaparthi, Jan 28, at analytics vidya.

[8].Granik, M., & Mesyura, V. (2017, May). Fake news detection using naive Bayes classifier. In 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON) IEEE. [9].Aphiwongsophon, S., & Chongstitvatana, P. (2018, July). Detecting Fake News with Machine Learning Method. In 2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTICON) IEEE.

[10].Gilda, S. (2017, December). Evaluating machine learning algorithms for fake news detection. In 2017 IEEE 15th Student Conference on Research and Development (SCOReD) IEEE.

[11].Zhang, J., Cui, L., Fu, Y., & Gouza, F. B. (2018). Fake news detection with deep diffusive network model. arXiv preprint arXiv:1805.08751.

[12].Xu, K., Wang, F., Wang, H., & Yang, B. (2018, June). A First Step Towards Combating Fake News over Online Social Media. In International Conference on Wireless Algorithms, Systems, and Applications (pp. 521-531). Springer, Cham.

[13].Tschiatschek, S., Singla, A., Gomez Rodriguez, M., Merchant, A., & Krause, A. (2018, April). Fake news detection in social networks via crowd signals. In Companion of the The Web Conference 2018 on The Web Conference 2018 (pp. 517-524). Fake news detection using machine learning DEPT OF ECE 32 [14].S. I. Manzoor, J. Singla and Nikita, "Fake News Detection Using Machine Learning approaches: A systematic Review," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 2019, pp. 230-234, doi: 10.1109/ICOEI.2019.8862770.

[15]. A. J. Keya, S. Afridi, A. S. Maria, S. S. Pinki, J. Ghosh and M. F. Mridha, "Fake News Detection Based on Deep Learning," 2021 International Conference on Science & Contemporary Technologies (ICSCT), 2021, pp. 1-6, doi: 10.1109/ICSCT53883.2021.9642565.

[16]. S. Patil, S. Vairagade and D. Theng, "Machine Learning Techniques for the Classification of Fake News," 2021 International Conference on Computational Intelligence and Computing Applications (ICCICA), 2021, pp. 1-5, doi: 10.1109/ICCICA52458.2021.9697267.

[17]. N. F. Baarir and A. Djeffal, "Fake News detection Using Machine Learning," 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH), 2021, pp. 125-130, doi: 10.1109/IHSH51661.2021.9378748.

[18]. S. Chaturvedi, A. Saxena, A. Gupta and F. Alam, "Detecting Fake News using Machine Learning Algorithms," 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), 2021, pp. 212-216, doi: 10.1109/ICAC3N53548.2021.9725560.

[19]. N. L. Siva Rama Krishna and M. Adimoolam, "Fake News Detection System using Logistic Regression and Compare Textual Property with Support Vector Machine Algorithm," 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), 2022, pp. 48-53, doi: 10.1109/ICSCDS53736.2022.9760768.

[20]. K. Rajesh, A. Kumar and R. Kadu, "Fraudulent News Detection using Machine Learning Approaches," 2019 Global Conference for Advancement in Technology (GCAT), 2019, pp. 1-5, doi: 10.1109/GCAT47503.2019.8978436.

