A Study on Fake News Detection Using Naïve Bayes, SVM, Neural Networks and LSTM

Research · July 2019

CITATION
1

READS
2,505

1 author:

Poonam Tijare
CMR Institute of Technology
7 PUBLICATIONS **8** CITATIONS

SEE PROFILE

A Study on Fake News Detection Using Naïve Bayes, SVM, Neural Networks and LSTM

Prannay S Reddy, Dept. of Computer Sci. & Engineering, CMR Institute of Technology, Bangalore.

E-mail: prannaysreddy@gmail.com

Diana Elizabeth Roy, Dept. of Computer Sci. & Engineering, CMR Institute of Technology, Bangalore.

E-mail: dianaelizabethroy@hotmail.com

P. Manoj, Dept. of Computer Sci. & Engineering, CMR Institute of Technology, Bangalore.

E-mail: manojreddy156@gmail.com

M. Keerthana, Dept. of Computer Sci. & Engineering, CMR Institute of Technology, Bangalore.

E-mail: keerthanamidadala@gmail.com

Poonam V Tijare, Dept. of Computer Sci. & Engineering, CMR Institute of Technology, Bangalore.

E-mail: poonam.v@cmrit.ac.in

Abstract--- Accounting to the expeditious digitization across all channels and mediums, the menace of fake news has been burgeoning at a colossal scale. Majority of the countries all across the world are trying to combat this challenge. This paper explores the application of Natural Language Processing and Machine Learning techniques to identify fake news accurately. Pre-processing tools are used to clean the data and apply feature extraction on them. Then a fake news detection model is built using four different techniques. Finally, the paper investigates and compares the accuracy of techniques which are Naive Bayes, Support Vector Machine (SVM), neural network and long short-term memory (LSTM) to find the best fit for the model.

I. Introduction

Over the last decade, there have encounters of flux in misinformation that spread like wildfires. The surge in fake news was noticed during the 2016 presidential elections that happened in the US that determined the fate of these elections. In many cases, it is seen that the sharing of hoax news has been more than that of accurate news. In a massive market like India, the scope of fake news propaganda has been artfully misused by many groups. Researches indicate that Facebook and WhatsApp are the platforms that are utilized for spreading fake news. An approximate of one in two Indians have agreed to have received fake news during the 2019 Lok Sabha elections.

According to research at Ohio University by M Laeeq Khan, the majority of the population do not verify the authenticity and veracity of a piece of information before forwarding it on social media. Khan conducted his work in the US and later extended it to Indonesia to test his hypothesis about spreading misinformation on a medley of online platforms. "Fake news and spreading of it could be rightly termed as the major issues of our time. No other study can explain the vital role of people in halting the spread of misinformation", says Laeeq. It was found that gender, age, and social class do not play a pivotal role in spreading misinformation, but rather the literacy regarding the internet and social media that mattered the most. The research which comprised of 396 participants stated that the information verification attitude of the people was taken for granted. It was found that the simple habit of googling the veracity of information before sharing it could prove crucial in fighting against the issue of fake news.

It is also seen that spam messages and fake news have striking similarities. They use manipulative ways to win over the reader's opinions. Most of them also have grammatical mistakes and they also use a similar restricted set of words among them. Since both the media share such similar properties, we can use similar approaches to detect fake news accurately. One way to tackle fake news is to manually classify news as real or fake. Even though that seems like the simplest solution it is not practical with the jillions of news that get produced to manually label it. Hence, there is a need to look for a pragmatic technical solution to do the same. The proposed method in this research is to exploit the advancement in machine learning. To do the same, the classification model has been trained with various machine learning algorithms to label the data. The results from the study indicate neural networks to be the case in which news highest accuracy is achieved.

ISSN 1943-023X

Received: 29 May 2019/Accepted: 17 June 2019

II. Literature Survey

In the paper by Shu A et al., they investigated how news can be classified as true or not by focusing on a few attributes that are repeatedly encountered in fake news. In their opinion, these characteristics were based on "psychology and social theories, existing algorithms from a data mining perspective, evaluation metrics and representative datasets". This paper also analyses the different challenges one will encounter while studying this topic. [1]

The paper is written by Rubin et al. deals with the domain of fake news which is composed of satirical news. Satire news intentionally provides hints revealing its own deception. While fake news wants the readers to believe a false fact, satire news must eventually be understood as a jest. This paper provides an in-depth view of the features of humor and satire news along with the style of the authors reporting. The paper has considered the news articles from twelve contemporary news topics in four different domains which are civics, science, business, and soft news. The paper proposes a Support Vector Machine based algorithm which can detect satire news based on features like Absurdity, Humor, Grammar, and Punctuation. The models achieved an accuracy of 90% and a recall of 84%. The aim is to reduce the negative impact of satire news on readers. [2]

In the paper by Kelly Stahl et al., they have considered past and current techniques for fake news identification in text formats while elucidating how and why news fake exists in any case. This paper incorporates a discussion on how the writing style of a paper can also impact on its classification. They had implemented their project using Naïve Bayes Classifier and Support Vector Machines methods. They had looked into the semantic analysis of the text for classification. [3]

In this paper by Marco L. Delia Vedov et al., they say that "we propose a novel ML fake news detection method which, by combining news content and social context features, outperforms existing methods in the literature, increasing their already high accuracy by up to 4.8%". The proposed model was then tested on a real-time application and they achieved high accuracy by testing it on a FB messenger chatbot. The accuracy achieved by them is close to 82%. [4]

In the research paper written by Mykhailo Granik et al., they have explored a simple approach to detect fake news using Naïve Bayes classifier. Their approach was conducted on Facebook news posts rather than news articles on the internet. The paper achieved an accuracy of 74% in the classification process. It also went on to state that simple machine learning models like the Naïve Bayes can achieve a moderate accuracy and in future, more artificial intelligence techniques could be used to tackle the menace of fake news. [5]

The paper by Namwon Kim et al. detects fake news based on unified key sentence information. The model extracts the key sentences of the article to the question from the article and then later combining the word vector for each key sentence. They make use of a Korean dataset for their analysis. It performs an efficient matching operation for the word vectors obtained using bidirectional LSTM. They have achieved an accuracy ranging from 64 to 69%. Their future plan is to develop a more advanced model which applies the model independently to each key sentence. [6]

In this paper by Hadeer Ahmed et al., in this project an elaborate study was done using n-gram while being compared with different feature extraction methods. Many feature extraction techniques and quite a few machine learning techniques were utilized in this study. The paper reflected the highest output while implementing unigram as the extraction method and SVM as the classification method. [7]

In this paper by Saranya Krishnan et al., they have proposed a generalized framework to predict tweet credibility. First, they have extracted the essential features and user features through the Twitter API. If a tweet has an image or image URL, the reverse image search is performed to check whether the same image has been tagged with different information in the past. In addition, if any URL is present in any tweet, it will be cross-checked against the fake news sources to see whether it is a part of a fake news websites dataset. All these features are then used by the data mining algorithms to classify tweets as fake or real. [8]

In this paper by Stefan Helmstetter et al., tweets were assimilated in real time. On arrival, these tweets were background checked to verify if they were classified properly. Then the incoming tweets were tested against a classification algorithm. Despite having a noisy dataset a decent result was obtained. [9]

In this paper by Akshay Jain, he suggested a classification methodology that could be used in real time data that gets generated on Facebook. It could also be used in other social media giants like Twitter and WhatsApp. He has utilized a simple Naive Bayes classification to classify the incoming data as trustworthy or non-trustworthy. [10]

Received: 29 May 2019/Accepted: 17 June 2019

III. Methodology

A. Dataset Description

The corpus of data implemented in this project had around 33000 articles of data. These articles mainly constituted news about US politics. The dataset obtained on Kaggle was noisy and required cleaning. The main features included in each row of the data were id, heading, author, content, classification of being fake or true. The dataset has the following features:

Table 1: Dataset Description

ID
HEADING
AUTHOR
CONTENT/TEXT
FAKE OR REAL CLASSIFIFCATION

B. Feature Extraction and Pre-Processing

To start off with the implementation, the data is obtained in raw format which is part of the dataset. This data needs to be pre-processed before we can implement it in the project. The process includes stop-word removal followed by making the entire document in lower case for uniformity. Also, any of the special characters that can cause an anomaly in the document are removed in this process. Stop words are words that are not relevant and have little meaning lexically. These words are most often ignored to not cause any discrepancies to the process of classification. In a sentence like "There is a Bengal tiger.", the first three words 'there', 'is' and 'an' are stop words and have no significant meaning. These are the words that are usually excluded and some of the examples are: who, of, a, what, etc. Doc2Vec model was introduced in the year 2014 and adds upon the previous Word2VecModel. In Doc2Vec, a feature vector is used to represent a single "document". Whereas, in the Word2Vec model, a feature vector is used to represent a single "word". Doc2vec is superior and works better when we have a corpus of documents, this justifies the choice of this particular feature extraction model.

C. Models

1) Naïve Bayes

Naïve Bayes is a conditional probability model which can be used for labeling. The goal is to find a way to predict the class variable (B) using a vector of independent variables (A), i.e., finding the function f: A-->B. In probability terms, the goal is to find P(B|A), i.e., the probability of B belonging to a certain class A. B is generally assumed to be a categorical variable with two or more discrete values. It is a mathematically simple way to include contributions of many factors in predicting the class of the next data instance in the testing set. The limitation of Naïve Bayes is that they assume that all features are not dependent on each other. The Naïve Bayes rule is based on the theorem formulated by Bayes:

$$P(c|x) = \frac{P(x|C)P(c)}{P(x)} \qquad (1)$$

2) Support Vector Machine

A support vector machine (SVM), which can be used interchangeably with a support vector network (SVN), is also considered to be a supervised learning algorithm. SVMs work by being trained with specific data already organized into two different categories. Hence, the model is constructed after it has already been trained. Furthermore, the goal of the SVM method is to distinguish which category any new data falls under, in addition, it must also maximize the margin between the two classes [13]. The optimal goal is that the SVM will find a hyperplane that divides the dataset into two groups. The kernel used in this application is RBF as it is best suited for large applications like a corpus of news articles. The Radial Basis function on two samples x *and* x' is given by:

$$K(x,x') = \exp\left(-\frac{||x-x'||^2}{2\sigma^2}\right) \qquad (2)$$

Where $||x - x'||^2$ represents the squared Euclidean distance and σ is a free parameter.

3) Long Short-Term Memory

LSTM, which stands for Long short term memory is an extension of the previously famous RNN (Recurrent Neural Network). In addition to RNN, LSTM's also have memory over the long run. It comprises of three gates

Received: 29 May 2019/Accepted: 17 June 2019

namely input, gate, output gate and forget gate. The forget gate is used to forget features that have little value or weight. As the algorithm keeps running, it learns what is important and what is not by assigning weights accordingly. This characteristic made it the best fit for this paper and will help in making relationships on the large corpus of data as the news dataset keep increasing in real-world scenarios. The LSTM is inherently nothing but a neural network. Unlike these interconnected networks, they have looped neurons. Here, i_t , f_t , and O_t represent the equations of the input, forget, and output gates respectively. The weights are represented by *w* and the sigmoid function is represented by σ [14]:

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \quad (3)$$

$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f) \quad (4)$$

$$O_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \quad (5)$$

4) Neural Network

Artificial Neural Networks in which the association between the units do not form a cycle are called Feed forward neural networks. Feed forward neural networks were the first type of artificial neural network invented and are simpler than recurrent neural networks. These networks are named so because the information moves linearly in the network through the input layer, then the hidden layer and ultimately through the output layer. The paper implements one feed-forward neural network models using Keras. The papers neural network implementation uses three hidden layers. In the Keras implementation layers of size 256, 256, and 90 are selected along with dropout layers. The ReLU, which is also known as Rectified Linear Unit is used or the "activation function" as it is the most suited for challenges that this paper tries to solve. [12] In the below equations, $f(x)$ and $f'(x)$ represent the equation and the derivative of the Rectified Linear Unit (ReLU).

$$f(x) = \begin{cases} 0 \ for \ x < 0 \\ x \ for \ x \geq 0 \end{cases} \quad (6)$$

$$f'(x) = \begin{cases} 0 \ for \ x < 0 \\ 1 \ for \ x \geq 0 \end{cases} \quad (7)$$

D. Classification Process

Fig 1 is a schematic representation of the classification process. The first step of the process is to pre-process the data which include stop word removal and lower casing of the entire characters. Also, special characters are removed from the corpus. The Doc2Vec process is used, and a vector is formed representing the documents involved. The ultimate step of the classifier is to predict the class of the given news article into real or fake. There is a total of 4 machine learning models being tested here to arrive at the best choice. The data is partitioned into train and test. The testing data has 26000 corpus of news articles and the training data has 7000 corpus of news articles.

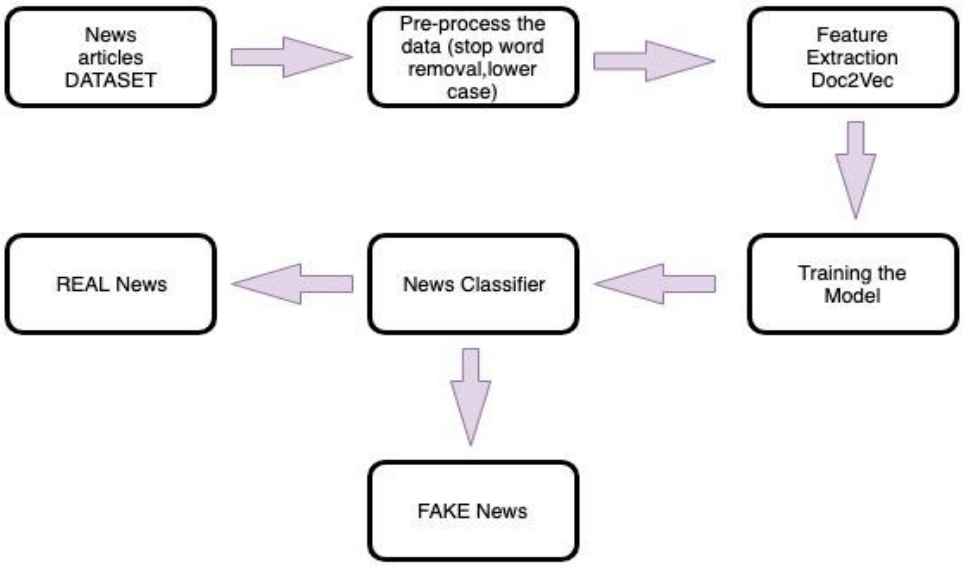

Fig. 1: Classification Process

Received: 29 May 2019/Accepted: 17 June 2019

IV. Results

Table 1 shows the results arrived at by calculating the accuracies of the various models mentioned above. The accuracy, precisions and f1 score can be computed with the help of confusion matrices. A single confusion matrix was created for each model. The values shown are the averaged values over successive trials.

Table 2: Results of Algorithm

Model	**Accuracy**
Naïve Bayes	71.84%
Support Vector Machine	87.37%
LSTM	94.27%
Keras Based Neural Network	90.62%

Based on the results in Table 1, the graph in Fig 2 is constructed by taking different algorithms on X-axis and accuracy on the Y-axis. It is inferred that LSTM provides us with the highest accuracy followed by Keras Neural Network, SVM and finally Naïve Bayes. The Keras based Neural network has a good accuracy that almost matches LSTM. LSTM provides such high results because the text is inherently a serialized object.

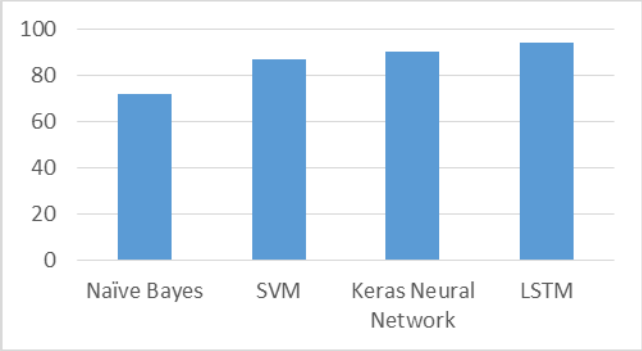

Fig. 2: Comparison of Algorithm Results

V. Conclusion & Future Work

In this paper, we have presented a model for fake news detection through different machine learning techniques. Furthermore, the paper investigated the four methods and compared their accuracies. The model that achieves the highest accuracy is LSTM and the highest accuracy score is 94%.

Fake news detection is an emerging research area which has a scarce number of datasets. There are no data on real-time news or regarding the current affairs. The current model is run against the existing dataset, showing that the model performs well against it.

In our future work, news article data can be considered related to recent incidents in the corpus of data. The next step then would be to train the model and analyze how the accuracies vary with the new data to further improve it.

References

[1] Shu K., Sliva A., Wang S., Tang J., Liu H., "Fake News Detection on Social Media: A Data Mining Perspective", ACM SIGKDD Explorations Newsletter, 2017, 19(1), 22-36.

[2] Rubin, V., Conroy N., Chen Y., Cornwell S., "Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News", Proceedings of the Second Workshop on Computational Approaches to Deception Detection Association for Computational Linguistics, 2016.

[3] Kelly Stahl., "Fake news detection in social media", B.S. Candidate, Department of Mathematics and Department of Computer Sciences, California State University Stanislaus, 2018.

[4] Marco L. Delia Vedova, Stefano Moret, Eugenio Tacchini, Massimo Di Pierro," Automatic Online Fake News Detection Combining Content and Social Signals", Research Gate May 2018.

[5] Mykhailo Granik, Volodymyr Mesyura, "Fake News Detection Using Naive Bayes Classifier", 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON).

Received: 29 May 2019/Accepted: 17 June 2019

[6] Namwon Kim, Deokjin Seo, Chang-Sung Jeong, "FAMOUS: Fake News Detection Model based on Unified Key Sentence Information", 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)
[7] Hadeer Ahmed, Issa Traore, Sherif Saad, "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques", Research Gate May 2017.
[8] Saranya Krishnan, Min Chen.," Identifying tweets with fake news", 2018 IEEE Conference on Information Reuse Integration for Data Science.
[9] Stefan Helmstetter, Heiko Paulhem, "Weakly Supervised Learning for Fake news detection on Twitter", 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).
[10] Akshay Jain, "Fake News Detection", IEEE 2018 International Students' Conference on Electrical, Electronics and Computer Sciences.
[11] Quoc L., Mikolov T., "Distributed Representations of Sentences and Documents", Arxiv Archive May 2014.
[12] Saxena, R., "How the Naive Bayes Classifier works in Machine Learning", http://dataaspirant.com /2017/02/06/naive-bayes-classifier-machine-learning/, Retrieved: April 15, 2019.
[13] Brambrick, Aylien, "Support Vector Machines: A Simple Explanation", https://www.kdnuggets.com /2016/07/support-vector-machines-simple-explanation.html, Retrieved: April 20, 2019.
[14] Hochreiter, S., Jrgen, S., "Long short-term memory", Institute of Bioinformatics Publications Newsletter, October 1997.
[15] Goldberg, Y., "A Primer on Neural Network Models for Natural Language Processing", Arxiv Archive 2015.

Received: 29 May 2019/Accepted: 17 June 2019