Twitter Sentiment Analysis

Raghul E

*Computer Science and Engineering Department*

*SRM Institute of Science and Technology*

Chennai

rahulelango1999@gmail.com

***ABSTRACT---* Sentiment analysis in social media is often used by companies to understand or analyse the feelings or sentiments sharing by people on social media about their product. It uses the natural language processing technique to analyse the data and bring out the results. Analysing of sentiments of people has become more important as it can be used to extract their reviews try to implement in their upcoming projects or update their product if there are more negative opinion. The natural language toolkit has made this project implement possible as it has various tools to analyse the textual data and come up with approximate results through various steps such as tokenization, eliminating stop words, and comparison with various emotional words to output whether the sentiment analysed is positive or negative or neutral one. Here I used twitter to analyse the sentiments as it is opted by most people to express their opinion on any topic. The GetOldTweets3 is a package which I used for importing large number of data for analysing the overall sentiment. Twitter API can also be used for this purpose but it has certain restrictions. It also keeps track of the huge number of tweets about a particular topic within a certain time periods using that we can get to know the overall sentiment of people within that specified time or date.**

1. INTRODUCTION:

Twitter has become most widely used platform for people to share their opinion, feelings, reviews, etc. Twitter has evolved a lot over the years with the addition of useful features to categories tweets such as search option, using which we can search anyone’s tweets based on hashtags or keywords related

to the tweet. The ability to categorise tweets efficiently has helped me build this project. As people are more inclined towards social media, it has become more important for organisation to analyse how people react to its company or product but this requires lot of time to analyse each and every tweet. People’s emotion/feeling is highly important for organisation to understand about their product’s success. This is where twitter sentiment analyse is implement.

Twitter sentiment analysis can be used by companies to understand peoples feeling/reaction towards something. It can be used to collect numerous tweets tweeted by people based on hashtags/keywords. Using that data, the tweets are analysed using machine learning (natural language processing) which would use certain algorithm to analyse the sentiment and it would output the common emotion. The output can be utilised to understand whether the tweets are positive/negative/neutral statement.

The analysed data would show series of values of each emotion used in a detailed bar graph which can be used for easier understanding of the inputted data that is loaded for analysis. While classifying the tweets based on token words, we can encounter few problems such as words which has two meanings or words that can be used to express both the emotion (positive and negative) depending upon the way it is used. This can affect the accuracy of the analysis which may lead to wrong interpretation of the output.

1. LITERATURE SURVEY:

In ‘Sentiment Analysis of Twitter data’ paper published in the year 2016 by Kiruthika M, Sanjana Woonna, Priyanka Giri, they have used the Twitter API to fetch the data and that data is cleaned (pre-processed) and it was used to categories the movie review into 3 (positive, negative and neutral).The key factor is that it has an UI for better understanding and the resultant data is shown in the form of tabular column with the name of the movie with its rating.

In ‘Machine Learning-Based Sentiment Analysis of Twitter Accounts’ paper published in the year 2018 by Ali Hasan, Sana Moin, Ahmad Karim and Shahaboddin Shamshirband, they have used sentiment analysis concept to analyse the political views by applying supervised machine learning algorithms such as Naïve Bayes and Support Vector Machines. The key differentiating factor is that they have used outputs obtained from 3 different analyzer (text blob, Sentiwordnet, WSD) to compare them based on the tweets twitted by peoples towards each political party.

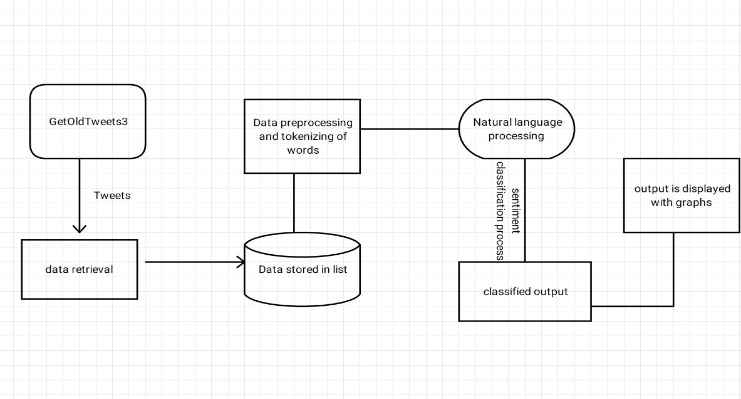
In ‘Sentiment Research on Twitter Data’ paper published in the year 2019 by A. Brahmananda Reddy, D.N.Vasundhara,P.Subhash, they have implemented the twitter analysis by categorising the inputted data into 7 categories such as strongly positive, positive, Weakly positive, neutral, strongly negative, negative and weakly negative. The twitter data is analysed using naïve bayes approach and the resultant data is displayed in pie chart for better understanding of the tweet’s sentiment. The key thing to note is the approach used to classify the emotions into 7 types.

The ‘Sentiment Analysis on Twitter’ paper was published by Akshi Kumar and Teeja Mary Sebastian in the year 2012, where they used the traditional twitter API to fetch the data and POS tagging for analysing tweets using R language.The key feature used is calculating the sentiment score using linear equation.

The ‘Opinion Mining and Sentiment Analysis on Online Customer Review’ paper was published by Santosh Kumar K L , Jayanti Desai , Jharna Majumdar in the year 2016, it uses the naïve bayes classification to text classification and opinion mining.Key things to note is the implemenatation of naïve bayes technique in opinon mining from online retail reviews.

1. METHODOLOGIES:

Twitter sentiment analysis is coded using python language with jupyter notebook as idle, it is implemented using natural language processing which is a part of machine learning. NLP uses natural language toolkit to analyse and classify the data which is fetched from the GetOldTweets3 package, it has the ability to interact with the machine and human using natural language and thus understanding the data and extracting valuable meaning from it. The NLP is mostly used by many organisations to extract information from the social media which are related to their products or brand. The idea behind NLP for sentiment analysis is that it scans for target words from the data and it classifies them into positive/negative/neutral type. Before that pre-processing is done to ensure the data is cleaned so that it does not hinder during the analysis.

This project uses ntlk i.e., natural language processing toolkit these are the package which is helpful in building it. The packages which are used are matplotlib for plotting and data visualisation, GetOldTweets3 to retrieve the tweets, stopwords (nltk) which is used to eliminate the stop words in the token, SentimentIntensityAnalyser to output the polarity and the sentiment of the data.

The GetOldTweets3 package is which is used to get the tweets from twitter.com.Earlier, the twitter API was used for same purpose but now it has certain limitations and it is a bit long process to gain access and some features such as number of the tweets and the tweets of specified date requires money to retrieve tweets. This package is free to use can it can get as much as tweets we want and it does not have any restriction on date, that means we can get old tweets of any given time and date. This package is very efficient and easy to use, it allows the user to specify the tweets by username, hashtag, language and location. The GetOldTweets3 package is an upgraded package of GetOldTweets.

1. DATASET:

Twitter sentiment analysis requires a dataset which is fetched from the GetOldTweets3 package. For a successful loading of the dataset there are few prerequisites which has to be inputted such as the search element/keyword/hashtag based on which the tweets have to be fetched then, the number of tweets required along with the date interval and language. Using these inputted values, the package would fetch all the data from official twitter website and it will be loaded for pre-processing and analysing of sentiment.

1. IMPLEMENTATION:

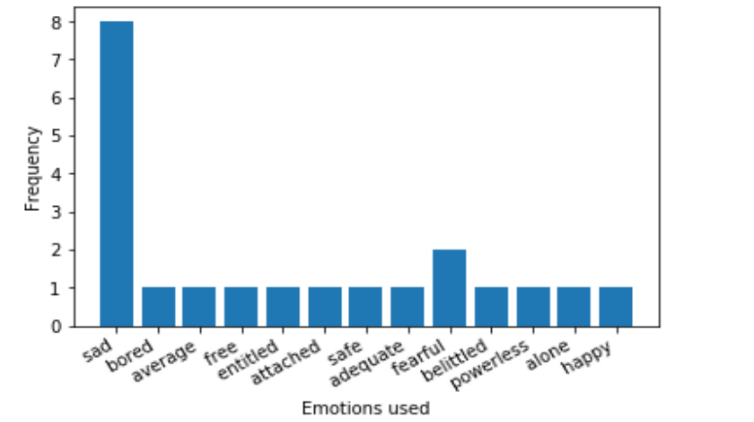
The first step in twitter sentiment analysis is loading the data for that the user has to input the specific tweet he/she wants using its username/hashtag/keywords and the number of tweets to be fetched using that the package would extract the data(tweets) from the twitter and stores them as a list. This list has to be pre-processed before moving into sentiment classification. In the pre-processing part, the data is first converted into lower case (case sensitive) and then all the punctuation and special characters are removed as they don’t involve in adding any emotion. Next, the data is analysed and all the words which don’t add any emotional meaning into the sentence are removed.

The cleaned data is passed into a function which would output the overall positive/negative/neutral percent of entire tweets using sentiment intensity analyser and prints the highest. Then, the cleaned data is split into individual words which are called as tokens, these tokens are then analysed with another data set which has a dictionary various emotional words with their respective emotion, target words are then categorised with the corresponding emotion and they are numbered based its frequency. The emotions along with their frequency are mapped on a bar graph which gives a further insight about the tweet.

The above figure is a data flow diagram for twitter sentiment analysis, where it shows how the data is passed and what happens during the execution of sentiment analysis. The first step in sentiment analysis is fetching the tweets from GetOldTweets3 package which is retrieved and stored as a list in, then that data is pre-processed before further execution .Once the pre-processing is done the data is tokenised and natural language processing is done in which the data is analysed using various packages. Finally, the classified data is sent for displaying the output where the final analysed value will be in the form of graphs which would provide easier understanding about the data.

1. RESULT AND DISCUSSION:

The result of twitter sentiment analysis will be a bar graph consisting of various emotions used in x axis and their frequency in y axis. We can configure the code to output only type of emotion(positive/negative/neutral) used and its percentage, using all these data we can analyse each aspect of the data such as its type of emotion used, number of times each emotion is encountered and the most used emotion.



The above sample results are displayed to provide an insight about how the result would be displayed. The first figure is a bar graph which has emotions used on x axis and its frequency on y axis, so using this graph we can easily conclude that the most used emotion is sad thus the overall opinion on that particular topic is negative which can be confirmed from the second figure. The second figure is a pie chart which comprises of positive, negative and neutral values with their percentage of frequency. This figure can be used to visualise how all these emotions are used, which emotion is used the most/least with values.



The third figure shows 2 values that is the most commonly used emotion and its percentage.

Using these results companies can understand the overall people’s emotions towards a particular product. This can help companies to know about their product whether it is a success or it needs any changes.

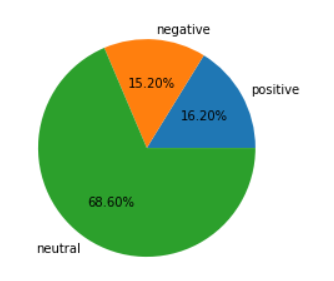
1. CONCLUSION AND FUTURE

SCOPE of WORK:

Twitter sentiment analysis can help organisation in many ways for a better implementation of their products as they can rely on people’s opinion. In this way both people and the organisation can be benefited, so understanding people’s emotion/sentiment has become a primary key for success for many companies. Twitter has become a reliant option in the social media cloud for public opinion and feedback and it has various features to scrap data for analysing and understanding. This project can be implemented using various ML and NL techniques.

The project can be improvised in the future by converting it into a website with interactive UI elements for simplified implementation. Further many features can be added such as the ability to analyse any text pasted into the field, and to compare the sentiment for tweets of different topic. The use of cloud storage can help users their old analysed dataset for future reference.

Acknowledgement:

****

I would like to express my gratitude and thanks to my parents and the members of organisation (Expert’s Hub) for giving me proper guidance and knowledge required to complete this project.

References:

1. A. Brahmananda Reddy, D.N.Vasundhara,P.Subhash “Sentiment Research on Twitter Data”, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8,Issue-2S11,September2019,<https://www.ijrte.org/wp-content/uploads/papers/v8i2S11/B11810982S1119.pdf>
2. Ali Hasan, Sana Moin, Ahmad Karim and Shahaboddin Shamshirband, “Machine Learning-Based Sentiment Analysis forTwitterAccounts”,Math.Comput.Appl.**2018**,23,11;doi:10.3390/mca23010011, <https://www.mdpi.com/2297-8747/23/1/11/pdf>
3. Kiruthika M, Sanjana Woonna, Priyanka Giri, “Sentiment Analysis of Twitter Data”, (IJIET Volume Issue 4 April 2016),

<http://ijiet.com/wp-content/uploads/2016/04/37.pdf>

1. Sentiment Analysis APIs Benchmark

<https://monkeylearn.com/blog/sentiment-analysis-apis-benchmark/>

1. Natural language toolkit documentation - <https://www.nltk.org/>
2. Matplotlib library documentation - https://matplotlib.org/
3. GetOldTweets3 package and its documentation/guide -

<https://github.com/Mottl/GetOldTweets3>

1. Methods to scrape twitter data-

<https://towardsdatascience.com/how-to-scrape-tweets-from-twitter-59287e20f0f1>

1. Twitter swot analysis –

<http://fernfortuniversity.com/term-papers/swot/1433/914-twitter.php>

1. Realtime sentiment analysis with Vader –

<https://towardsdatascience.com/almost-real-time-twitter-sentiment-analysis-with-tweep-vader-f88ed5b93b1c>

1. SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods-

<https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-016-0085-1>

1. Akshi Kumar and Teeja Mary Sebastian ‘Sentiment Analysis on Twitter’ IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3, July 2012 , <https://ijcsi.org/papers/IJCSI-9-4-3-372-378.pdf>
2. Santosh Kumar K L , Jayanti Desai , Jharna Majumdar ‘Opinion Mining and Sentiment Analysis on Online Customer Review’ 2016 IEEE International Conference on Computational Intelligence and Computing Research. <https://www.researchgate.net/publication/316902743_Opinion_mining_and_sentiment_analysis_on_online_customer_review>