

MINING TWITTER DATA FOR SENTIMENT ANALYSIS

T.S.S.Rushitha, SK. Mubasshir, SK.MD.Gulab Shaanaz, M.Somendra Sai, B.Bhuvan Sai
SRM UNIVERSITY AP, ECE.

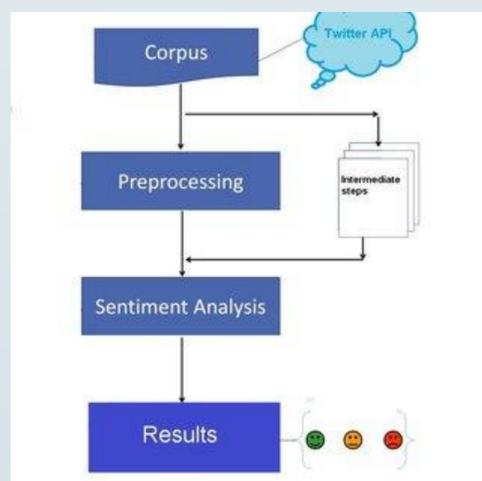


Abstract

In today's world, people are using social media platforms to share their ideas, experiences and post their opinions on a particular topic. Twitter, being one among several popular social media platforms, is a place where people often choose to express their emotions and sentiments about any topic so if an analysis is carried out on tweets, the emotions or feelings of a large group of people can be obtained. Sentiment analysis is an approach for studying people's emotions and sentiments in which the opinion of people can be analyzed and classified into positive, negative, or neutral. In this, tweets were extracted from Twitter by using twitter API. Later, to analyze the sentiments we performed pre-processing and cleaning steps to remove unwanted data from the dataset. VADER sentiment analysis is performed which gives the final output i.e., polarity of each word. The outcome of the analysis is depicted for positive, negative and neutral remarks about the opinions using visualization techniques such as pie chart and word cloud.

Objective

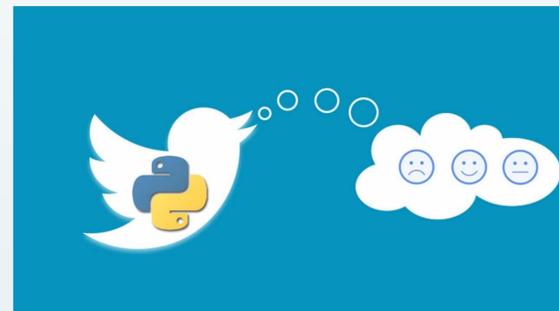
The main objective of this project is to perform Sentimental Analysis of people's opinions related to current issue of the globe,covid-19. The number of tweets on COVID-19 are increasing at an unusual rate by including positive, negative and neutral tweets, so if an analysis is carried out on tweets, the emotions or feelings of a large group of people towards this pandemic situation can be obtained.



Methodology

TWITTER DATA :

- Tweets related to COVID hashtags will be extracted from twitter through twitter API.



DATA PRE-PROCESSING :

- Data Preprocessing is the process of transforming the given data into usable or meaningful format for the model.
- To analyze the sentiments, pre-processing is necessary to remove unnecessary data like; removing all URLs (e.g. www.xyz.com), hash tags (e.g. #topic), targets (@username), hashtags, usernames and mentions, numbers and emojis, unnecessary spaces, punctuations marks and symbols.

SENTIMENT ANALYSIS :

- Sentiment analysis was done for the live tweets retrieved using twitter API. It is performed by importing sentiment intensity analyzer from NLTK package in python to get polarity scores i.e., positive, negative or neutral.



- The sentiments are based on polarity of the text. If the polarity is < 0 sentiment of the tweets is considered as Negative, if the polarity is 0 or 0.5 then the sentiment of the tweet is considered neutral and if polarity is >0.5 it is considered as positive.

Results

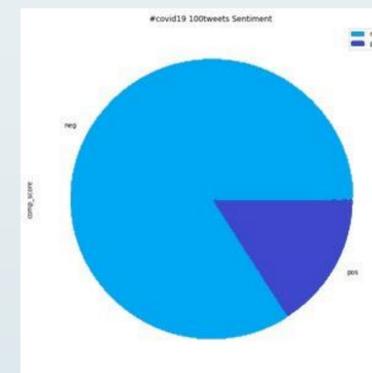
The results are classified and are visualized using pie chart and word cloud.

TEXT	SENTIMENT	COMPOUND	COMP_SCORE
Who will explain to the families that are suffering right now due to #COVID19	{'neg': 0.137, 'neu': 0.752, 'pos': 0.111, 'compound': -0.1531}	-0.1531	NEGATIVE
Wish <u>God</u> God Speed Recovery from #COVID19God Bless You to quickly be back in action	{'neg': 0.0, 'neu': 0.631, 'pos': 0.369, 'compound': 0.765}	0.7650	POSITIVE
Watch closely. Here's Matt Hancock wears his mask outside in the fresh air, walking alone	{'neg': 0.09, 'neu': 0.807, 'pos': 0.103, 'compound': 0.0772}	0.0772	NEGATIVE

Tabular representation which contains various columns depicting sentiment. Columns contain tweets, its respective score and percentage.

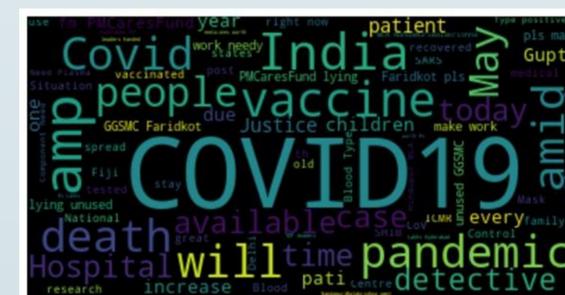
PIE CHART :

- To plot a pie chart we will use matplotlib, to know the sentiment of extracted tweets.



WORD CLOUD :

- Word Cloud gives more sense of the most frequent words used in tweets.



Conclusion

- Sentiment analysis is one of the most important applications of natural language processing (NLP) today. It can understand the emotions in people's words and speech, understand the true trend of comments. The methodology of this project can be used for sentiment analysis in various fields like product reviews, movie reviews. For future work, a system can be developed that can work efficiently and more accurate than the current system and also the project can be expanded which can help people to determine the quality of service from the analysis of tweets of the others experience. we can also expand this by showing different emotions along with three sentiments. The system can also be expanded for more than one language, such as Hindi, Punjabi, Urdu. In our current work we have used three class labels such as positive, negative and neutral. To be more specific, the labels can be extended to five class labels. They are very positive, positive, neutral, very negative, and negative. We have applied Multinomial Naive Bayes and VADER in the current work for sentimental analysis and found this process to be successful. In future more classifiers can be tried with advanced machine learning approaches like deep learning.

References

1. P. Arora and P. Arora, "Mining Twitter Data for Depression Detection," 2019 International Conference on Signal Processing and Communication (ICSC), 2019, pp. 186-189, doi: 10.1109/ICSC45622.2019.8938353.
2. V. Prakruthi, D. Sindhu and D. S. Anupama Kumar, "Real Time Sentiment Analysis Of Twitter Posts," 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), 2018, pp. 29-34, doi: 10.1109/CSITSS.2018.8768774.
3. S. Dhawan, K. Singh and P. Chauhan, "Sentiment Analysis of Twitter Data in Online Social Network," 2019 5th International Conference on Signal Processing, Computing and Control (ISPCC), 2019, pp. 255-259, doi: 10.1109/ISPCC48220.2019.8988450.
4. L. Mandloi and R. Patel, "Twitter Sentiments Analysis Using Machine Learning Methods," 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-5, doi: 10.1109/INCET49848.2020.9154183.
5. A. J. Nair, V. G and A. Vinayak, "Comparative study of Twitter Sentiment On COVID - 19 Tweets," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021, pp. 1773-1778, doi: 10.1109/ICCMC51019.2021.9418320.
6. C. Kaur and A. Sharma, "Social Issues Sentiment Analysis using Python," 2020 5th International Conference on Computing, Communication and Security (ICCCS), 2020, pp. 1-6, doi: 10.1109/ICCCS49678.2020.9277251.
7. K. S. Madhu, B. C. Reddy, C. Damarukanadhan, M. Polireddy and N. Ravinder, "Real Time Sentimental Analysis on Twitter," 2021 6th International Conference on Inventive Computation Technologies (ICICT), 2021, pp. 1030-1034, doi: 10.1109/ICICT50816.2021.9358772.