

# Fake News Detection

Prof. Kiran Naikwadi<sup>1</sup>, Prof. Kamlesh Patil<sup>2</sup>, Prof. Jyoti Dhanke<sup>3</sup>, Dr. Maharudra Kapse<sup>4</sup>

<sup>1</sup> Department of Engineering Sciences and Allied Engineering, Bharati Vidyapeeth's College of Engineering for Women-Pune, India

E-mail: [kiran.naikwadi@bharativedyapeeth.edu](mailto:kiran.naikwadi@bharativedyapeeth.edu)

<sup>2</sup> Department of Information Technology, Bharati Vidyapeeth's College of Engineering for Women – Pune, India

E-mail: [kamlesh.patil@bharativedyapeeth.edu](mailto:kamlesh.patil@bharativedyapeeth.edu)

<sup>3</sup> Department of Engineering Sciences, Bharati Vidyapeeth's College of Engineering, Lavale – Pune, India

E-mail: [jyoti.dhanke@bharativedyapeeth.edu](mailto:jyoti.dhanke@bharativedyapeeth.edu)

<sup>4</sup> Department of Humanities and basic Sciences, Bharati Vidyapeeth's Jawaharlal Nehru Institute of Technology – Pune, India

E-mail: [kmaharudra@gmail.com](mailto:kmaharudra@gmail.com)

**Abstract:** Fake news is inaccurate information that is intentionally disseminated for a specific purpose. If allowed to spread, fake news can harm the political and social spheres, so several studies are conducted to detect fake news. This research considers previous and current methods for fake news detection in textual formats while detailing how fake news exists in the first place. Detecting fake news on social media becomes a challenging problem which turns out to be very difficult to manually analyze because more and more online news is increasing on social network. Although a lot of fake news detection researches have shown some significant results and improvements by using different classification algorithms and feature extraction methods, it still has some gaps to meet the important necessities in classifying news. To address this problem, this paper investigates a fake news detection model using machine learning.

**Keyword:** Fake news detection, Machine learning, Fake news detector, Social media

## I. INTRODUCTION

Over the last few years, online social networks have seen both the number of users and the amount of information shared explosively rise. Users may use these sources of messages to connect, share, discover and disseminate information<sup>[1]</sup>. The main objective of this paper is to analyse comparative study of fake news detection using machine learning and neural network approaches. It also aims to know the ideas and process of fake news detection system and explore which methods are more efficient and effective for news classification<sup>[2]</sup>.

hardware's are cheaper and bigger datasets are available easily<sup>[3]</sup>. Due to the large dissemination capabilities of the social media the news can reach to millions of users within minutes. Even before a latest breaking news comes out through a conventional media it is out through the microblogs. The news that comes through the social media has the ability to deceive people and the majority of the population believe in these news even after knowing that they are not put forward by any trusted source. At times fake news mentions truth injected within incorrect context, and also the language used in fake news is similar to genuine news as they are created in the intention to be trusted<sup>[4]</sup>.



Fig.- Fake News. [Source:- Google]

As recently data mining algorithms have started to work much better on lots of classification problems as

## II. LITERATURE SURVEY

### [1] Smart System For News Detection Using Machine Learning.

In this paper System Uses Twitter's Social Network as our case study. To identify the document, numerous techniques were suggested, including rule-based, neural network, decision trees, and machine learning. There are also several machine learning-based tricks and classifications. The basic idea behind these strategies is to identify news types using a qualified classifier that can predict some of the predefined classes associated with a news category automatically. Nave Bayes employs the idea of chance. The parameter in Nave

Bayes was taught by training the module with the Bayesian rule of probability. The performance of a system that represents a text document as a bag of words with each word considered independent of the others is primarily degraded.

## **[2] Comparative Study of Fake News Detection Using Machine Learning and Neural Network Approaches.**

The main objective of this paper is to analyse comparative study of fake news detection using machine learning and neural network approaches. It also aims to know the ideas and process of fake news detection system and explore which methods are more efficient and effective for news classification. We have divided the system into four main phases: data acquisition, data pre-processing, feature extraction and system modeling. After that, we have evaluated and compared the performance results using different evaluation metrics and analyse which classification model achieves better result than other models.

## **[3] Fake News Identification and Classification Using DSSM and Improved Recurrent Neural Network Classifier.**

As recently data mining algorithms have started to work much better on lots of classification problems as hardware's are cheaper and bigger datasets are available easily. The major drawback of these baseline classifiers is that it does not capture semantics in text and semantic features would be very useful for fake news identification. Deep Neural Networks are trained using new methods to outperform traditional methods of identification of fake news. Our system describes a simple fake news detection method based on deep learning algorithms like improved RNN and DSSM.

## **[4] Analysis on Fake News Detection Methodologies.**

It is difficult to detect whether a generated news content is fake or not. Even the used language style can attempt to distort the truth. Due to the large dissemination capabilities of the social media the news can reach to millions of users within minutes. Even before a latest breaking news comes out through a conventional media it is out through the microblogs. The news that comes through the social media has the ability to deceive people and the majority of the population believe in these news even after knowing that they are not put forward by any trusted source .

## **[5] Fake News Detection Using Machine Learning Approaches.**

This paper proposes a methodology to create a model that will detect if an article is authentic or fake based on its words, phrases, sources and titles, by applying supervised machine

learning algorithms on an annotated (labeled) dataset, that are manually classified and guaranteed. Then, feature selection methods are applied to experiment and choose the best fit features to obtain the highest precision, according to confusion matrix results.

## **[6] Fake news detection based on news content and social contexts: a transformer-based approach.**

This paper evaluate their system by conducting experiments on real-world datasets: (i) NELA-GT-19 that consists of news articles from multiple sources and (ii) Fakeddit that is a multi-modal dataset containing text and images in posts on the social media website Reddit. While the social contexts used in this model are from Reddit, consisting of up votes, down votes, and comments on posts, the same model can be generalized to fit other social media datasets. The same method is also generalizable for any other news dataset.

## **[7] Defensive Modeling of Fake News Through Online Social Networks.**

In this paper, For detection and controlling of misinformation (rumor) in OSN, an susceptible-verified-infected-recovered (SVIR) model is proposed, which is inspired by the epidemic modeling of virus spreading in population. This model is based on different types of epidemic classes and has two layers of control mechanism to control the rumor in the social network. This model assumes that all users are susceptible that means anyone may turn a victim of misinformation or untrusted message. For protection, initially, the users are authenticated using a verified class.

## **[8] Fake News Detection using Machine Learning Algorithms.**

To develop the 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%.

## **[9] Fake News Detection Using Machine Learning Ensemble Methods.**

In this paper author described there are a number of computational techniques that can be used to mark certain articles as fake on the basis of their textual content. Majority of these techniques use fact checking websites such as “PolitiFact” and “Snopes.” There are a number of repositories maintained by researchers that contain lists of websites that

are identified as ambiguous and fake. However, the problem with these resources is that human expertise is required to identify articles/websites as fake. More importantly, the fact checking websites contain articles from particular domains such as politics and are not generalized to identify fake news articles from multiple domains such as entertainment, sports, and technology.

#### **[10] Media-Rich Fake News Detection: A Survey**

Fake news detection topic has gained a great deal of interest from researchers around the world. There are numerous social science studies have been done on the impact of fake news and how humans react to them. Fake news can be any content that is not truthful and generated to convince its readers to believe in something that is not true. For instance, when a fake news story titled “Did Palestinians Recognize Texas as Part of Mexico?” broke out on social media multiple news sites and blogs including Reddit.com picked up on this story, the tone of the title of this story convinces its readers to believe that Palestinians recognized Texas as part of Mexico.

### **III. RESEARCH GAP**

A significant research gap in the current state-of-the-art is that it focuses primarily on fake news detection rather than early fake news detection. The seminal works on early detection of fake news usually detect the fake news after at least 12 h of news propagation, which may be too late. An effective model should be able to detect fake news early, which is the motivation of this research. Another issue that we want to highlight here is the scarcity of labelled fake news data (news labelled as real or fake) in real-world scenarios. Existing state-of-the-art works generally use fully labelled data to classify fake news. However, the real-world data is likely to be largely unlabelled. Considering the practical constraints, such as unavailability of the domain experts for labelling, cost of manual labelling, and difficulty of choosing a proper label for each news item, we need to find an effective way to train a large-scale model. One alternative approach is to leverage noisy, limited, or imprecise sources to supervise labelling of large amounts of training data. The idea is that the training labels may be imprecise and partial but can be used to create a strong predictive model. This scheme of training labels is the weak supervision technique.

### **IV. EXISTING SYSTEM**

In the 21st century, the majority of the tasks are done online. Newspapers that were earlier preferred as hard-copies 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. The research in this paper focuses on detecting the fake news by reviewing it in two stages: characterization and disclosure. In the first stage, the basic concepts and principles of fake news are highlighted in social media. During the discovery stage, the current methods are reviewed for detection of fake news using different machine learning algorithms.

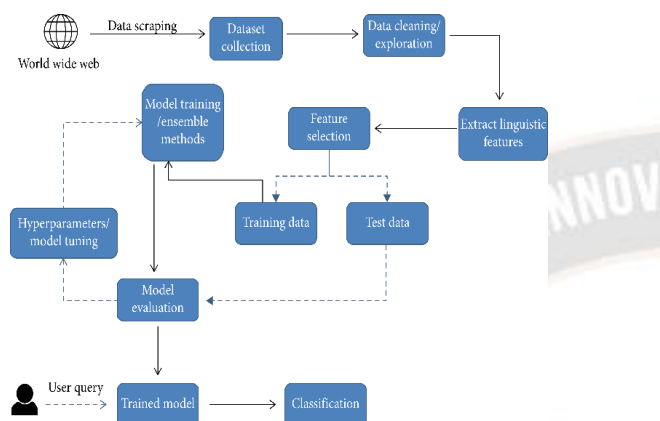
Due to the complexity of fake news detection in social media, it is evident that a feasible method must contain several aspects to accurately tackle the issue. This is why the proposed method is a combination of Naïve Bayes classifier, Support Vector Machines, and semantic analysis. The proposed method is entirely composed of Artificial Intelligence approaches, which is critical to accurately classify between the real and the fake, instead of using algorithms that are unable to mimic cognitive functions. The three-part method is a combination between Machine Learning algorithms that subdivide into supervised learning techniques, and natural language processing methods. Although each of these approaches can be solely used to classify and detect fake news, in order to increase the accuracy and be applicable to the social media domain, they have been combined into an integrated algorithm as a method for fake news detection. In the future, various models and methods can be improved, such as using bidirectional encoder representations from transformers (BERT), applying LSTM sequences to sequences, implementing bigrams and trigrams in training traditional ML and neural network models. Many current studies show that researchers can analyze statistical data on a given topic and predict the positive or negative aspects of that topic.

### **V. PROPOSED SYSTEM**

However, high time to analyze dynamic data. Promoting the Future can identify the source of false news and stop the spread of false news on social media and social media. It will also be able to track and locate sources of fraudulent information so that we can stop people trying to talk to the public about this malicious intent. Basing fake news detection only on supervised models on text have shown not to be enough in all the cases. In order to solve this problem, most of the research focus on additional information such as author information. I think the most successful approach would be



automatic fact checking model, that is, compelling the model with some kind of knowledge base, the purpose of the model would then be to extract information for the text and verify the information in the database. The problem with this approach would be that the knowledge base would need to be constantly and manually update to stay up to date.



**Fig:-** Workflow for training algorithms and classification of news articles.

[Source:-Iftikhar Ahmad,<sup>1</sup> Muhammad Yousaf, <sup>1</sup>Suhail Yousaf,<sup>1</sup>and Muhammad Ovais Ahmad].

The corpus collected from the World Wide Web is preprocessed before being used as an input for training the models. The articles' unwanted variables such as authors, date posted, URL, and category are filtered out. Articles with no body text or having less than 20 words in the article body are also removed. Once the relevant attributes are selected after the data cleaning and exploration phase, the next step involves extraction of the linguistic features. Linguistic features involved certain textual characteristics converted into a numerical form such that they can be used as an input for the training models. These features include percentage of words implying positive or negative emotions; percentage of stop words; punctuation; function words; informal language; and percentage of certain grammar used in sentences such as adjectives, preposition, and verbs. The learning algorithms are trained with different hyper parameters to achieve maximum accuracy for a given dataset, with an optimal balance between variance and bias. Each model is trained multiple times with a set of different parameters using a grid search to optimize the model for the best outcome.

Basing fake news detection only on supervised models on text have shown not to be enough in all the cases. In order to solve this problem, most of the research focus on additional information such as author information. I think the most successful approach would be automatic fact checking

model, that is, compelling the model with some kind of knowledge base, the purpose of the model would then be to extract information for the text and verify the information in the database. The problem with this approach would be that the knowledge base would need to be constantly and manually update to stay up to date.

## VI. CONCLUSION

This study proposed a fake news stance detection model, based on the headline and the body of the news irrespective of the previous studies which only considered the individual sentences or phrases. The easy access of social media to everyone has obvious advantages but also it has some disadvantages, such as the rapid quick spread of fake news. It is a very tedious job to check every news item manually, so, in order to overcome this problem, researchers are developing algorithms to detect fake news automatically. Fake news identification is a contextual problem in which the meaning of the same words may be different depending on the context. Various feature extractors have been built to efficiently solve this problem. The proposed system describes a personalized based news recommendation from social media. The online news population dataset also available on machine learning UCI repository. During the initial research process, the system's output is assessed using this dataset, and accuracy is calculated. However, there is still room for development by introducing a hybrid model that uses a range of feature selection approaches. It is pertinent to say that dimensionality reduction approaches can reduce the number of features while preserving the high performance of classifiers.

In the future, various models and methods can be improved, such as using bidirectional encoder representations from transformers (BERT), applying LSTM sequences to sequences, implementing bigrams and trigrams in training traditional ML and neural network models. In future work, we would like to expand the research outlined in this thesis .Incorporating more user-independent features will be one important addition. These include demographics, sex, age, place, to learn more about the readers, and their related patterns of reading. We plan on investigating whether a social media post's dissemination pattern will help us better identify fake news. We would also like to see if the user features could help us in identifying users who are more likely to believe and spread fake news.

## REFERENCE

- [1] Smart System For News Detection Using Machine Learning. (2021) Puja Sunil Erande<sup>1</sup>, Monika Dhananjay Rokade<sup>2</sup>, ijariie.

- [2] Comparative Study of Fake News Detection Using Machine Learning and Neural Network Approaches. (2021). May Me Me Hlaing 1□ and Nang Saing Moon Kham 2.
- [3] Fake News Identification and Classification Using DSSM and Improved Recurrent Neural Network Classifier (2019). Shrutika S. Jadhav & Sudeep D. Thepade. Applied Artificial Intelligence.
- [4] Analysis on Fake News Detection Methodologies (2020). Vaishnavi R, Anitha Kumari S. Blue Eyes Intelligence Engineering & Sciences Publication.
- [5] Fake News Detection Using Machine Learning Approaches (2021). Z Khanam1, B N Alwasel1, H Sirafi1 and M Rashid2. IOP Publishing.
- [6] Fake news detection based on news content and social contexts: a transformer-based approach (2021). Shaina RazalChen Ding1. Springer.
- [7] Defensive Modeling of Fake News Through Online Social Networks (2020). Gulshan Shrivastava , Member, IEEE, Prabhat Kumar, Senior Member, IEEE, Rudra Pratap Ojha , Pramod Kumar Srivastava , Senthilkumar Mohan , and Gautam Srivastava. IEEE.
- [8] Fake News Detection using Machine Learning Algorithms (2020). Uma Sharma, Sidarth Saran, Shankar M. Patil. Ijert.
- [9] Fake News Detection Using Machine Learning Ensemble Methods. (2020). Iftikhar Ahmad,<sup>1</sup> Muhammad Yousaf,<sup>1</sup>Suhail Yousaf,<sup>1</sup> and Muhammad Ovais Ahmad. Hindawi.
- [10] Media-Rich Fake News Detection: A Survey (2019). Shivam B. Parikh and Pradeep K. Atrey. University at Albany