International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 2061-2064 www.ijaem.net ISSN: 2395-5252

Detecting Fake News

[1]Student,Harshini, [2]Student, Shalma Aishal Rodrigues, [3]Student,Shivani Purushotham, [4]Student, Sweekritha Jeevandas Hegde, [5]Professor, Mrs. Geethalaxmi,

Canara Engineering College Mangalore, Karnataka
Canara Engineering College Mangalore, Karnataka
Canara Engineering College Mangalore, Karnataka
Canara Engineering College Mangalore, Karnataka
Dept.of ISE, Canara Engineering College Mangalore, Karnataka

Submitted: 15-06-2022 Revised: 25-06-2022 Accepted: 27-06-2022

ABSTRACT: As a result of occurrence of different technologies and media platforms, information has become more widely available. The news gets spread to a great amount of people, resulting in increased use of internet media categories including LinkedIn, Instagram, Meta, and others. Spreading bogus news has a number of negative implications. Phishers frequently use news features to their gain, such as raising cash through click baits. False news is propagating at a rapid and increasing rate, owing to the growth of social media and communications. Incorrect information discovery is a relatively new subject of study that has drawn a lot of interest. We present a model that uses a passive aggressive method, term frequency – inverse document frequency (TF-IDF), and an Optical Character Recognition to detect bogus news in this paper (OCR). Detection of fake news is decided by the project's outcome.

KEYWORDS: Benign, Malignant, Pituitary, MRI.

I. INTRODUCTION

The growing popularity of social media & mobile technology with this information is accessible at one's fingertips.Mobile apps and social media like Facebook and Twitter have overthrown traditional media in the field of information and news. With the convenience and speed that digital media offers, people expresspreference towards socialmedia. Not only has it empowered consumers withfaster access but it has additionally given benefit looking for parties a solid stage to catch a more extensive crowd.

With a lot of information or news, the one question occurred whether the given news or information is True or Fake. Fake news is commonly distributed with an intent to mislead or make an inclination to get political or monetary benefits. Let's consider the example - In the recent elections of India, there has been a lot of discussion in regards to the credibility of different news reports preferring certain applicants and the political thought processes behind them. In this growing interest, exposing fake news is paramount in preventing its negative impact on people and society.

Different strategies include the investigation of the spread of fake news interestingly with real news. Specifically, this approach analyses fake news articles propagates differently on the internet relative to a true article. The reaction that an article gets can be separated at a theoretical level to arrange the article as real or fake. The hybrid approach can also be used to investigate the social responsibility of an article alongside investigating the text-based features to examine whether an article is deceptive or not.

The method of fake news detection based on different algorithms called the Naive Bayes classifier and passive aggressive classifier helps to examine how this particular method works for the particular problem with a manually labelled (fake or real) dataset and to support the idea of using machine learning to detect fake news.

II. IMPLEMENTATION

Data: In this project, we have extracted our data from a famous dataset vendor known as Kaggle. The Kaggle website provides us with large number of datasets rightfrom the oldest to the trending data set. The dataset consists of large number of rows of data from various articles on newspaper.There are lots of pre-processing to be made on the dataset before training the data. A whole training dataset has the following attributes:

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 2061-2064 www.ijaem.net ISSN: 2395-5252

- Id: Unique id for news article
- Title: The title of the news article
- Text: The text of the article; incomplete in some cases
- Label: A label that marks the article as possibly erratic

-Fake
-Real

Pre-Processing and Feature Extraction: Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating our model. In prior to the training and testing of data using various models,the data should be pre-processed and it has to be sorted and subjected according to certain conditions such as removing stop words, tokenization, a lower casing and punctuation removal.This process will make the data to reduce its size and remove the irrelevant data. This is the step to remove the waste data. The generic function in the program will help to eliminate the punctuation as well as the non-character objects.

The stop words in the context will contain insignificant words used in the language that will make trouble and disturb the fake news detection process. There may be even non-character words such as special characters, which must be eliminated in order to reduce the size of the actual data. Then the common words such as a, as, about, an, are, at, be, by, for, from, how, in, is, of, on, or, that, these etc. This process will produce a comma separated list of words which is given as an input tothe next process. This is done using Term frequency-inverse document frequency (TF-IDF) method. Our goal is to produce a vector form of our dataset of each article.

Passive Aggressive Classifier: Passive Aggressive algorithms are online learning algorithms. Such an algorithm remains passive for a correct classification outcome, and turns aggressive in the event of a miscalculation, updating and adjusting. Unlike most other algorithms, it does not converge.Its purpose is to make updates that correct the loss, causing very little change in the norm of the weight vector.

Fig.iImplementation of Passive Aggressive classifier

Passive-Aggressive algorithms are generally used for large-scale learning. In online machine learning algorithms, the input data comes in sequential order and the machine learning model is updated step-by-step, as opposed to batch learning, where the entire training dataset is used at once. This is very useful in situations where there is a huge amount of data and it is computationally infeasible to train the entire dataset because of the sheer size of the data. We can simply say that an online-learning algorithm will get a training example, update the classifier, and then throw away the example.

Naive Bayes Classifier:

In order to acquire an accuracy rate of our data, we applied a naive bayes classifier. In particular, we took the help of scikit-learn implementation of gaussian naïve bayes. Knowing that it is the simplest methods towards classification, through which a probabilistic method was used, having the hypothesis that each feature is briefly liberated given the class label. Since it is impossible to stop the spread of fake news at once for all, we need to consider the ways in which we can prevent them. We will count the number of times a word appears in the headline, given that the news is fake. Change that to a probability, and then calculate the probability that the headline is fake, as compared to the headline being real.

Fig.ii Implementation of Naive Bayes classifier

Optical Character Recognition (OCR): We are using Tesseract optical character recognition

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 2061-2064 www.ijaem.net ISSN: 2395-5252

engine. It's an open-source OCR (Optical character recognition) engine that can recognize more than 100 languages with Unicode support. Also, it can be trained to recognize other languages. An OCR engine can save time by digitizing documents rather than manually typing the content of the document.

We should installPython-tesseract,which is a wrapper for Tesseract OCR engine. Also, we need to install a Python imaging libraryPillow. It takes the image name and language code as parameters. We use a pytesseract method which returns the unmodified output as a string from Tesseract OCR. Additionally, we have added two helper methods. The print data method prints string output, and the output file method writes the string output to a text file.

Testing: Testing is the process by which the framework monitors the behaviour to ensure that the set requirements are met. Procedure testing is carried out by doing out experiments and determining whether the framework's behaviour is positive or negative.

Types of Testing:

Unit Testing: Naive bayes algorithm was separately tested than passive aggressive algorithm to see if they were able to give the accuracy better than the existing system. These were tested separately so that, we could compare between each other.

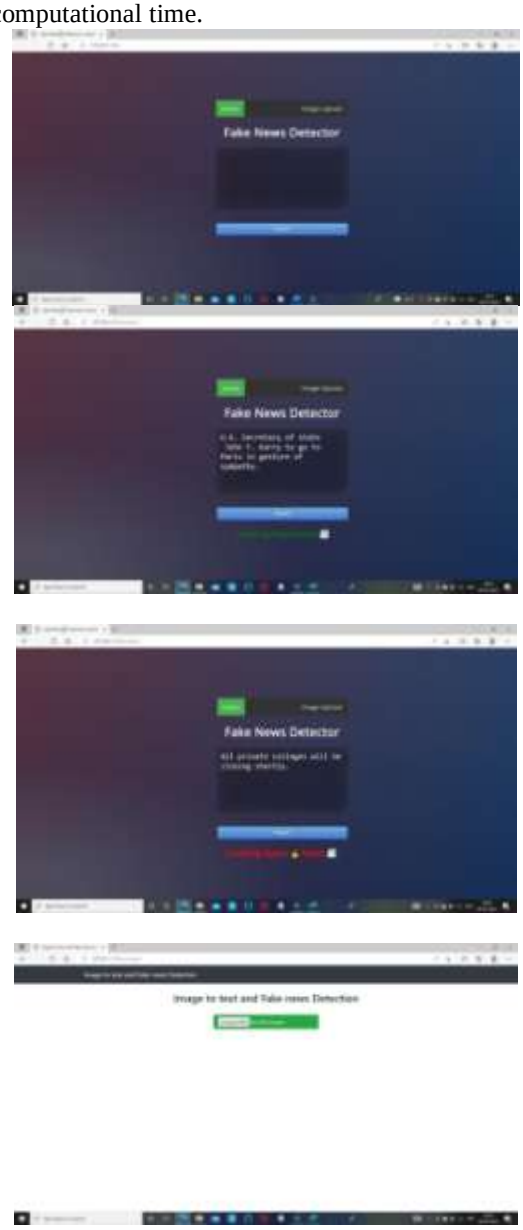
System Testing: Under System Testing technique, the entire system is tested as per the requirements. It is a Blackbox type testing that is based on overall requirement specifications and covers all the combined parts of a system. Here we tested if the endintegrated code could run on any system, we saw that the integrated codecan run on anysystem having python version 3.6 or more, and we never faced any error.

Compatibility Testing: In general, we used it on windows and python version 3.6. And we used the libraries like Sci-Kit, NumPy, pandas etc. And also, there was no specific requirement like we need to use this particular version of python or Sci-Kit library of particular version, so we worked in a closed loop. We were successful at getting greater efficiencywith these.

Usability Testing: This project could be easy for python and data science programmer, not meant for general purpose. Application is usable for data science engineers to pick the model for future research in the area of fake news classification.

III. RESULT

We have developed a website that detects fake news and differentiates between fake and real news.The website includes a home page, detection page and implementation page. The proposed model had obtained an accuracy of 93% and yields promising results without much errors and less computational time.



International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 2061-2064 www.ijaem.net ISSN: 2395-5252

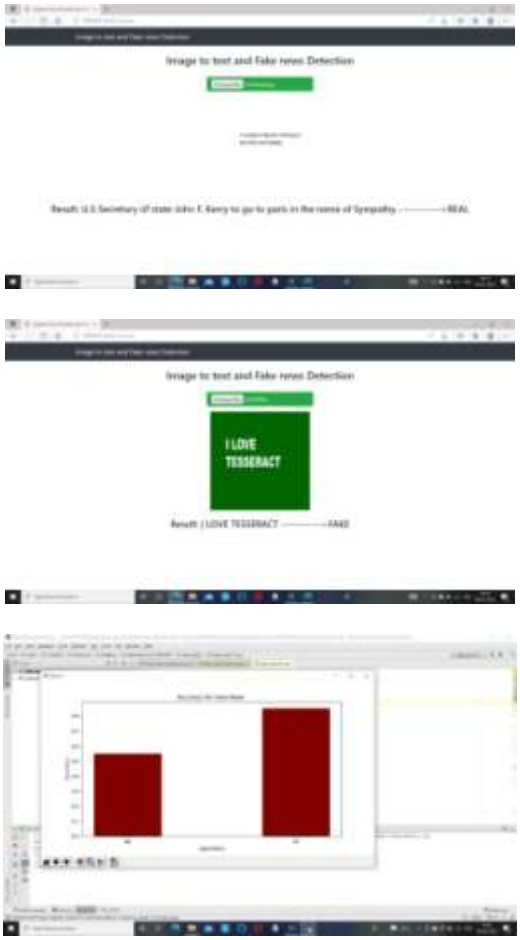


Fig.iiiFake news detection result

IV. CONCLUSION

News stories must be explicitly categorized and we require a great understanding and mastered in identifying oddities. Since it takes a while to verify news so we have opted this project using ML models and a set of methods. It is important for having accessibility to inaccurate news, or we should know that whatever society tells us is not always accurate. As a result, we need to be careful. This is how we can guide people to have more knowledge so that they will not get convinced to believe in false data. In this project feature extraction method like Term frequency inverse document frequency (TF-IDF Vectorizer) has been used and also different classification algorithms like Passive aggressive Classifier and naïve bayes classifier have been used to classify the news as fake or real. By using the classification algorithms, we got highest accuracy with Passive aggressive classification algorithm and with TF-IDF feature extraction with 0.93 accuracy. So has we have found out that passive aggressive is the best system, hence we have used it with our front end for this project. In future we can also use deep learning methods and sentiment analysis to classify the news as fake or real which may get high accuracy and we can extract further useful text like publication of the news, URL domain etc., We can use more data for training purposes - In machine learning problems usually availability of more data significantly improves the performance of a learning algorithm.

REFERENCES

[1]. Mayur Bhogade, Bhushan Deore, Abhishek Sharma, Omkar Sonawane and Prof. Manisha Singh proposed a Review Paper on Fake News Detection.
[2]. Pranav Ashtaputre, Ashutosh Nawale and Rohit Pandit proposed A Machine Learning Based Fake News Content Detection Using NLP.
[3]. Iftikhar Ahmad and Muhammad Yousaf proposed Fake News Detection Using Machine Learning Ensemble Methods.
[4]. Muhammad Syahmi Mokhtar, Yusmadi Yah Jusoh, Novia Admodisastro, Noraini Che Pa and Amru Yusrin Amruddin proposed Fakebuster: Fake News Detection System Using Logistic Regression Technique in Machine Learning.
[5]. Dr Jimmy Singla, Syed Ishfaq Manzoor and Nikita proposed Fake News Detection Using Machine Learning approaches: A systematic Review.
[6]. Pradeep K. Atrey and Shivam B. Parikh proposed Media-Rich Fake News Detection: A Survey.
[7]. Manisha Gahirwal, Sanjana Moghe, Tanvi Kulkarni, Devansh Khakhar and Jayesh Bhatia proposed Fake News Detection based on Natural language processing.
[8]. Akshay Jain and Amey Kasbe proposed Fake News Detection.
[9]. Sholka Gilda proposed evaluating machine learning algorithms for fake news detection.
[10]. Niall J. Conroy, Victoria L. Rubin, and Yimin Chen proposed Automatic Deception Detection: Methods for finding fake news.