Fake News Detection

Akshay Jain
Department of Electronics and Communication Engineering
Maulana Azad National Institute of Technology
Bhopal, India
akshay2014jain@gmail.com

Amey Kasbe
Department of Electronics and Communication Engineering
Maulana Azad National Institute of Technology
Bhopal, India
amey.kasbe@gmail.com

***Abstract* — Information preciseness on Internet, especially on social media, is an increasingly important concern, but web-scale data hampers, ability to identify, evaluate and correct such data, or so called "fake news," present in these platforms. In this paper, we propose a method for "fake news" detection and ways to apply it on Facebook, one of the most popular online social media platforms. This method uses Naive Bayes classification model to predict whether a post on Facebook will be labeled as REAL or FAKE. The results may be improved by applying several techniques that are discussed in the paper. Received results suggest, that fake news detection problem can be addressed with machine learning methods.**

Keywords— Machine Learning; Naïve Bayes Classifier; Web Scrapping.

I. INTRODUCTION

Modern life has become quite convenient and the people of the world have to thank the immense contribution of the internet technology for communication and information sharing. There is no doubt that internet has made our lives easier and access to surplus information viable [3].

But this information can be generated and manipulated by common folks in bulk and the spread of such data is reckless due to the presence of social media. Platforms like Facebook and Twitter have allowed all kinds of questionable and inaccurate "news" content to reach wide audiences without proper monitoring. Social media users bias toward believing what their friends share and what they read regardless of authenticity allow these fake stories to propagate widely through and across multiple platforms and increase their credibility[4]. Google and Facebook have now begun testing out new tools to help users better spot and flag fake news sites. Google is now barring hoax sites from its advertising platform and is testing fact-checking labels in Google News, and Facebook is implementing a new system for users and fact checkers to report suspicious stories [7].

In this domain, computational machine learning algorithms have proven useful where data volumes overwhelm human analysis abilities.

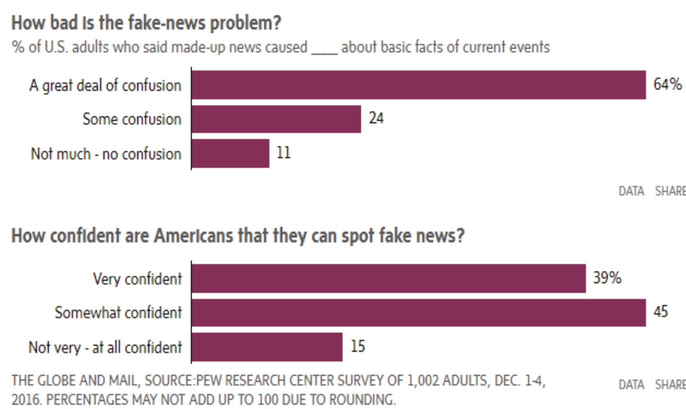

Fig 1: Graphical representation of fake-news problem[7]

This paper describes a simple fake news detection method based on one of the machine learning algorithms – naïve Bayes classifier. The goal of the research is to examine how naïve Bayes works for this particular problem, given a manually labeled news dataset, and to support (or not) the idea of using artificial intelligence for fake news detection. Further, this technique can easily be applied to social platforms like Facebook and Twitter by adding recent news and enhancing the dataset on a regular basis. The difference between this paper and other papers on the similar topics is that in this composition naive Bayes classifier was specifically used for fake news detection; we have tested the difference in accuracy by taking different length of articles for detecting the fake news; also a concept of web scrapping was introduced which gave us an insight into how we can update our dataset on regular basis to check the truthfulness of the recently updated Facebook posts..

Web Scrapping is a technique employed to extract large amounts of data from different websites and to store as desired [6]. This extraction of data is used to withdraw truthful information from reliable sources which in turn will update the dataset in real time. These sources can be news websites which rely on journalistically trained "gatekeepers" to filter out low-quality content.

978-1-5386-2663-4/18/$31.00 ©2018 IEEE

II. NAIVE BAYES CLASSIFIER

In machine learning, naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes theorem [8]. It predicts relationship probabilities for each class such as the probability that given record or data point belongs to a particular class. Naive Bayes classifier assumes that all the features are unrelated to each other. Presence or absence of a feature does not influence the presence or absence of any other feature. It is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle of the number of times a particular event has occurred [1].

Naive Bayes has been studied extensively since the 1950s. It was introduced under a different name into the text retrieval community in the early 1960s and remains a popular baseline method for text categorization which is the problem of judging documents as belonging to one category or the other with word frequencies as the features. With appropriate pre-processing, it is viable in this domain with more advanced methods including support vector machines [9] resulting in improved accuracy.

III. BAYES THEOREM AND ITS MATHEMATICAL IMPLEMENTATION FOR FAKE NEWS DETECTION.

Bayes Theorem works on conditional probability, which is the probability that an event will happen, given that a certain event has already occurred. Using this concept we can calculate the probability of any event based on the likelihood of another event [10].

Below is the formula for calculating the conditional probability.

$$\mathrm{P(H|E)} = \frac{\mathrm{P(\,E\mid H\,) * P(H)}}{P(\,E\,)} \qquad (1)$$

Where,

- P(H) is the probability of hypothesis H being true. This is known as the prior probability.
- P(E) is the probability of the evidence(regardless of the hypothesis).
- P(E|H) is the probability of the evidence given that hypothesis is true.
- P(H|E) is the probability of the hypothesis given that the evidence is there.

The concept we use to classify fake news is that fake news articles often use the same set of words while true news will have a particular set of words. It is quite observable that few sets of words have a higher frequency of appearing in fake news than in true news and a certain set of words can be found in high frequency in true news. Of course, it is impossible to claim that the article is fake just because of the fact that some words appear in it, hence it is quite unlikely to make a system perfectly accurate but the appearance of words in larger quantity affect the probability of this fact [1].

The formula for calculating the conditional probability of the fact, that news article is fake given that it contains some specific word looks as following:

$$P(F|W) = \frac{P(W|F) * P(F)}{P(W|F) * P(F) + P(W|T) * P(T)} \qquad (2)$$

where,

- $P(F|W)$ is the probability that a news article is fake given that word W appears in it.
- $P(W|F)$ is the conditional probability that word W will appear in the texts of fake news.
- $P(F)$ is the probability that the news article will be a fake news article.
- $P(W|T)$ is the conditional probability that the word W will appear in the texts of true news.
- $P(T)$ is the probability that the text will be a true article.

The dispute of calculating the conditional probability of finding a specific word in fake news articles and in true news articles can be resolved such that in a given training set, which contains lots of news articles, labeled as true or fake, one can define the probability of finding a specific word in any fake news article as the ratio of fake news articles, which contain this word to the total number of fake news articles. The probability of finding a specific word in true news articles can be defined similarly [1].

IV. DATASET

The dataset used to test the efficiency of the model is produced by GitHub, containing 11000 news article tagged as real or fake. It has 6335 rows and 4 columns. The 4 columns consist of index, title, text and label. News categories included in this dataset include business; science and technology; entertainment; and health. The authenticity of this dataset lies in the fact that it was checked by journalists and then labeled as “REAL” or “FAKE” [5].

Title involves the minimal information required to understand the news article similar to the heading of the newspaper which describes the content within. Text entails a detailed description of the news article embedded with peculiarities like location, details, people involved and their background etc. Label is basically a tag which tells whether the news articles are ‘Fake’ or ‘Real’.

V. WEB SCRAPPING.

Web Scraping (also termed Screen Scraping, Web Data Extraction, and Web Harvesting) is a technique employed to extract large amounts of data from websites. The data which is extracted from any web source is saved to a local file in our computer or to a database in a table (spreadsheet) format[6]. The data available on these websites do not offer the functionality to save a copy for personal use. The only option then is to manually copy and paste the data - a very tedious job which can take many hours or sometimes days to complete. Web Scraping is the technique of automating this process so that instead of manually copying the data from websites, the Web Scraping software will perform the same task within a fraction of the time [6].

Here we are using web scraping technique to obtain news articles from websites of trusted news agencies. They are labeled as “REAL”. In this way, we can update our database by keeping track of the recently happening events and can also check the truthfulness of the freshly posted content on Facebook using the model. Basically, the use of web scraping is to modernize the dataset we have with newly happening events.

VI. IMPLEMENTATION DETAILS

- Out of the four mentioned fields in the data set first we use title as the only source of information followed by text for using Naïve Bayes classifier.
- Titles of the news articles were retrieved from the dataset and a dataframe was created. .
- Every news article labeled as “REAL” and “FAKE” were tagged as 1 and 0 respectively so as to apply Naïve Bayes classifier.
- The dataset was randomly shuffled, and then it was divided into two unequal subsets: training dataset, test dataset. Training dataset was used to train the naive Bayes classifier. Test dataset was used to get the impartial estimation of how well the classifier performs on new data in terms of AUC score. The training dataset contains 75% of the total dataset while rest 25% was allotted to test dataset.
- Firstly we generate our vocabulary by using the bag-of-words concept. This approach is simple and is a commonly used way to represent text for use in machine learning, which ignores structure and only counts how often each word occurs. CountVectorizer allows us to use the bag-of-words approach by converting a collection of text documents into a matrix of token counts. .
- Next, when we transform our training data, fit our model, make predictions on the transform test data, and compute the AUC score for the title.
- Consider the classification procedure of the naive Bayes classifier. While iterating through the words of the news articles that are being classified, a corner case is possible: some specific words might not be present in the training dataset at all. In such cases, the current implementation just ignores these words stating them as neither Real nor Fake. This serves as a major drawback of this model. But while applying this model to social media platforms like Facebook and Twitter a subtle change is made. If these words were not present in training data which will be obtained using web scrapping technique we will mark it as "FAKE".
- Now for the second model, instead of Title as the primary source implementation of the model is done with Text as the source of vocabulary building. The same process is repeated and we train the model and calculate the AUC score for the Text model.
- One problem with our previous bag-of-words approach is that word order is disregarded. One way we can add some context is by adding sequences of word features known as n-grams. Now for both Title and Text cases, we will transform our training data, fit our model, make predictions on the transformed test data, and compute the AUC score resolving the above-mentioned problem [11].
- Compare the four AUC scores received and determine the best model and parameters suited for better implementation of the model.

VII. RECEIVED RESULTS

We have a probabilistic, a binary classifier logistic regression. Before presenting the ROC curve (Receiver Operating Characteristic curve), the concept of confusion matrix must be understood.

When we make a binary prediction, there can be 4 types of outcomes:

- We predict FAKE while we should have the class is actually FAKE: this is called a True Negative.
- We predict FAKE while we should have the class is actually TRUE this is called a False Negative.
- We predict TRUE while we should have the class is actually FAKE this is called a False Positive.
- We predict TRUE while we should have the class is actually TRUE: this is called a True Positive.

To get the confusion matrix, we go over all the predictions made by the model and count how many times each of those 4 types of outcomes occurs. Since to compare two different models it is often more convenient to have a single metric rather than several ones hence we compute two metrics from the confusion matrix, which we will later combine into one: True positive rate (TPR), aka sensitivity, hit rate, and recall is defined as,

$$\frac{TP}{TP + FN} \qquad (3)$$

Intuitively this metric corresponds to the proportion of positive data points that are correctly considered as positive, with respect to all the positive data points. In other words, the higher TPR, the fewer positive data points we will miss [15]. False positive rate (FPR), aka. Fall-out is defined as:

$$\frac{FP}{FP + TN} \qquad (4)$$

Intuitively this metric corresponds to the proportion of negative data points that are mistakenly considered as positive, with respect to all the negative data points. In other words, the higher FPR, the more negative data points will be misclassified.

To combine the FPR and the TPR into one single metric, we first compute the two former metrics with many different thresholds for the logistic regression and then plot them on a single graph, with the FPR values on the abscissa and the TPR values on the ordinate. The resulting curve is called ROC curve, and the metric we consider is the AUC of this curve [12].

The obtained AUC scores are presented in the table given below:

S.No	DATA	AUC SCORE
1	Title	0.806
2	Text	0.912

Table 1 : AUC scores without n_grams

The obtained AUC score when the concept of n-grams is used are presented in the table given below:

S.No	DATA	AUC SCORE
1	Title	0.807
2	Text	0.931

Table 2 : AUC scores with n_grams

VIII. CONCLUSION

It can be seen from the results obtained from implementation of simple Naïve Bayes shown in Table 1 that the AUC scores have increased when the amount of data existing under a particular tag increases as seen in the case of Title and Text. As Title consisted of smaller version of news articles and Text was a descriptive version of the same.

Now on comparing the results of different methodologies where the concept of n_grams was introduced in the second model we can see that the AUC scores have improved with n_grams as visible from Table 2 because the number of Vectors in the second model have increased hence providing better judgment capacity to the second model.

IX. WAYS OF IMPROVING MODEL

- Use of more data for training purposes - In machine learning problems usually availability of more data significantly improves the performance of a learning algorithm. The dataset, which was described in this article, contains only around 11000 articles [5]. This number is quite small, and a dataset with larger number of news articles from different sources would be of a great help for the learning process as news from different sources will involve larger vocabulary and greater content.

- Use of dataset with greater length of the news articles - The news articles that were presented in the GitHub dataset were not long and contained smaller volume of data, as in the experiment we have proved by using the title and the text case that accuracy is dependent on more number of words or description of the news article. Training a classifier on a dataset with larger news articles should improve its performance considerably.

- Use stemming - In information retrieval, The Porter stemming algorithm (or 'Porter stemmer') is a method of removing the commoner morphological and inflexional endings from words in English [13]. It can be greatly useful since words like win and winning will be treated similar after this process.

- Omitting stop words from the news articles - Stop words are the words, which are common to all types of articles like is, the, am, are, that, they, and etc. These words are so common, that they don't really disturb the correctness of the information in the news article and no significant output can be achieved from them, so it makes clear sense that while performing fake news analysis it is important for us to get rid of these

REFERENCES

[1]. Fake News Detection Using Naive Bayes Classifier by Mykhailo Granik, Volodymyr Mesyura. *Available :* http://ieeexplore.ieee.org/document/8100379/

[2]. Automatically Identifying Fake News in Popular Twitter Threads by Cody Buntain Available: http://ieeexplore.ieee.org/abstract/document/7100738/

[3]. Essay: The Advantages and Disadvantages of the Internet. Available: https://www.ukessays.com/essays/media/the-disadvantages-of-internet-media-essay.php

[4]. Essay: The Impact Of Social Media. Available: https://www.ukessays.com/essays/media/the-impact-of-social-media-media-essay.php

[5]. Dataset acquired. GitHub. Available: https://github.com/chen0040/keras-fake-news-generator-and-detector

[6]. Web Scrapping explanation. Available: https://www.webharvy.com/articles/what-is-web-scraping.html

[7]. Article what is 'fake news,' and how can you spot it? Available : https://www.theglobeandmail.com/community/digital-lab/fake-news-quiz-how-to-spot/article33821986/

[8]. Wikipedia article about Naïve Bayes. Available: https://en.wikipedia.org/wiki/Naive_Bayes_classifier

[9]. A proposed way of implementation. Available: https://www.datacamp.com/community/tutorials/scikit-learn-fake-news

[10]. Study about Bayes theorem. Availabel: http://dataaspirant.com/2017/02/06/naive-bayes-classifier-machine-learning/

[11]. A video lecture about understanding sentiment analysis and the use of n_grams concept. Available: https://www.coursera.org/learn/python-text-mining/lecture/MJ7g3/demonstration-case-study-sentiment-analysis

[12]. Discussin about AUC concept. Available: https://stats.stackexchange.com/questions/132777/what-does-auc-stand-for-and-what-is-it

[13]. Stemming concept. Available: https://tartarus.org/martin/PorterStemmer/