JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A SUPERVISED LEARNING ESTIMATOR FOR DETECTING FAKE NEWS ARTICLES USING NATURAL LANGUAGE PROCESSING

[1]Swetha Nalanagula, [2]Disha Bhattacharya

[1]Computer Science and Engineering,
[1]St.thomas College of Engineering & Technology, Kolkata, India.

ABSTRACT: With the Internet everywhere these days, everyone relies on a wide range of online news sources. It has become increasingly common for news to spread rapidly among millions of people in a very short amount of time as Facebook, Twitter, and other social media platforms have gained popularity. Initially, fake news spreads across social media platforms and later finds its way onto traditional TV and radio news. The results of this study are presented and discussed in this paper. In order to identify whether the news is false or true, Naive Bayes classifier and SciPy tools are used in conjunction with natural language processing and Naive Bayes algorithms to extract quotes from the news articles and calculate the likelihood percentage based on the extracted quotes.

Keywords: Machine Learning, Natural Language processing, Fake News, Articles, Social Media

INTRODUCTION:

Today, most of our lives are spent online, including using social media platforms like Facebook, Twitter, etc., and there are fewer people who try to avoid social media sites. Rather than being spread through newspapers, most fake news today is spread on social media platforms and then on other media channels as well. As soon as the news is posted on social media, it spreads quickly, and finding the source of the news is difficult, for example, where the news came from and who started it. False news spreads throughout the country, and many people believe it is real news rather than fake news, which will negatively impact each individual in society. By spreading fake news, social media not only benefits but also affects people in one way or another. It is unlikely that fake news will spread spontaneously through newspapers or physical documents, rather than through social media. There are times, however, when false information or news is deliberately spread on social media for financial reasons and in particular in politics, for reasons of spontaneity and speed. In addition, social media allows news to spread quickly and spontaneously. There are also estimates of millions of fake news posts spreading through Twitter, Facebook, and other social media outlets. Many scientists believe that machine learning algorithms and techniques can be used to address fake news issues. Natural Language Processing (NLP) is a tool whose purpose is to process text and classify it using a variety of algorithmic and technical techniques. We discussed the Naive Bayes algorithm here as one such method. The probability of words is calculated by counting how many words there are, how many verbs there are, and how many unique words there are. We present and discuss how to calculate the probability score and identify real and fake news using the Naive Bayes classifier.

FRAMEWORK:

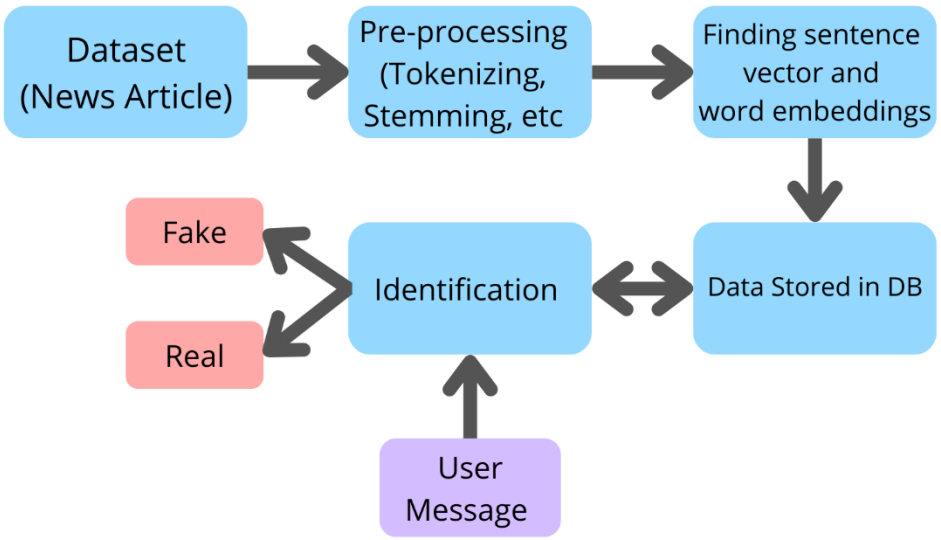
As we all know, fake news has become one of the biggest problems of our lives not just back then but now as well. It has caused a lot of problems and has affected every individual and every business in one way or another. Before, we didn't have any technology to determine what news was fake and what news was real. In other words, they just had people manually seek out news from various sources to see whether it is real or fake. Many times, when we receive the news, we don't even realize that its news, and we assume its real news without realizing it. So, as the days passed, technology has become more prevalent. In addition, with these new emerging technologies, it has become easier to detect fake news than it was in the past. Let's now discuss the characteristics of fake news.

FAKE NEWS:

It has been shown that fake news can be detected in many different ways. One way to identify fake news and debug it is to check the facts. Fake news usually has many mistakes in grammar. The fake news usually highlights the text with different colours. They will try to emphasize the text or the news and affect the readers' opinion. Text sources will also not always be accurate. As well as their case, URLs may contain special symbols and numeric characters as well as special symbols and numerals, which may also be displayed in the URL. It is not always true that all the news that shows such types of content is true. As well as clicking the highlighted buttons, attackers also attempt to catch the attention of the readers or viewers.

PROPOSED SYSTEM:

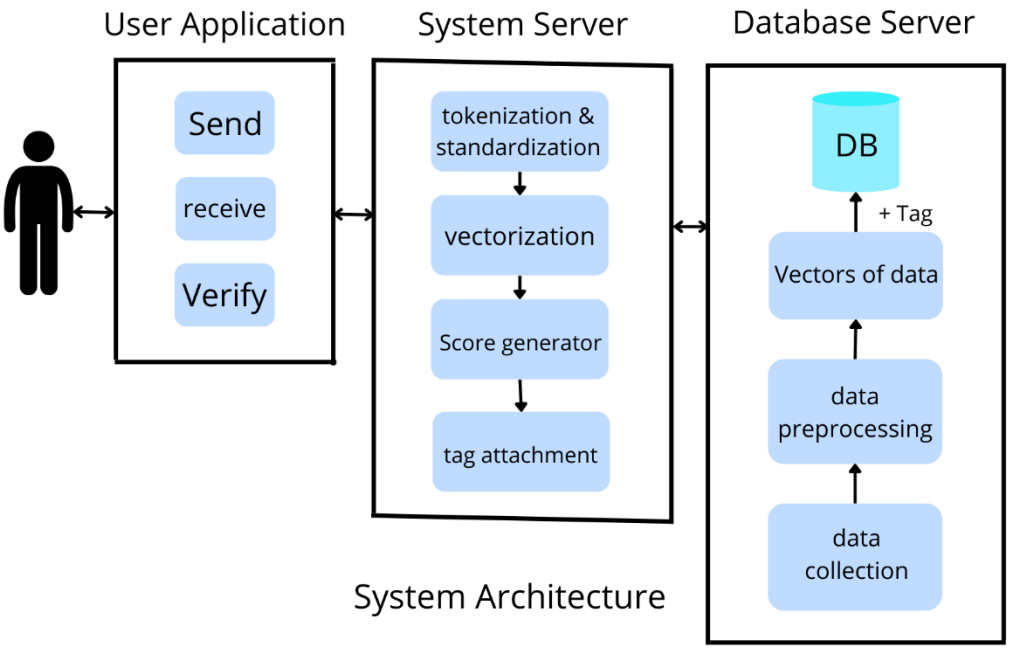
It has been shown that the Naive Bayes algorithm is one of the methods of supervised learning. In this paper, we use Python libraries such as Sci-kit as the main tools. Python is an open-source IDE that you can download and install. Python comes with many extensions and libraries that allow you to work with emerging technology, Machine Learning. Sci-kit libraries contain machine learning algorithms from all over the world. With Django, you can build web pages using HTML, CSS, and Javascript. Implementation on the client-side will be accomplished using Sci-Kit.



Workflow Model

PROCESS FLOW:

As a result of this procedure and methodology, the following steps are involved:
Step 1: First, the dataset must be loaded.
Step 2: The second step includes pre-processing the date from the dataset such as stopping words, tokenizing, etc.
Step 3: Feature selection includes features such as Tf-IDFs and Count Features.
Step 4: Using a classicization model such as the Naive Bayes classifier is step 4.
Step 5: We are going to compare the new data with the old data, and we will determine whether the text is true or false, and find out whether it has a probability score greater than 0. The text will be considered fake news if the probability score exceeds 0 and the text will be considered true otherwise.



FAKE NEWS DETECTION ALGORITHM:

A fake news detection algorithm works as follows: The text is initialized, counted, and tokenized in the document as part of the scoring system. A Naive Bayes algorithm will be used to remove stop words, and common words from the text or paragraph using the algorithm. After detecting quotes, it will eliminate the stop words and common words. As a result, the Naive Bayes classifier will analyze the document and determine the probability and score of the text. Our model is initially trained with the training datasets before it is developed and finally tested with the testing datasets. Depending on the score, the text in the document will be classified as fake news or real news.

ALGORITHM

1 For each doc in Documents

2 pCount <- sigma(number of paragraphs)

3 Tokenize doc by paragraph

4 FakeRank <- 0

5 For each paragraph in doc

6 qScore <- (number of quotes)

7 if qScore > 0 then

8 A-score <- 0

9 For each quoteset

10 quote_cl classify.naivebayes(quoteset_attribution_space, d)

```
11          If quote_cl
12                  A-score <- A-score +1
13          Else
14                  A-score ← A-score-1
15          return A-score
16      FakeRank <- FakeRank + A-score
17 If FakeRank >=0 then
18      docLabel <- real
19 If FakeRank < 0 then
20      docLabel-fake
```

CLASSIFICATION USING NAÏVE BAYES CLASSIFIER:

Based on the Bayes theorem, naive Bayes is a supervised learning algorithm that can be used to solve classification problems. By calculating the number of words, verb count, etc., it predicts using a probability based classifier.

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)}$$

In order to generate the likelihood table, first convert the dataset into frequency tables, then find the probability of each feature given. Finally, calculate the probability.

CONCLUSION:

As of now, the spread of fake news is one of the biggest problems which we face. This paper shows how to detect fake news using this algorithm effectively. In today's world, we are so confused about any news we receive that it is hard to tell whether the news we receive is fake or not. Therefore, this fake news detection system was developed in order to eliminate the issues associated with fake news. Our system uses Processing (NLP) and Machine Learning algorithms like Naive Baye's classifiers to classify news inputs as real or fake and then gives the percentage of news that is detected. The model will be trained initially by taking the use of the appropriate training datasets and then tested by using the appropriate testing datasets. In accordance with the scores and probability calculated from the algorithm, a news story will be deemed either fake or real.

REFERANCES:

[1] M. Balmas, «When Fake News Becomes Real: Combined Exposure to Multiple News Sources and Political Attitudes of Inefficacy, Alienation, and Cynicism," Communic. Res., vol. 41, no. 3, pp. 430 454, 2014.

[2] P. R. Brewer, D. G. Young, and M. Morreale, "The Impact of Real News about "Fake News": Intertextual Processes and Political Satire," Int. J. Public Opin. Res., vol, 25, no. 3,

[3] Quinlan, J. R. (1986)."Induction of decision trees". Machine Leaming.

[4] I. Rish, "An empirical study of the naive bayes classifier," in IJCAI 2001 workshop on empirical methods in artificial intelligence, pp. 41-46, 2001

[5] Tolles, Juliana; Meurer, William J (2016). "Logistic Regression Relating Patient Characteristics to Outcomes"

[6] C. Silverman and J. Singer-Vine, "Most Americans Who See Fake News Believe It, New

Survey Says," BuzzFeed News, 06-Dec-2016.

[7] D. Berkowitz and D. A. Schwartz, "Miley, CNN and The Onion," Journal. Pract., vol. 10, no.1, pp. 1-17, Jan. 2016.

[8] MG Yadav, R Nennuri, " Data Mining based Modern and Advanced Design and Development Applications", International Journal of Pure and Applied Mathematics 119-16, 4651-4658

[9] KS M Geetha Yadav , S Srihitha, C Tejaswi, Clustering of Bigdata in Application Review

Analysis", International Joural of Advanced Science and Technology 29 (special issue)

[10] C. Kang and A. Goldman, "In Washington Pizzeria Attack, Fake News Brought Real Guns,"New York Times, 05-Dec-2016.