



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Special Issue 2, April 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Fake News Detection Using Machine Learning

Bhuvanesh A^[1], Archita Ritu M A^[2], Naveen Krishnan M^[3], Neha G S^[4], Vijay M^[5]

UG Student, Department of Computer Science Engineering, Velammal Engineering College, Chennai, Tamil Nadu, India ^[1]

UG Student, Department of Computer Science Engineering, Velammal Engineering College, Chennai, Tamil Nadu, India ^[2]

UG Student, Department of Computer Science Engineering, Velammal Engineering College, Chennai, Tamil Nadu, India ^[3]

UG Student, Department of Computer Science Engineering, Velammal Engineering College, Chennai, Tamil Nadu, India ^[4]

UG Student, Department of Computer Science Engineering, Velammal Engineering College, Chennai, Tamil Nadu, India^[5]

ABSTRACT: Fake news detection typically involves analysing text to identify misleading or false information. There are several approaches to detecting fake news, including: Natural Language Processing (NLP) techniques such as sentiment analysis, topic modelling, and named entity recognition. Machine learning algorithms that use supervised or unsupervised learning to classify news articles as fake or real based on features such as vocabulary, syntax, and semantics. Network analysis to identify the source and spread of fake news through social media platforms and online communities. Overall, the key to detecting fake news is to carefully analyse the content of an article and look for patterns or signals that suggest that the information may not be accurate or trustworthy. By combining different techniques and approaches, you can increase your chances of accurately detecting fake news and promoting reliable information.

KEYWORDS: Natural Language Processing (NLP), Machine learning algorithms, Network analysis

I. INTRODUCTION

Today, with the rise of digital media, it is considered as an issue that the society has to address. Recently the social networks online has developed extremely where they have to handle with fake news in terms of business and political use which has been emerging enormously as well as broadcasting through online. By these words the users of social network have been affected since they are fake news that created the remarkable effects among the group of people. It is more important in developing the trust among the people using social network. So, it is very important to detect the fake news at the earliest [1]

To detect fake news there are several techniques used such as, **sentiment analysis** is the process of detecting positive or negative sentiment in text. It's often used by businesses to detect sentiment in social data, gauge brand reputation, and understand customers, **named entity recognition (NER)** — sometimes referred to as entity chunking, extraction, or identification — is the task of identifying and categorizing key information (entities) in text. An entity can be any word or series of words that consistently refers to the same thing and **topic modelling** is an unsupervised machine learning technique that's capable of scanning a set of documents, detecting word and phrase patterns within them, and automatically clustering word groups and similar expressions that best characterize a set of documents.

II. RELATED WORK

Fake news often has the aim of damaging the reputation of a person or entity, or making money through advertising revenue. Although false news has always been spread throughout history, the term "fake news" was first used in the 1890s when sensational reports in newspapers were common. Social media for news consumption is a critical sword. On the one hand, it is low cost, easy access, and rapid publishing of information lead people to seek out and



consume news from social media. On the other hand, it enables the wide spread of ‘fake news’, i.e., low quality news with intentionally false information. for example, 62 percent of U.S. adults get news on social media in 2016, while in 2012, only 49 percent reported seeing news on social media. It was also found that social media now defeat television as the major news source. It was estimated that over 1 million tweets are related to fake news ‘Pizzagate’ by the end of the presidential election[2]. Not only in social media also in news organizations: Journalists and news organizations use fake news detection techniques to verify the accuracy of news stories and to fact-check claims made by politicians or public figures. This involves analysing the content of news articles and checking the validity of sources and information, Government agencies may use fake news detection to identify and combat disinformation campaigns or propaganda efforts aimed at influencing public opinion. This can involve monitoring social media and news sources for false or misleading information and taking action to counter it and in Researchers use fake news detection techniques to study the prevalence and impact of fake news in society. This involves analysing large datasets of news articles and social media posts to identify patterns of behaviour and language associated with fake news. Always check the source or “author” of the content you are viewing. A reliable and reputable source: provides in-depth, evidence-based information that is authoritative, whether news is unbiased and offers a balanced approach whether news is well known and has built a reputation of being trustworthy over time Look for: **Authority** - who is the author and where did the item originate? Check on the individual, company or institution, to see where their biases (personal inclinations/favouritisms) are. Is the author’s name listed? What are their credentials that make them an expert in the topic? **Accuracy** - information that is accurate and free from errors is considered more reliable. One or two typos are ok but information that is presented in a sloppy manner with too many spelling or grammatical mistakes comes across as unprofessional. Are there links to other resources to back up the claims one source is making? For statistical references and data check where they come from. Numbers can be easily changed and manipulated to show personal opinions. Fact and opinion are different. **Objectivity** - check to see if the information is presented with the least amount of prejudice or personal bias. Is it an opinion or is it trying to sell you something? **Timeliness** - when was the information first published? Is the content you are looking at up-to-date, or the website updated regularly, or is it something that is old made to look like new information? [3].

III. METHODOLOGY

Logistic Regression analysis predicts the odds of an outcome of a categorical variable based on one or more predictor variables. A categorical variable is one that can take on a limited number of values, levels, or categories, such as "valid" or "invalid". A major advantage of Logistic Regression is its predictions are always between 0 and 1, unlike Linear Regression. Regularization is extremely important in logistic regression modelling. Without regularization, the asymptotic nature of logistic regression would keep driving loss towards 0 in high dimensions. Consequently, most logistic regression models use one of the following two strategies to dampen model complexity such as L2 regularization and early stopping, that is, limiting the number of training steps or the learning rate [4].

The **random forest algorithm** is an extension of the bagging method as it utilizes both bagging and feature randomness to create an uncorrelated forest of decision trees. Feature randomness, also known as feature bagging or “the random subspace method”, generates a random subset of features, which ensures low correlation among decision trees. This is a key difference between decision trees and random forests [5].

True Positive (TP): <ul style="list-style-type: none"> Reality: A wolf threatened. Shepherd said: "Wolf" Outcome: Shepherd is a hero. 	False Positive (FP): <ul style="list-style-type: none"> Reality: No wolf threatened. Shepherd said: "Wolf." Outcome: Villagers are angry at shepherd for waking them up.
False Negative (FN): <ul style="list-style-type: none"> Reality: A wolf threatened. Shepherd said: "No wolf." Outcome: The wolf ate all the sheep. 	True Negative (TN): <ul style="list-style-type: none"> Reality: No wolf threatened. Shepherd said: "No wolf." Outcome: Everyone is fine.

Fig 1.0

Figure 1.0 shows summarization of true positive, false positive, false negative, true negative. A true positive is an outcome where the model correctly predicts the positive class. Similarly, a true negative is an outcome where the model correctly predicts the negative class. A false positive is an outcome where the model incorrectly predicts the positive class. And a false negative is an outcome where the model incorrectly predicts the negative class.



IV. EXPERIMENTAL RESULTS

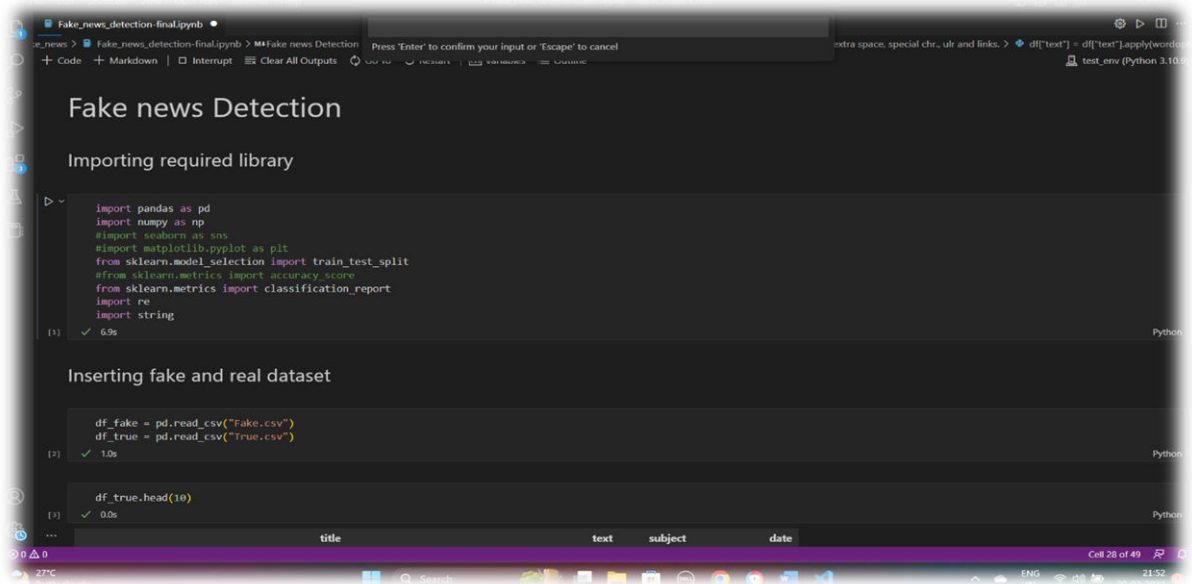


Fig. 2a

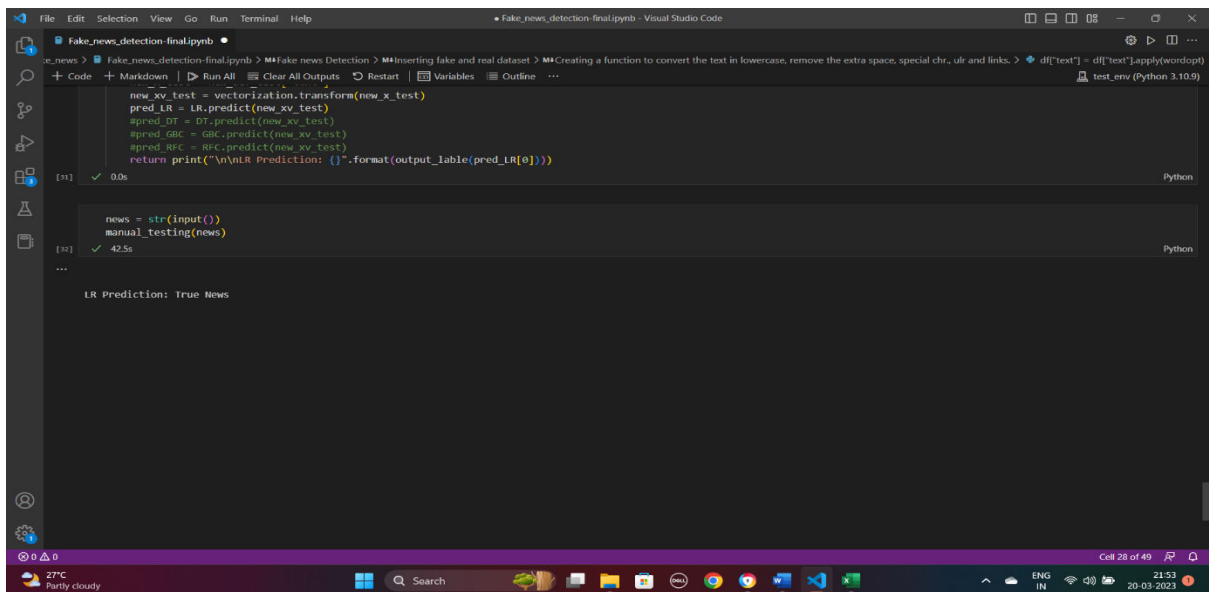


Fig. 2b

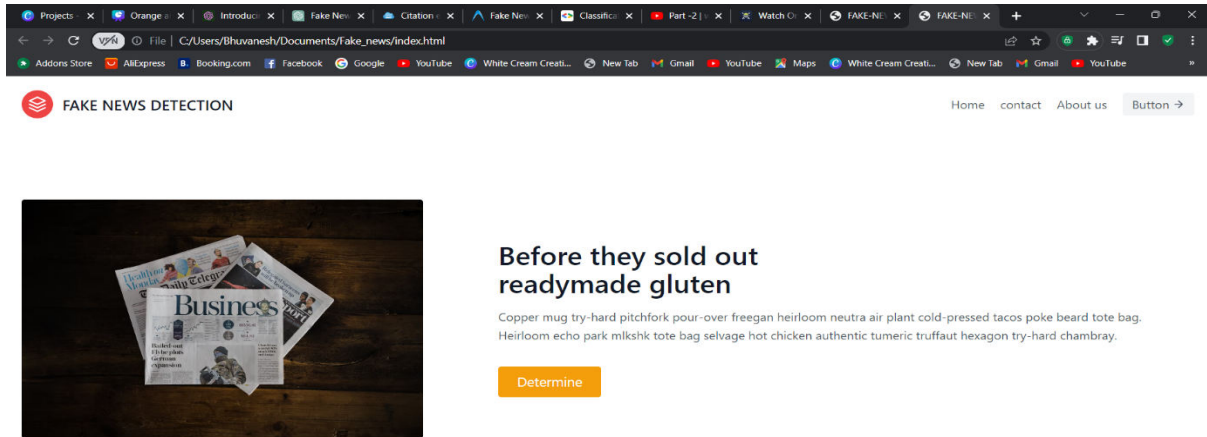


Fig. 3a

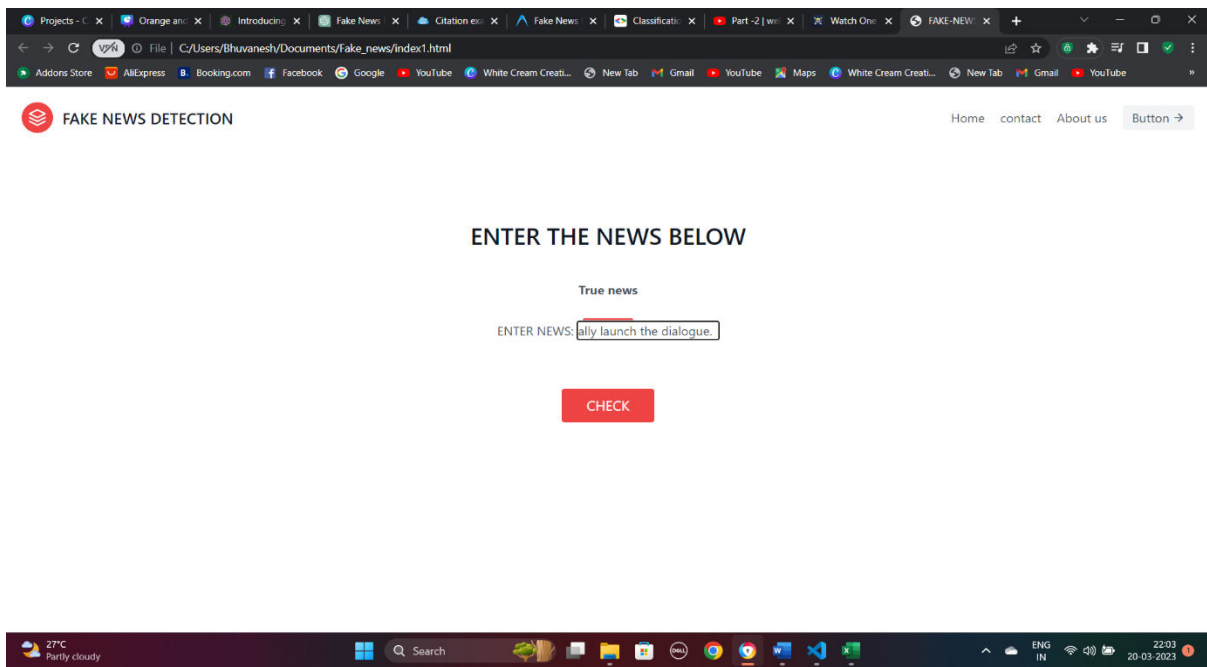


Fig. 3b

Figures 2a, 2b shows the source code of the project. Figures 3a, 3b shows the website of the project.

V. CONCLUSION

We have implemented fake news detection using machine learning. Our algorithm logistic regression, random forest, true positive, false positive, true negative and false negative successfully detected true and false news from the processed data.



REFERENCES

- [1] Dheeraj Kumar, 2022, Fake News Classification Using Fuzzy Based Deep Convolutional Neural Networks (FDNN) on Social Media Data, AIJR Publisher
- [2] Palash Sukla Das, Dipan Acharjee, Partha Pritam Deb, 2022, Fake News Detector on social media Using Machine Learning, AIJR Publisher
- [3] <https://settlement.org/ontario/daily-life/communication/ethnic-and-cultural-media/what-is-fake-news-and-how-to-stop-spreading-misinformation/>
- [4] [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/), 2022, ML concepts, Logistic Regression, google developers
- [5] Ronny Hänsch, 2017, Handbook of Random Forests Theory and Applications for Remote Sensing, World Scientific Book



SJIF: 8.423



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INTERNATIONAL CENTRE



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details