WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Fake News Detection Using Machine Learning

Shravani B V[1], Shreelekha Joshi[2], Varapradha K[3], Vidya Lakshmi C[4], Mrs. Shwetha Shree A[5]

[1]Dept of CSE (of Aff.), Ballari Institute of Technology and Management (of Aff.), Ballari,India.
[2]Dept of CSE (of Aff.), Ballari Institute of Technology and Management (of Aff.), Ballari,India.
[3]Dept of CSE (of Aff.), Ballari Institute of Technology and Management (of Aff.), Ballari,India.
[4]Dept of CSE (of Aff.), Ballari Institute of Technology and Management (of Aff.), Ballari,India.
[5]Assistant Professor Dept of CSE (of Aff.), Ballari Institute of Technology and Management (of Aff.) Ballari,India.

ABSTRACT—

One of the catastrophe in the present times is fake news. The most harmful tool for swaying public opinion and facts is fake news, which has great potential for doing so. The proposed project use NLP approaches to spot "fake news," or false news items that come from questionable sources. The creation of a model based on the Passive Aggressive Classifier method can be used to identify bogus news. In response, we are trying to create a system begun to address the issue. To identify bogus news, the proposed project will make use of datasets that are trained using the count vectorizer and its efficacy will be evaluated by applying machine learning techniques (such as Support Vector Machine or Naive Bayes Classifier). We will also implement the system inside the web application to view and classify the news as original or fake.

Index Terms—NLP, datasets, machine learning algorithms.

I. INTRODUCTION

According to a survey, it is estimated that 86% of the internet users are being tricked by fake news nowadays. The primary goal of the fake news is to delude the readers, thereby making it non-trivial to simply identify based just on the available news content. This was driving force to implement a system to recognize fake news. The best way to combat fake news, which is a serious problem for both academia and business, is to use human fact-checking. Regardless, the real-time nature of fake news on social media makes it even harder to recognise online fake news. However, due to its low efficiency and reliability, trained human fact-checking could only be partially helpful. Additionally, fact-checking by human beings can be extremely costly and laborious. The spread of this false information has a variety of negative effects, including the emergence of political beliefs that are biassed in one direction or another. By clicking on obstacles and utilising alluring news headlines, spammers also legitimise advertisements. Most smartphone users prefer to read the news on online discussion groups. News-related websites act as trustworthy sources and disseminate breaking news. In what way to distribute headlines and content across so many internet platforms, including Twitter, Facebook pages, microblogs, WhatsApp groups, and social networking sites, is currently an issue. The public faces the danger of being harmed in order to spread these stories and create headlines. It is vital to put an end to rumors and focus on real, established challenges, especially in developing countries like India. Since the development of the media, fake news has proliferated all over the world. As a result, people now distrust fake news. In the current digital era, fake news is a ubiquitous challenge which is difficult to solve in today's digital age since there are so many places where misleading information can spread. Hence, in order to automate this procedure, we must employ the technique of Machine Learning (ML). The primary goal of the proposed work is to recognize bogus news using machine learning in order to ensure credibility, get the rewards of the actual news, and deserve the truth. In order to identify bogus news, several hierarchical classification techniques can be applied.

II. LITERATURESURVEY

In-depth comparisons of similar works are provided in this chapter. The majority of earlier research has focused on applying deep learning and machine learning algorithms to discern between true and incorrect information. "Fake news detection using naive Bayes classifier," 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), by V. Mesyura and M. Granik [1], pp. 900–903, doi: 10.1109/UKRCON.2017.8100379. In this paper, the author uses a straight forward method to identify fake news using Naïve Bayes Classifier. After being turned into a software system, the technique was evaluated using a set of Facebook news posts. The author's classification accuracy on the test set was roughly 74%, which is a decent result given the model's relative simplicity. The author also notes that there are a number of ways in which these outcomes could be enhanced. The findings of this study suggest that the problem of false news identification can be resolved using artificial intelligence approaches. The study showed that even very simple artificial intelligence algorithms, such as the naive Bayes classifier, can yield effective outcomes when used to tackle a pressing problem like the detection of bogus news. The results of this study further suggest that artificial intelligence methods could be successfully used to deal with this important problem.

Nikita, J. Singla, and S. I. Manzoor [2], "Fake News Detection Using Machine Learning approaches: A systematic Review," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 2019, pp. 230–234; doi: 10.1109/ICOEI.2019.8862770. The author of this research notes that utilising a variety of machine learning techniques, bogus news and posts can be detected with apparent success. However, the characteristics and features of fake news in social media networks are constantly changing, making it difficult to categorise. Numerous research projects will use deep learning techniques in a variety of applications and multi-task learning in the classification, as a result of the implementation of deep learning research and applications in the past.

2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), pp. 1–5, "Fake News Detection," by A. Jain and A. Kasbe [3], doi: 10.1109/SCEECS.2018.8546944.

The author claims in this study that it is clear from the outcomes of the application of straightforward Naive Bayes that the AUC scores have risen as the volume of data associated with a given tag has grown, as is the case with the title and text. As Title was a condensed version of news articles, and the text was a similar descriptive version. We can see that the AUC scores have increased with n_grams as seen by comparing the outcomes of different approaches where the idea of n_grams was introduced in the second model

N. F. Baarir and A. Djeffal [4], "Fake News detection Using Machine Learning," 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH), 2021, pp. 125-130, doi: 10.1109/IHSH51661.2021.9378748.

In this paper, the author used Support Vector Machine for the classification. Initially the author studied about the impact of the fake news and the methodologies for the detection of fake news. Then they developed a solution which uses dataset. The data is preprocessed using cleaning, stemming, TF-IDF. Support Vector Machine methodology was used to create a model which classifies the news. The author gave the future scope to expand this work with large dataset.

M. Kumar Jain, D. Gopalani, R. Kumar and Y. Kumar Meena [5], "Machine Learning based Fake News Detection using linguistic features and word vector features," 2020 IEEE 7th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), 2020, pp. 1-6, doi: 10.1109/UPCON50219.2020.9376576.

The author concentrated only on the text field of the news. Experiments done on the text field gave better conclusions. This proposed methodology uses a group of 70 stylometric traits which produced more required results than the current methodologies. Along with lemmatization and stemming, BOW TF-IDF was also used. The author suggests for running the models which includes both stylometric features and word vector characteristics to obtain desired outcome.

V. Kumar, A. Kumar, A. K. Singh and A. Pachauri [6], "Fake News Detection using Machine Learning and Natural Language Processing," 2021 International Conference on Technological Advancements and Innovations (ICTAI), 2021, pp. 547-552, doi: 10.1109/ICTAI53825.2021.9673378.

The author uses the principles of AI &ML and executes the binary categorization of news. The author believes that people are dependent on the social media news and are misleaded, so author wants to make it possible for the users to identify the news as fake or real. With the increase in the users of the social media sites such as Facebook, Instagram, Whatsapp, Twitter, the news spreads quickly within seconds among the numerous media users.

E. Z. Mathews and N. Preethi [7], "Fake News Detection: An Effective Content-Based Approach Using Machine Learning Techniques," 2022 International Conference on Computer Communication and Informatics (ICCCI), 2022, pp. 1-7, doi: 10.1109/ICCCI54379.2022.9741049.

The aim of this paper is to address the issue of fake news by impulsive detection of false news based on the news content. Using the metrics like accuracy, precision are used to evaluate the performance of the methodology. The author defines the fake news as the information that is false or deceptive and regularly used to spread ideas, usually for commercial or political gain.

III. ARCHITECTURE AND SYSTEM

DESIGN

Finding evidence to back up false information can be difficult and time-consuming because it is meant to make readers believe it. The limitations in earlier work are fixed by using a machine learning model, an approach to check the credibility of news, publications, and social media. In the class of group of online learning algorithms, passive aggressive classifier is member which is capable enough of handling larger datasets and adjust itself to new circumstances is faces. The passive aggressive algorithm can adjust its weights in response to fresh data because it is an online learning algorithm. The passive aggressive classifier's regularization parameter C enables a trade-off between the size of the margin and the quantity of misclassifications. After each iteration, the passive aggressive classifier evaluates a fresh instance to see if it was correctly classified or not, and if not, modifies its weights accordingly. If the event is accurately classified, there won't be a noticeable change in weight. However, if this instance is incorrectly classified, the passive aggressive method modifies its weights accordingly in order to classify subsequent instances more accurately. The degree of weight modification by the Passive Aggressive method depends on the regularization parameter C and its confidence in the categorization of that particular occurrence. The passive aggressive classifier separates a set of training data into two groups: a training set and a test set, much like other supervised learning algorithms. Items must be appropriately categorized into one of two groups in order to do so. The training set is used by the passive aggressive classifier. Once it has figured out how to accomplish this, the model is tested using data from the test set, and its accuracy is assessed.

A. BLOCK DIAGRAM OF PROPOSED WORK

The overall block diagram of the proposed work illustrates how it would operate using news data as input and feature extraction from data after preprocessing. After that, the data should be separated into training and testing sets. Determine whether the information is accurate next. Large-scale learning typically employs passive-aggressive algorithms, which are uncommon among "online-learning algorithms". In contrast to batch learning, which uses the entire training dataset at once, online machine learning techniques use sequential input data and update the machine learning model one step at a time.

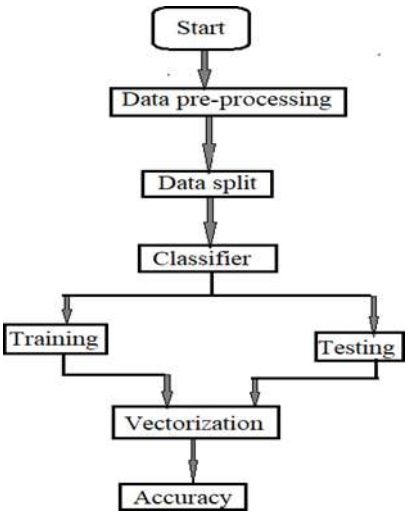

Fig.1. Block diagram of the proposed work.

This is especially helpful when there is a large amount of data and training the complete dataset would be computationally impossible due to the magnitude of the data. In a nutshell, an online learning algorithm will acquire a training example, update the classifier, and then discard the example. Finding fake news on a social media platform like Twitter, where new information is being published every second, would be a very good illustration of this. It would be excellent to use an online-learning algorithm to dynamically read Twitter data continuously because the amount of data would be enormous. In that they don't require a learning rate, passive-aggressive algorithms resemble Perceptron models in certain ways. They do, however, have a regularization parameter.

Passive: If the prediction is accurate, make no modifications to the model.

Aggressive: If the prediction is inaccurate, changes to be done to the model.

IV. PROPOSED METHODOLOGY

A. DATA PRE-PROCESSING

Pre-processing is the process of modifying or altering data via a series of steps. Our data is modified before being provided to the algorithm. Data processing is the process of converting messy records into orderly record sets, especially when it is carried out by a computer. When information is from many sources, analysis is done. Next, a legible format (graphs, papers, etc.) is created from the raw file. It turns knowledge from unprocessed data. Data processing companies require qualified personnel with the ability to process data using a range of technologies. Our project initially contains a column called "class" with the values 0 for inaccurate information and 1 for accurate information. The two distinct true and false CSV files are done into a single CSV file. Ten rows of data are saved in a separate file with a randomly generated order for testing reasons. After removing any columns that are not necessary for prediction and scanning for any null values, the pertinent rows are then deleted. With our feature, you can alter the capitalization and get rid of extra spaces, special characters, URLs, and links.

B. FEATURE EXTRACTION

Feature extraction is often carried out when the original data is more different and cannot be used directly for learning modeling. The required form is then given to the original raw data. Feature extraction is the process of creating new, more precise features from raw data that capture the majority of it information. One of the many feature extraction methods that we use is the TF-IDF vectorizer.

C. TF-IDF VECTORIZER

Term frequency-inverse document frequency (TF-IDF). The TF-IDF measure is usually applied to information retrieval and text mining. Search engines frequently use iterations of the TF-IDF measure algorithm to evaluate and rank the similarity of documents. A statistical study of a word's importance to a document in a group is conducted by TF-IDF measure. The frequency of a term in the collection affects how important a word is (data-set), even if it counteracts the importance boost brought on by the term's frequency in the text.

D. BUILDING THE MODELS

Three models are created here, and the best model is selected for deployment. The utilized models are

1. Support Vector Machine(SVM)
2. Naïve Bayes Algorithm
3. Passive Aggressive Classifier SUPPORT VECTOR MACHINE

The problems of classification and regression are solved by Support Vector Machine which is the mostly used supervised learning methods. However, the most of its applications involve Machine Learning Classification issues. The aim of the SVM algorithm is to find a hyper plane in an N-dimensional space that uniquely classifies the data points. To swiftly categorize new data points in the future, the SVM technique aims to define the best decision boundary or line that can divide n-dimensional space into classes. This optimum choice limit is referred to as the "hyper plane". SVM builds the hyper plane by choosing the most extreme points and vectors. The SVM technique is built on support vectors, which are utilised to represent these extreme circumstances.

NAÏVE BAYES ALGORITHM

The Naive Bayes algorithm, a supervised learning technique built on the Bayes theorem, is used to address classification problems. Its primary use is text classification with a huge training set. The naive bayes technique is one of the most basic and powerful classification method, in which that enables rapid development of machine learning designs that are capable of producing reliable predictions. The algorithm formulates predictions on the basis of the chance that an particular event occur since it is a classifier that is based on probabilities. Article categorization, sentiment analysis, and spam filtration are a few uses for Naive Bayes algorithms.

PASSIVE AGGRESSIVE CLASSIFIER

When you first begin using machine learning, you only employ supervised and unsupervised learning techniques to address problems. Thus, supervised and unsupervised machine learning are thought to be the only two types of machine learning, according to the majority of practitioners. An online learning technique known as a passive aggressive classifier trains a system incrementally by sequentially feeding it instances, either one at a time or in small groupings known as mini-batches. In order for online learning to continue learning as new data sets are collected, a machine learning model is created and implemented. We may therefore draw the conclusion that an algorithm like the passive aggressive classifier is most useful for systems that receive data in a continuous stream.

D. MODEL DEPLOYMENT

The Flask web app's home page offers the user the opportunity to choose an input method right away. The user is given the relevant page after choosing the input type. After the user enters the news text, the input data is transmitted to the backend. The appropriate method is used to make a prediction based on user input after the input has been evaluated. NewsCatcher API is used for identifying the entered news's source and the article where that news was published.

V. CONCLUSION AND FUTURE WORK

Online news articles and services like Facebook and Twitter are gradually taking the place of newspapers, which were once preferred as hardcopies. One such useful resource is the WhatsApp forwards. The issue of fake news, which is becoming worse, makes things more difficult and attempts to affect people's opinions about using digital technology. When someone is duped by the real news, they could start to think that their perceptions on particular subjects are true. To remedy the situation, we developed our Fake News Detection system, which accepts human text input and classifies news as authentic or fake.

REFERENCES

[1] 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.

[2] Nikita and S. I. Manzoor, J. Singla, "Fake News Detection Using Machine Learning approaches: A systematic Review," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 2019, pp. 230-234, doi: 10.1109/ICOEI.2019.8862770.

[3] A. Jain and A. Kasbe, "Fake News Detection," 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), 2018, pp. 1 -5, doi: 10.1109/SCEECS.2018.8546944.

[4] A. Djeffal and N.F. Baarir, "Fake News detection Using Machine Learning," 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH), 2021, pp. 125-130, doi: 10.1109/IHSH51661.2021.9378748.

[5] M. Kumar Jain, D. Gopalani, Y. Kumar Meena and R. Kumar, "Machine Learning based Fake News Detection using linguistic features and word vector features," 2020 IEEE 7th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), 2020, pp. 1-6, doi: 10.1109/UPCON50219.2020.9376576.

[6] V. Kumar, A. Kumar, A. K. Singh and A. Pachauri, "Fake News Detection using Machine Learning and Natural Language Processing," 2021 International Conference on Technological Advancements and Innovations (ICTAI), 2021, pp. 547

-552, doi: 10.1109/ICTAI53825.2021.9673378.

[7] E. Z. Mathews and N. Preethi, "Fake News Detection: An Effective Content-Based Approach Using Machine Learning Techniques," 2022 International Conference on Computer Communication and Informatics (ICCCI), 2022, pp. 1-7, doi: 10.1109/ICCCI54379.2022.9741049