The Seybold Report

ISSN 1533-9211

DOI 10.5281/zenodo.6671081

MACHINE LEARNING ALGORITHMS USED TO DETECT FAKE NEWS

SUDHAKAR .M

Department of Computer Science, Bharath Institute of Higher Education and Research, Chennai, India.
Email: sudhakarmtech@gmail.com

Dr. KALIYAMURTHIE K.P

Department of Computer Science, Bharath Institute of Higher Education and Research, Chennai, India.
Email: kpkaliyamurthie@gmail.com

Abstract

Fake news has been there before the introduction of the Internet. The spreading of fake news will create a severe issue in the world. With the help of technology, information has changed from paper to online. Political people and traders are currently using social media to spread fake news, which will mislead the people and cause many problems in the countries. The purpose of establishing this paper is to detect whether the information is fake or authentic. We used three different machine learning algorithms to achieve this task: logistics regression, Support Vector classifier and Naïve-Bayes and many extraction methods to detect fake or real news. We used the tf-idf extraction method.

Keywords: Fake News, Social Media, Machine Learning, Detection System, Classification, TF-IDF.

INTRODUCTION

These days' fake news is continuously growing, and sometimes it is challenging to find harder and harder what is true and what is fake. After two weeks of the Russia and Ukraine war, many flows of false information happens. Some people began to spread the war is spoof a video of women and men having fake blood applied to their faces and shared on social media, and many people viewed their posts. People have hired them to act so that everyone has mercy on their country. Here the below picture shows an example of fake news. But this video is not related to the war; this video was taken during the Ukraine Tv series in 2020.

Fig: 1 Fake image during Russia and Ukraine war

The Seybold Report
ISSN 1533-9211
DOI 10.5281/zenodo.6671081

The second post shows two Ukraine men holding wooden guns, and it went viral.

Fig: 2 Fake guns were given to civilian

This picture was taken during the training period of volunteers of Ukraine civilians. This type of fake news can create a problem inside and outside the community. The following picture shows that Ukraine's vice president's wife joined the Ukraine armed force to fight against the Russian armed force. This is fake because Ukraine doesn't have a vice president.

Fig: 3 fake pics of Ukraine VP

The graph shows the Google search trends of fake news about the Covid-19 pandemic—the misleading information spreading everywhere. Due to Covid-19, mass death happened in Iron, but this fake news was shared. During Covid-19, social media content was analyzed from 87 countries and found much misleading information shared on social media. WHO guided the public in reporting the fake news about the Covid-19? The government and other organizations in the world taken an effort to show the people how to handle or identify fake news.

ISSN 1533-9211

DOI 10.5281/zenodo.6671081

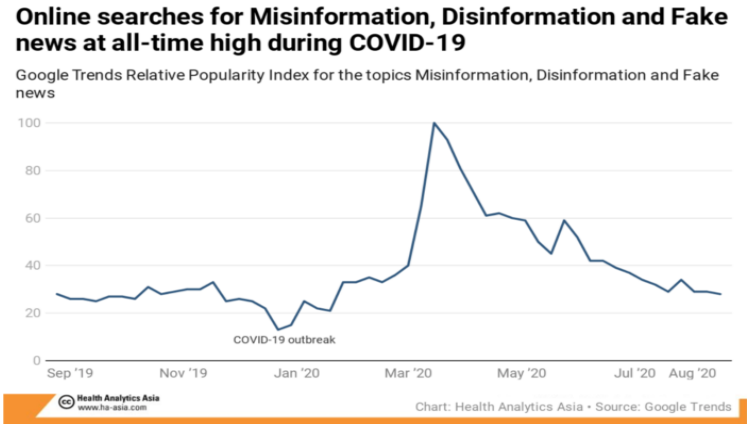


Fig: 4 Google search fake news about Covid-19.

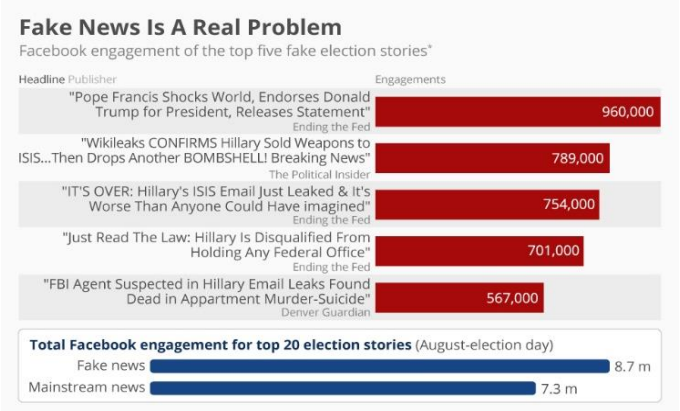


Fig: 5 Engagement is measured based on shares, comments, likes

RELATED WORK

Detecting Fake News in Social Media Network

Fake news and hoaxes have been there since before the advent of the Internet. The authors suggest that the most widely agreed-upon definition of fake news is that counterfeit or twisted journals, articles, blogs, and news are generated to purposefully confuse the readers or the audience between what is right and what is wrong. The authors[2] focus on profiting through clickbait. Clickbait are thumbnails that attract users and entice curiosity with flashy headlines or designs to click links to increase advertisements revenues. In this research, the authors analyze the prevalence of fake news considering the advances in communication made possible by the drastic rise of social networking sites. Their work aims to develop a solution that users can utilize to detect and filter out sites containing false and misleading information. In conclusion, they use simple and carefully selected title and post features to identify fake posts accurately. The experimental results show a 99.4% accuracy using the logistic classifier.

The Seybold Report
ISSN 1533-9211
DOI 10.5281/zenodo.6671081

Detecting Fake News using Machine Learning and Deep Learning Algorithms

Social media provides a platform to broadcast news to the network exponentially and is an excellent source of information. But this does not imply that everything we see today in the transmission or social media is true. Words, photos, and videos are "photoshopped" into being delusional or subjective to their propaganda. As a standout amongst the most well-known ongoing news sources, Twitter also ends up among the most dominant news radiating mediums. This paper[3] proposes a model for recognizing forged news messages from Twitter posts by figuring out how to anticipate precision appraisals because of computerizing forged news identification in Twitter datasets. As a result, the authors compared five well-known Machine Learning algorithms, like Support Vector Machine, Naïve Bayes Method, Logistic Regression and Recurrent Neural Network models, separately to demonstrate the efficiency of the classification performance on the dataset. In conclusion, results obtained from different learning algorithms were compared, and the result showed that SVM and Naïve Bayes classifier outperformed the other algorithms.

Which machine learning paradigm for fake news detection?

It is essential to understand the sense of realization of right and wrong, the When, what and why of any news in this social era. Fake news detection/classification is gradually becoming of paramount importance to society to avoid the so-called reality vertigo and protect the less educated persons. Various machine learning techniques have been proposed to address this issue. This article presents a comprehensive performance evaluation of eight machine learning algorithms for fake news detection/classification.

Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques

Fake news significantly impacts our socio-economic life, especially in the political world. Counterfeit news detection is an emerging research area gaining interest but involves some challenges due to limited available resources (i.e., datasets and published literature). In this paper, the author utilizes a fake news detection model that uses n-gram analysis and machine learning techniques. As a result, the author investigates and compares two different feature extraction techniques and six different machine classification techniques. Experimental evaluation [4] yields the best performance using Term Frequency-Inverted Document Frequency (TF-IDF) as a feature extraction technique and Linear Support Vector Machine (LSVM) as a classifier, with an accuracy of 92%.

METHODS USED TO DETECT THE FAKE NEWS

Figure 6 shows the entire workflow of the detection of fake news. All the experimental data was collected from Twitter in the selected areas. We used normalization to manipulate the retrieved raw data. The next step we used was to remove the replication of data. In the final step, we used the machine learning method for classification.

ISSN 1533-9211

DOI 10.5281/zenodo.6671081

The primary source of our data is coming from Twitter. We have given and created a Twitter developer account; after that, we have to start posting on Twitter. We can download many tweets based on the keywords, Id, name and text; only the user verified from Twitter. Twitter will take a long process to verify the user, and if the user is confirmed, it will show as the blue tick on the user profile. Our dataset was collected from the message 948373. Twitter has two types of user accounts; the first one is the older version which consists of two character attributes it has 9 to 10 digits, and the second one is a new account with 18 digit numbers. We have to convert verified user columns to labels. If the user is confirmed, it will indicate 0; otherwise, it will display 1.

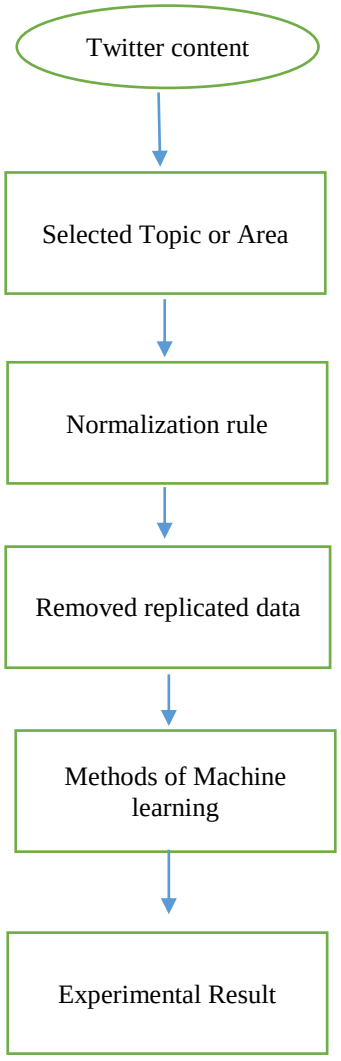


Fig. 6 Methodology

Here punctuations are removed, and all of the datasets are converted into lowercase letters. Various punctuations and emotions users used in the tweets it will not help in the detection

ISSN 1533-9211

DOI 10.5281/zenodo.6671081

methods. Empty columns have been removed. Repeated post, characters, tags, and URL has been removed from our dataset.

Once our dataset is ready, it will give us a clear understanding and visualization of the graphs. This graph will show us a clear understanding of our dataset's categorization.

Here we used the tokenization function. This one will help us separate more important body messages into smaller words.

```
0    [house, dem, aide, we, didnâ, t, even, se, com...
1    [ever, get, the, feling, your, life, circles, ...
2    [why, the, truth, might, get, you, fired, octo...
3    [videos, 15, civilians, kiled, in, single, us,...
4    [print, an, iranian, woman, has, ben, sentence...
Name: text, dtype: object
```

Url will describe the location of the destination message, or it will not like the destination message. Maybe this Url doesn't have value or has one or more values. Four types of messages will show to us here,

1. True Positive is the number of correct messages classified as believable messages.
2. True Negative is the number of correct messages classified as unbelievable messages.
3. False Positive is the number of incorrect messages classified as believable messages.
4. False Negative is the number of incorrect messages classified as unbelievable messages.

TF-IDF technique will help us to compute the weight of each word. This technique is primarily used in the retrieval of information and text mining process. With the help of TF, we can summarize how often a given text appears in the given document. We are not going to use the entire dataset for further analysis. Machine Learning Algorithms,

- Logistic Regression

This algorithm is used to predict the probability and categorize the dependent variable, and this dependent variable is based on binary. There are two types of binary variable codes 0 and 1. If the code is 1 means success and 0 means failure or no. It is a linear regression model used in the cost function and can be defined as a sigmoid function. The cost function is between 0 to 1, so linear functions failed to represent the value as per the logistic regression hypothesis. So this is classified into binomial, multinomial and ordinal.

$$\mathbf{0 \leq h\Theta(x) \leq 1}$$

Equation: 1 logistic regression hypothesis

- Naïve Bayes

It is a simple probabilistic classifier model, and it is a simple technique for constructing a classifier model that assigns class labels to vectors of feature values and problem instances. Based on the calculated overall probability, we can get the approximate value and detect whether the news is fake or accurate.

The Seybold Report
ISSN 1533-9211
DOI 10.5281/zenodo.6671081

- P (A|B) = P (B|A) · P (A) / P (B), (1)
- P (A) = PRIOR PROBABILITY
- P (A|B) = POSTERIOR PROBABILITY FINDING PROBABILITY:
- P (A|B1) = P (A1||B1). P (A2||B1). P (A3||B1) (2) P (A|B2) =P (A1||B2).
- P (A||B2). P (A3||B2) (3)
- If the probability is 0
- P (Word) = Word count +1/ (total number of words+ No. of unique words)

- Long Short-Term Memory

To learn long-term dependencies, we can use RNN based particular type of LSTM. LSTM will provide the best solution for the vanishing gradient problem. LSTM cell is replaced by LSTM-RNN hidden layer. It preserves the error that can be backpropagated based on time and lower layer. The input will be in both directions simultaneously in bi-directional LSTM.

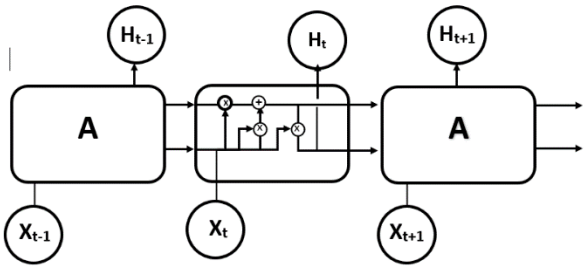


Fig. 7 LSTM Architecture

- Support Vector Machine

It will provide an accurate classification of linear data. If the given data is in non-linear form, we can use the data with the help of the kernel trick, and we can avoid complex transformations into a linear model. This algorithm will develop a hyper plane. The hyper plane in N-dimension space is dependent on the dimension of input fed during the training period of the model. The hyper plane will provide a boundary line between different groups. The line will also help us point the maximum distance from the data points of each group and categorize the new input.

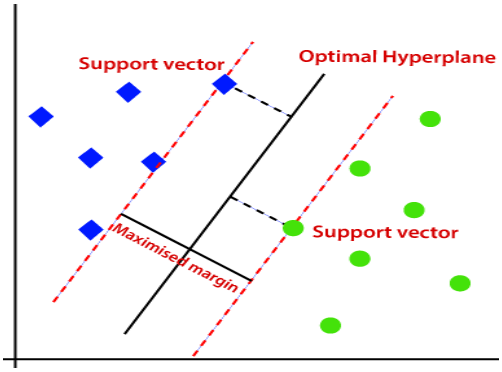
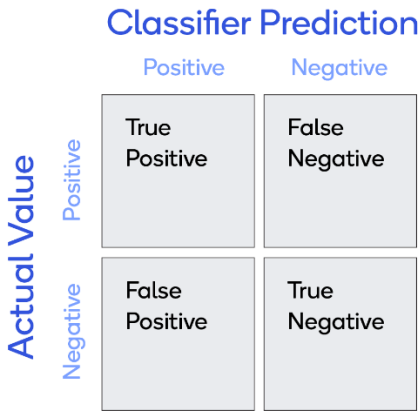


Fig. 8 SVM

The Seybold Report
ISSN 1533-9211
DOI 10.5281/zenodo.6671081

Training & Testing Model

The data must go for training once the network builds for each algorithm.



Fig; 9 Confusion matrix

We trained the algorithm to learn features from the dataset. Later, testing is performed to evaluate the performance of our built model using performance metrics, accuracy, precision score and recall.

RESULTS

We analyzed tweets based dataset. The dataset was extracted from Twitter and got the results. The reviews can be divided into two-part, one is reliable, and another is unreliable based on the text information. We used real-time tweets using some keywords to preprocess the tweets that we extracted from the dataset. Logistic regression is the first algorithm we used to perform the dataset. This one is the baseline model for machine learning and big data. This model is less prone to overfitting. This method will provide 95.2%, as we expected.

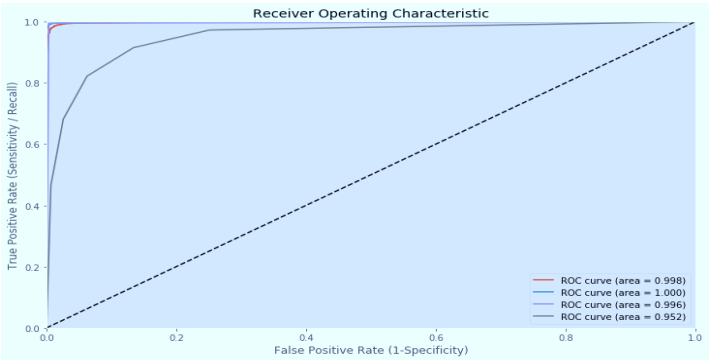


Fig. 10 Logistic Regression ROC curve

The Seybold Report
ISSN 1533-9211
DOI 10.5281/zenodo.6671081

Naïve Bayes is the following algorithm that we used to experiment with the dataset. Here we considered each variable independently, and its performance was better than expected. Unfortunately, due to zero frequency [7], the classifier achieved only 73.0%. Fig. 11 is the description of our ROC curve for this model.

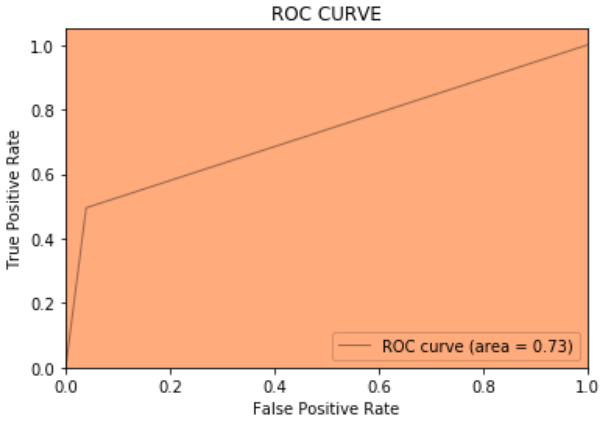


Fig. 11 Naïve Bayes ROC Curve

Natural networks always have an issue with large datasets and storage. We used Long, Short-Term Memory to overcome this problem to avoid this error. It uses memory blocks and gates for its functionality. Early Stopping was also used to restrict the model from overfitting. After running for five epochs, we got an accuracy of 65.0%, the lowest among all our classifiers. This indicated that this type of data doesn't need neural networks, and it should be trained using a less complicated structure. Following is the figure that depicts the ROC curve of our LSTM model.

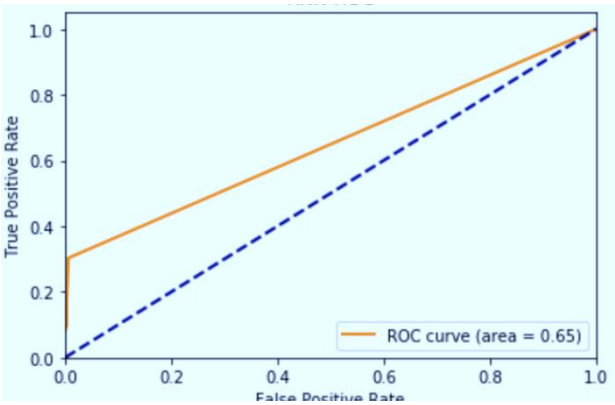


Fig. 12 ROC Curve for Long Short-Term Memory

The latest algorithm which was performed on tweets to detect fake news was the Support Vector Machine. It uses deciding boundaries to separate two classes with the most considerable margin, also called the best hyperplane. Linear, SVM method was used to

ISSN 1533-9211

DOI 10.5281/zenodo.6671081

analyse the data extracted. SVM achieved a staggering 98% accuracy which is the second-best accuracy achieved after logistic regression. Based on considerable research, SVM works wonders with linear and non-linear data and is more reliable in giving higher accuracy results. Below fig. 13 depicts one of the best ROC curves in our analysis.

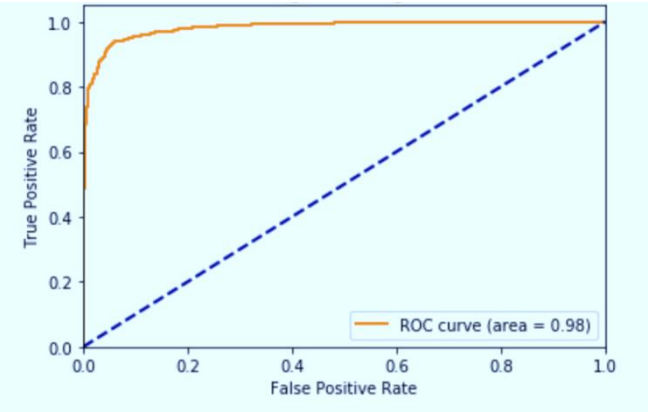


Fig. 13 ROC Curve for Support Vector Machine

Algorithms	Precision	Recall	F1 Score	Accuracy
Logistic Regression	95.2%	95.2%	95.2%	95%
Naïve Bayes	79%	73%	71%	74%
Long Short-Term Memory	65%	62%	35%	54%
Support Vector Machine	98%	98%	98%	98%

Comparison Results

CONCLUSION

Online social network is widespread of fake news for various purposes. This one will create more problems worldwide, affect the country's economy, change the political effects everywhere, and create panic among the people in the nation. To detect and avoid this type of fake news, we need a robust algorithm. The models present there have manually labelled each tweet are real or fake news. This method will take a long time to predict whether the information is fake or real. We solved this problem; we collected the actual tweets and then applied the preprocessing of the data we collected. We used four machine learning algorithms to detect whether the news is fake or real. As a result, two algorithms provided a better solution for it. Logistic regression provided 95% accuracy, and SVM provided 98%. Many social media companies fight against fake news.

The Seybold Report
ISSN 1533-9211
DOI 10.5281/zenodo.6671081

REFERENCES

[1] D. Katsaros, G. Stavropoulos and D. Papakostas, "Which machine learning paradigm for fake news detection?," 2019 IEEE/WIC/ACM International Conference on Web Intelligence (WI), Thessaloniki, Greece, 2019, pp. 383-387.

[2] Abdullah-All-Tanvir, E. M. Mahir, S. Akhter and M. R. Huq, "Detecting Fake News using Machine Learning and Deep Learning Algorithms," 2019 7th International Conference on Smart Computing & Communications (ICSCC), Sarawak, Malaysia, Malaysia, 2019, pp. 1-5.

[3] M. Aldwairi and A. Alwahedi, "Detecting fake news in social media networks," Procedia Comput. Sci., vol. 141, pp. 215–222, 2018.

[4] H. Ahmed, I. Traore, and S. Saad, "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2017.

[5] Google Trends. 2020. Google Trends. [online] Available at: <https://trends.google.com/trends/?geo=US> [Accessed 1 April 2020].

[6] Help.twitter.com. 2020. About Verified Accounts. [online] Available at: <https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts>.

[7] Harder, Better, Faster, Stronger. 2020. The Zero Frequency Problem (Part I). [online] Available at: <https://hbfs.wordpress.com/2014/09/23/the-zero-frenquency-problem-part-i/>.

[8] G. Cauwenberghs and T. Poggio, "Incremental and decrementai support vector machine learning," Adv. Neural Inf. Process. Syst., no. x, 2001.

[9] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, "Truth of varying shades: Analyzing language in fake news and political fact-checking," EMNLP 2017 - Conf. Empir. Methods Nat. Lang. Process. Proc., pp. 2931–2937, 2017

[10] Lewis D.D. (1998) Naive (Bayes) at forty: The independence assumption in information retrieval. In: Nédellec C., Rouveirol C. (eds) Machine Learning: ECML-98. ECML 1998. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), vol 1398. Springer, Berlin, Heidelberg.

[11] O. Aborisade and M. Anwar, "Classification for Authorship of Tweets by Comparing Logistic Regression and Naive Bayes Classifiers," 2018 IEEE International Conference on Information Reuse and Integration (IRI), Salt Lake City, UT, 2018, pp. 269-276.

[12] Joshkelliott. (2020, March 24). One hantavirus death in China sparks 'hysteria' over old disease. Retrieved from <https://globalnews.ca/news/6724399/hantavirus-china-death-coronavirus/>.