Machine Learning Based Detection of Deceptive Tweets on Covid-19

Amisha Sinha, Mohnish Raval, S Sindhu

Abstract: Social media plays a vital role in connecting people around the world and developing relationships. Social media has a huge potential audience and the circulation of any information does impact huge populations. With the surge of Covid-19, we can see a lot of fake news and tweets circulating about remedies, medicines, and general information related to pandemics. In this paper, we set out a machine learning-based detection of deceptive information around Covid-19. With this paper, we have described our project, which could detect whether a tweet is fake or real automatically. The labeled dataset is used in the process which is extracted from the Xiv repository. Data set has tweets, upon which various methods are applied for cleaning, training, and testing. Pre-processing, Classification, tokenization, and stemming/removal of stop words are performed to extract the most relevant information from the dataset. For classification, we have used two classification techniques - TF-IDF and Bags of words. To achieve better accuracy, we have used two other methodologies - SVM and Random Forest and have achieved an F1-score of 94 using SVM.

Keywords: Artificial Intelligence, Fake News, Social Media, SVM

I. INTRODUCTION

With the commencement of the pandemic due to the novel coronavirus, social media has acted as the foremost way of passing the information from one source to the other. With the large and epiphanic flow of data, thecertitude of the latter is indiscernible. With the pandemic extent, it is the need for an hour to stratify the large data into# and gospel. Our project uses various machine learning techniques to analyze the tweets into “Real” and “Fake”. The focus of the project is to help in analyzing the data and forbid instances of spreading false information. The model created can be assessed in real-time for tracking the source of false information. This will help the government of India to take appropriate actions against the culprits and draconianically deplete wrong tweets circulating on the internet.

Manuscript received on June 14, 2021.
Revised Manuscript received on June 19, 2021.
Manuscript published on June 30, 2021.

* Correspondence Author

Amisha Sinha*, Department of Information Technology, SRM Institute of Science and Technology, Chennai (Tamil Nadu), India. Email: amishasinha84@gmail.com

Mohnish Raval, Department of Information Technology, SRM Institute of Science and Technology, Chennai (Tamil Nadu), India. Email: mohnishraval40@gmail.com

S Sindhu, Assistant Professor, Department of Information Technology SRM Institute of Science and Technology, Chennai (Tamil Nadu), India. Email: sindhus2@srmist.edu.in

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEJESP). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
This dataset consists of labeled data in two columns with one column having the tweets and the other column marked as Label with data as either '0' or '1'. Here, '0' implies a real tweet and '1' implies a potential fake tweet. The dataset consists of 2140 tweets with 1120 real and 1020 fake tweets. It has a vocabulary size of 63032 with a number of URLs and unique usernames.

IV. METHODOLOGY

The Methodology used in our project is based on basic natural language processing. The methodology consists of the following components-like collection of labeled data, processing of dataset, applying classification techniques such as tf-IDF and bow, visualization of the dataset, tokenizing the tweets in the dataset, finding the stems in the tokenized words, and removal of stopwords.

A. Pre-processing

In this work, we mainly focus on the polarity of the tweet and the tweet itself. So, we have pre-processed the tweets by removing unwanted details, stemming, tokenization, and removal of stop words. This involves the removal of mentions, retweets, URLs, and non-alphanumeric characters. The main purpose of data cleaning is to refine the quality of the dataset which is critical for modeling processes. The quality of the dataset determines how the model will perform in a specific circumstance. Mentions consist of the tagged tweets with the person's username and retweets are a kind of reply to a tweet that plays no role in determining the nature of the tweet. Similarly, the removal of URLs is crucial as they are links to other websites and do not provide any meaning. Non-alphanumeric characters provide no meaning to the context and can be removed. These steps of data cleaning ensure structured data which is a salient step before proceeding with the dataset.

B. Tf-idf technique

TF-IDF is a method used to fetch information supporting the frequency of a term (TF) and its inverse document frequency. It acknowledges a keyword in any given content and based on the number of times it appears in a document, it earmarks its importance. Taking a hypothetical situation like we have a group of words that says “Crocin is the answer to COVID-19” and we are keeping that as the reference want to rank the order of priority of other words in the document, can be started simply by eliminating the sentences without this keyword and often, the maximum frequency definition of the word in the document is used to measure the importance, and the count of the word in a given document is referred to as term frequency, and the process is known as term frequency. Talking about inverse document frequency, within a document, the usage of the words in the document are not equally important. The words like “the”, “a” can be seen more frequently in any document, irrespective of its relevance. Prepositions, supporting verbs are to be removed. The relevance of a keyword within a document is expressed as a corpus and can be checked through the tf-idf algorithm. Rule.

\[
\text{tfidf} = \text{tf} \times \text{idf}
\]

\[
\text{idf}(t) = \log \left( \frac{n+1}{\text{df}(d, t) + 1} \right) + 1
\]

Fig. 1. TF-IDF Formula

TF-IDF score is the degree of importance of a word in a sentence. The tf score and idf score are used to obtain information from the text provided by the user.

In our project, we have made sure that the collection of corpora is checked and for that purpose, we have used this approach.

C. Bag of words

Bag of words is used for text which converts a particular text into a bag of words i.e. it keeps track of the frequency in which such terms appear in a specific piece of content. It is easy to comprehend and implement the algorithm. The bag of words models creates a frequency matrix of the words in the given content. It converts the text into a vector of fixed size. Major steps involved here are Vocabulary determination (1) where all the words from the document are restored, counting (2), where all the words along with their frequency is determined in vector form. Tokenizer() and tokenizer.texts_to_matrix() are python functions that are reused to carry out the tasks. The bag of words is useful and widely used due to its simplicity and also for the fact that its computation is inexpensive.

Fig. 2. Snapshot of TF-IDF Result

“Fig. 2.” shows the snapshot of code, where we have achieved the Score of 0.8 using this technique.

Fig. 3. Snapshot of BOW Result
D. Tokenization

Tokenization involves splitting sequences of words (Sentences) into keywords, symbols, and other components referred to as tokens. These tokens can be used as input for the processes. Tokens are mostly separated by whitespace or any line breaks. Only words can be included as tokens (any character with a continuous stream of words). To implement this in any dataset, a Natural language toolkit is used which has both word and sentence tokenization functions. Word_tokenizer() is used to split a sentence into tokens. Python offers various tokenization functions which include:

1. Sentencetokenization which is imported from the NLTK library.
2. PunktSentenceTokenizer which is used when there are large chunks of data and tokenization is not easy using the regular sentence tokenizer.
3. Tokenizesentenceofdifferentlanguages.
4. Word_tokenizer
5. TreebankWordTokenizer which separates the sentence using punctuation and white spaces.
6. PunktWordtokenizer
7. Using Regular expression which uses regexp_tokenizer() to tokenize based on the expression.

Tokenization leads to the creation of tokens from a series of words which is useful in finding patterns that are used in further steps of stemming and lemmatization.

E. Stemming and Removal of Stop Words

Stemming reduces morphological variants of a word. For instance, it reduces all the words like [likes, liked, likely] to the root word ‘like’. It is beneficial in reducing akin words. Some algorithms to implement stemming are Potter’s stemmer, Lovins, Dawson, Krovetz, Xerox, and N-gram. The stemming method is useful in reducing the number of tokens and assimilating them to a domain vocabulary library. It is also very useful to fetch results in a search engine like Google.

```
if gram > 1:
    w = []
    for i in range(len(words) - gram + 1):
        w += ["":".join(words[i:i + gram])"]
    return w
else:
    return words

if stem:
    stemmer = PorterStemmer()
    words = [stemmer.stem(word) for word in words]
    return words
```

Fig.4. Snapshot- Stemming & Removal of Stop words

Using the Hinge loss function, we strive to enhance the margin in the SVM algorithm as shown in “Fig.6”.

\[
h_\theta(x) = \begin{cases} 
1 & \text{if } \theta^T x \geq 0 \\
0 & \text{otherwise}
\end{cases}
\]

Fig.6. Hinge Loss Function

On the condition that the value obtained out of prediction and that outcome of calculation come to a difference at all, then the loss is calculated as zero; else we calculate the loss using the hinge loss function. In order to balance the maximization of margin and loss, we add cost function, as in “Fig.7”

V. MODULE DESCRIPTION

We have used the Support Vector Machine (SVM) algorithm in our Random Foresting project.

A. Support Vector Machine (SVM)

SVM is an algorithm that has been facile to embrace but extremely efficient and versatile in its application. It can be used for classification as well as regression. The framework of this algorithm is illustrated in “Fig.5”. The space is depicted as a multi-dimensional space of various groups. In “Fig.5”, the hyperplane must be chosen such that the distance between the support vectors is maximized, referred to as the maximum margin, whereas support vectors are the points closest to the hyperplane and points of two classes are shown in green and blue.
Machine Learning Based Detection of Deceptive Tweets on Covid-19

\[
\text{Cost}(h_S(x), y) = \begin{cases} 
\max(0, 1 - \theta^T x) & \text{if } y = 1 \\
\max(0, 1 + \theta^T x) & \text{if } y = 0 
\end{cases}
\]

Fig.7. CostFunction

For Non-Linear SVM algorithm, the function used is shown in “Fig.8”

\[
\text{Hypothesis : } h_S(x) = \begin{cases} 
1 & \text{if } \theta^T f \geq 0 \\
0 & \text{otherwise}
\end{cases}
\]

\[
\text{Cost Function : } J(\theta) = \frac{1}{n} \sum_{i=1}^{n} y_i [\theta^T f(i)] + \frac{1}{2} \| \theta \|^2
\]

Fig.8. SVMFunction

The only difference between the function used for linear and non-linear SVM algorithms is that “f” is used instead of ‘x’, where f is a function of x. This is called the Kernel function. There are various Kernel functions among which “Gaussian function” is the most popular one as it can be used on the condition where there is no prior understanding of the data.

In our project we have used SVM algorithm, a snapshot of the code and the accuracy is shown in “Fig.9”.

In a nutshell, it comes to the conclusion that the SVM algorithm works well on the clear margin of separation between classes and is effective where the dimensional frequency is greater than that of samples.

B. Random Forest

This supervised algorithm due to its simplicity and its dual purpose to serve classification and regression because of which it has been the most prevailing technique.

The concept of the decision tree is used in the latter which has three cardinal parts: Root node, Leaf node, and branches. At every node, the decision is either the left sub-node or the right node depending on the condition fulfilled in terms of “True” or “False”. The verdict taken is determined from the features of the dataset. A minute change in the data can produce varied consequences due to its tactfulness. The muddle of overfitting is also eloquently abated due to use of multiple decision trees rather than a single decision tree.

In our project we have used Random Forest algorithm, a snapshot of the code and the accuracy is shown in “Fig.11”.

Fig.9. Snapshot - SVM algorithm

Fig.10. RandomForestWorking

Fig.11. Snapshot - RandomForest
VI. SYSTEM ARCHITECTURE

![Flowchart-SystemArchitecture](image)

Fig.12. Flowchart-System Architecture

VII. PERFORMANCE OF FINAL METHOD

In our project, we have used two classifiers: tf-idf and bag of words and two algorithms-SVM (Support Vector Machine) and Random Forest.

Table 1: Shows the summary of the accuracies obtained from all the algorithms used:

| S.No | Algorithms Implemented | Accuracy (in %) |
|------|------------------------|----------------|
| 1    | Support Vector Machine (SVM) | 94%            |
| 2    | Random Forest           | 92%            |
| 3    | Bayes Theorem (TF-IDF)  | 88.88%         |
| 4    | Bayes Theorem (BOW)     | 77.77%         |

VIII. CONCLUSION AND FUTURE WORK

In our model, we achieved an accuracy of 94 percent using the SVM method from the various classification methods as shown in Table 1. Perforce, we conclude that the SVM approach has been found to be the most rational. We bring forward this substructure to assist in the identification of misleading info on any probable public health like the 3rd and 4th waves of coronavirus, which are anticipated analogous to writing this survey. This same framework can be used even in the detection of any misleading news, other than global health as well. For our future work, we would like to look at how other pre-trained models work with the same dataset and also our framework performance with other datasets. It'll be riveting to see how our system executes against supplementary assimilated Fake News dataset.

REFERENCES

1. Ayan Basak, Saikat Dutta and Sourya Dipta Das (2020) - “Heuristic-driven ensemble framework for covid-19 fake news detection” (razorthinkinc, USA): IIT Madras, India
2. Fayez Gebali, (Life Senior Member, IEEE) , Kinfu Li, (Senior Member, IEEE) and Mohamed K. Elhadi (2020) - “Detecting Misleading Information on COVID-19” IEEE-Department of Electrical and Computer Engineering, University of Victoria, Canada
3. Cody Buntain and Jennifer Golbeck - “Automatically Identifying Fake News in Popular Twitter Threads”: University of Maryland
4. Srishri Agrawal, Ruchika Arora, Pronika Chawla et al. (2020) - “Fake News Detection Using ML” - (IJRET): MIRIRS, Faridabad
5. Md. Sayeed Al-Zaman - Article on “COVID-19-Related Social Media Fake News in India”: Department of Journalism and Media Studies, Jahangirnagar University, Savar, Dhaka 1342, Bangladesh
6. Anjali Jain, harsh khater and avinashshakya - (2019) - “A smart system for fake news detection using machine learning” : Dr. Ajabdul kalam university, lucknow, India
7. Aphiwongsophon & Chongstitvatanva (2018) - “Detecting Fake News with Machine Learning Method.” (2018): Computer, Telecommunications and Information Technology (ECTICON) https://doi.org/10.1109/ECTICon.2018.8620051
8. Aswini Thota, Priyanka Talik, Simrat Ahiwaliaatel. (2019)- “Fake News Detection Using Machine Learning Algorithms”: Bharati Vidyapeeth College of Engineering, Pune, India
9. Uma Sharma, Sidarth Saran, Shankar M. Patil (2021) - “Fake News Detection using Machine Learning Algorithms”: Bharati Vidyapeeth College of Engineering, Pune, India
10. Markines, B.Cattuto, C. & Menczer, F. (2009). - “Social spam detection. In Proceedings of the 5th International Information Retrieval and Workshop on the Web”
11. H. Gupta, M.S.Jamal, S. Madisettyand M. S. Desarkar (2018) - “A framework for real-time spam detection in Twitter,” 10th International Conference on Communications Systems & Networks (COMSNETS), Bengaluru, pp. 380-382
12. Aayush Ranjan - “Fake News Detection Using Machine Learning” (2018): Department of Computer Science & Engineering, Delhi Technological University
13. C. Buntain and J. Golbeck - (2017) “Automatically Identifying Fake News in Popular Twitter Threads,” - IEEE International Conference on Smart Cloud, New York, NY, 2017, pp. 210-215.
14. Cadet Metz - (2016) - “The bittersweet sweepstakes to build an AI that destroys fake news.”
15. http://www.fakenewschallenge.org/
16. IT. Granskogen and J. A. Gulla - (2017) - “Fake news detection: Network data from social media used to predict fakes,” CEUR Workshop Proc., pp. 59-64
Machine Learning Based Detection of Deceptive Tweets on Covid-19

AUTHORS PROFILE

Amisha Sinha, B. Tech with specialization in Information Technology from SRM Institute of Science & Technology, Kattankulathur, 2020 graduate. Software Engineer. Core Interest: Machine Learning, Data Science, Algorithm development.

Mohnish Raval, B. Tech with specialization in Information Technology from SRM Institute of Science & Technology, Kattankulathur, 2020 graduate. Application development Analyst. Core Interest: Machine Learning, Data Science, Algorithm development.

S Sindhu, Assistant Professor (OG) at SRM Institute of Science & Technology, Kattankulathur. M.E with specialization in Computer Science from Anna University, 2011. Research Interest: Data Mining, Big data.