



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 4, April 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.512



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

Fake News Detection System using Machine Learning

Dr.Saylee Gharge¹, Juveriya Khan², Poonam Masaye³, Dipti Ramane⁴, Riddhi Sakpal⁵

Professor, Dept. of Electronics & Telecommunication Engineering, University of Mumbai Vivekanand Education Society's Institute of Technology, Mumbai, Maharashtra, India¹

Student, Dept. of Electronics & Telecommunication Engineering, University of Mumbai Vivekanand Education Society's Institute of Technology, Mumbai, Maharashtra, India²³⁴⁵

ABSTRACT: In today's life most of the people are using internet to access news because it is easy and cheap. But it results in wide spreading of fake news very fast. Social media and news outlets publish fake news to increase readership or as part of psychological warfare. Fake news and lack of trust in the media are growing problems with huge ramifications in our society. Obviously, a purposely misleading story is "fake news" but lately blathering social media's discourse is changing its definition. Some of them now use the term to dismiss the facts counter to their preferred viewpoints. It affects so much to people. The purpose of the work is to come up with a solution that can be utilized by user to detect false and misleading information. The experimental results show a 98% accuracy using Random Forests Classifier.

KEYWORDS: ML, NLP, Fake News, Social Media, Random Forests Classifier.

I. INTRODUCTION

Social media plays an important role in today's life. It enables people to learn information, share ideas, discover new feature, research purpose. It allows to debate over issues such as democracy, education and health, etc. Also it has made communication easier, helps in connecting people and developing relationship. Social media makes communication easier not only in personal life but in business life also. Social networking sites can be used for managing events, creating activities or groups through which people can showcase their ideas, opinions, issues and make wider audience aware of them. Allows user to take part in activities and discussions on sites, create or modify content and upload. It supports user's acting, working and thinking together. In our life Internet has become compulsory. Accessing Social media is now very easy than it was before. People can access social media by many ways but most use mobile apps and computer web browsers. Internet access has become cheap. Earlier people use to read newspaper but now there is no doubt many people uses internet for getting news rather than newspaper, radio, etc. It is very easy to access news through digital devices but publishing news creates so many issues. News widespread very fast through internet. Some publishers uses misleading information about their interests. Difference in misinformation and disinformation is that misinformation is sometimes an honest mistake but disinformation creates confusion and lead audience to believe a lie. There are so many effects of fake news, it causes harm to people and they can get infected by online fake news easily. Fake news has been existing for a long time. There is rapid increase in the spread of fake news in the last decade. Fake news for various commercial and political purposes has been appearing in large numbers and widespread due to development in social network. Content publishers using concept of click bait. Click baits are nothing but text, link or phrases which are designed to attract attention of a user to click on that link, listen to the content, etc. Upon clicking link user is directed to a web page whose content is below their expectations. There is spread of inaccurate information on social media due false headlines. Over-the-top and crazy headlines makes impossible promises. People find click baits to be irritation which results that most of such individuals spends a very short time visiting such sites. There are many cases that people getting hurt or killed because of rumors on the internet.

In the 2016 US elections, various kind of fake news about the candidates spread in online networks, which may have effect on election result. One recent case is the spread of novel corona virus, where fake reports spread over the Internet about the origin, nature and behavior of virus [1]. Fake news creates problem in various domains like sports, health, science, financial marketed. It is very important to detect whether given new is real or fake rather than putting blinding

trust over it. For improving trustworthiness of information in online social networks is to identify fake news timely. The researchers [2] viewed about 16000 and 3000 accounts and pages from Twitter and Facebook respectively to learn how fake news gets polarized in India. This research indicates a “strong and coherent” proliferation of right-wing ideology, while networks which promote left-wing ideology based fake news were found to be less organized and effective. Another research [3] by the BBC resulted that nearly 72% of Indian citizens are not able to differentiate between real facts from made-up ones. Altogether, these concludes that we need to expose people in India to the digital literacy to overcome the consequences of fake news in the country. Many studies have focused on detection and classification of fake news on social media platforms such as Facebook and Twitter. Experiments results that machine learning algorithms have ability to detect fake news.

This project includes ML and NLP techniques for making model. Model has been proposed for incorporate the network structure information into model learning. The comparison is across traditional and nontraditional methods. Normalization method is used for cleaning data before using machine learning method to classify data. This paper reviews various machine learning approaches in detection fake and fabricated news. Method in this paper uses Naïve Bayes classification model to predict whether given news is real or fake. The technique facilitate training of machine learning algorithms in efficient manner. The objective of this project is to create a model for detecting fake news which is detected using machine learning algorithms. The result of this project validate the improved performance using confusion matrix, accuracy score.

II. RELATED WORK

The problem of fake news has been a major concern amongst various backgrounds. It is exponentially being regarded as a part of security related with information technology. It is considered to be a threat to information security. The easy access & exponential growth of the information available on social media networks has made it to intricate to distinguish between false & true information. The easy dissemination of information by way of sharing has added to exponential growth of its falsification. The credibility of social media networks is also at stake where the spreading of fake information is prevalent. Thus, it has become a research challenge to automatically check the information source, content & publisher for categorizing as it is false or not. Especially, some publishers, companies have included malicious code as the part of their content which they provide on their web pages.

It is leading those who visit such sites to click the links & download the malware without their knowledge. On the other side from the risk of having malware introduced into their information management systems, some of the modern firms also have to deal with the challenge of having their employees manipulated into giving out their credentials [4]. Data integrity has also been considered as being one the information security implications associated with fake news. So the companies are putting in all the necessary measures that would help secure sensitive information from being accessed by unauthorized persons. The available literature has described many automatic detection techniques of fake news detection ranging from using catboats for spread of misinformation to use of click baits available in social media networks including Facebook, Instagram which enhance sharing & liking of posts which in turn spreads falsified information. It is also apparent that the idea of using click baits to lure non-suspecting users to visit web pages has played significant role. As more & more users share their link to the site containing information that is seemingly related to an organization, other investors gain interests in these operations & its share price increases considerably. However, this result in worse consequences as a majority of the individuals who buy the shares based on the misinformation end up being disappointed.

The advancement in today's technology have authorized journalism & newspapers to spread over the all social networking sites. Because the networking sites have become the primary method for communication of people. For quick sharing of the all of these information exactness of the news is essential. As fake news has majorly influenced everyday lives & the social requests of many individuals & caused some negative impacts. The main point to study this is revealing the advantages that AI uses for the knowledge about fake news. Finally, the contemporary literature indicates that there have been ethical concerns about the fake news which is especially regarding the involvement of individuals having background in journalism. Journalists are expected to furnish readers with information whose accuracy have been determined to the last detail. When professionals engage in activities that are intended to misguide their readers for the sake of increasing web traffic & online ad revenues, that's why it raises a concern as to whether such people are keen on complying with the code of conduct associated with their career [5].

III. METHODOLOGY

In the following, we describe our proposed block diagram:

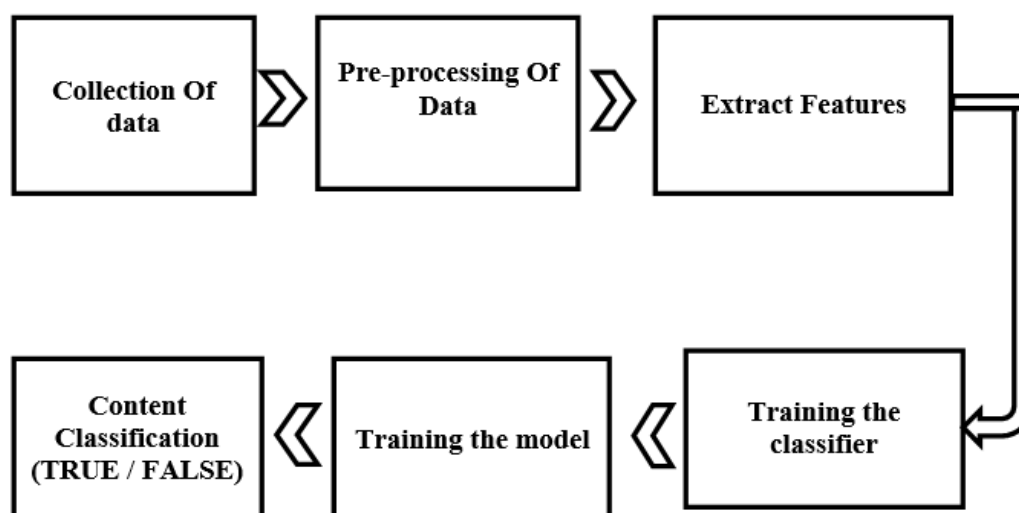


Fig.1

COLLECTION OF DATA:

First step includes importing some of the required libraries including pandas, numpy, seaborn, matplotlib, sklearn, string and re .The next step includes inserting fake and real dataset.

PRE-PROCESSING OF DATA:

This step includes insertion of a column called “class” for fake and real news dataset to categories fake and real news. Here, 0 is assigned for fake and 1 is assigned for real news. Last 10 rows and from both the dataset are removed for testing purpose. We have calculated the shape of dataset .Then next step was to merge the testing data frame in single datasheet and save in a csv file. The main fake and real data frame are merged. The columns such as “title”, “subject”, “date” are not required for detecting the fake news ,so we dropped these columns. Then we randomly shuffled the data frame. This blocks also includes creating a function to convert the text in lowercase, remove the extra spaces, special characters, URL and links. Then we have define dependent and independent variables as x& y and spitted the dataset into training and testing set.

EXTRACT FEATURES:

This block includes initialization of tf-idf vectorizer and fitting and transforming of data set.

TRAINING THE CLASIFIER:

The classifier we have used is Random Forests classifier. Random forests is supervised learning algorithm. It is also the most flexible and easy to use algorithm. A forest is comprised of trees. It is said that the more trees it has, the more robust a forest is. Random forests creates decision trees on randomly selected data samples, gets prediction from each tree and selects the best solution by means of voting. It also provides a pretty good indicator of the feature importance. Random forests has a variety of applications, such as recommendation engines, image classification and feature selection.

This classifier works in four steps:

- Select random samples from a given dataset.
- Construct a decision tree for each sample and get a prediction result from each decision tree.
- Perform a vote for each predicted result.
- Select the prediction result with the most votes as the final prediction.

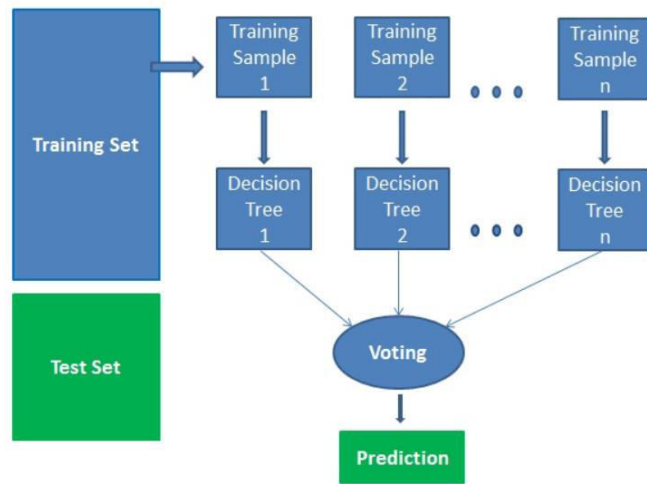


Fig.2

The above diagram from datacamp.com shows typical working of RFC classifier [6].

TRAINING THE MODEL:

We generated a model on the selected training set features, perform predictions on the selected test set features, and compare actual and predicted values.

CONTENT CLASSIFICATION:

This blocks takes a manual input from user to predict the given news is fake or real and accordingly gives output as fake/real news.

IV. EXPERIMENTAL RESULTS

TABLE 1. FOR MULTINOMIAL NB TF-IDF

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| fake | 0.99 | 0.98 | 0.99 | 5916 |
| real | 0.98 | 0.99 | 0.98 | 5309 |
| accuracy | | | 0.98 | 11225 |
| macro avg | 0.98 | 0.98 | 0.98 | 11225 |
| weighted avg | 0.98 | 0.98 | 0.98 | 11225 |

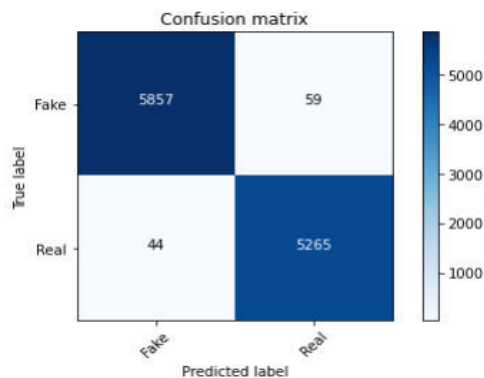


Fig 3. Confusion Matrix for Multinomial nb tf-idf

TABLE 2. FOR RFC

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| fake | 0.99 | 0.99 | 0.99 | 5916 |
| real | 0.99 | 0.99 | 0.99 | 5309 |
| accuracy | | | 0.99 | 11225 |
| macro avg | 0.99 | 0.99 | 0.99 | 11225 |
| weighted avg | 0.99 | 0.99 | 0.99 | 11225 |

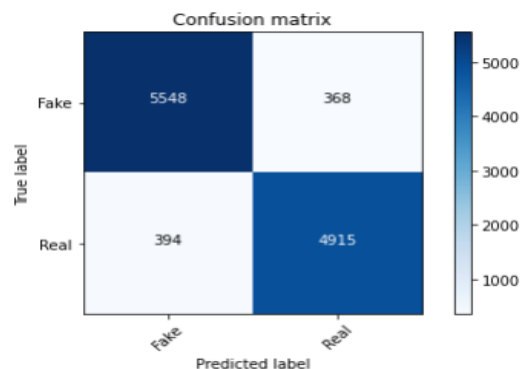


Fig 4. Confusion Matrix for RFC

Accuracy of Datasets in Percentages:

1. News.csv

Test size=0.1, Accuracy: 99.3%

Test size= 0.2, Accuracy: 99.1%

Test size=0.3, Accuracy: 99.1%

Test size= 0.4, Accuracy: 98.7%

Test size=0.5, Accuracy: 98.4%

2. True.csv & Fake.csv

Test_size=0.33, Accuracy: 93.35%

Test_size=0.40, Accuracy: 93.17%

Test_size=0.29, Accuracy: 93.63%

Test_size=0.20, Accuracy: 94.55%

Test_size=0.10, Accuracy: 93.53%

Test_size=0.50, Accuracy: 98.71%

1. URL: <https://www.hindustantimes.com/cities/mumbai-news/maharashtra-reports-another-daily-spike-in-covid-19-cases-with-63-294-infections-101618152957886.html>

News present on that URL:



Entered URL:



Output:



2.URL:https://twitter.com/NoncebaMhlauli/status/1345720298461659138?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1345720298461659138%7Ctwgr%5E%7Ctwcon%5Es1&ref_url=https%3A%2F%2Fwww.timeslive.co.za%2Fpolitics%2F2021-01-04-rumours-that-were-going-into-level-5-are-fake-news-says-presidency%2F

News present on that URL: Entered URL:



Output:



V. CONCLUSION

The explosive growth in fake news and its erosion to democracy, justice, and public trust has increased the demand for fake news detection and intervention. So from analysing the two data sets according to their test size and accuracy, we can conclude that as the test size differs so as accuracy of output getting fake or real news differ. The outcomes of this Project shows the capability of ML to be fruitful in this Task. We have tried to build a model that helps in catching many intuitive indications of real and fake news as well in the visualization of the classification decision.



REFERENCES

- [1] J. Hua and R. Shaw, "Corona virus (covid-19) "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](#)
- [2] Fake News Detection Using Machine Learning
Lilapati Waikhoma and Rajat Subhra Goswami. Department of Computer Science & Engineering, NIT, Arunachal Pradesh, India. International Conference on Advancements in Computing & Management (ICACM-2019)
- [3] <https://scroll.in/latest/838693/fake-news-is-a-concern-for-83-of-indian-media-consumers-reveals-bbc-study>
- [4] Marco L. Della Vedova, Eugenio Tacchini, Stefano Moret, Gabriele Ballarin, Massimo DiPierro, Lucade Alfaro, "Automatic Online Fake News Detection Combining Content and Social Signals", ISSN 2305-7254, 2017
- [5] Saranya Krishanan, Min Chen, "Identifying Tweets with Fake News", 2018 IEEE International Conference on Information Reuse and Integration for Data Science.
- [6] <https://www.datacamp.com/community/tutorials/random-forests-classifier-python>
- [7] The 9th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2018). Detecting Fake News in Social Media Networks. Monther Aldwairi, Ali Alwahedi. College of Technological Innovation, Zayed University, Abu Dhabi 144534, UAE..
- [8] International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072 © 2020, IRJET | Impact Factor value: 7.529 | ISO 9001:2008 Certified Journal | Page 5780



**Impact Factor:
7.512**



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com



www.ijirset.com

Scan to save the contact details