

# Fake News Detection: A Review

Affreen Parveen<sup>1</sup> Disha Gite<sup>2</sup> Disha Shetty<sup>3</sup>

<sup>1,2,3</sup>Department of Information Technology

<sup>1,2,3</sup>Vasantdada Patil Prathishtans College of Engineering, Mumbai, Maharashtra, India

**Abstract**— Social media is a boon for the people who want to share information, news, opinions, etc. with the world. It is considered one of the fastest mediums to transfer information. On one click, the entire world can access the information. But it also has some cons. Information can be manipulated and uploaded on social media. It can be used to mislead people which can be dangerous for the human race. This manipulated information can spread quickly and can make it difficult to control the situation. Machine learning algorithms may help to identify truth and detect fake information. The results may not be 100% accurate but it helps to maximize the proximity.

**Keywords:** Fake News, Machine Learning, Deep Learning

## I. INTRODUCTION

The rate of Internet use is growing rapidly in India. Around 627 million people use the Internet in India as conducted by the survey in 2019. From that, 36 percent of people use social media which is 230 million people. Nowadays, news spreads faster through social media than TV or newspapers. Hence, to get fresh news, people rely on social media.

Political leaders and public figures choose social media as a way to convey a message or information to the people. But many times wrong information is also circulated. This is done to defame someone, induce fear among people, to grab attention, etc. (mostly for political purposes). The main aspect of news on social media is that anyone can register as a new publisher with free of cost (eg. Anyone can create an account on social media claiming to be a media organizer). Not only traditional news but also corporations are increasing on social media.

Fake news is generally fabricated news or news developed by taking real-life references. Fake news spreads like wildfire on social media creating chaos.

Fake news can be detected using machine learning techniques. It includes data preprocessing, feature extraction, and classification using machine learning algorithms. Naive Bayes algorithm is used to find the accuracy of the news. The Decision Tree algorithm is used to predict the values. Confusion Matrix displays the data along with the prediction to compare the results.

## II. LITERATURE REVIEW

Fake news detection can be done by using different techniques like using Machine Learning algorithms, Deep Learning methods, and other such techniques.

For the detection of fake news, the dataset must be prepared by collecting the data accurately and then preprocessing the dataset and applying the required data mining algorithms.

Akshay Jain and Amey Kasbe studied the detection of fake news using the Naive Bayes classifier [1]. They used web scraping techniques to extract a large amount of data. Naive Bayes classifier was applied over the dataset obtained by first considering only the titles of the news articles and

then by considering the entire article and the results were analyzed. On analysis, the results obtained by using the entire news article were better than that by considering only the news titles.

A.Lakshmanarao, Y.Swathi, T. Srinivasa Ravi Kiran performed an analysis on a dataset by applying different machine learning algorithms namely Support Vector Machine, Decision Tree classification, K-Nearest Neighbor, Random Forest classification algorithms [5]. The results of all the algorithms were compared and tabulated. On comparison, the Random forest gave better accuracy than the other algorithms.

Authors of [2] have studied the types of fake news. They have described the data which is circulated as fake news. It can be visual-based, user-based, knowledge-based, style-based, or even stance based. Before detecting the fake news we have to understand how the fake news is created. It may be a written article, an image, video, audio, etc. It should also be studied that among these, which one is used the most for the fake news and made viral.

[3] The authors studied that the fake news articles and spam messages have similar properties of grammatical mistakes, they use similar and limited amounts of words and are often emotionally manipulative. The dataset was processed and filtered and shuffled, which was divided into three subsets namely, training dataset, validation dataset, and test dataset. They used Naive Bayes Classifier for detection using the prepared dataset. The results obtained were not significant. They concluded some ways to improve accuracy by increasing the size of the dataset and using news articles having a longer length.

[4] The authors took a recent dataset from BuzzFeed news articles and classified stories as "non-factual", "mostly false", "mixture of true and false", "true". The algorithms used were K-Nearest Neighbor, Naive Bayes, Random Forest, Support Vector Machine, and XGBoost.

Deep Learning algorithms can be used for fake news detection, but using Deep Learning algorithms is more complicated than using Machine Learning algorithms.

The authors [6] have preprocessed the data and two algorithms, Naive Bayes and SVM are applied for detection. They then selected the most accurate result from the two algorithms.

[7] The authors have proposed a framework to automatically detect fake news articles. They have suggested building a machine learning model which is based on Logistic regression, SVM, Random forest. This model will focus only on text information in the news articles.

[8] The authors have summarised the types of data in news classifying them into text, multimedia, hyperlinks and audio. The paper also classifies fake news into various types such as visual based, user based, post based, network based, knowledge based, style based and stance based. Methods of Fake news detection can be Linguistic features

based methods, Deception modeling based methods, Clustering based methods, Predictive modeling based methods, Content cues based methods and Non-text cues based methods.

### III. DETECTION METHODS

#### A. Naive Bayes Algorithm

Naive Bayes is mainly used for prediction and forecasting data. Naive Bayes algorithm uses a large dataset and calculates the probability of the required data by comparing it with the historical dataset having similar features. In fake news detection, the Naive Bayes algorithm can be used to detect the accuracy of the news. In this method, a large dataset is considered which contains similar fake news articles or a dataset consisting of words commonly found in fake news articles. This algorithm provides the probability of the required news being fake or real or genuine news.

Data preprocessing methods should be applied to the dataset. The dataset is filtered and checked for missing values. The next step is to apply feature extraction methods on the prepared dataset. The dataset is then classified into a test dataset and train dataset. The Naive Bayes algorithm is then applied and the probability of the fake news article is calculated.

#### B. Decision Tree Algorithm

Decision Tree algorithm is a type of supervised learning method. It is used to predict the value of a target variable by learning decision rules deduced from historical data. The decision tree algorithm uses a tree structure where an attribute is represented by an internal node and the class label is represented by the leaf of the tree. In this method, prediction of a data is done by comparing the root of the decision tree with the features of the data to be predicted and after comparison the corresponding branch is considered. The dataset is thus split based on the required features. In fake news detection, the decision tree algorithm can be used to predict fake news articles based on its features.

#### C. Confusion Matrix

The confusion matrix is mainly used for calculating the accuracy of the data. This method is used for visualization of the performance of an algorithm. It is used for analyzing the performance of any classification model that is applied to a dataset. This method is used to detect whether the algorithms applied to the fake news dataset give the desired accuracy or not. Based on the results the news article can be tagged as fake news or real news with better accuracy.

### IV. CONCLUSION

This paper concludes that Machine Learning algorithms combining with Data Mining algorithms can be used to detect fake news with maximum accuracy. A simple data mining algorithm is used to calculate the probability of the given data. Naive Bayes algorithm is an easy way to calculate it. It is performed on large datasets. This data is classified into two categories and then the algorithm is applied. This algorithm is popularly used as it can be performed on large datasets.

The Decision Tree algorithm uses a tree structure to predict the values. Confusion Matrix calculates the accuracy of the information obtained after performing any algorithm on it. This algorithm is applied at the end to determine accuracy. This helps to get the best results.

As time passes technology changes, new techniques arrive which eases the way. Therefore there is always scope for improvement. And we should keep on working with this. Machine learning algorithms have helped us to overcome such complex issues (fake news).

### REFERENCES

- [1] Fake News Detection by Akshay Jain, Amey Kasbe, Maulana Azad National Institute of Technology, Bhopal, India, IEEE 2018 International Students' Conference on Electrical, Electronics, and Computer Sciences.
- [2] Fake News Detection Using Machine Learning approaches: A systematic Review, Syed Ishfaq Manzoor, Dr Jimmy Singla, Nikita, Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019).
- [3] Fake News Detection Using Naive Bayes Classifier, Mykhailo Granik, Volodymyr Mesyura, 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering.
- [4] Supervised Learning for Fake News Detection, Julio C. S. Reis, Andre Correia, Fabricio Murai, Adriano Veloso, and Fabricio Benevenuto, Universidade Federal de Minas Gerais, 2019 IEEE.
- [5] An Efficient Fake News Detection System Using Machine Learning, A.Lakshmanarao, Y.Swathi, T. Srinivasa Ravi Kiran, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8, Issue-10, August- 2019.
- [6] A smart system for fake news detection using machine learning, Anjali Jain, Avinash Shakya, Harsh Katter, Amit Kumar Gupta, 2019 IEEE 2nd International conference on ICICT.
- [7] Detecting Fake News Articles, Jun Lin, Glenna Tremblay-Taylor, Guanyi Mou, Di You, Kyumin Lee Computer Science and Engineering Louisiana State University Baton Rouge, LA, USA.
- [8] Media-Rich Fake News Detection: A Survey, Shivam B. Parikh and Pradeep K. Atrey, 2018 IEEE Conference on Multimedia Information Processing and Retrieval.