

e-ISSN:2582 - 7219



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 4, Issue 7, July 2021



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 5.928



9710 583 466



9710 583 466



ijmrset@gmail.com



www.ijmrset.com



Fake News Detection: An Overview

Miss. Puja Sunil Erande¹, Prof. Monika Rokade²

PG Student, Sharadchandra Pawar College of Engineering, Junnar Pune, India¹

Assistant Professor, Sharadchandra Pawar College of Engineering, Junnar Pune, India²

ABSTRACT: Information is readily available, thanks to the increased use of mobile technology and social media. In the delivery of news, mobile apps and social media platforms have surpassed traditional media. With the increased use of online media platforms such as Facebook, Twitter, and others, news disseminated swiftly across a huge number of clients with limited ability to spend time. The two methodologies used to investigate the truthiness of the information are machine learning and knowledge-based approach and approach. Public and private opinions on a wide range of topics are regularly discussed and disseminated via various online platforms. The majority of approaches are used, such as controlled AI. The dissemination of fake news has a wide range of consequences, from creating one-sided attitudes to influencing political race results in favour of specific candidates. Additionally, spammers use engaging news elements to generate revenue through click bait notices. In this research, we will use Artificial Intelligence, Natural Language Processing, and Machine Learning and Deep Learning to perform a parallel grouping of various internet news pieces. The project's outcome determines the identification of false news for social networks using machine learning and deep learning, as well as the legitimacy of the news website that is being published.

KEYWORDS: Fake News, News articles, Internet, Social media, Classification, Artificial Intelligence, Machine Learning, Deep Learning.

I. INTRODUCTION

Because of the increased popularity of social media and mobile technologies, information is now at one's fingertips. In the sphere of information and news, mobile apps and social media sites such as Facebook and Twitter have supplanted traditional media. People choose social media because of the convenience and speed that digital media provides. Not only has it provided customers with faster access, but it has also provided benefit-seeking parties with a reliable platform to attract a larger audience.

With so much information or news, the question of whether the information or news is true or false arose. Fake news is frequently disseminated with the goal of misleading or inducing a desire for political or financial gain. Consider the following scenario: During India's recent elections, there was much debate about the veracity of various news reports favouring specific candidates and the political thought processes underlying them. Exposing fake news is critical in preventing its negative influence on people and society, given this increased interest.

Machine learning and Deep learning algorithms such as Logistic Regression, Random Forests, Decision Trees, Support Vector Machines, Stochastic Gradient Descent, and others are employed by false news detection systems. A simple method of detecting fake news based on one of the AI algorithms known as the Naive Bayes classifier was used to investigate how well this method works for the specific problem with a manually labelled (fake or real) dataset and to support the idea of using machine learning to detect fake news.

II. LITERATURE REVIEW

According to [1] Evaluating Machine Learning algorithms for Fake News Detection. In this article, the author introduced the concept of the importance of NLP in stumbling across incorrect information. They have used time frequency-inverse document frequency (TF-IDF) of bigrams and probabilistic context-free grammar detection. Shloka Gilda introduced the concept of the importance of NLP in stumbling over incorrect information. They used Bi-Gram Count Vectorizer and Probabilistic Context-Free Grammar (PCFG) to detect deceptions. They examined the data set in more than one class of algorithms to find out a better model. The count vectorizer of bi-grams fed directly into a stochastic gradient descent model which identifies noncredible resources with an accuracy of 71.2%.



According to [2] Fake News Detection on Social Media: A Data Mining Perspective. In this paper to detect fake news on social media, a data mining perspective is presented that includes the characterization of fake news in psychology and social theories. This article looks at two main factors responsible for the widespread acceptance of fake messages by the user which is naive realism and confirmatory bias. It proposes a general two-phase data mining framework that includes 1) feature extraction and 2) modeling, analyzing data sets, and confusion matrix for detecting fake news.

According to [3] Media Rich Fake News Detection: A Survey. Social networking sites read news mainly in three ways: The (multilingual) text is analyzed with the help of computational linguistics, which semantically and systematically focuses on the creation of the text. Since most publications are in the form of text, a lot of work has been done on analyzing them. Multimedia: Several forms of media are integrated into a single post. This can include audio, video, images, and graphics. This is very attractive and attracts the viewer's attention without worrying about the text. Hyperlinks allow the author of the post to refer to various sources and thus gain the trust of viewers. In practice, references are made to other social media websites, and screenshots are inserted.

According to [4] Fake News Detection using Naive Bayes classifier. This article describes a simple method of fake news detection based on one of the artificial intelligence algorithms called the Naive Bayes classifier. The goal of the research is to examine how this particular method works for the particular problem with a manually labeled (fake or real) dataset and to support the idea of using machine learning to detect fake news. The difference between this article and articles on similar topics is that this article is extensively based on a Naive Bayes classifier which is used for the classification of fake news and real news; In addition, the developed system was tested on a relatively new data set, which provided the opportunity to evaluate its performance against the most recent data.

According to [5] Using a naive Bayes classifier, a basic strategy for detecting fake news has been developed. This method was turned into a software system and put to the test on a collection of Facebook news posts. On the test set, we achieved a classification accuracy of around 74%, which is a good result considering the model's relative simplicity. The findings can be improved in a number of ways, which are detailed in the article. According to the findings, the challenge of detecting false news can be solved using artificial intelligence methods.

According to [6] We propose ways to detect such people (Spammers) on the Twitter social network as part of our research (a popular OSN). Followers/Followees, URLs, Spam Words, Replies, and HashTags are some of the capabilities we use at the tweet and user level. We used three learning algorithms in our research: Naive Bayes, Clustering, and Decision Trees. Furthermore, a novel integrated technique is provided to improve Spammer identification by "combining" the advantages of the three learning algorithms stated above. Total Accuracy, Spammers Detection Accuracy, and Non-Spammers Detection Accuracy are used to measure spam detection improvement.

According to [7] sentiment analysis is a vastly used term to classify user's opinion using Natural Language Processing (NLP) and Machine Learning (ML) Approaches. Various researchers have used different methods for aspect based classification and polarity based classification [1], [3], [8], etc. Product review based sentiment analysis is similar to the proposed sentiment analysis approach. Figure 1 summarizes the basic model of the sentiment classification task.

Mahdieh Labaniet. al. [8] proposed a multivariate filter method for feature selection which is used for various text classification approach. This method focuses on the reduction of redundant features using minimal-redundancy and maximal-relevancy concepts. The proposed method takes into account document frequencies for each term, while estimating their usefulness. It not only selects the features with maximum relevancy but also the redundancy between them is taken into account using a correlation metric. Results obtained using this approach are better than state-of-the-art filter methods.

Asriyanti Indah Pratiwi and Adiwijaya [9] proposed feature selection and classification based on information gain for document sentiment analysis. Information Gain Classifier (IGC) is used to extract the various features from movie review dataset. Authors proposed IG-DF-FS based hybrid method called a combination of Information Gain + Document Frequency Feature Selection etc.

Haoyue Liu et. Al. [10] proposed a system of feature selection for imbalanced data. If the dataset is imbalanced, it has a problem of bias-to-majority. This issue is solved using Weighted Gini Index (WGI) approach. The WGI approach calculates an impurity reduction score for each feature and features with a high score are considered as important.

Monika Rokade and Yogesh Patil [11] proposed a system deep learning classification using anomaly detection from network dataset. The Recurrent Neural Network (RNN) has classification algorithm has used for detection and



classifying the abnormal activities. The major benefit of system it can works on structured as well as unstructured imbalance dataset.

The MLIDS A Machine Learning Approach for Intrusion Detection for Real Time Network Dataset has proposed by Monika Rokade and Dr.YogeshPatil in [12]. The numerous soft computing and machine learning classification algorithms have been used for detection the malicious activity from network dataset. The system depicts around 95% accuracy ok KDDCUP and NSLKDD dataset.

Monika D. Rokade and Yogesh Kumar Sharma [13] proposed a system to identification of Malicious Activity for Network Packet using Deep Learning. 6 standard dataset has sued for detection of malicious attacks with minimum three machine learning algorithms.

Sunil S. Khatal and Yogeshkumar Sharma [14] proposed a system Health Care Patient Monitoring using IoT and Machine Learning for detection of heart and chronic diseases of human body. The IoT environment has used for collection of real data while machine learning technique has used for classification those data, as it normal or abnormal.

Data Hiding In Audio-Video Using Anti Forensics Technique For Authentication has proposed by Sunil S.Khatal and Yogeshkumar Sharma [15]. This is a secure data hiding approach for hide the text data into video as well as image. Once sender hide data into specific objects while receivers does same operation for authentication. The major benefit of this system can eliminate zero day attacks in untrusted environments.

Sunil S.Khatal and Yogesh Kumar Sharma [16] proposed a system to analyzing the role of Heart Disease Prediction System using IoT and Machine Learning. This is the analytical based system to detection and prediction of heart disease from IoT dataset. This system can able to detect the disease and predict accordingly.

III. PROPOSED METHODOLOGY

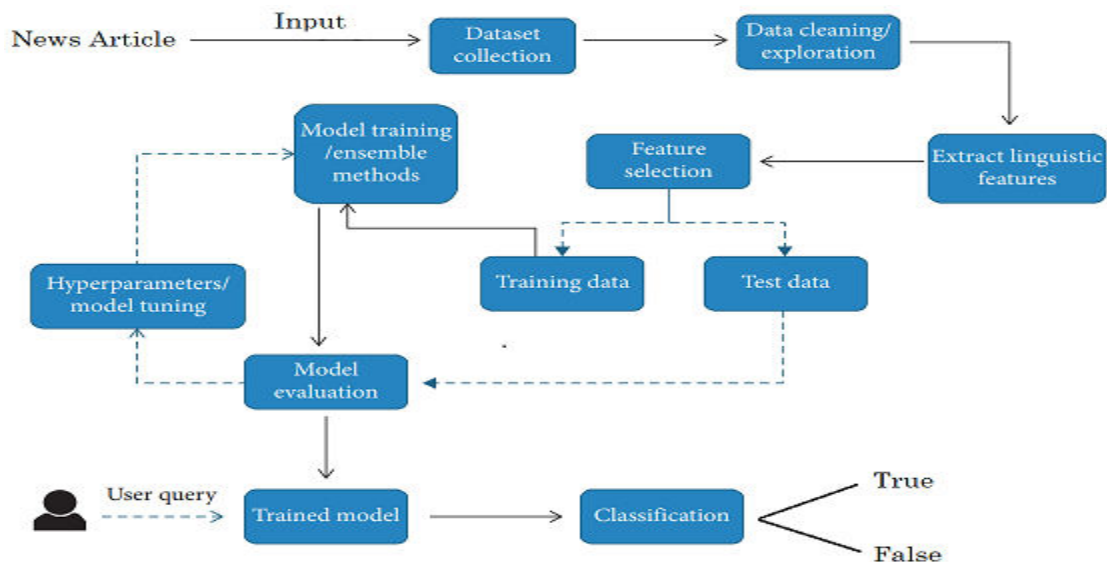


Figure -1: Flow of the module in Fake News Detection System

The Bogus News Detection System will aid in the management of fake news on social media. This manner, we can assist people in making more informed decisions without forcing them to consider what others are attempting to persuade them to believe. A Fake News Detection system relieves the load of manually verifying the news's legitimacy and saves time.



IV. RESULTS AND DISCUSSION

Multiple instances of both unsupervised learning and supervised learning algorithms have been utilised to classify text in false news detection systems. The majority of the literature reviews concentrate on specialised subjects, the most important of which is politics. As a result, the method is best suited to a specific type of article's domain and does not produce ideal results when used to articles from various fields. It's difficult to develop a general algorithm that performs well across all news spaces because articles from different topics have different literary structures. In comparison to other algorithms, we found that the Random Forests algorithm using a basic term frequency-inverse document frequency vector produces the best results. Our research looked at a variety of text qualities that can be used to identify between fake and authentic information, and we used these properties to train a variety of machine learning and deep learning algorithms.

V. CONCLUSION

Manually classifying news stories necessitates in-depth knowledge and expertise in spotting textual irregularities. We explored the topic of categorising false news stories using machine learning models and ensemble techniques because manually verifying a single article takes a long time.

It's critical that we have a system in place to detect fake news, or at the very least be aware that not everything we read on social media is accurate. That is why we must always be critical thinkers. This manner, we can assist individuals in making better selections, and they won't be persuaded to believe what others are attempting to persuade them to believe.

REFERENCES

- [1]Mugdha, Shafayat Bin Shabbir, SayedaMuntahaFerdous, and Ahmed Fahmin. "Evaluating Machine Learning Algorithms For Bengali Fake News Detection." 2020 23rd International Conference on Computer and Information Technology (ICCIT).IEEE, 2020.
- [2] Shu, Kai, et al. "Fake news detection on social media: A data mining perspective." ACM SIGKDD explorations newsletter 19.1 (2017): 22-36.
- [3] Parikh, Shivam B., and Pradeep K. Atrey. "Media-rich fake news detection: A survey." 2018 IEEE conference on multimedia information processing and retrieval (MIPR).IEEE, 2018.
- [4] Granik, Mykhailo, and VolodymyrMesyura. "Fake news detection using naive Bayes classifier." 2017 IEEE first Ukraine conference on electrical and computer engineering (UKRCON).IEEE, 2017.
- [5] M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), 2017, pp. 900-903, DOI: 10.1109/UKRCON.2017.8100379.
- [6] Gupta, Arushi, and RishabhKaushal. "Improving spam detection in online social networks." 2015 International conference on cognitive computing and information processing (CCIP).IEEE, 2015.
- [7] Data extraction performance with pos tag pattern of dependency relation in aspect-based sentiment analysis. In2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP) 2018 Mar 26 (pp. 1-6). IEEE.
- [8]Labani M, Moradi P, Ahmadizar F, Jalili M. A novel multivariate filter method for feature selection in text classification problems.Engineering Applications of Artificial Intelligence. 2018 Apr 1;70:25-37.
- [9] Pratiwi AI. On the feature selection and classification based on information gain for document sentiment analysis. Applied Computational Intelligence and Soft Computing. 2018;2018.
- [10] Liu H, Zhou M, Lu XS, Yao C. Weighted Gini index feature selection method for imbalanced data. In2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC) 2018 Mar 27 (pp. 1-6). IEEE.
- [11] Monika D.Rokade ,Dr.YogeshkumarSharma,"Deep and machine learning approaches for anomaly-based intrusion detection of imbalanced network traffic."IOSR Journal of Engineering (IOSR JEN),ISSN (e): 2250-3021, ISSN (p): 2278-8719
- [12] Monika D.Rokade ,Dr.YogeshkumarSharma"MLIDS: A Machine Learning Approach for Intrusion Detection for Real Time Network Dataset", 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), IEEE
- [13]Monika D.Rokade, Dr.Yogesh Kumar Sharma. (2020). Identification of Malicious Activity for Network Packet using Deep Learning. *International Journal of Advanced Science and Technology*, 29(9s), 2324 - 2331.



- [14] Sunil S.Khatal ,Dr.Yogeshkumar Sharma, “Health Care Patient Monitoring using IoT and Machine Learning.”, **IOSR Journal of Engineering (IOSR JEN)**, ISSN (e): 2250-3021, ISSN (p): 2278-8719
- [15]Sunil S.Khatal ,Dr.Yogeshkumar Sharma, “Data Hiding In Audio-Video Using Anti Forensics Technique ForAuthentication ”, IJSRDV4I50349, Volume : 4, Issue : 5
- [16]Sunil S.KhatalDr.Yogesh Kumar Sharma. (2020). Analyzing the role of Heart Disease Prediction System using IoT and Machine Learning. *International Journal of Advanced Science and Technology*, 29(9s), 2340 - 2346.



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor:
5.928

ISSN

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY



9710 583 466



9710 583 466



ijmrset@gmail.com

www.ijmrset.com