Implementation Paper on Analyzing COVID-19 Vaccines on Twitter Dataset Using Tweepy and Text Blob

Abhishek Akshay Chaudhri[1], S.S. Saranya[2], Sparsh Dubey[3]

[1]Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur Campus, Tamil Nadu, India. E-mail: aa8937@srmist.edu.in

[2]Associate Professor, Department of Computer Science and Engineering, School of Computing, SRM Institute of Science and Technology, Kattankulathur Campus, Tamil Nadu, India. E-mail: saranyas6@srmist.edu.in

[3]Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur Campus, Tamil Nadu, India.E-mail: ss3742@srmist.edu.in

ABSTRACT

This paper depends on utilizing twitter for sentiment analysis of the views of people on COVID-19 vaccines and eagerness to have the jab. An ever increasing number of individuals have begun posting on the web about whether they are in favor of these vaccines or not. Many mediums are there which give these analysis among which one such medium is Twitter which has gotten very mainstream as of late. Twitter is a famous microblogging website where clients are permitted a constraint of 280 characters; this sort of limitation causes the clients to be compact just as expressive simultaneously. Thus, it turns into a rich hotspot for opinion examination and belief mining. In this examination, we will investigate the eagerness of having different vaccine by people and their recommendation to others after having the same. In the wake of applying different models and machine learning algorithms to tweets information, we have discovered that it is for sure conceivable to anticipate the vaccine is being favored or not on a large scale.

KEYWORDS

Sentimental Analysis, Tweepy, Textblob, Python.

Introduction

Sentiment analysis can be defined as detailed study of NLP which includes the computational studying of opinions, sentiments and feelings expressed in textual format. The idea with which we initialized this research was to be able to tell the average polarity of the views of individuals regarding getting the vaccine shots getting released by various medical organisations during this difficult time of COVID-19, so that they can have a better view and mindset of it's pros and cons and related side effects. By this research, people will be able to make a clear mindset and decision on, if they want to have vaccine or not. Sentimental Analysis or Opinion Mining focuses at learning people's views, attitudes and feelings towards an entity. The entity can represent individual, events or issues. So the best place for people to approach and publish about anything and how they feel of that product is one and only Twitter platform. It can provide us the attitude, emotions and assessments. An enormous amount of researches have been performed in the field of sentimental analysis. But majority of them are aligned on classification of formal and bigger pieces of textual data like reviews. With the mass popularity of social network platforms and microblogging websites and a large amount of data accessible from these resources, research projects on sentimental analysis have witnessed a gradual domain shift. Twitter is the bestsource for analysing the data of all types because:

- The Application Program Interface is neat and comes with rich developing tools.
- The data it provides is full of information's and has a data-structure that is perfect for for analysis.
- Twitter data is easily and freely available to all those with proper usage rights.

Implementation Techniques

In this research carried by us in analysing the views of people regarding vaccine shots, we have made use of Python and it's libraries namely Textblob, MatplotLib and apart from that we have also made use of Twitter API called as Tweepy.

TextBlob: Textblob is a python library that is used for analysing the data present in textual format. It provides an API for performing various natural language processes such as classification, sentimental analysis and speech recognition.

It works on the framework of NLTK. It returns us the levels of polarity and subjectivity. Polarity lies between $(-1,+1)$ with -1 meaning negative sentiments and +1meaning positive sentiments. Polarity is reversed by negation words. Subjectivity lies between $(0,1)$. Subjectivity deals with the information in detection if its fact based or personal. Higher the subjectivity means higher amount of personal and factual data.

Intensity is the other factor it returns us which determines the if the words modifies the next words e.g. very beautiful.
TextBlob performs various functions and steps in the backend to give us a result in the form of polarities.

1.) Data Cleaning: As the basic and initial process, TextBlob performs the cleaning of data. It consists of various aspects:
    a.) Removing @, #, and URLs from the tweets
    b.) Removing dates from the tweets
    c.) Extraction of interesting and non-trivial data from non-structured data.
2.) Tokenisation: It refers to the process of breaking down the sentences in to separate words. Token basically refers to a word and it's a pretty straightforward process with TextBlob.

Implementation of Tokenisation

```
from textblob import TextBlob
document = ("In computer science, artificial intelligence (AI), \
sometimes called machine intelligence, is intelligence \
demonstrated by machines, in contrast to the natural intelligence \
displayed by humans and animals. Computer science defines AI \
research as the study of \"intelligent agents\": any device that \
perceives its environment and takes actions that maximize its")
text_blob_object = TextBlob(document)
document_sentence = text_blob_object.sentences

print(document_sentence)
print(len(document_sentence))

Result = [Sentence("In computer science, artificial intelligence (AI),"
"sometimes called machine intelligence, is intelligence "
"demonstrated by machines, in contrast to the natural intelligence"
"displayed by humans and animals."), Sentence("Computer science defines AI"
"research as the study of "intelligent agents": any device that "
"perceives its environment and takes actions that maximize its)]
```

3.) Lemmatization: Since the words in a sentence may contain various forms of it, they need to be reduced to their root forms. This process of reducing them to their root forms is referred as Lemmatization. For performing lemmatization using TextBlob, we have to use Word object from it.

MatPlot Library: It is a library used for plotting using the python language and is an numerical extension of NumPy. It gives and object oriented API for making plots in application.

Tweepy: It is an open source platform that communicates with Twitter API by enabling Python.

Database Designing

1.) Data Training
2.) Data Testing
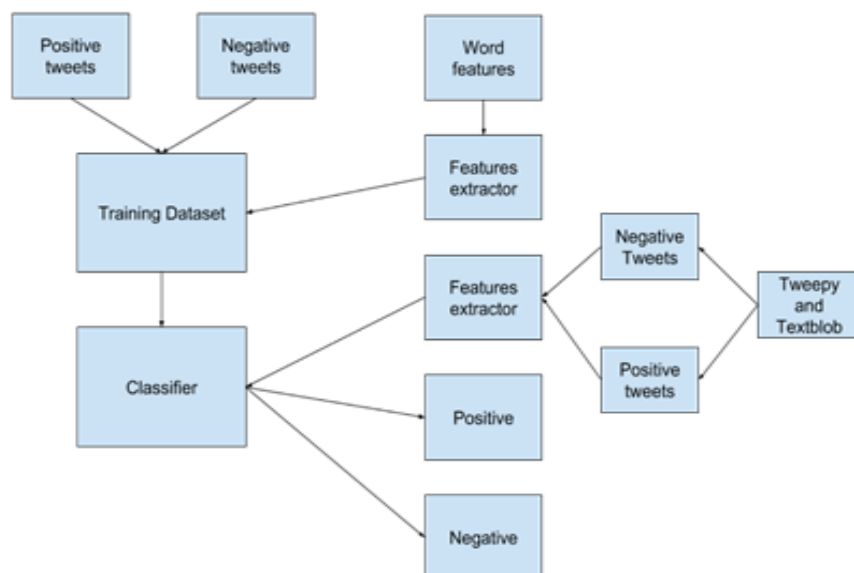
In the code that we have prepared, we have initialised the Tweepy using the 4 keys namely, consumer key, consumer secret key, access token, access token secret. Then we have created some variables that are initialised from zero. Then tweets are iterated and based on the polarity, each of the above variables are given increase in their values.
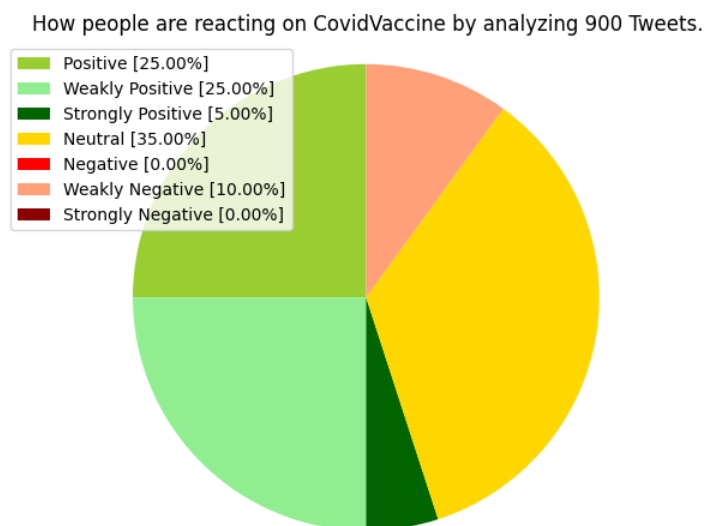
As we have to present the average sentiments of people, we have created some variables which will take the above variables as arguments. Polarity mean is then calculated.

Percentage of Polarity is Measured Using the Following Formula

```
def percentage(self, part, whole):
temp = 100 * float(part) / float(whole)
return format(temp, '.2f')
```

Then we have made use of Matplot library of python to create the plots using the percentages in the form of pie charts

How people are reacting on CovidVaccine by analyzing 900 Tweets.

Positive [25.00%]
Weakly Positive [25.00%]
Strongly Positive [5.00%]
Neutral [35.00%]
Negative [0.00%]
Weakly Negative [10.00%]
Strongly Negative [0.00%]

Pic. 1. Pie chart showing the percentage of people lying in each category

```
General Report:
Weakly Positive


Detailed Report: |
25.00% people thought it was positive
25.00% people thought it was weakly positive
5.00% people thought it was strongly positive
0.00% people thought it was negative
10.00% people thought it was weakly negative
0.00% people thought it was strongly negative
35.00% people thought it was neutral
```

Pic. 2. Final verdict based on overall average calculation

Result

After analysing around 900 tweets, we came to a result that average number of people are**Weekly positively** in favour of having the covid-19 vaccine shots.

References

[1] Shulong, T., Yang, L., Huan, S., Ziyu, G., Xifeng, Y., Jiajun, B., &Xiaofei, H. (2014). Interpreting the Public Sentiment Variations on Twitter. *IEEE Transactions on Knowledge and Data Engineering, 26*(5), 1158–1170. https://doi.org/10.1109/tkde.2013.116

[2] Yu, X., Liu, Y., Huang, X., & An, A. (2012). *Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain. IEEE Transactions on Knowledge and Data Engineering, 24*(4), 720–734. https://doi.org/10.1109/tkde.2010.269

[3] Wang, X., Wei, F., Liu, X., Zhou, M., & Zhang, M. (2011). Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. *In Proceedings of the 20th ACM international conference on Information and knowledge management*, 1031-1040. https://doi.org/10.1145/2063576.2063726

[4] Jose, R., &Chooralil, V. S. (2016). Prediction of election result by enhanced sentiment analysis on twitter data using classifier ensemble Approach. *In international conference on data mining and advanced computing (SAPIENCE)*, 64-67. https://doi.org/10.1109/sapience.2016.7684133

[5] Prakruthi, V., Sindhu, D., & Kumar, S.A. (2018). Real time sentiment analysis of Twitter posts. *In 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, 29-34. https://doi.org/10.1109/csitss.2018.8768774

[6] Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter sentiment analysis: The good the bad and the omg!.*In Proceedings of the International AAAI Conference on Web and Social Media*, *5*(1).

[7] Ren, F., & Wu, Y. (2013). Predicting user-topic opinions in twitter with social and topical context. *IEEE Transactions on affective computing*, *4*(4), 412-424.

[8] Dhanush, M., Ijaz Nizami, S., Patra, A., Biswas, P., &Immadi, G. (2018). Sentiment analysis of a topic on twitter using tweepy. *International Research Journal of Engineering and Technology, 5*(5), 2881.

[9] Hasan, A., Moin, S., Karim, A., &Shamshirband, S. (2018). Machine learning-based sentiment analysis for twitter accounts. *Mathematical and Computational Applications*, *23*(1), 11.