International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
Impact Factor (2018): 5.426

Detection of Fake News Using Binary Classification

Abhyudya Bajpai[1], Amit Kumar Tripathi[2]

[1]B.Tech in CSE, SRM Institute of Science and Technology, Delhi-NCR, India, abhyudyabajpai10[at]gmail.com
[2]B.Tech in CSE, SRM Institute of Science and Technology , Delhi-NCR, India, amitkumartripathi9887[at]gmail.com

Abstract: ***The idea behind this project is to detect the accuracy of the fake news using Binary Classification such as Multinomial Naïve Bayes, Passive Aggressive classifier. Here the two datasets are provided i.e., test dataset and train dataset. Test data is later matched with groups of train dataset and accuracy is found using Binary classification. This helps in determining whether given news is fake or real. It delivers maximum accuracy and helps to identify fabricated news. The data is pruned by removing stop words and common English words by using vectorizer.***

Keywords: Fake news, Binary Classification, Multinomial Naïve Bayes algorithm, Passive Aggressive Classifier algorithm, outliers, TFIDF Vectorizer

1. Introduction

In the modern world where an individual can get live updates on what is happening round the globe in less than a minute, it is hard to rely on the authenticity of the news. People often get misled and become victim to fake or deceptive news. Sometimes this counterfeit news leads to rage and anger between groups of different belief. Today any individual can post news on social media platforms like twitter or Facebook communities and other microblogging websites. Websites specifically target the audience based on their likings, a person having interest in sports will receive sports news more often than a person who has a keen interest in politics. These websites bombard individual's news feed column with news inclined to respective interests. The circulation of fake news catches more fire than the real news. Several instances were noted where fake news caused havoc among the crowd. If we look at the news which went viral at the time of demonetization in India, people were told that newly printed currency notes will be equipped with some geo-location chip to track the circulation of the money in the market. This rumor looked so convincing that it caused confusion even for the legit sources. Recently, misinformation regarding coronavirus (COVID-19) pandemic is circulating in the form of social media messages about unverified home remedies, fake and deceptive advisories and conspiracy theories. Several arrests were made involving people spreading fake and unverified news. On 7 March 2020, Prime Minister of India Mr. Narendra Modi appealed to citizens not to believe any rumors related to the residing pandemic. The Press Information Bureau conducted a fact check on 24 March 2020 that stories from anonymous sources about a financial emergency being imposed in the country are fake. To counteract this measure, around 400 Indian Scientists worked together to debunk the misleading information about the virus. Various growing news channels hunt for publicity and forget about their real value i.e., delivering actual facts.

The main objective of this research paper is to detect the deceptive news, that follows the classic text base classification. A machine learning model is to be prepared that classifies the news into two broad categories namely "Real" and "Fake" news respectively. The textbox is to be provided upon implementation, the news or the url can be put in the textbox whose authenticity is to be examined. The proposed machine learning model delivers the legitimacy of the news. As a result, the user is informed about the type of the news as well as the percentage accuracy of the given news.

2. Literature Survey

In the paper by Syed Ishfaq Manzoor, they studied how news can be differentiated between true and false. This paper reviews various Machine learning algorithms as well as Deep Learning approaches that are applied in detection of fake and fabricated news. They categorized the forms in which the deceptive news is spread; it can range from using chatbots to spread false information to click baits used for spreading rumors. This paper also uncovers the challenge that the ever changing characteristics and features of fabricated news is posing in social media.[1]

In the paper, the authors have classified every post on twitter as binary classification Problem. The Classification is purely on the basis of anonymity of source of the post/tweet. The Authors used comparatively small manually sampled data sets using twitter API. It was concluded with 15 percent news identified as fake, nearly 45 percent real and rest was undecided or unclassified. Even though the model used SVM and Random Forest algorithms, the efficiency was not compelling. [2]

The authors in the paper have listed several methods of detection of fake news. Various Detection Methods introduced:

1. Linguistics basis

Deception modeling

2. Clustering
3. Predictive modelling
4. Content cue based methods

The authors have acquired accuracy of these models ranging from 63 to 70 percent only.[3]

The authors in [9] reviewed and analyzed about 14 million posts/messages that were retweeted around 400 thousand times on Twitter at time of the 2016 U.S. Presidential campaign later came to be known as election by bots. This had a major influence in driving the US citizens in choosing the right and deserving candidate for the President run. The methods to classify the tweets massively spread by bots were described.

According to the authors of [6] paper the fabrication of information and misleading information spreading is not new. The authors have provided the reference of columnist of Guardian Nougayrède who quoted "The use of propaganda is ancient, but never before has there been the technology to so effectively disseminate it". Counterfeited information, deceptive information as well as propaganda-based information has been prevailing features since Roman times. The authors in [6] studied in detail the principles, methodologies and algorithms employed for classification of falsified news articles, authors and papers from social media networks and estimating the corresponding reach and performance. The paper also carefully mentioned the research challenges and hurdles encountered in the undiscovering of characteristic features of fake news and wide connections among news published in various online forums, authors and subjects. The Authors of the research paper role out concept of automatic fake news inference model termed as FakeDetector which is based on textual classification that builds a model deep diffusive network to learn how the news articles are represented, authors and subjects in the time. FakeDetector encompasses two major components: feature representation learning, and credibility label inference, together that composes deep diffusive network model named FakeDetector.

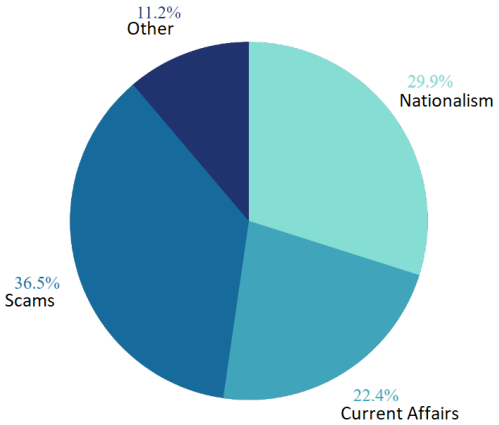
The Authors in [7] have discussed fake news that prevails in click baiting format in a tabloid manner. They managed to explain that the rapid source of rumor and misinformation online is click baiting. The authors have explained potential ways to automatically detect fake news disguised as clickbait in the form of deception. Content cues that contains semantic and lexical level of analysis were successfully implemented by the authors.

Last year tech giants like Twitter and Facebook came forward to stop the spread of misinformation regarding untested remedies and conspiracy theory behind the novel corona virus involving high profile figures like Microsoft's CEO Bill Gates. Twitter's new labels now provide links to additional information in cases where it is plausible that people could be confused or misguided, the company said in a recent blog post. Warnings might also get added in case of a tweet conflicts under guidance from public health experts before a user accesses it. A similar approach is taken by Facebook, third-party fact-verifying partners that include Reuters, debunk and rate content that goes viral on the site with certain labels. Lastly, YouTube also announced that it would start showing information view panels with third-party, fact-checked posts and articles for US related video search results.

Recently, BBC conducted a research in India and surprisingly found that the elevating slope of nationalism in the Eastern country is a major and most crucial factor of fake news in public forums, and vice versa. Crumbling trust in mainstream media news has propelled people to spread stories/articles from alternative sources, which are often from untrustworthy sources.

Around eighty participants from Kenya, Nigeria and India allowed BBC to access their mobile phones for a week-long period during which the researchers investigated participants' media utilization as well as how they shared news articles via WhatsApp and Facebook.

Percentage of messages shared on WhatsApp by topic



3. Methodology

We cannot train the text present in the dataset directly when working with machine learning algorithms. The reason why we need to change from text to numbers is because, machines as advanced as they may be are not capable enough of interpreting words and sentences in the same way as we humans do, in order to make documents palatable for computers, they must first be converted into some numerical structure or representation. We may want to perform specification of data present in dataset, so each item is an "input" and class label as "output" for our predictive algorithm. Algorithms prefer working with vectors of numbers as input; therefore dataset must be converted into fixed-length vector input of integers. A simple yet an effective model that can be used to convert the data in machine learning is known as the Bag-of-Words Model. The model is quite easy to use as it does not care about the order of information in the words rather focuses on the count of occurrence of words in a file.

This is achieved by assigning a unique number to each word. After that the document we have can be encoded as a fixed-length vector with the vocabulary of known words as the length. A count or frequency of each word in the encoded document can be used to fill the value in each position in the vector. The model where we are concerned only with encoding schemes that represent what words or

the degree to which they are present in the encoded documents without any information about order is called the bag of words model.

The proposed machine learning model uses two techniques to convert the news into machine meaningful vectored array of numbers.

3.1 CountVectorizer

The CountVectorizer lays out a smart way to both tokenize by creating a collection of text files and constructs a vocabulary of known words; it encodes new documents using that vocabulary. CountVectorizer is a helpful tool provided by the scikit-learn library. It is used to convert a given script into a vector array on the account of the frequency (count) of each word that comes in the whole text body. The value of each cell comprises of the count of the word in that particular text sample. Inside CountVectorizer, each word is given a particular index value.

3.2 TFIDFVectorizer

Word counts in CountVectorizer is a great way to parse the dataset, but it is less efficient at some places. One main issue residing with word count is that few words like "the" can appear multiple times and their large occurrences will make the vector array little less meaningful and this might affect the efficiency of the model.

This problem can be dealt by calculating word frequencies, and till date the most effective method is called TF-IDF. The acronym stands for "Term Frequency - Inverse Document" Frequency that are the integral components of the final scores allocated to each word.

- **Term Frequency**: This analyzes how many times a given word appears within a document.
- **Inverse Document Frequency**: This tabulates and confronts words that appear most in the documents.

In a layman term TF-IDF are word count scores which try to map out words that are more interesting, e.g., frequent in a document but not across documents.

4. Algorithm and Implementation

4.1 Approach

The complexity of the fake news can vary depending upon the topic, this makes the classification of fake news difficult at hand. With continuous enhancement in the machine learning algorithm nothing stays for a long time, the algorithm today might solve the issue but as the data gets enormously large algorithm at hand becomes less effective and need of new methodology becomes necessary. The only way to deal with data complexity with time is to upgrade the algorithms on the fly. There are several algorithms to use for the purpose of detection of fabricated news like Naïve Bayes, Support Vector Machines, Logistic Regression, Decision Tree Classifier, Random Forest Classifier. After careful efficiency analysis of the mentioned models in the literature survey a model is been selected with two highly efficient algorithms i.e., Multinomial Naïve Bayes classifier and Passive Aggressive Classifier for the prediction of fake and real news.

4.1.1 Naïve Bayes

Naïve Bayes Classifier is a supervised learning algorithm that uses Bayes' theorem along with the assumption of conditional independency between each pair of features provided the value of the class label. It works on the probability of the variable.

$$P((X|Ci) = \prod P(xk|Ci) = P(x1|Ci) \times P(x2|Ci)\ n\ k=1 \times \ldots \times P(xn|Ci)$$

The classification is made by taking out the maximum posterior, the maximal P(Ci|X) on applying the above assumption according to Bayes theorem. This hypothesis significantly drops the computational cost by counting just the class distribution. The Classifier has proven efficient in many real-world scenarios, mostly document classification and spam filtering. The Classifier requires comparatively small training data to construct the necessary arguments. Naive Bayes classifiers can be at times amazingly fast compared to other sophisticated algorithms. The decoupling of the conditional feature distributions class implies that each distribution can easily be independently estimated as a single dimensional distribution leading to the reduction of curse of dimensionality issues.

4.1.2 Passive Aggressive Classifier

The Passive-Aggressive machine learning algorithms are a kind of algorithms which are lesser known by beginners and even to some intermediate Machine Learning enthusiasts. Though, it can be very useful and efficient for certain real-world situations. Passive-Aggressive Classifiers are often used where large-scale learning is required. It is among few of the '**online-learning algorithms**'. In online machine learning designs, the input data is added in a sequential order with machine learning model updating step-by-step, in contrast to batch learning, where the whole training dataset is provided at once. This is very useful in scenarios where there is enormous amount of data in the dataset and it becomes computationally infeasible to train the entire dataset due to the large size of the data. Online learning algorithm works in a way that it gets a training dataset, updates the classifier, and then throw away the dataset.

This application is useful in predicting fake news from a data set that continuously keeps expanding such a social media platforms Twitter and Facebook, using an online-learning algorithm would be suitable here. Passive-Aggressive Classifier is related to a Perceptron model, in the manner that they do not require a learning rate. Although, they do need a regularization argument.

Passive Aggressive Algorithm is made up from two words mainly:

- **Passive:** When the prediction is found to be correct and there is no need to the change the data in the dataset model.
- **Aggressive:** When the prediction is found to be incorrect, some changes are bound to be made in the model to make the prediction correct.

Passive Aggressive Classifier parameters:

- C : It denotes the regularization parameter, and adds a penalty to the model when it makes an incorrect prediction.
- max_iter : It indicates the maximum number of iterations the model will make on the training dataset.
- tol : This is the stopping criterion. When set to None, the model stops if this becomes true (loss> prior_loss-tol).

4.2 Workflow

The model first extracts the data from the dataset and cleans the data that is to be trained. Later CountVectorizer and TFIDFVectorizer is used to convert the text data and vector array of numbers. The dataset is split into train dataset and test dataset in the ratio of 1:4. The model is trained through machine learning algorithms i.e., Multinomial Naïve Bayes and Passive Aggressive Classifier.

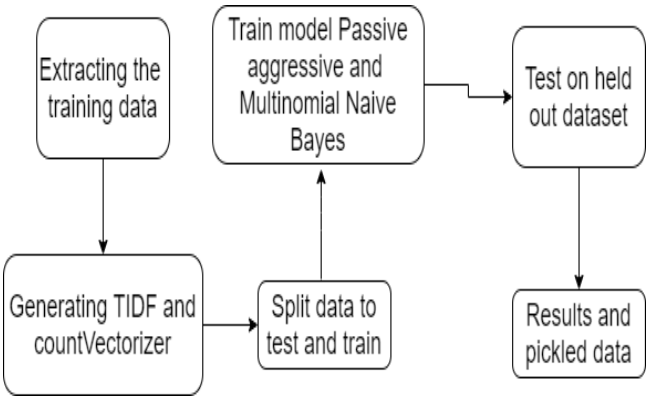


Figure 1: Model workflow

Confusion matrix is plotted for the same. The accuracy is calculated for both the algorithms. The algorithm with highest accuracy is ideal.

S.No.	Comparing Different Methodologies used earlier		
	Article	**Accuracy**	**Methodology used**
1	R. V. L, C. Yimin, and C. N. J (2016)	76%	NLP
2	Helmstetter, S., & Paulheim, H (2018)	77%	SVM
3	Y. Seo, D. Seo, and C. S. Jeong (2018)	86.55%	CNN
4	**Jain A., Khatter H., Shakya A. (2019)**	**93.50%**	**Naïve Bayes, SVM**

5. Conclusion

The model is successfully able to classify fake and real news with 0 and 1 label respectively. The accuracy of the model with Multinomial Naïve Bayes algorithm was found 94.5% and with Passive Aggressive the accuracy was further enhanced and clocked 99.5%. Though the spread of fake news through different channels will not steer down but preventive measures can be taken to avoid oneself from being misguided. In the near future esteemed news organizations and the government are likely to come together to counteract the negative influence these fabricated news cause to innocent readers. Fake news will persist as long as there is no governing body to verify the news from its source. The truthfulness can be measured on various grounds as to which crowd is the news addressed to and is it published in more than one renowned newspaper organization. In the study it was found that most of the fake news present in the datasets was political. Therefore, it can be concluded that there are chances of coming a across a fake news article which is related to politics. Efficiency of the model can further be improved by training the model with dataset holding latest news.

References

[1] Syed Ishfaq Manzoor,Jimmy Singla,Nikita,"Fake News Detection Using Machine Learning approaches: A systematic Review" , IEEE 3rd International Conference on Trends in Electronics and Informatics (ICOEI), April 2019
[2] Helmstetter, S., & Paulheim, H ,"Weakly supervised learning for fake news detection on Twitter", International Conference on Advances in Social Networks Analysis and Mining (ASONAM), August 2018
[3] Parikh, S. B., & Atrey, P. K., "Media-Rich Fake News Detection: A Survey", IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), April 2018
[4] Chaitra K Hiramath;G. C Deshpande,"Fake News Detection Using Deep Learning Techniques" IEEE 1st International Conference on Advances in Information Technology (ICAIT), Feb 2020.
[5] Della Vedova, M. L., Tacchini, E., Moret, S.,Ballarin, G., DiPierro, M., & de Alfaro, L., "Automatic Online Fake News Detection Combining Content and Social Signals", 22nd Conference of Open Innovations Association, May 2018.
[6] Zhang, J., Cui, L., Fu, Y., & Gouza, F. B. (2018). Fake news detection with deep diffusive network model. arXiv preprint arXiv:1805.08751.
[7] Chen, Y., Conroy, N. J., & Rubin, V. L. (2015, November). Misleading online content: Recognizing clickbait as false news. In Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection (pp. 15-19). ACM.
[8] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436.
[9] Shao, C., Ciampaglia, G. L., Varol, O., Flammini, A., & Menczer, F. (2017). The spread of fake news by social bots. arXiv preprint arXiv:1707.07592, 96-104.
[10] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, “The rise of social bots,” Communications of the ACM, vol. 59, no. 7, pp. 96–104, 2016.

[11] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017
[12] Tacchini, E., Ballarin, G., Della Vedova, M. L., Moret, S., & de Alfaro, L. (2017). Some like it hoax: Automated fake news detection in social networks. arXiv preprint arXiv:1704.07506
[13] Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. Journal of Big Data, 2(1), 1.
[14] Haiden, L., & Althuis, J. (2018). The Definitional Challenges of Fake News
[15] Yash Khivasara; Yash Khare; Tejas Bhadane. (2020). Fake News Detection System using Web-Extension. 2020 IEEE Pune Section International Conference (PuneCon).