



# Hybrid System-Based Fake News Detection

<sup>1</sup>Amol Parde,<sup>2</sup>Dr. Rachna K. Somkunwar

<sup>1</sup>Research Scholar,<sup>2</sup>Associate Professor

<sup>1</sup>Department of Computer Engineering,

<sup>1</sup>D.Y. Patil Institute of Technology, Pimpri, Pune, India

*Abstract:* Fake news significantly impacts our social lives, especially in the political sphere. Fake news identification is an exciting subject of study but has presented some challenges due to the limited tools (e.g., datasets published in literature). External knowledge has been incorporated recently to offer helpful background information and aid in detecting news. However, current knowledge-enhanced efforts include directly incorporating all knowledge contexts via static entity embeddings, which may lead to noisy and irrelevant knowledge. Our proposed Hybrid System will obtain the best results. Our System detects fake news using Term Frequency, Term Frequency Inverted Document frequency and Kullback Leibler Divergence (TF-TDF-KLD). Our hybrid System is more efficient than the existing one, with 92.5 % accuracy.

*Index Terms* - Fake News Detection, TF-IDF-KLD, feature extraction, social lives

## I. INTRODUCTION

Over the past few years, online advertising has dramatically influenced the opinions and decisions of consumers. The most important source of information for e-commerce customers seeking to learn more about the products they wish to buy is opinions and online reviews[1]. The emergence of social networks and the advancement of Internet communication technologies have made it feasible for regular people to post comments and publish news online. While this is very convenient, it also creates an atmosphere that encourages the production and dissemination of false information. Fake news may be detrimental to people, societies, and even nations. Lately, it has come to light that there are spam opinions in addition to customer comments and product reviews. Fake articles and news are just another increasingly common form of spam. Some of the most popular places to find fake news or rumors are social networking platforms like Facebook, Twitter, Google Plus, and others. Although the issue of false news is not new, identifying fake news is thought to be challenging since individuals often trust misleading information and have little control over disseminating incorrect content [2]. Fake news has drawn more attention in the past few years, particularly in the wake of the 2016 US election. It's challenging for individuals to recognize fake news. One could

argue that someone can only manually identify bogus news if they have extensive expertise. Even with the experience, it can be challenging to establish whether the article's material is accurate. False news is easy to create and spread since the internet and social media are readily available. While the purpose and impact of fraudulent reviews are more accessible to measure and monitor, the purpose and effects of disseminating incorrect information are more challenging to measure and understand precisely. For instance, it is evident that fraudulent evaluations impact online retailers, customers, and product owners, yet it is difficult to pinpoint the precise parties affected by misinformation. This is because measuring the news's dissemination is crucial, as it has shown to be complex and resource-intensive [3] to identify such entities.

Artificial intelligence (AI) is essential for detecting fake news because it can meticulously examine linguistic and contextual nuances that human censors might overlook [4-5]. Recent advancements in artificial intelligence (AI) and natural language processing (NLP) have increased interest in the identification of fake news, leading to the development of numerous creative research methodologies in this field [6-7]. The process becomes more challenging due to the wide range of issues covered by the extensive online content. This has caused researchers to concentrate on creating algorithms for automatic fake news detection. Thus, preserving the accuracy of information on the internet depends on this technological innovation [8].

The main contributions of our work is summarized as follows:

The main contributions of our work is summarized as follows:

1. We suggested a state-of-the-art method for forecasting false news using TF-IDF-KLD.
2. We assess the suggested method with our News dataset from Kaggle and the dataset from the event registry.

## II. EXISTING WORK

Fake news makes articles hard to read since it persuades the reader to trust false information. Machine learning algorithms, a subset of artificial intelligence algorithms, can be used to help solve the challenge of fake news identification. This work used LSTM to accomplish stance detection that is utilized to identify false information. The goal was to develop a method, like agreeing or disagreeing with known sources, to identify publications from unidentified sources[9].

The research of identifying fake news is still in its early stages because, in terms of societal attention, this is, at most, a relatively recent trend. We examine a few of the works that are listed here. Three categories can typically be used to classify fake news. Fake news is the first group, comprising writers of wholly untrue pieces. False satire news falls into a different category. This type of false news primarily aims to put viewers at ease. The third group comprises poorly written news articles that include some accurate information but

are not entirely correct. To put it briefly, it's news that uses political people, for instance, to reveal a whole fabrication. Typically, this kind of news aims to further a particular agenda or set of biased opinions [10]. Rubin et al. [11] discuss three types of fake news. Every one of them presents unsatisfactory or erroneous facts. Authors also assess the various forms of misinformation as well as the benefits and drawbacks of employing different techniques for predictive modeling and text analysis.

The ease with which fraudulent and authentic articles can be distinguished was demonstrated by Horne et al.'s [12]. Their findings show that fake news titles have more nouns and verbs and fewer stop words and nouns. The following traits have been retrieved and are divided into three groups: -Complexity features quantify the readability and complexity of the text. The quantity of emotive and informal terminology in the scriptures, for example, are examples of how psychology illustrates and examines the mental processes and individual desires that underlie them. - Stylistic elements, such as the quantity of verbs and nouns, reflect the writer's style and the syntax of the text. These features were used to create an SVM classification model. The authors used Burfoot and Baldwin's satirical dataset in addition to a dataset of actual news from BuzzFeed and other websites to validate the concept [13].

Authors suggested some techniques for spotting false information from four different perspectives [14]. Based on the review, they suggested some possible study projects. Specifically, list and essential theories that are related across different fields in order to promote cross-disciplinary studies on fake news. In order to investigate false news, this survey will help professionals in the social sciences, political science, computer and information sciences, and journalism work together more effectively.

The approach to detecting false news involves looking at news content and information in social networks from several angles. Authors used techniques from data mining, machine learning, natural language processing, information retrieval, and social search. The difficulties in detecting fake news automatically, effectively, and efficiently are also explained in this work [15].

The three publicly accessible datasets WELFake, FakeNewsNet, and FakeNewsPrediction are used in this paper to present a reliable method for false news detection. In order to reduce over fitting and encourage model generalization, the further refined these algorithms using regularization and hyper parameter optimization by integrating Fast Text word embeddings with a variety of Machine Learning and Deep Learning techniques. Notably, a hybrid model that combined Long Short-Term Memory and Convolutional

Neural Networks and was enhanced with Fast Text embeddings outperformed other methods in terms of classification performance across all datasets [16].

This research presented a proactive approach to identify bogus news on the internet using an LSTM + GloVe classifier-implemented stance detection model. Rather than looking into the facts, the easiest way to handle this issue is to compare the opinions of reliable sources regarding these kinds of statements. The reliability of this hybrid classifier-based system is higher [17].

### III. BASIC CONCEPTS

False or misleading information (misinformation, including disinformation, propaganda, and hoaxes) presented as news is known as fake news or information disorder. The goal of fake news is frequently to harm someone or something's reputation or to profit from advertising. Fake news has gained popularity across a range of platforms and media outlets. Compared to traditional newspapers and periodicals, over 60% of Americans get their news from social media, according to research from the Pew Research Center[18]. Due to social media's widespread use, people can readily obtain misinformation and fake news. It has been shown that false information circulated quickly on social media during the Italian elections of 2012 and on Facebook during the US presidential campaign of 2016[19]

#### 1. Types of Fake news[20]:

- Parody or satire ("no intention to cause harm but has potential to fool")
- Faulty connection, or "when the content isn't supported by the headlines, images, or captions" false material (also known as "misleading use of information to frame an issue or an individual")
- Incorrect context ("when genuine content is shared with false contextual information")  
fake or fabricated sources are used to create impostor content ("when genuine sources are impersonated").
- "When real information or imagery is manipulated to deceive," as in the case of a "doctored" photo, this is known as manipulated content.
- Falsified material ("new content is 100% false, intended to mislead and cause harm")

## 2. Knowledge-Based Fake News Detection & Manual Fact Checking:

Fact-checking is a common method used by knowledge-based observers to identify fake news. The goal of fact checking, which was first developed in journalism, is to evaluate the veracity of news by contrasting known facts with information taken from news that needs to be confirmed, such as claims or statements. Classical fact-checking, sometimes referred to as manual fact-checking, and how it can be integrated into automated systems to identify false information (also known as automated fact-checking)[14].

In general, there are two types of manual fact-checking: (I) expert-based and (II) crowd-sourced.

- Manual fact-checking by experts. Expert-based fact-checking uses subject-matter experts as fact-checkers to confirm the information provided in the news. Expert-based fact-checking is less expensive and does not scale well when the volume of news items to be verified increases. It is typically carried out by a limited number of very reliable fact-checkers, is simple to manage, and produces very accurate results.
  - Fact-checking by experts websites. In order to better serve the public, a number of websites have recently appeared that enable fact-checking by experts[21].

## IV. PROPOSED SYSTEM

### 1. Term Frequency (TF)

Term Frequency is a method that uses the number of terms in the documents to categorize the similarities between them. Each document represents a vector of identical length that contains the word counts. Every vector is then normalized so that the sum of its elements is added to it. The likelihood of such a word appearing in the papers increases as the word count increases. For example, a term is displayed as one if it appears in a certain document and as zero otherwise. Word classes, thus, represent every text[1].

### 2. TF-IDF

The TF-IDF is one weighing metric that is often used to collect information and intuitively understand language. It acts as a barometer for the importance of a term in the document of a dataset. The rise in word appearances in the document that corresponds with term relevance is at odds with the frequency of a word in the corpus. One of the main features of IDF is its capacity to increase unusual frequency while decreasing term frequency. For example, words like "the" and "then" are often employed in the wordings, and their

frequency count is higher when we utilize TF solely. Consequently, the meaning of these notions is diminished when employing IDF scales[1].

### 3. KLD

Relative entropy, often known as Kullback-Leibler (KL) divergence, is a statistic used to compare two data distributions. Information contained in two probability distributions is contrasted in the information theory concept.

System architecture flow diagram is shown in figure 1. First we will give input of news report, then news data is preprocessed. After preprocessing the data, features are extracted using TF-IDF-KLD. As a result, the feature subsets chosen by the suggested approach are more suited for the classification problem. Selectivity efficiency is increased and random selection is avoided by creatively incorporating KLD during the selection phase. Lastly, we verify the effectiveness of particular feature subset selections. output of features are trained through the model classifier then it is classified to get the validity of fake news.

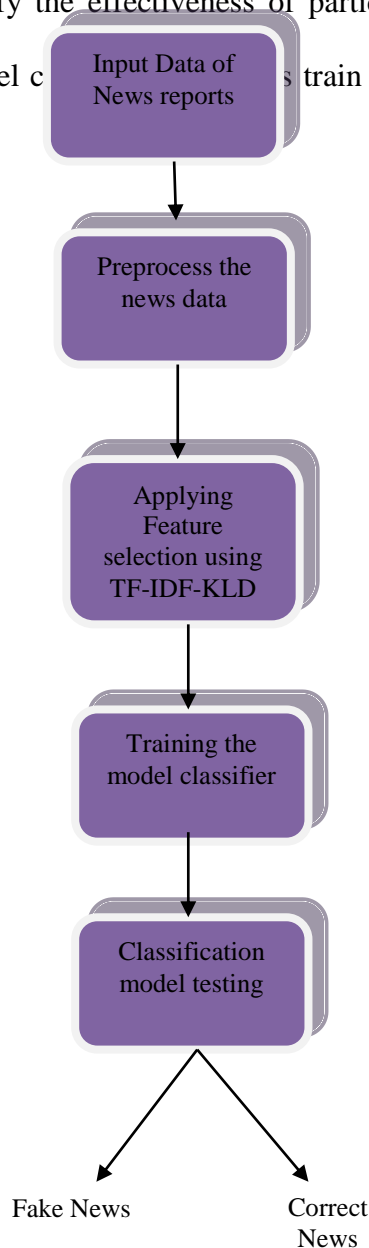
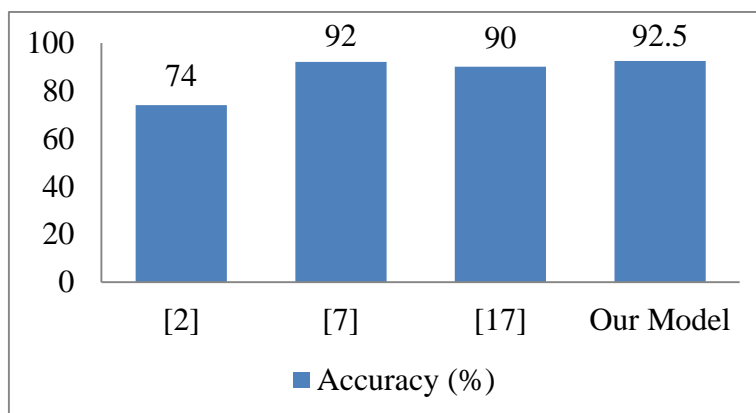


Fig 1: Proposed System Flow diagram

## V. RESULTS

The proposed System is checked for the accuracy. Various news contents are collected from various internet sources and the System is verified for the accuracy to detect the fake news content.

The graph as shown in fig. 2 shows the accuracy of the System as compared to existing systems.



**Fig. 2: Fake news detection system accuracy comparison**

The overall model accuracy is found to be 92.0% and is better, compare to existing systems.

## VI. CONCLUSION

As social media grows in popularity, an increasing number of people are getting their news from these platforms rather than from traditional news sources. In this work, we introduced a model for false news identification through the several extraction techniques. Furthermore, we examined distinct extraction techniques. We run our model on an already-existing dataset and show that it outperforms the original approach adopted by the authors of the dataset with the accuracy 92.5%.

## REFERENCES

- [1] Asha, J., et al. "Fake site detection using Machine Learning Algorithm and N-Gram Analysis." Proceedings of the 1st International Conference on Artificial Intelligence, Communication, IoT, Data Engineering and Security, IACIDS 2023, 23-25 November 2023, Lavasa, Pune, India. 2024.
- [2] Fake News Detection Using Naive Bayes Classifier by Mykhailo Granik, Volodymyr Mesyura. Available : <http://ieeexplore.ieee.org/document/8100379>.
- [3] Automatically Identifying Fake News in Popular Twitter Threads by Cody Buntain
- [4] C. Martel and D. G. Rand, "Misinformation warning labels are widely effective: A review of warning effects and their moderating features," *Current Opinion Psychol.*, vol. 54, Dec. 2023, Art. no. 101710.

- [5] S. Wang, "Factors related to user perceptions of artificial intelligence (AI)-based content moderation on social media," *Comput. Hum. Behav.*, vol. 149, Dec. 2023, Art. no. 107971.
- [6] K. Węcel, M. Sawiński, M. Stróżyna, W. Lewoniewski, E. Księżniak, P. Stolarski, and W. Abramowicz, "Artificial intelligence—Friend or foe in fake news campaigns," *Econ. Bus. Rev.*, vol. 9, no. 2, pp. 41–70, 2023.
- [7] A. Altheneyan and A. Alhadlaq, "Big data ML-based fake news detection using distributed learning," *IEEE Access*, vol. 11, pp. 29447–29463, 2023.
- [8] S. D. M. Kumar and A. M. Chacko, "A systematic survey on explainable AI applied to fake news detection," *Eng. Appl. Artif. Intell.*, vol. 122, Jun. 2023, Art. no. 106087.
- [9] Rachna Somkunwar, Anil Kumar Gupta, Faizan Shikalgar, Shubham Maral, Arpit Paithankar, Rohan Kapdi, "Fake News Detection: A Survey", © May 2021| IJIRT | Volume 7 Issue 12 | ISSN: 2349-6002
- [10] Apuke, O.D.; Omar, B. Fake news and COVID-19: Modelling the predictors of fake news sharing among social media users. *Telemat. Inform.* 2021, 56, 101475.
- [11] Rubin, V.L. Disinformation and misinformation triangle: A conceptual model for “fake news” epidemic, causal factors and interventions. *J. Doc.* 2019, 75, 1013–1034.
- [12] Horne, Benjamin, and Sibel Adali. "This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news." *Proceedings of the international AAAI conference on web and social media*. Vol. 11. No. 1. 2017.
- [13] Horne, B.; Adali, S. This Just In: Fake News Packs A Lot In Title, Uses Simpler, Repetitive Content in Text Body, More Similar To Satire Than Real News. *Proc. Int. AAAI Conf. Web Soc. Media* 2017, 11, 759–766.
- [14] Gruppi, Maurício, Benjamin D. Horne, and Sibel Adali. "NELA-GT-2020: A large multi-labelled news dataset for the study of misinformation in news articles." *arXiv preprint arXiv:2102.04567* (2021).
- [15] Zhou, Xinyi, and Reza Zafarani. "A survey of fake news: Fundamental theories, detection methods, and opportunities." *ACM Computing Surveys (CSUR)* 53.5 (2020): 1-40.
- [16] Hashmi, Ehtesham, et al. "Advancing fake news detection: Hybrid deep learning with fasttext and explainable ai." *IEEE Access* (2024).



- [17] Somkunwar, Rachna K., et al. "A Novel Hybrid Approach for Fake News Detection." 2021 5th International Conference on Information Systems and Computer Networks (ISCON). IEEE, 2021.
- [18] Zhang, Jiawei, Bowen Dong, and S. Yu Philip. "Fakedetector: Effective fake news detection with deep diffusive neural network." 2020 IEEE 36th international conference on data engineering (ICDE). IEEE, 2020.
- [19] Gottfried, Jeffrey, and Elisa Shearer. "News use across social media platforms 2016." (2016).
- [20] Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov. "Political effects of the internet and social media." *Annual review of economics* 12 (2020): 415-438.
- [21] Wardle, Claire. "Fake news. It's complicated." *First draft* 16 (2017): 1-11.