

**A Project/Dissertation Review-ETE Report**

**on**

**Fake News Detection**

**Submitted in partial fulfillment of the  
requirement for the award of the degree of**

**BACHELOR OF TECHNOLOGY**



**Under The Supervision of**

**Name of Supervisor/Guide : Mrs. Urvashi Sugandh**

**Submitted by:**

**Sourish Keshav 19SCSE1010766**

**Om Rastogi 19SCSE1010052**

**SCHOOL OF COMPUTING SCIENCE AND ENGINEERING  
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**GALGOTIAS UNIVERSITY, GREATER NOIDA**

**INDIA**

**12,2021**

## **Table of Contents**

**Title :**

**1.Abstract**

**2. Introduction**

**3.Literature Survey**

**4. Architecture**

**5.Conclusion**

## **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. The advent of the World Wide Web and the rapid adoption of social media platforms (such as Facebook and Twitter) paved the way for information dissemination that has never been witnessed in the human history before. With the current usage of social media platforms, consumers are creating and sharing more information than ever before, some of which are misleading with no relevance to reality. Automated classification of a text article as misinformation or disinformation is a challenging task. Even an expert in a particular domain has to explore multiple aspects before giving a verdict on the truthfulness of an article. In this work, we propose to use machine learning ensemble approach for automated classification of news articles. Our study explores different textual properties that can be used to distinguish fake contents from real. By using those properties, we train a combination of different machine learning algorithms using various ensemble methods and evaluate their performance on 4 real world datasets. Experimental evaluation confirms the superior performance of our proposed ensemble learner approach in comparison to individual learners.

## **Introduction**

Fake news is defined as “particular news articles that originate either on mainstream media (online or offline) or social media and have no factual basis, but are presented as facts and not satire”. The importance of combatting fake news is starkly illustrated during the current COVID-19 pandemic. Social networks are stepping up in using digital fake news detection tools and educating the public towards spotting fake news. At the time of writing, Facebook uses machine learning algorithms to identify false or sensational claims used in advertising for alternative cures, they place potential fake news articles lower in the news feed, and they provide users with tips on how to identify fake news themselves . Twitter ensures that searches on the virus result in credible articles and Instagram redirects anyone searching for information on the virus to a special message with credible information.

These measures are possible because different approaches exist that assist the detection of fake news. For example, platforms based on machine learning use fake news from the biggest media outlets, to refine algorithms for identifying fake news . Some approaches detect fake news by using metadata such as a comparison of release time of the article and timelines of spreading the article as well where the story spread .

The purpose of this research paper is to, through a systematic literature review, categorize current approaches to contest the wide-ranging endemic of fake news.

## Literature Survey

Fake news is not a new concept. Before the era of digital technology, it was spread through mainly yellow journalism with focus on sensational news such as crime, gossip, disasters and satirical news . The prevalence of fake news relates to the availability of mass media digital tools . Since anyone can publish articles via digital media platforms, online news articles include well researched pieces but also opinion-based arguments or simply false information . There is no custodian of credibility standards for information on these platforms making the spread of fake news possible. To make things worse, it is by no means straightforward telling the difference between real news and semi-true or false news .

The nature of social media makes it easy to spread fake news, as a user potentially sends fake news articles to friends, who then send it again to their friends and so on. Comments on fake news sometimes fuel its 'credibility' which can lead to rapid sharing resulting in further fake news .

Social bots are also responsible for the spreading of fake news. Bots are sometimes used to target super-users by adding replies and mentions to posts. Humans are manipulated through these actions to share the fake news articles .

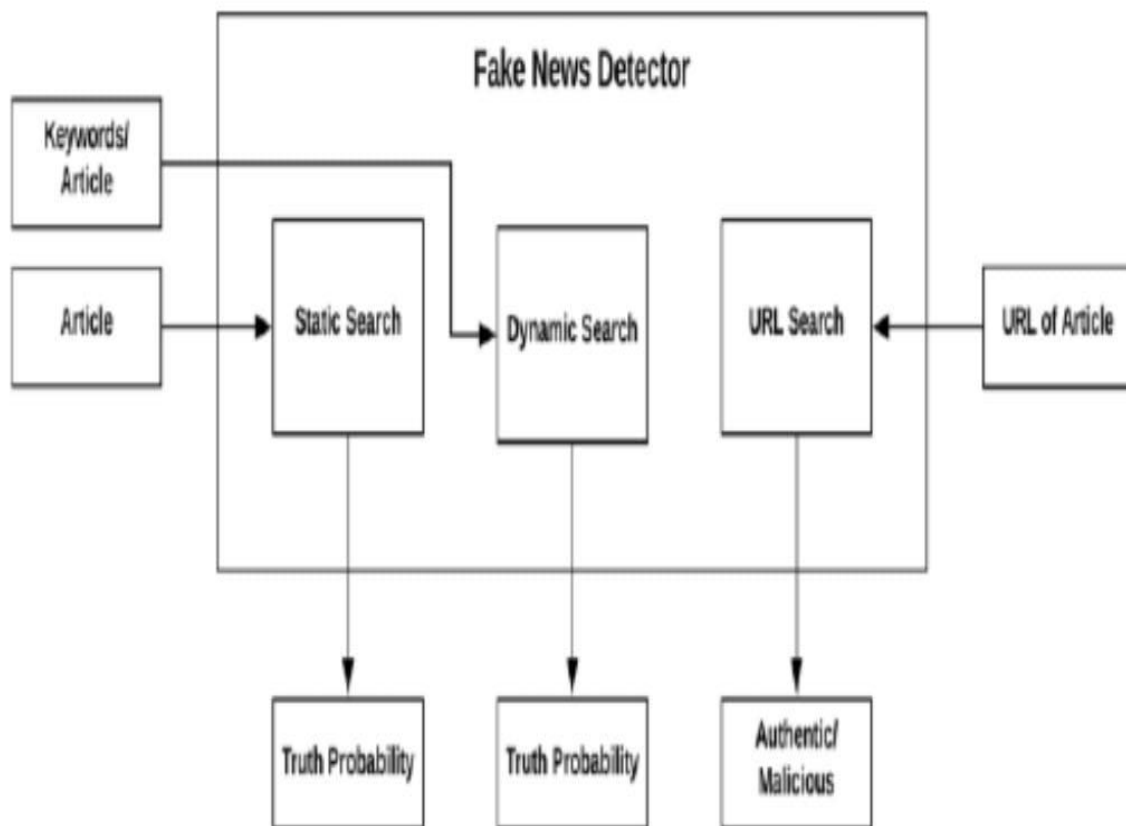
Clickbait is another tool encouraging the spread of fake news. Clickbait is an advertising tool used to get the attention of users. Sensational headlines or news are often used as clickbait that navigate the user to advertisements. More clicks on the advert means more money .

Fortunately, tools have been developed for detecting fake news. For example, a tool has been developed to identify fake news that spreads through social media through examining lexical choices that appear in headlines and other intense language structures . Another tool, developed to identify fake news on Twitter, has a component called the Twitter Crawler which collects and stores tweets in a database . When a Twitter user wants to check the accuracy of the news found they can copy a link into this application after which the link will

be processed for fake news detection. This process is built on an algorithm called the NER (Named Entity Recognition).

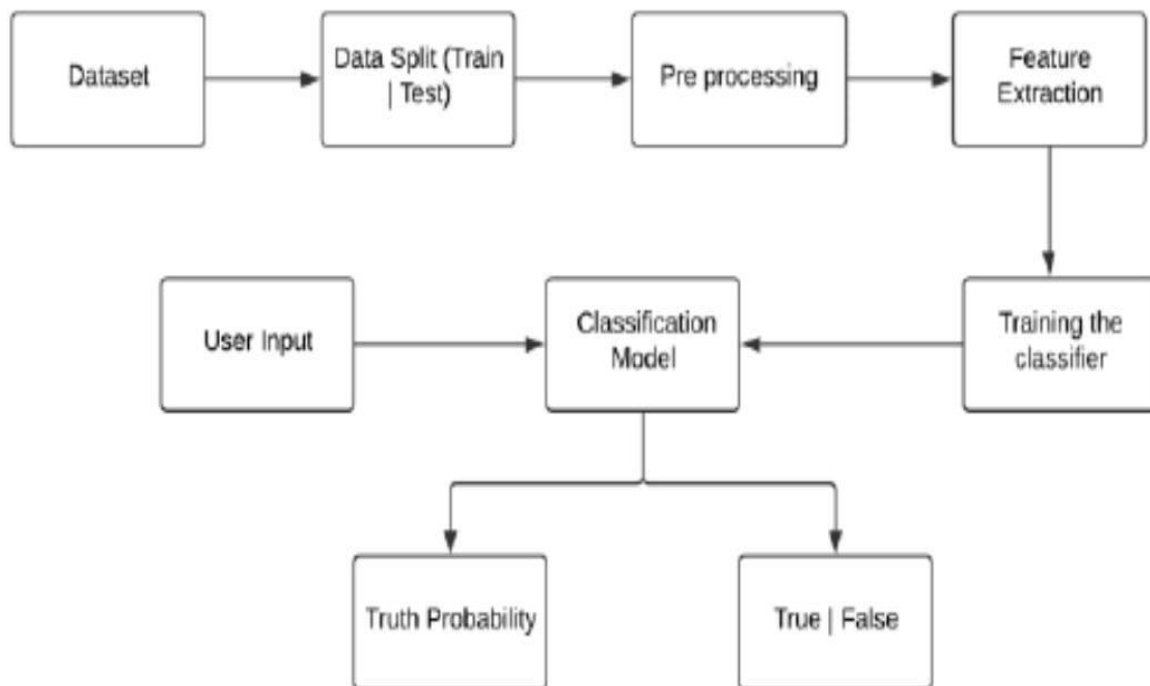
There are many available approaches to help the public to identify fake news and this paper aims to enhance understanding of these by categorizing these approaches as found in existing literature.

## System Design



## System Architecture

---



## Algorithms :

### 1.) K-Nearest Neighbors (KNN):-

KNN is an unsupervised machine learning model where a dependent variable is not required to predict the outcome on a specific data. We provide enough training data to the model and let it decide to which particular neighborhood a data point belongs. KNN model estimates the distance of a new data point to its nearest neighbors, and the value of K estimates the majority of its neighbors' votes; if the value of K is 1, then the new data point is assigned to a class which has the nearest distance. The mathematical formulae to estimate the distance between two points can be calculated as follows

$$\text{Euclidean distance} = \sqrt{\sum_{i=1}^k (x_i - y_i)^2},$$

$$\text{Manhattan distance} = \sum_{i=1}^k |x_i - y_i|,$$

Minkowski distance

$$= \left( \sum_{i=1}^k |x_i - y_i|^q \right)^{1/q}.$$

### 2.) Support Vector Machine:-

Support vector machine (SVM) is another model for binary classification problem and is available in various kernels functions . The objective of an SVM model is to estimate a hyperplane (or decision boundary) on the basis of feature set to classify data points . The dimension of hyperplane varies according to the number of features. As there could be multiple possibilities for a hyperplane to exist in an N-dimensional space, the task is to identify the plane that separates the data points of two classes



with maximum margin. A mathematical representation of the cost function for the SVM model is defined as given in and shown in

$$J(\theta) = \frac{1}{2} \sum_{j=1}^n \theta_j^2,$$

such that

$$\theta^T x^{(i)} \geq 1, \quad y^{(i)} = 1,$$

$$\theta^T x^{(i)} \leq -1, \quad y^{(i)} = 0.$$

### 3.) Performance Metrics:-

To evaluate the performance of algorithms, we used different metrics. Most of them are based on the confusion matrix. Confusion matrix is a tabular representation of a classification model performance on the test set, which consists of four parameters: true positive, false positive, true negative, and false negative

## **Conclusion**

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. 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 intend 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.

## REFERENCE

1. A. Douglas, "News consumption and the new electronic media," *The International Journal of Press/Politics*, vol. 11, no. 1, pp. 29–52, 2006. View at: [Publisher Site](#) | [Google Scholar](#)
2. J. Wong, "Almost all the traffic to fake news sites is from facebook, new data show," 2016. View at: [Google Scholar](#)
3. D. M. J. Lazer, M. A. Baum, Y. Benkler et al., "The science of fake news," *Science*, vol. 359, no. 6380, pp. 1094–1096, 2018. View at: [Publisher Site](#) | [Google Scholar](#)
4. S. A. García, G. G. García, M. S. Prieto, A. J. M. Guerrero, and C. R. Jiménez, "The impact of term fake news on the scientific community scientific performance and mapping in web of science," *Social Sciences*, vol. 9, no. 5, 2020. View at: [Google Scholar](#)
5. A. D. Holan, 2016 Lie of the Year: Fake News, Politifact, Washington, DC, USA, 2016.
6. S. Kogan, T. J. Moskowitz, and M. Niessner, "Fake News: Evidence from Financial Markets," 2019, <https://ssrn.com/abstract=3237763>. View at: [Google Scholar](#)
7. A. Robb, "Anatomy of a fake news scandal," *Rolling Stone*, vol. 1301, pp. 28–33, 2017. View at: [Google Scholar](#)
8. J. Soll, "The long and brutal history of fake news," *Politico Magazine*, vol. 18, no. 12, 2016. View at: [Google Scholar](#)
9. J. Hua and R. Shaw, "Corona virus (covid19) "infodemic" and emerging issues through a data lens: the case of China," *International Journal of Environmental Research and Public Health*, vol. 17, no. 7, p. 2309, 2020. View at: [Publisher Site](#) | [Google Scholar](#)
10. N. K. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: methods for finding fake news," *Proceedings of the Association for Information Science and Technology*, vol. 52, no. 1, pp. 1–4, 2015. View at: [Publisher Site](#) | [Google Scholar](#).