



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

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 10, October 2024

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 8.524**

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)

# Fake News Detection System using Machine Learning

**Prof. Jayashri Bhoj<sup>1</sup>, Vivek Kumar<sup>2</sup>, Raghav Kumar<sup>3</sup>, Abhishek Kumar<sup>4</sup>, Shubham Kumar<sup>5</sup>**

Associate Professor, Department of Computer Engineering, Sandip Institute of Technology and Research Centre  
Nashik, Maharashtra, India<sup>1</sup>

Department of Computer Engineering, Sandip Institute of Technology and Research Centre, Nashik,  
Maharashtra, India<sup>2 3 4 5</sup>

**ABSTRACT:** In our modern era where the internet is ubiquitous, everyone relies on various online resources for news. Along with the increase in the use of social media platforms like Facebook, Twitter, etc. news spread rapidly among millions of users within a very short span of time. The spread of fake news has far-reaching consequences like the creation of biased opinions to swaying election outcomes for the benefit of certain candidates. Moreover, spammers use appealing news headlines to generate revenue using advertisements via click-baits. In this paper, we aim to perform binary classification of various news articles available online with the help of concepts pertaining to Artificial Intelligence, Natural Language Processing and Machine Learning. We aim to provide the user with the ability to classify the news as fake or real and also check the authenticity of the website publishing the news.

**KEYWORDS:** Internet, Social Media, Fake News, Classification, Artificial Intelligence, Machine Learning, Websites, Authenticity.

## I. INTRODUCTION

As an increasing amount of our lives is spent interacting online through social media platforms, more and more people tend to hunt out and consume news from social media instead of traditional news organizations.[1] The explanations for this alteration in consumption behaviours are inherent within the nature of those social media platforms: (i) it's often more timely and fewer expensive to consume news on social media compared with traditional journalism, like newspapers or television; and (ii) it's easier to further share, discuss, and discuss the news with friends or other readers on social media. For instance, 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 [1]. It had been also found that social media now outperforms television because the major news source. Despite the benefits provided by social media, the standard of stories on social media is less than traditional news organizations. However, because it's inexpensive to supply news online and far faster and easier to propagate through social media, large volumes of faux news, i.e., those news articles with intentionally false information, are produced online for a spread of purposes, like financial and political gain. It had been estimated that over 1 million tweets are associated with fake news —"Pizzagate" by the top of the presidential election. Given the prevalence of this new phenomenon, —"Fake news" was even named the word of the year by the Macquarie dictionary in 2016 [2]. The extensive spread of faux news can have a significant negative impact on individuals and society. First, fake news can shatter the authenticity equilibrium of the news ecosystem for instance; it's evident that the most popular fake news was even more outspread on Facebook than the most accepted genuine mainstream news during the U.S. 2016 presidential election. Second, fake news intentionally persuades consumers to simply accept biased or false beliefs. Fake news is typically manipulated by propagandists to convey political messages or influence for instance, some report shows that Russia has created fake accounts and social bots to spread false stories. Third, fake news changes the way people interpret and answer real news, for instance, some fake news was just created to trigger people's distrust and make them confused; impeding their abilities to differentiate what's true from what's not. To assist mitigate the negative effects caused by fake news (both to profit the general public and therefore the news ecosystem). It's crucial that we build up methods to automatically detect fake news broadcast on social media [3].

Internet and social media have made the access to the news information much easier and comfortable [2]. Often Internet users can pursue the events of their concern in online form, and increased number of the mobile devices makes this process even easier. But with great possibilities come great challenges. Mass media have an enormous influence on

the society, and because it often happens, there's someone who wants to require advantage of this fact. Sometimes to realize some goals mass-media may manipulate the knowledge in several ways. This result in producing of the news articles that isn't completely true or maybe completely false. There even exist many websites that produce fake news almost exclusively. They intentionally publish hoaxes, half-truths, propaganda and disinformation asserting to be real news – often using social media to drive web traffic and magnify their effect. The most goals of faux news websites are to affect the general public opinion on certain matters (mostly political). Samples of such websites could also be found in Ukraine, United States of America, Germany, China and much of other countries [4]. Thus, fake news may be a global issue also as a worldwide challenge. Many scientists believe that fake news issue could also be addressed by means of machine learning and AI [5]. There's a reason for that: recently AI algorithms have begun to work far better on many classification problems (image recognition, voice detection then on) because hardware is cheaper and larger datasets are available. There are several influential articles about automatic deception detection. In [6] the authors provide a general overview of the available techniques for the matter. In [7] the authors describe their method for fake news detection supported the feedback for the precise news within the micro blogs. In [8] the authors actually develop two systems for deception detection supported support vector machines and Naive Bayes classifier (this method is employed within the system described during this paper as well) respectively. They collect the info by means of asking people to directly provide true or false information on several topics – abortion, execution and friendship. The accuracy of the detection achieved by the system is around 70%. This text describes an easy fake news detection method supported one among the synthetic intelligence algorithms – naïve Bayes classifier, Random Forest and Logistic Regression. The goal of the research is to look at how these particular methods work for this particular problem given a manually labelled news dataset and to support (or not) the thought of using AI for fake news detection. The difference between these article and articles on the similar topics is that during this paper Logistic Regression was specifically used for fake news detection; also, the developed system was tested on a comparatively new data set, which gave a chance to gauge its performance on a recent data.

They often have grammatical mistakes. They are often emotionally coloured. They often try to affect readers' opinion on some topics. Their content is not always true. They often use attention seeking words and news format and click baits. They are too good to be true. Their sources are not genuine most of the times [9].

## II. LITERATURE SURVEY

Mykhailo Granik et. al. in their paper [3] shows 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 were collected 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 achieved classification accuracy of approximately 74%. Classification accuracy for fake news is slightly worse. This may be caused by the skewness of the dataset: only 4.9% of it is fake news.

Himank Gupta et. al. [10] gave a framework based on different machine learning approach that deals with various problems including accuracy shortage, time lag (BotMaker) and high processing time to handle thousands of tweets in 1 sec. Firstly, they have collected 400,000 tweets from HSpam14 dataset. Then they further characterize the 150,000 spam tweets and 250,000 non- spam tweets. They also derived some lightweight features along with the Top-30 words that are providing highest information gain from Bag-of-Words model. 4. They were able to achieve an accuracy of 91.65% and surpassed the existing solution by approximately 18%.

Marco L. Della Vedova et. al. [11] first proposed a novel ML fake news detection method which, by combining news content and social context features, outperforms existing methods in the literature, increasing its accuracy up to 78.8%. Second, they implemented their method within a Facebook Messenger Chabot and validate it with a real-world application, obtaining a fake news detection accuracy of 81.7%. Their goal was to classify a news item as reliable or fake; they first described the datasets they used for their test, then presented the content-based approach they implemented and the method they proposed to combine it with a social-based approach available in the literature. The resulting dataset is composed of 15,500 posts, coming from 32 pages (14 conspiracy pages, 18 scientific pages), with more than 2, 300, 00 likes by 900,000+ users. 8,923 (57.6%) posts are hoaxes and 6,577 (42.4%) are non-hoaxes.

Cody Buntain et. al. [12] develops a method for automating fake news detection on Twitter by learning to predict accuracy assessments in two credibility-focused Twitter datasets: CREDBANK, a crowd-sourced dataset of accuracy assessments for events in Twitter, and PHEME, a dataset of potential rumours in Twitter and journalistic assessments of their accuracies. They apply this method to Twitter content sourced from BuzzFeed’s fake news dataset. A feature analysis identifies features that are most predictive for crowd-sourced and journalistic accuracy assessments, results of which are consistent with prior work. They rely on identifying highly retweeted threads of conversation and use the features of these threads to classify stories, limiting this work’s applicability only to the set of popular tweets. Since the majority of tweets are rarely retweeted, this method therefore is only usable on a minority of Twitter conversation threads.

In his paper, Shivam B. Parikh et. al. [13] aims to present an insight of characterization of news story in the modern diaspora combined with the differential content types of news story and its impact on readers. Subsequently, we dive into existing fake news detection approaches that are heavily based on text-based analysis, and also describe popular fake news datasets. We conclude the paper by identifying 4 key open research challenges that can guide future research. It is a theoretical approach which gives illustrations of fake news detection by analysing the psychological factors.

### III. METHODOLOGY

This paper explains the system which is developed in three parts. The first part is static which works on machine learning classifier. We studied and trained the model with 4 different classifiers and chose the best classifier for final execution. The second part is dynamic which takes the keyword/text from user and searches online for the truth probability of the news. The third part provides the authenticity of the URL input by user.

In this paper, we have used Python and its Sci-kit libraries [14]. Python has a huge set of libraries and extensions, which can be easily used in Machine Learning. Sci-Kit Learn library is the best source for machine learning algorithms where nearly all types of machine learning algorithms are readily available for Python, thus easy and quick evaluation of ML algorithms is possible. We have used Django for the web-based deployment of the model, provides client-side implementation using HTML, CSS and Javascript. We have also used BeautifulSoup (bs4), requests for online scrapping.

#### A. SYSTEM DESIGN-

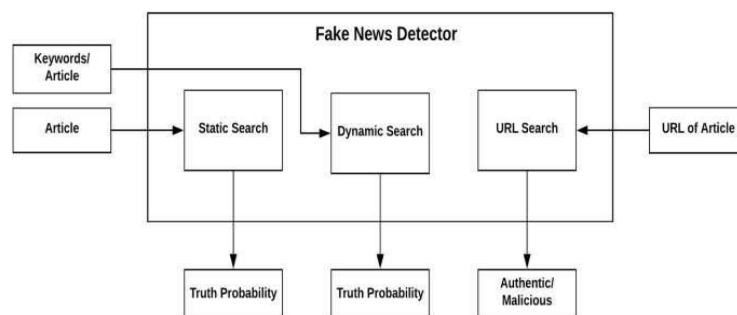


Figure 1: System Design

#### B. System Architecture-

##### i) Static Search-

The architecture of Static part of fake news detection system is quite simple and is done keeping in mind the basic machine learning process flow. The system design is shown below and self-explanatory. The main processes in the design are-

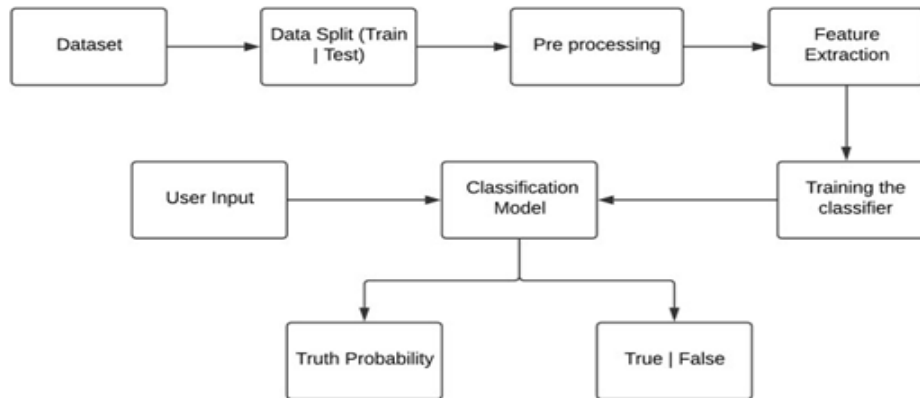


Figure 2: System Architecture

ii) Dynamic Search-

The second search field of the site asks for specific keywords to be searched on the net upon which it provides a suitable output for the percentage probability of that term actually being present in an article or a similar article with those keyword references in it.

iii) URL Search-

The third search field of the site accepts a specific website domain name upon which the implementation looks for the site in our true sites database or the blacklisted sites database. The true sites database holds the domain names which regularly provide proper and authentic news and vice versa. If the site isn't found in either of the databases then the implementation doesn't classify the domain it simply states that the news aggregator does not exist.

#### IV. CONCLUSION

In the 21st century, the majority of the tasks are done online. Newspapers that were earlier preferred as hardcopies are now being substituted by applications like Facebook, Twitter, and news articles to be read online. Whatsapp's forwards are also a major source. The growing problem of fake news only makes things more complicated and tries to change or hamper the opinion and attitude of people towards use of digital technology. When a person is deceived by the real news two possible things happen- People start believing that their perceptions about a particular topic are true as assumed. Thus, in order to curb the phenomenon, we have developed our Fake news Detection system that takes input from the user and classify it to be true or fake. To implement this, various NLP and Machine Learning Techniques have to be used. The model is trained using an appropriate dataset and performance evaluation is also done using various performance measures. The best model, i.e. the model with highest accuracy is used to classify the news headlines or articles. As evident above for static search, our best model came out to be Logistic Regression with an accuracy of 65%. Hence we then used grid search parameter optimization to increase the performance of logistic regression which then gave us the accuracy of 75%. Hence we can say that if a user feed a particular news article or its headline in our model, there are 75% chances that it will be classified to its true nature.

The user can check the news article or keywords online; he can also check the authenticity of the website. The accuracy for dynamic system is 93% and it increases with every iteration.

We intent to build our own dataset which will be kept up to date according to the latest news. All the live news and latest data will be kept in a database using Web Crawler and online database.

#### V. FUTURE SCOPE

Incorporation of Multimodal Data: Current systems primarily rely on textual content to detect fake news. Future work could integrate multimodal data sources such as images, videos, and audio clips alongside textual analysis. Analyzing

media content in combination with the text would help identify deepfakes or manipulated content, further improving the system's detection capability.

**Real-time Detection and Analysis:** As social media platforms enable rapid dissemination of information, fake news can spread within minutes. Future systems should aim to detect and flag fake news in real-time, providing live updates and alerts to users and media outlets to prevent the spread of misinformation.

**Enhanced Natural Language Understanding:** Advanced natural language processing (NLP) techniques, such as contextual embeddings (e.g., BERT, GPT), could be leveraged to improve the understanding of complex and nuanced language. Future systems could also better comprehend sarcasm, irony, or political bias in news articles, which is currently challenging for machine learning algorithms.

**Cross-lingual Fake News Detection:** Fake news is a global issue that transcends language barriers. Future research could focus on developing systems that work across multiple languages and dialects. Cross-lingual models, along with machine translation technologies, could enhance the ability to detect fake news in different languages.

**User Behavior and Social Network Analysis:** Future systems could consider user behavior and social network structures to improve fake news detection. By analyzing how fake news spreads through networks and identifying the most influential nodes, algorithms could predict and prevent misinformation from going viral.

**Explainable AI for Trustworthiness:** As machine learning models become more complex, it's crucial to develop systems that provide transparency and explainability in their predictions. Future research can focus on creating explainable AI (XAI) models that allow users to understand why a particular article was flagged as fake. This will improve user trust and system adoption.

**Collaboration with Human Fact-Checkers:** Automation of fake news detection is a significant step forward, but collaboration with expert human fact-checkers can further enhance accuracy. Future systems could leverage human-in-the-loop approaches, where machine learning models and human evaluators work together to verify complex claims more effectively.

**Addressing Evolving Misinformation Techniques:** Fake news creators constantly adapt their strategies to bypass detection systems. Future work should focus on creating adaptive learning models that evolve with these techniques. Incorporating adversarial learning and continuous model retraining could help stay ahead of sophisticated misinformation tactics.

**Ethical and Privacy Considerations:** As fake news detection systems become more pervasive, ethical and privacy considerations will need to be addressed. Future research could explore how to balance the detection of misinformation with user privacy and prevent censorship or misuse of detection tools. Ensuring fairness, accountability, and transparency in these systems will be a critical focus area.

**Scalability and Integration with Social Platforms:** Future systems should aim to scale effectively and integrate seamlessly with social media platforms, news websites, and search engines. Building large-scale solutions that can handle massive amounts of data, while maintaining high accuracy, will be vital for widespread adoption and impact.

## REFERENCES

1. Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu, —Fake News Detection on Social Media: A Data Mining Perspective| arXiv:1708.01967v3 [cs.SI], 3 Sep 2017
2. Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu, —Fake News Detection on Social Media: A Data Mining Perspective| arXiv:1708.01967v3 [cs.SI], 3 Sep 2017
3. M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kiev, 2017, pp. 900-903.
4. Fake news websites. (n.d.) Wikipedia. [Online]. Available: [https://en.wikipedia.org/wiki/Fake\\_news\\_website](https://en.wikipedia.org/wiki/Fake_news_website). Accessed Feb. 6, 2017
5. Cade Metz. (2016, Dec. 16). The bittersweet sweepstakes to build an AI that destroys fake news.

6. Conroy, N., Rubin, V. and Chen, Y. (2015). —Automatic deception detection: Methods for finding fake news | at Proceedings of the Association for Information Science and Technology, 52(1), pp.1-4.
7. Markines, B., Cattuto, C., & Menczer, F. (2009, April). —Social spam detection |. In Proceedings of the 5th International Workshop on Adversarial
8. Information Retrieval on the Web (pp. 41-48) Rada Mihalcea , Carlo Strapparava, The lie detector: explorations in the automatic recognition of deceptive language, Proceedings of the ACL-IJCNLP
9. Kushal Agarwalla, Shubham Nandan, Varun Anil Nair, D. Deva Hema, —Fake News Detection using Machine Learning and Natural Language Processing, | International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-6, March 2019
10. H. Gupta, M. S. Jamal, S. Madisetty and M. S. Desarkar, "A framework for real-time spam detection in Twitter," 2018 10th International Conference on Communication Systems & Networks (COMSNETS), Bengaluru, 2018, pp. 380-383
11. M. L. Della Vedova, E. Tacchini, S. Moret, G. Ballarin, M. DiPierro and L. de Alfaro, "Automatic Online Fake News Detection Combining Content and Social Signals," 2018 22nd Conference of Open Innovations Association (FRUCT), Jyväskylä, 2018, pp. 272-279.
12. C. Buntain and J. Golbeck, "Automatically Identifying Fake News in Popular Twitter Threads," 2017 IEEE International Conference on Smart Cloud (SmartCloud), New York, NY, 2017, pp. 208-215.
13. S. B. Parikh and P. K. Atrey, "Media-Rich Fake News Detection: A Survey," 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Miami, FL, 2018, pp. 436-441
14. Scikit-Learn- Machine Learning In Python Dataset- Fake News detection William Yang Wang. " liar, liar pants on \_re": A new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648, 2017.
15. Shankar M. Patil, Dr. Praveen Kumar, —Data mining model for effective data analysis of higher education students using MapReduce | IJERMT, April 2017 (Volume-6, Issue-4).
16. Aayush Ranjan, — Fake News Detection Using Machine Learning |, Department Of Computer Science & Engineering Delhi Technological University, July 2018.
17. Patil S.M., Malik A.K. (2019) Correlation Based Real-Time Data Analysis of Graduate Students Behaviour. In: Santosh K., Hegadi R. (eds) Recent Trends in Image Processing and Pattern Recognition. RTIP2R 2018. Communications in Computer and Information Science, vol 1037. Springer, Singapore.
18. Badreesh Shetty, "Natural Language Processing (NLP) for machine learning | at towardsdatascience, Medium.
19. NLTK 3.5b1 documentation, Nltk generate n gram Ultimate guide to deal with Text Data (using Python) – for Data Scientists and Engineers by Shubham Jain, February 27, 2018
20. Understanding the random forest by Anirudh Palaparthi, Jan 28, at analytics vidya.
21. Understanding the random forest by Anirudh Palaparthi, Jan 28, at analytics vidya.
22. Shailesh-Dhama,—Detecting-Fake-News-withPython |, Github, 2019
23. Aayush Ranjan, — Fake News Detection Using Machine Learning |, Department Of Computer Science & Engineering Delhi Technological University, July 2018.
24. What is a Confusion Matrix in Machine Learning by Jason Brownlee on November 18, 2016 in Code Algorithms From Scratch.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details