

# FAKE NEWS DETECTION ON SOCIAL MEDIA NETWORKS

Shivam Mishra<sup>1</sup>, Priyanka Prasad<sup>2</sup>, Priyanka Sodhi<sup>3</sup>, Dr. D. R. Ingle<sup>4</sup>

<sup>1,2,3</sup>Computer Engineering, Bharati Vidyapeeth College of Engineering, Navi Mumbai, India

<sup>4</sup>Professor, Head of Department of Computer Engineering, Bharati Vidyapeeth College of Engineering, Navi Mumbai, India

**Abstract** - The Internet has offered us the privilege to access the latest news of the events happening worldwide within minimum time and effort. But as said "Every coin has two sides", so does the internet. On one hand, the internet has revolutionized news consumption around the globe by enabling each and every person to gain information in just a few clicks on their device, but on the other hand, the very blessing becomes a curse as the internet makes it easier for the illegitimate media stream to exploit the advantages of internet to spread fake news.

"Fake News" is incorrect, untruthful, and untrustworthy information that has the immense power of influencing public opinion and creating an extremely negative impact on individuals. According to a report, 69% of Indians were infested with fake news during Coronavirus Lockdown 2020. Hence, it is crucial to make an effort to detect fake news and prevent it from circulating on the internet.

In this paper, we present "Fakebox", a system designed on machine learning and natural language processing techniques for detecting fake news. To build "Fakebox", we used a dataset set of a combination of fake and real news from different domains like technology, politics, entertainment, sports, etc., to train the model using Scikit-learn and NLTK libraries in Python. We extracted features from the dataset using a text representation model that is Bag-of-Words for a comprehensive understanding of the dataset which has resulted in the model giving up to 85% accuracy.

**Key Words:** Fake News, Natural Language Processing, Artificial Intelligence, Machine Learning, Bag of Words

## 1. INTRODUCTION

Today, in the age of technology, while people spend most of their time interacting on social media platforms, they tend to consume hundreds of information from random sources every day. According to a report, there are 4.66 billion active users on the internet; more and more people are adding up due to various benefits of the internet like easy user accessibility, low cost, freedom to express one's opinion, etc.

The change in news consumption from traditional newspapers to digital media has happened mainly because of the power of the internet to reach the intended audience

with the latest news in real-time. But the quality of news available online is still questionable. It is because anyone on the internet is free to create content and spread it all over the internet and there is no regulating authority to verify and control the information, unlike the traditional newspapers. The information which is available online seems so appealing that people fall prey to it and make opinions that might be detrimental to society.

Fake News is misleading information that is deliberately designed to manipulate the users with intention of spoiling the reputation of an individual, company, etc. It has the potential to break the balance of the news ecosystem by hurting the sentiments of some people which has even resulted in some brutal riots in our history. One of the major examples of fake news can be considered as when India had introduced a new 2,000 Rupee currency note for the public, as a part of 2016 Indian banknote demonetization, multiple fake news reports about banknotes embedded with "spying technology" or "Nano-GPS Chip" went viral on various social media platforms. "Fake News" was even named as Word of the Year by the Collins Dictionary in 2017.

## 1.1 SIGNIFICANCES OF FAKE NEWS DETECTION

Over the latter, a long time, the quick and hazardous advancement of social media has seen broad development within the amount of fake news. These days, fake news is annoying, prominent, diverting, and all over the online world. It has significant impacts on both people and society. Therefore, it is noteworthy for building a successful discovery framework for fake news recognizable proof. The basic characteristics of fake news can be summarized as follows:

- **The volume of fake news:** Without any confirmation method, anybody can effortlessly type in fake news on the internet. There are various web pages that are built up intentionally to spread fake news and stories, such as denverguardian.com, ABCnews.com.co, and so on. The domain names of those websites often match with genuine news organization's websites and are intentionally made to convey deceptions, publicity, and disinformation, frequently for financial or political gain. Hence, an enormous sum of fake articles is dispersed through the web, indeed without the user's awareness.

- **The variety of fake news:** There are a few near definitions of fake news, such as rumors, parody news, fake audits, deception, fake notices, scheme speculations, untrue articulation by politicians, etc., which influence every angle of people's lives. With the expanding popularity of social media, fake news can rule the public's opinions, interface, and choices. In addition, fake news alters the way that individuals connected with genuine news. A few fake news are made intentioned to delude and confuse social media clients, particularly youthful students and old people who are purge of self-protection consciousness. Recently, a story shared on Facebook utilized selective TV evaluations information to create the deluding claim that Cable News Organize (CNN) was not one of the 10 most observed cable systems in 2018.
- **The velocity of fake news:** Websites that create fake news tend to be short-lived. For illustration, numerous dynamic fake news webpages amid 2016 U.S. elections did not exist after the campaign. As more consideration is paid to fake news in recent years, more fake news generators are nothing but a temporal streak in order to maintain a strategic distance from detection by the detection frameworks. Besides, most of the fake news on social media is centering on current events and hot issues to bring more attention to online clients. The real-time nature of fake news on social media makes distinguishing online fake news indeed more difficult. It is complicated to assess how numerous online clients are included with a certain piece of instant message, and it is difficult to tell when and how the far-reaching results of fake news halt.

In this paper, we propose our model that is "Fakebox" for fake news detection using machine learning and natural language processing techniques. In particular, we studied and developed methods and tools for detecting fake news, also proposing a methodology for that purpose and implementing an algorithm that classifies whether the news is fake or real.

## 2. LITERATURE SURVEY

In this paper, we mainly focus on designing a fake news detection system that assists users to find whether the news which they are reading is real news or fake news. Firstly, we would use the machine learning libraries such as Scikit-learn (<http://scikit-learn.org>) and NLTK (<http://nltk.org>) in Python since it has built-in methods that implement different classification approaches.

We will gather a dataset set of a combination of fake and real news from different domains like technology, politics, entertainment, sports, education, foreign affairs, etc. After obtaining a proper dataset, we will perform a data cleaning

and exploration process which will include elimination of various features such as the date on which the article was published, author's name, category of news, article containing empty body, etc. For text representation models, we would use the Bag-of-Words model to extract linguistic features such as percentage of words implying positive and negative emotions, informal language, punctuations, etc., from the dataset. After the normalization and filtration of the dataset are done, it will be divided into a 70:30 ratio as a training and testing dataset. The model will be trained with different hyperparameters to achieve maximum accuracy. After training the model, it will be evaluated using a testing dataset and will be available for online users to enter their query and get efficient and accurate results. The analysis will be done based on the contents of the news article.

This paper mainly focuses on helping the user by providing the facility to verify the credibility of the news article with ease. For an easy and relevant user experience, the website will be made keeping in mind the interaction-oriented design. The website will have features like an input field for entering the news article and contact page. In addition, the website will also provide links to legitimate news sources. The important goal of the system is to make people aware of fake news and prevent its spread at an early stage before it makes a detrimental impact on society.

## 3. CONCLUSIONS

The problem of fake news and its impact is at all time high now. This is mainly because of the advancement of technology and communication mediums which has enabled everyone to spread a piece of news without proper verification. There are hundreds of anonymously hosted websites which is especially used for spreading fake news on the internet. Once a fake news is spread, it is impossible to control its propagation and in no matter of time, it becomes a dangerous piece of article that has the potential to destroy the harmony of society and cause religious riots and spoil the image of a person or a company. This is why it is extremely important to make an effort to verify the content of news and prevent it's spreading.

Hence, in this paper, we proposed a model using machine learning and natural language processing techniques which aims to verify the credibility of the news article based on their characteristics and classify them as fake or real news. This would give the users a platform to check the authenticity of daily piece of news they read on the social media. Along with that, the model also aims to provide links to real and legitimate news sources that are worldwide known so that the users can be made aware and have access to authentic and legitimate channels for real news which also might help in decreasing their news consumption from non-credible sources.

## ACKNOWLEDGEMENT

A project bridges the gap between theoretical concepts and practical working. And we have explored this field with the help of our professors. We would like to express that it's our proud privilege to release the feelings of our gratitude to several people who helped us directly or indirectly to conduct this project work. We express our heart full indebtedness and owe a deep sense of gratitude to our Professor and Project Mentor Dr. Dayanand R. Ingle for his sincere guidance.

The study has indeed helped each one of us as we were able to understand a lot about different software and hardware aspects of our project which has helped us use our abilities to find the solution for a problem. It has also encouraged us to explore more such problems in-dept and find interesting solutions. We would also like to thank all the staff members of our computer engineering department of Bharati Vidyapeeth College of Engineering for their generous attitude and helping nature throughout the course of project.

## REFERENCES

[1] Shao, C., Ciampaglia, G. L., Varol, O., Flammini, A., & Menczer, F. (2017). The spread of fake news by social bots. arXiv preprint arXiv:1707.07592, 96-104.

[2] Hunt, E. (2016). What is fake news? How to spot it and what you can do to stop it. *The Guardian*, 17.

[3] Ruchansky, N., Seo, S., & Liu, Y. (2017, November). Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 797-806). ACM.

[4] Reis, J. C., Correia, A., Murai, F., Veloso, A., Benevenuto, F., & Cambria, E. (2019). Supervised Learning for Fake News Detection. *IEEE Intelligent Systems*, 34(2), 76-81.

[5] Pal, S., Kumar, T. S., & Pal, S. (2019). Applying Machine Learning to Detect Fake News. *Indian Journal of Computer Science*, 4(1), 7-12

[6] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22-36, 2017. View at: [Publisher Site](#) | [Google Scholar](#).

[7] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146-1151, 2018. View at: [Publisher Site](#) | [Google Scholar](#).

[8] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *Journal of Economic Perspectives*, vol. 31, no. 2, pp. 211- 236, 2017. View at: [Publisher Site](#) | [Google Scholar](#).

[9] V. L. Rubin, N. Conroy, Y. Chen, and S. Cornwell, "Fake news or truth? using satirical cues to detect potentially misleading news," in *Proceedings of the Second Workshop on Computational Approaches to Deception Detection*, pp. 7-17, San Diego, CA, USA, 2016. View at: [Google Scholar](#).