

Tweets Based Sentiment Analysis

Gajendra R. Wani

Department of Computer Science and I.T., Bhusawal Arts, Science and P.O.Nahata
Commerce College, Bhusawal

ABSTRACT

Nowadays Twitter, Facebook, WhatsApp are getting so much attention from people and also, they are getting very much popular among people. People share their views, opinion, experiences on social media platform. Twitter is one of such platforms where common people shares their reviews in short blogs. These reviews are also been used by other people to get feedback about anything such as movie, product, politics and much more. Mostly people share the unbiased opinions about anything they wanted, that's why one can consider these reviews as more generalized and real one. This paper contain implementation of NLTK and python-twitter APIs to classify tweets.

Keywords — Twitter, Sentiment, NLTK, Python, Machine Learning, Positive, Negative, Neutral

I. INTRODUCTION

In today's modernised era the data collected from social media also plays a very significant role. But, the data available on social networking sites can be unstructured and structured. Almost 80% of the data on the internet is unstructured. Thus, to convert unstructured data into structured data and to do some sort of analysis on that data is very important.

Sentiment analysis provides many opportunities to develop a new application in the industrial field. Sentiment analysis has a big effect, like government organization and big companies, their desire is to know about what people think about their product, their market value [1][7]. The computational study of people's opinions, attitudes, sentiments and emotions expressed in written language is known as sentiment analysis or opinion mining[2]. In this, the subjective information in the text is analysed and then mine the opinion. Sentiment analysis is the procedure in which information is extracted from the opinions, and emotions of people in regards to events, entities and their attributes. In decision making, the opinions of others play an important role in making choices with regards to online shopping, choosing events, and products. There are various Machine Learning algorithms which are currently employed in the classification of tweets into positive and negative classes based on their sentiments, such as Baseline, Naive Bayes Classifier, and Support Vector Machine etc[1][3][7]. The actual implementation of text sentiment analysis of tweets is done with the help of NLTK and python-twitter APIs. This paper aim is to analyze real time tweets and classify them into three categories as positive, negative and neutral sentiments.

II. RELATED WORK

There is numerous research work done in field of sentiment analysis. But the variations in tweets has always been a challenge for the analysis. Because of the impartial nature of tweets, it has been in lime light for sentiment analysis of products, movies, popularity or anything of that sort. The paper presented by Alec Go et al.[1] , 2009 was classifying tweets into positive and negative classes using distant supervision .They have presented an approach for automatically classifying the sentiment of tweets with respect to a query term. They have presented results of machine learning algorithms (Naive Bayes, maximum Entropy, and SVM) for classification. The paper presented by Svetlana et al.[4] was about sentiment analysis of small texts which are informal. They have used 2 test suits one was with tweets and another was with SMS. They have collected tweets from twitter API and worked on it. The messages which are not containing words with polarity, were cut down. The paper presented by Hassan Saif et al.[5] , 2012 was about semantic sentiment analysis of twitter. They have explained an approach of adding semantics as additional features into training set for sentiment analysis. For each extracted entity its semantic concepts were added as an additional feature and the correlation of the representative concepts with positive or negative sentiments were calculated.

III. PROPOSED WORK

The whole project work has been done in python language. Our objective is to find out polarity of the tweets. The polarity of tweets is classified as positive, negative or neutral. This polarity is obtained by searching the keywords in the tweets. For this purpose, the related sentiments to the

polarity is collected in the experimental dataset. Fig 1 shows the sample sentimental dataset.

```
'forced': 'powerless',
'forsaken': 'hated',
'framed': 'cheated',
'free': 'free',
'free & easy': 'happy',
'frightened': 'fearful',
'frisky': 'happy',
'frustrated': 'angry',
'full of anticipation': 'attracted',
'full of ennui': 'apathetic',
'fuming': 'angry',
'funereal': 'sad',
'furious': 'angry',
'gallant': 'fearless',
'genial': 'happy',
'glad': 'happy',
'gleeful': 'happy',
'gloomy': 'sad',
'glum': 'sad',
'grief-stricken': 'sad',
'grieved': 'sad',
'guilt': 'sad',
'guilty': 'singled out',
'happy': 'happy',
'hardy': 'fearless',
'heartbroken': 'sad',
'heavyhearted': 'sad',
'hesitant': 'fearful',
'high-spirited': 'happy',
'hilarious': 'happy',
'hopeful': 'attracted',
'horny': 'lustful',
'horrified': 'fearful',
'hot and bothered': 'lustful',
'humiliated': 'sad',
'humorous': 'happy',
```

Fig 1 Snapshot of sentiment dataset

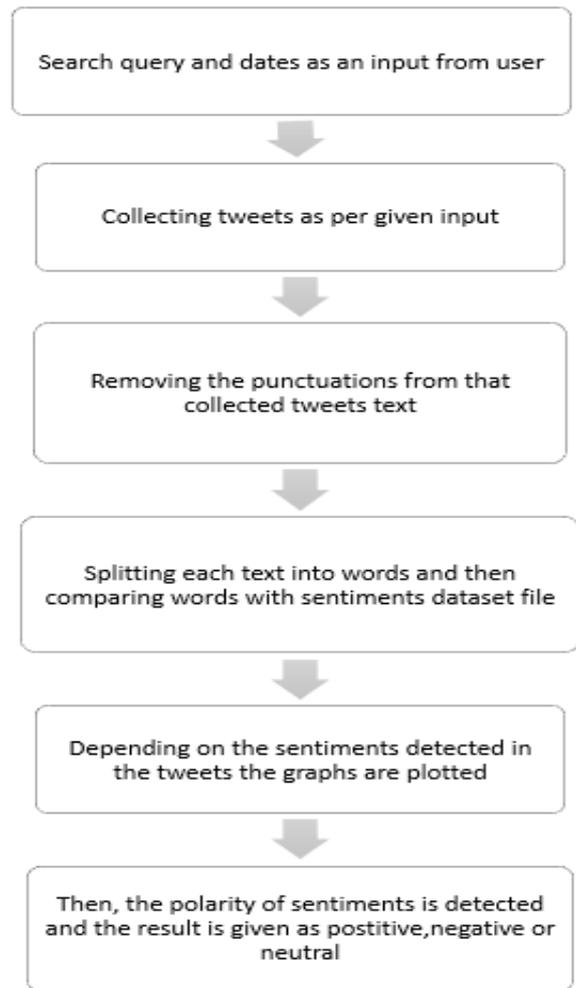


Fig 2 Identification of polarity of sentiments

Fig 2 represents pictorial view of propose approach.

Algorithm

- Step 1: User will provide search query and dates as an input.
- Step 2: All the tweets related to input will be collected in the text format.
- Step 3: Removing all punctuation from the collected tweets text.
- Step 4: This text will be then split into words.
- Step 5: All words from text will be converted into lower case.
- Step 6: These words are then compared with the sentiments in the dataset.
- Step 7: Then score will be calculated and based on this score polarity will be decided.
- Step 8: The score of each sentiment found is visualised in the form of graph.

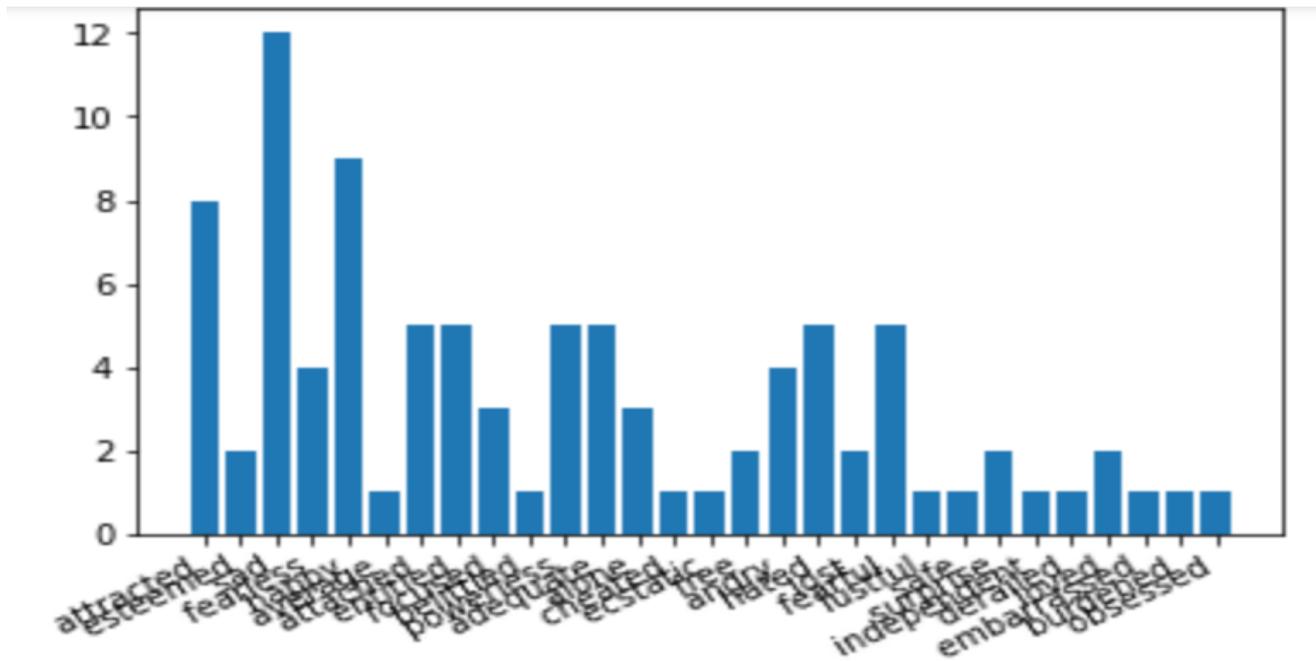
IV. RESULTS AND OBSERVATIONS

The sentiments are classified into 3 polarity positive, negative, and neutral.

Case 1: Positive sentiments

The given search query was “Women Education” and the dates was from 2020-01-01 to 2020-08-08 and the result given is positive sentiments. And the counter score as mention below.

```
{' sad': 12, ' happy': 9, ' attracted': 8, ' attached': 5, ' entitled': 5, ' powerless': 5, ' adequate': 5, ' hated': 5, ' fearful': 5, ' fearless': 4, ' angry': 4, ' focused': 3, ' alone': 3, ' esteemed': 2, ' free': 2, ' lost': 2, ' surprise': 2, ' loved': 2, ' average': 1, ' belittled': 1, ' cheated': 1, ' ecstatic': 1, ' lustful': 1, ' safe': 1, ' independent': 1, ' derailed': 1, ' embarrassed': 1, ' burdened': 1, ' obsessed': 1}}
```



Positive Sentiment

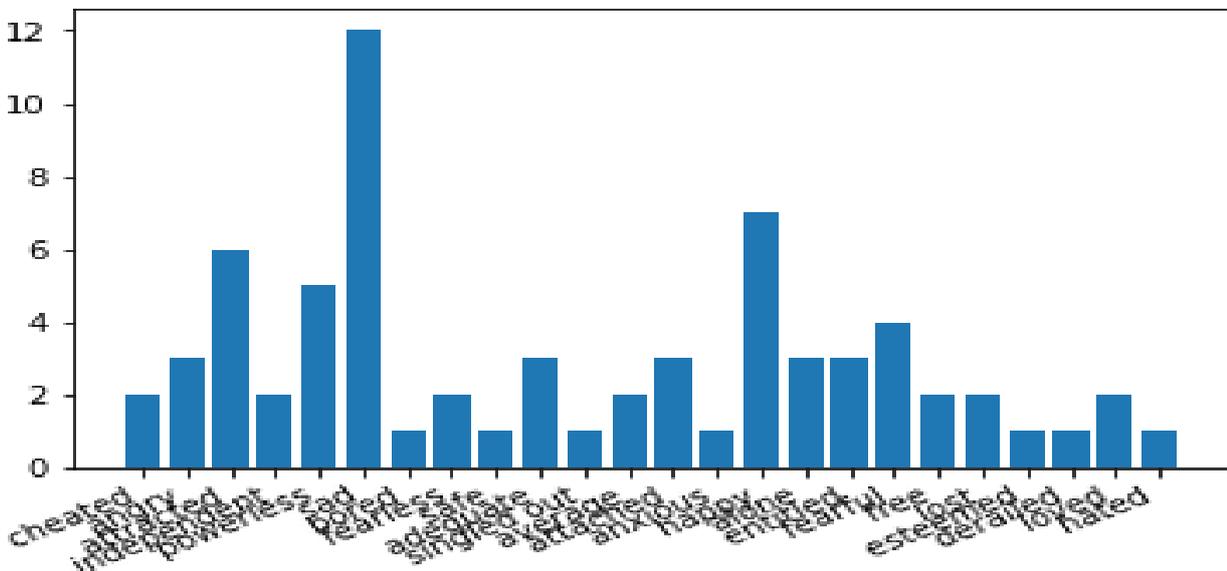
Case 2 : Negative sentiments

The given search query was “covid-19” and the dates was from 2020-01-01 to 2020-08-08 and the result given is negative sentiments. And the counter score as mention below. ({ sad': 12, ' happy': 7, ' attracted': 6, ' powerless': 5, ' fearful': 4, ' angry': 3, ' adequate': 3, ' attached': 3, ' alone': 3, ' entitled': 3, ' cheated': 2, ' independent': 2, ' fearless': 2, ' average': 2, '

free': 2, ' lost': 2, ' loved': 2, ' bored': 1, ' safe': 1, ' singled out': 1, ' anxious': 1, ' esteemed': 1, ' derailed': 1, ' hated': 1 }

Case 3 : Neutral Sentiment

When the counter score of positive will be equal to counter score of negative then it will give neutral polarity.



Negative Sentiment

V. CONCLUSION

In this paper, we discuss sentimental analysis to determine the polarity of the tweets among the specified time period. There is positive, negative or neutral polarity. Determination of the polarity plays a vital role in business. From the polarity, people's reactions can easily find out. People's reactions or feedback is really important to run the business or system. This approach can be applicable to the online feedback or review from customers relates to the product. According to the polarity, changes can be carried out.

The experimental result show that 'women education' is positive sentiment and 'Covid19' is a negative sentiment.

REFERENCES

- [1] A. Goel, J. Gautam and S. Kumar, "Real time sentiment analysis of tweets using Naive Bayes," 2016 2nd International Conference on Next Generation Computing Technologies (NGCT), Dehradun, 2016, pp. 257-261, doi:10.1109/NGCT.2016.7877424.
- [2] B. Nandi, M. Ghanti and S. Paul, "Text based sentiment analysis," 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, 2017, pp. 9-13, doi: 10.1109/ICICI.2017.8365326.
- [3] Alec Go, Richa Bhayani and Lei Huaug, Stanford university(2009), "Twitter Sentiment Classification using Distant Supervision ," in:The Third International Conference on Data Analytics.
- [4] Svetlana Kiritchenko ,Xiaodan Zhu ,Saif M. Mohammad Sentiment Analysis of Short Informal Texts. Journal of Artificial Intelligence Research 50(2014)723- 762.Pageno-725.
- [5] Saif, Hassan; He, Yulan and Alani, Harith (2012), "Semantic sentiment analysis of twitter," in: The 11th International Semantic Web Conference (ISWC 2012), 11-15 November 2012, Boston, MA, USA.
- [6] Bifet and E. Frank, "Sentiment Knowledge Discovery In Twitter Streaming Data", In proceedings of 13th International Conference of Discovery Science, Berlin,GermanySpringer,2010
- [7] Walaa Medhat a,* Ahmed Hassan b , Hoda Korashy b Sentiment analysis algorithms and applications:A survey . Ain Shams Engineering Journal. Page no-1098.