ISSN No:-2456-2165

Machine Learning for the Classification of Fake News

Swatej Patil
Computer Science and Engineering
G. H. Raisoni College of Engineering
Nagpur, India.

Dipti Theng
Computer Science and Engineering
G. H. Raisoni College of Engineering
Nagpur, India

Abstract:- Social networking services were designed to bring individuals from all over the world together and provide them a place to express their views and opinions. However, since their inception, social media platforms such as Facebook, Instagram, and Twitter have been misused for harmful purposes including publishing inaccurate information and promoting fake news. Surprisingly, due to its accessibility and wide range of topics, more individuals are turning to social media to consume news, rather than conventional news sources like newspapers and television. Recently, classification of fake news has caught the attention of many researchers, and there is an increasing demand for controlling the spread of fake news among these networking sites. In this manuscript, we have presented a method for classifying false information using TF-IDF vectorizer and Natural Language Processing. For training and evaluating the performance of the model we have used a dataset from Kaggle and Buzzfeed News. Our model shows promising results.

Keywords:- *Fake News, Machine Learning, Term Frequency, Inverse Document Frequency, Vectorizer, Natural Language Processing.)*

I. INTRODUCTION

Individuals from all over the world were connected via social networking sites, which also gave a forum for them to voice their ideas and opinions on a variety of subjects. Since their inception, social media platforms such as Facebook, Instagram, and Twitter have been used maliciously to disseminate incorrect information and propagate false news [1]. Although there is no universally recognised terminology for fake news, it may be defined as news items that are purposefully generated to mislead the public and have been shown to be untrue.

Several times, bogus news has wreaked havoc on the population and jeopardised their health. During the Covid-19 outbreak, there was a surge in the distribution of false information among people, with social media accounting for over half of the spread of fake news[3]. False information jeopardises people's health by speeding up phoney statistics that mention the virus's cure. Additionally, researchers at the disinformation research organisation Zignal Labs identified fabricated narratives stating that the vaccine contains tracking microchips and that authorities would make vaccinations mandatory to circulate[4].

Fake news was also linked to the 2016 US presidential election. Of the 30 million tweets created before the election, nearly 25% of the tweets were either clearly fake news or extremely biased information[5].The number of people using social media to consume news is increasing rapidly. More and more people are consciously and unconsciously disseminating fake news. There is a great demand for a model or system to classify and detect articles of false information.

II. LITERATURE SURVEY

This section gave an overview of the existing research on dealing with the fake news detection system and discussed the results.

The author in [6] developed a model called Article Abstraction using the Bidirectional Multi-Perspective Matching for Natural Language Sentence(BiMPM) method. They created their own Korean news Dataset. The output of the model is done by checking and comparing the news article with the fact DB. One downside is that BiMPM has some limitations because the longer the length of the input sentence, the poorer its performance, and it has difficulty making an accurate judgment when an unrecognized word or relationship between words appears.

Naive Bayes, Support Vector Machine, Decision Tree, Artificial Neural Network, and Logistic Regression were among the classifiers utilised by the author of [7]. The author used Kaggle datasets. Among all algorithms, logistic regression with count vectorizer has the highest accuracy. Even when combined with either of the vectorizers, the Artificial Neural Network has the lowest accuracy.

The author of [8] used Sentiment Analysis to create a false news detecting system. Their approach uses sentiment as a basic characteristic to predict the model's accuracy. They put their model to the test on four distinct datasets. The author compares several approaches for identifying false news and finds that tf-idf with cosine similarity outperforms tf-idf without cosine similarity, albeit the difference is not significant.

The author of [9] has concentrated on detecting false news that spreads via Twitter tweets. In addition to natural language processing, the presented approach integrates different machine learning and deep learning methods for the categorization of false news. The suggested approach takes into account not only the content of the tweets, but also additional attributes such as the number of likes, retweets, source, length, and verified. The random forest algorithm outperformed the other seven machine learning algorithms (Logistic Regression, Decision Tree, Random Forest, Naive Bayes, Gradient-Boosted Tree, Support Vector Machine, Multi Linear Perceptron) with an accuracy of 79 percent, precision of 85 percent, and F1-score of 83 percent.

The author of [10] took a straightforward solution to this problem, incorporating a Naive Bayes classifier into their false news detection algorithm. They utilised a dataset from

Buzzfeed News (which is the same dataset we use in our system) and were able to obtain a 74 percent accuracy rate. Spam messages and false news items have a lot in common, therefore the authors chose to use the Naive Bayes classifier to identify fake news articles because it works well with spam messages.

For the identification of false news linked to COVID19, the author in[11] has used Open Source Intelligence (OSINT). MedOSINT is a web-based tool that verifies news in three phases. The program begins by selecting a piece of news from the internet, then double-checking its validity with reputable news sources such as the World Health Organization (WHO), before storing the cases in a database. The dataset is then utilized by the CBR, which classifies all of the instances as true or false. It speeds up the categorization of news by allowing it to be kept in the database. CBR does not verify the accuracy of the information; instead, it characterizes it as true or false.

III. OVERVIEW OF THE DATASET

Datasets were obtained from Kaggle [12] and Buzzfeed News [13], however they needed to be preprocessed before we could use them in our model.The majority of the content is made up of a few common terms. As a result, we must eliminate punctuation and Stop words. Stop words are the terms that appear often in the majority of the texts in a corpus[14]. Cleaning and preprocessing the dataset will allow us to scan it more quickly and make the proposed methodology more convenient to use.

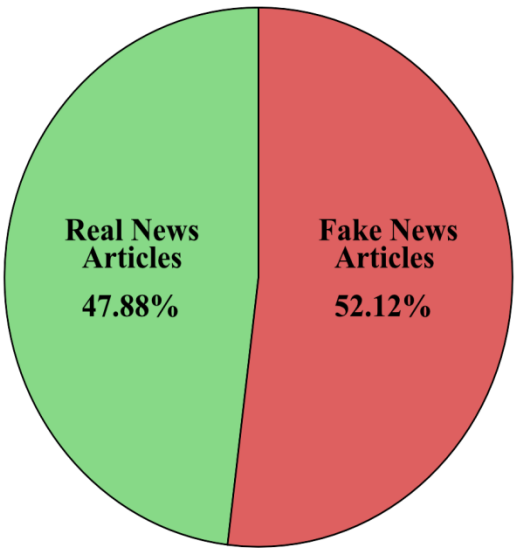

Figure 1: Dataset Description

IV. IMPLEMENTATION & METHODOLOGY

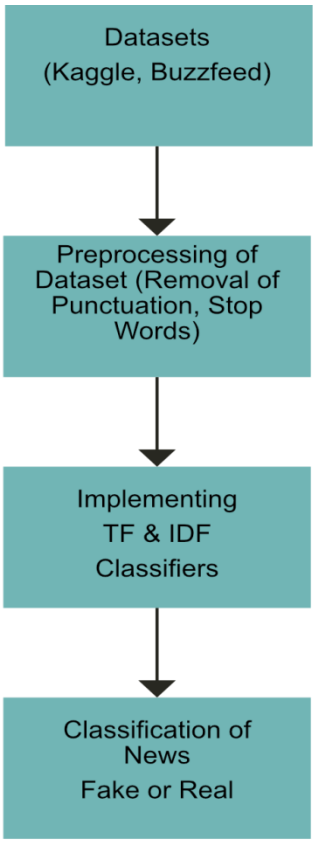

Figure 2: Flowchart for process of the fake news detection

A. *Term Frequency*

The term frequency (TF) indicates how frequently a phrase appears in a document[15]. In linguistic form, terms are synonymous with words or sentences.

B. *Inverse Document Frequency:*

The inverse document frequency (IDF) is a statistical weight for determining the significance of a phrase in a collection of text documents. The number of papers in which a phrase appears determines its document frequency DF[16].

$$Tf\ idf(t, d, D) = tf(t,d) \cdot idf(t, D)$$

Where:

$$tf(t,d) = \log(1 + freq(t,d))$$

$$idf(t, D) = \log(N/count(d\ \varepsilon\ D{:}t\ \varepsilon\ d\ ))$$

C. *Passive Aggressive Classifier:*

In Passive, if the information in the model or dataset is valid, the classifier does not make any modifications; in other words, the data in the datasets is insufficient to make any changes. Aggressive may be defined as the ability of the classifier to alter the model if the judgment made is incorrect. In passive, the information hidden in the example is

insufficient for updating; in aggressive, the information demonstrates that you are incorrect at least this time, and a better model should correct this error.

D. Natural Language Processing

Natural Language Processing (NLP) is a set of theoretically based computer approaches for evaluating and modelling naturally occurring text at one or more levels of linguistic analysis in order to achieve human-like language processing for a variety of activities and applications[17].

V. RESULTS & ANALYSIS

Without any additional algorithms, our system can execute with 92.72 percent accuracy. The model's accuracy does not alter much as a result of applying NLP.

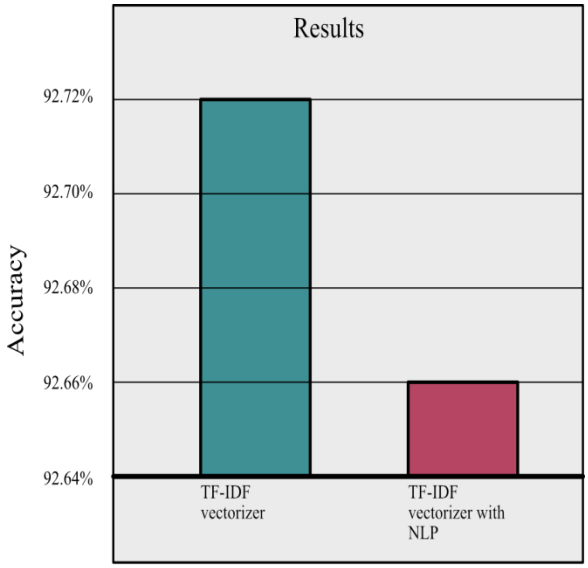

Figure 3: Comparison of the results obtained for TF-IDF vectorizer with and without NLP

Table 1: Results of the implementation

	Result Accuracy
TF-IDF vectorizer	92.72%
TF-IDF vectorizer with NLP	92.66%

VI. CONCLUSION AND FUTURE SCOPE

In recent years, an increasing number of individuals have turned to social media to get their daily news. However, the propagation of false news via social media is on the rise. We investigated the existing method for classifying false news in this paper and presented a solution based on machine learning. Machine learning algorithms may be utilised in a variety of sectors in the future, including the identification of erroneous information.

REFERENCES

[1.] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media. ACM SIGKDD Explorations Newsletter, 19(1), 22–36. https://doi.org/10.1145/3137597.3137600

[2.] 1.Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. Journal of Economic Perspectives, 31(2), 211–236. https://doi.org/10.1257/jep.31.2.211

[3.] Naeem, S. B., Bhatti, R., & Khan, A. (2020). An exploration of how fake news is taking over social media and putting public health at risk. Health Information & Libraries Journal, 38(2), 143–149. https://doi.org/10.1111/hir.12320

[4.] Lerman, R. (2020, December 18). Vaccine hoaxes are rampant on social media. Here's how to spot them. Washington Post. https://www.washingtonpost.com/technology/2020/12/18/faq-coronavirus-vaccine-misinformation/

[5.] Bovet, A. (2019, January 2). Influence of fake news in Twitter during the. . . Nature Communications. https://www.nature.com/articles/s41467-018-07761-2#Sec2

[6.] Kim, K. H., & Jeong, C. S. (2019). Fake News Detection System using Article Abstraction. 2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE). Published. https://doi.org/10.1109/jcsse.2019.8864154

[7.] Poddar, K., Amali D., G. B., & Umadevi, K. (2019). Comparison of Various Machine Learning Models for Accurate Detection of Fake News. 2019 Innovations in Power and Advanced Computing Technologies (i-PACT). Published. https://doi.org/10.1109/i-pact44901.2019.8960044

[8.] Bhutani, B., Rastogi, N., Sehgal, P., & Purwar, A. (2019). Fake News Detection Using Sentiment Analysis. 2019 Twelfth International Conference on Contemporary Computing (IC3). Published. https://doi.org/10.1109/ic3.2019.8844880

[9.] Madani, Y., Erritali, M., & Bouikhalene, B. (2021). Using artificial intelligence techniques for detecting Covid-19 epidemic fake news in Moroccan tweets. Results in Physics, 25, 104266. https://doi.org/10.1016/j.rinp.2021.104266

[10.] Granik, M., & Mesyura, V. (2017). Fake news detection using naive Bayes classifier. 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON). Published. https://doi.org/10.1109/ukrcon.2017.8100379

[11.] Martinez Monterrubio, S. M., Noain-Sánchez, A., Verdú Pérez, E., & González Crespo, R. (2021). Coronavirus fake news detection via MedOSINT check in health care official bulletins with CBR explanation: The way to find the real information source through OSINT, the verifier tool for official journals. Information Sciences, 574, 210–237. https://doi.org/10.1016/j.ins.2021.05.074

[12.] Fake News | Kaggle. (2021, October 14). Kaggle. https://www.kaggle.com/c/fake-news/data

[13.] BuzzFeedNews. (2017). GitHub - BuzzFeedNews/2017-12-fake-news-top-50: Data and analysis supporting the BuzzFeed News article, "These Are 50 Of The Biggest Fake News Hits On Facebook In 2017," published on Dec. 28, 2017. GitHub.

https://github.com/BuzzFeedNews/2017-12-fake-news-top-50
[14.] Munková, D., Munk, M., & Vozár, M. (2014). Influence of Stop-Words Removal on Sequence Patterns Identification within Comparable Corpora. ICT Innovations 2013, 67–76. https://doi.org/10.1007/978-3-319-01466-1_6
[15.] Qaiser, S., & Ali, R. (2018). Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. International Journal of Computer Applications, 181(1), 25–29. https://doi.org/10.5120/ijca2018917395
[16.] Ounis, I. (2009). Inverse Document Frequency. SpringerLink. https://link.springer.com/referenceworkentry/10.1007%2F978-0-387-39940-9_%20933?error=cookies_not_supported&code=fd3f39ab-fd8b-4c26-8932-fb2e0ec40c24
[17.] Liddy, E.D. 2001. Natural Language Processing. In Encyclopedia of Library and Information Science, 2nd Ed. NY. Marcel Decker, Inc.