Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 165 (2019) 377–383

Procedia Computer Science

www.elsevier.com/locate/procedia

INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING 2019, ICRTAC 2019

Analysis of Classifiers for Fake News Detection

VasuAgarwal[a], H.ParveenSultana[a] ,SrijanMalhotra[a] *, AmitrajitSarkar[b]

[a] Vellore Institute of Technology, Vellore, 632014, India
[b] Ara Institute of Canterbury, New Zealand

Abstract

As time flows, the amount of data, especially text data increases exponentially. Along with the data, our understanding of AI also increases and the computing power enables us to train very complex and large models faster. Fake news has been gathering a lot of attention worldwide recently. The effects can be political, economic, organizational, or even personal. This paper discusses the approach of natural language processing and machine learning in order to solve this problem. Use of bag-of-words, n-grams, count vectorizer has been made, TF-IDF, and trained the data on five classifiers to investigate which of them works well for this specific dataset of labelled news statements. The precision, recall and f1 scores help us determine which model works best.

Keywords: AI; Classifiers;Fake news; Machine learning; NLP;Prediction

Corresponding Author: Tel:- +91 9787111720

Email :srijan.malhotra@yahoo.in

1877-0509
10.1016/j.procs.2020.01.035

1. Introduction

The term “Fake News” was a lot less unheard of and not prevalent a couple of decades ago but in this digital era of social media, it has surfaced as a huge monster. Fake news, information bubbles, news manipulation and the lack of trust in the media are growing problems within our society. . However, in order to start addressing this problem, an in-depth understanding of fake news and its origins is required. Only then one can look into the different techniques and fields of machine learning (ML), natural language processing (NLP) and artificial intelligence (AI) that could help us fight this situation.“Fake news” has been used in a multitude of ways in the last half a year and multiple definitions have been given. For instance, the New York Times defines it as “a made-up story with an intention to deceive”.Measuring fake news or even defining it properly could very quickly become a subjective matter, rather than an objective metric.In its purest form, fake news is completely made up, manipulated to resemble credible journalism and attract maximum attention and, with it, advertising revenue [2]. Despite all these shortcomings, several entities have tried to categorize fake news in different manners.

1.1. Motivation

The widespread problem of fake news is very difficult to tackle in today’s digital world where there are thousands of information sharing platforms through which fake news or misinformation may propagate. It has become a greater issue because of the advancements in AI which brings along artificial bots that may be used to create and spread fake news [1]. The situation is dire because many people believe anything they read on the internet and the ones who are amateur or are new to the digital technology may be easily fooled. A similar problem is fraud that may happen due to spam or malicious emails and messages. So, it is compelling enough acknowledge this problem take on this challenge to control the rates of crime, political unrest, grief, and thwart the attempts of spreading fake news.

1.2. Outline

Text, or natural language, is one form which is difficult to process simply because of various linguistic features and styles like sarcasm, metaphors, etc. Moreover, there are thousands of spoken languages and every language has its own grammar, script and syntax. Natural language processing is a branch of artificial intelligence and it encompasses techniques that can utilize text, create models and produce predictions. The aim of this work is to create a system or model that can use the data of past news reports and predict the chances of a news report being fake or not.

2. Related Work

Various researchers have attempted solving this challenge in a multitude of ways to test which method works and get desirable results. A few studies have discussed fake news detection approaches from a data mining perspective, including feature extraction and model construction. A methodology of feature extraction (both news content features and social context features) combined with metric evaluation using precision, recall and f1 scores has proved to bear educated results but the problem is not that simple. Other parameters like bot spamming, click bait, source of news also affect the predictions [3]. These were some data mining and NLP oriented approached, but with more and more research and development in AI, researchers became interested in heavy neural network oriented approaches. A paper showed a method of ‘capture’, ’score’, ‘integrate’ and creates a model of recurrent neural networks for stance detection of fake news. They used recurrent neural network to capture temporal pattern of user activity around a specific article/text, then the user behaviour is used to extract source characteristics. All this information is used and integrated to form a model for classification of fake news [4].
The studies prove that even a simple straightforward network model can outperform complex models. So clearly, complexity of the model is not the optimum solution here and rather the right choice of parameters and data is essential.

Even the use of convolutional neural networks has been done while some others tried linguistically infused neural networks.
Another paper discusses linguistically-infused neural network model with Long-Short-Term-Memory (LSTM) and Convolutional Neural Network (CNN) to classify twitter posts. The linguistic part was introduced using the GloVe

library of pre trained vectors [5].So,it is evident that many attempts have been made but it is all a bit messy and scattered. There is a lot of room for development and research in this area especially because news statements have so many variables attached to them: sarcasm, abbreviation, metaphors, etc. However, efforts have been made to arrange reliable and vast data into a quality dataset. One such benchmark dataset has been used in this project.Fake news problem is growing at an alarming rate and it needs to be addressed more aggressively.

3. Data

Categorizing a news statement as "fake news" could be a very challenging and time-consuming task. For this reason, the use of an existing dataset, that has already collected and classified fake news, has been made. The data source used for this project is LIAR dataset. Given below is a brief description about the data files used for this study. The dataset has been cited in the paper [6] "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection, to appear in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017), short paper, Vancouver, BC, Canada, July 30-August 4, ACL [6].
The original dataset contained 13 variables/columns for train, test and validation sets. For the sake of simplicity, only two variables from this original dataset have been chosen for this classification task. The other variables may be used later to achieve more detailed analysis. The two columns that have been used are: 'Statement', which is the actual news statement itself, and 'Label', which refers to the statement being true/false.The procedure used for reducing number of classes in the dataset:

- True, mostly-true, half-true become 'True'.
- Barely-true, false, pants-fire becomes 'False'.

An important point to note here is that this dataset comes with 3 separate TSV files for train, validation and test. So,it is not required to manually split the data into test, train, validation.

4. Model

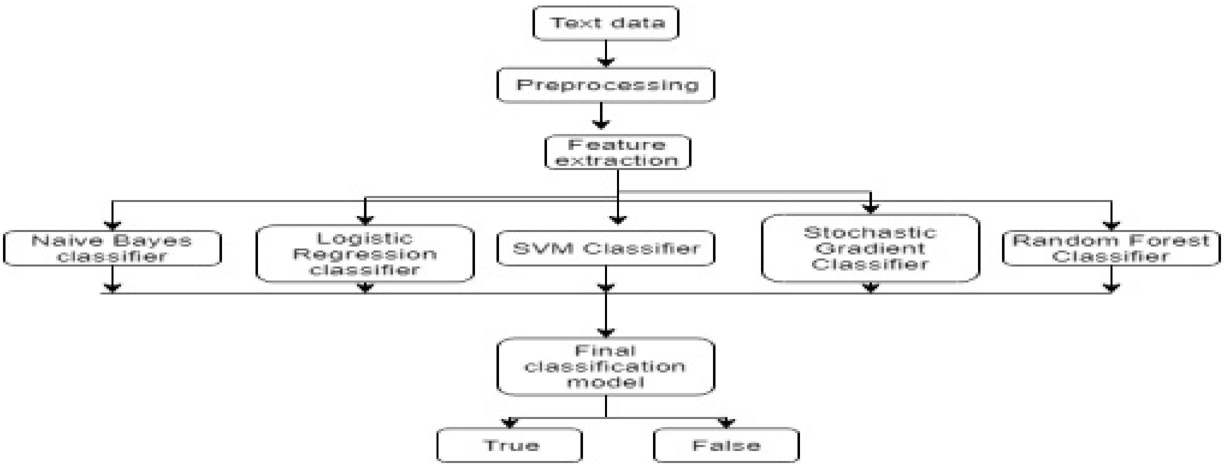

Fig1: Classifier prediction model

The performance of a classifier may vary based on the size and quality of the text data (or corpus) and also the features of the text vectors. Common noisy wordscalled 'stopwords' are less important words when it comes to text feature extraction, they don't contribute towards the actual meaning of a sentence and they only contribute towards feature dimensionality and may be discarded for better performance.
This helps in reducing the size/dimensionality of the text corpus and add text context for feature extraction. Also, lemmatization is used to convert words to their core meaning and this result in multiple word conversion into a single discrete representation [7].

In this step, first the test, train and validation data has been read and then tokenization and stemming has been done as shown in Fig.2. After this, a textbook exploratory data analysis for understanding the dataset and cleaning it has been conducted.Feature extraction and assortment of selection methods has been facilitated by the python's "scikit-learn" library. For feature selection, the use of methods like simple bag-of-words and n-grams and then term frequency like tf-idf weighting has been done.

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right) \quad (1)$$

N-grams are permutations of word combinations. They help in providing context to the text by combining nearby words and making a single feature out of them.

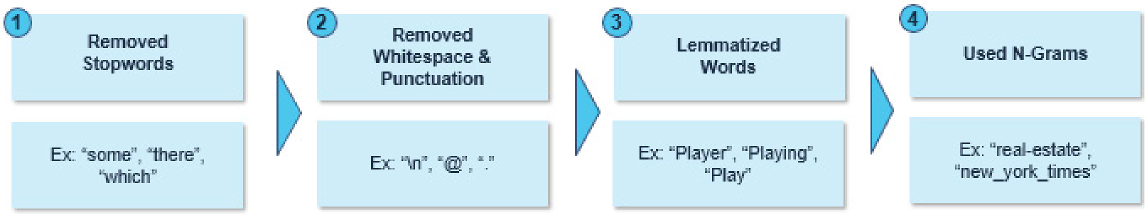

Fig 2: Pre-processing of text data

The extracted features are then fed into different classifiersNaïve Bayes, Logistic Regression, Linear SVM, Stochastic Gradient Classifier and Random Forest Classifiers from the sci-kit learn library have been used as shown in Fig.1. Each of the extracted features have been used in all of the classifiers.An extensive research on a similar approach has been done in the paper [8] and the authors have discussed the use of count vectorizers and TF-IDF vectorizers for the NLP task of classifying fake news using machine learning. After fitting and training the model, comparison of the 'f1' scores is done and confusion matrix is referred, in order to make an educated decision as shown in Fig.3.

After fitting all the classifiers, two best performing models were selected as candidate models for fake news classificationParameter tuning done by implementing GridSearchCV methods on these candidate models is an efficient and reliable approach, and it helps one chose best performing parameters for these classifiers. Finally, the best performing model was used for prediction task.

The data is rarely evenly distributed in the dataset. So, in such cases one may use to measure the performance of a classifier. True positives are the correct predictions of the classifier and false positives are the incorrect predictions. Using these numbers makes the task of calculating precision, recall and f1 scores effortless.

	Predicted class		
Actual Class		Class = Yes	Class = No
	Class = Yes	True Positive	False Negative
	Class = No	False Positive	True Negative

Fig3: Model of confusion matrix

The respective precision, recall and f1 scores can be calculated using the follow formulae:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive+False positive}} \tag{2}$$

$$\text{Recall}= \frac{\text{True positive}}{\text{True positive+False Negative}} \tag{3}$$

$$\text{F1 score} = 2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \tag{4}$$

The final selected, best performing model is ‘SVM’, which was persisted using the ‘pickle’ library in python. Using this model,and predictions can be made on a new news statement.

5. Result

The results show that SVM and logistic regression classifier have the best performance on this dataset in the model, with SVM having a slightly better performance than logistic regression classifier. The same can be perceived from the f1 scores.Also, the training data is largely based on US politics and economics news so it has been observed in our test cases, that the news statements related to US politics have been correctly classified and fake news was detected. But the test cases which have news related to technology have been wrongly predicted.

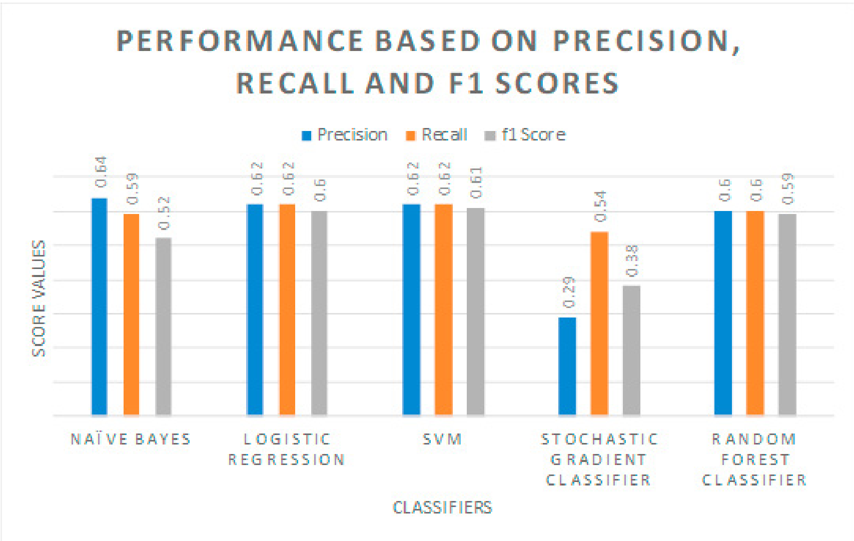

Fig4: Performance graphs of classifiers

Table 1: Test Cases for News

News Statement	Prediction	Reality
Says American polling shows Russian President Vladimir Putin has an 80 percent approval rating.	True	True
The Obama administration leaked information, deliberately or otherwise, that led to the identification of the Pakistani doctor that helped us in achieving our goals and killing bin Laden.	False	False
The percentage of black children born without a father in the home has risen from 7 percent in 1964 to 73 percent today, due to changes from President Lyndon Johnsons Great Society.	True	False
About 106,000 soldiers had a prescription of three weeks or more for pain, depression or anxiety medication.	True	True
India becomes the world's greatest exporter of rice.	True	False
Google enters e-commerce business, gives Amazon the chills	True	False
The suicide rates in US show that house wives and CEOs are on top of the list	True	False

The above Table 1shows a quick glance at some of our test cases which reveal how training news data domain can highly influence the model performance. Although the predictions for news related to US politics and economics are not always correct, one can see in the last three test cases (news which are not related to US politics/economics, our training data), the predictions are almost always incorrect. So, it is clearly visible how much the quality and quantity of training data affects this fake news detection model.If the model is trained with a more diverse dataset with news from various different domains, obtaining a much more robust and accurate classifier is not too far-fetched. Also, more technical improvements, such as hyper parameter tuning and better feature selection, can be used which have been discussed in the future scope of this paper.

6. Conclusion, Limitation and Scope for Future work

The fake news challenge is perilous and is spreading rapidly like a wildfire as it becomes easier for information to reach the mass in various flavours. Reports have shown that, just like in the last US presidential elections, fake news can have a huge impact in politics and thereafter on the people like a domino effect. With the help of artificial intelligence, we can control and limit the spread of such misinformation more quickly and efficiently as compared to manual efforts. The work in this project proposes a stacked model which fine tunes the informational insight gained from the data at each step and then tries to make a prediction.

Although many attempts have been made to solve the problem of fake news, any significant success is yet to be seen. With huge amounts of data collected from social media websites like Facebook, Twitter, etc., the best models improve every day. With the use of deep neural networks, the future work in this field seems a lot more promising.

The limitations that come packaged with this problem is that, the data is erratic and this means that any type of prediction model can have anomalies and can make mistakes.

For future improvements, concepts like POS tagging, word2vec and topic modelling can be utilized. These will give the model a lot more depth in terms of feature extraction and fine-tuned classification.

Word2Vec: The Word2Vec technique converts text to features while maintaining the original relationships between words in a corpus. It is a combination of techniques and is one of the best feature extraction techniques in NLP. It generally uses a model of pretrained vectors (like GloVe) and then transfer learning can be used to obtain a superior model.

Topic Modelling: News can contain a vast range of topics. Just the classification based on labels is not enough if realistic results are desired. For this reason, an advanced technique called topic modelling can come in handy. Topic modelling categories each piece of text into topics and using this one can make more accurate predictions. The most popular topic modelling technique used in NLP is "Latent Dirichlet Allocation", also known as LDA.Use of LDA can add another layer of depth to the fake news classification task.

References

[1] 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.

[2] Hunt, E. (2016). What is fake news? How to spot it and what you can do to stop it. The Guardian, 17.

[3] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, *19*(1), 22-36.

[4] Ruchansky, N., Seo, S., & Liu, Y. (2017, November). Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 797-806). ACM.

[5] Volkova, S., Shaffer, K., Jang, J. Y., &Hodas, N. (2017, July). Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 647-653).

[6] Wang, W. Y. (2017). " liar, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648*.

[7] Reis, J. C., Correia, A., Murai, F., Veloso, A., Benevenuto, F., & Cambria, E. (2019). Supervised Learning for Fake News Detection. IEEE Intelligent Systems, 34(2), 76-81.

[8] Pal, S., Kumar, T. S., & Pal, S. (2019). Applying Machine Learning to Detect Fake News. Indian Journal of Computer Science, 4(1), 7-12.