2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) | 978-1-6654-1451-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/Confluence51648.2021.9377048

Sentimental Analysis of Twitter Comments on Covid-19

Supriya Raheja
Department of Computer Science & Engineering
Amity University
Noida, India
supriya.raheja@gmail.com

Anjani Asthana
Department of Computer Science & Engineering
Amity University
Noida, India
anjanirocks21@gmail.com

***Abstract*— The rise of social media offers users to share their state of mind and opinions in all aspects of their life. Vast amount of textual content is available and needs techniques to use of information meaningfully by separating and examining them. Sentimental Analysis mines the opinion of human being for a viewpoint. In this Pandemic condition of Corona, whole world is sharing their opinions on the social media. This paper shows a analysis of sentiments to get the opinion of people if they are positivity during this situation or not. The paper is using the technique of polarity to know the opinion is positive, negative, or nonpartisan [1]. Three main keywords "COVID", "Corona virus" and "COVID-19" are used to check the polarity.**

Keywords—*Sentimental analysis, mining, twitter comments, Covid-19, coronavirus*.

I. Introduction

Corona virus the uncontrollable pandemic situation came in all over the word. It is a infection which majorly effects respiratory conditions of human beings. It was first recognized in December 2019 in Wuhan, China, and has since outspread universally. In March 2020, W.H.O. has declared it as pandemic. Due to world-wide spread, almost all countries have applied lockdown [2]. Lockdown means nobody is supposed to go outside their homes and to interact others. During March-May, almost entire word was house arrest and were working from homes. Some of the studies reported the psychological effect of lockdown on human behavior like anxiety, depression, frustration etc. All these behaviors badly impact the health of a human being [3]. While using social media, their comments reflect their behavior. So, by analyzing the social comments, we can identify the human behavior either positive, negative, and neutral [1]. Keeping this in mind, this research work, presenting a sentimental analysis of twitter comments to know the human opinion on Covid-19.

Due to the large number of posts on online media and knowing one's perspective about it is a very tedious task. Sentimental Analysis is a technique which analysis and does the classification of emotions using text analytics. So, this work is making use of sentiment analysis. In the paper, we have shown the textual analyses of Twitter information to know the public views tracking the advancement of fear, which is directly related with the vast spread of Coronavirus disease [4].

Rest of the paper is composed as follows: Section 2 discusses the literature review. Section 3 presents the chosen methodology for the work. Section 4 comprises the outcomes and analysis of twitter comments followed by conclusion in section 5.

II. Literature Review

Vast usage of social media by people for expressing their opinions in all aspects of life produces lot of information on Web. Analyzers and analysts are constantly dealing with this, how they can transform this massive information available on web to the useful information. Sentiment Analysis is one of the fields which helps the analyzers to watch the assessment of individuals on a particular watchword [5]. Several researchers are working in this field. Aliza et al. presented the design of a sentiment mining by extracting many tweets. They classified customers' reviews through tweets into notions like positive and negative [1].

Hamid and Johirul shown the application of sentiment knowing for Twitter information. They explored the sentiment mining for various topics from politics to humanity and concluded the impact of sentiment mining [6].

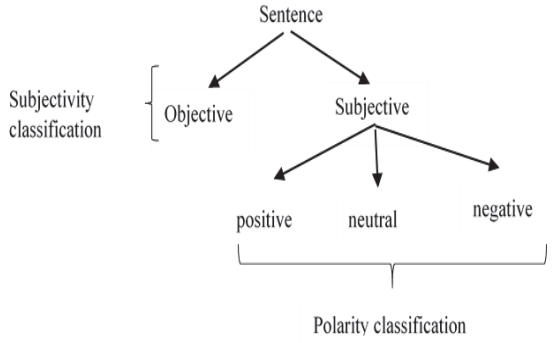
Jim et al. identified user sentiment based on the pandemic using Coronavirus specific tweets. They concluded insights into the progress of fear-sentiment over time as coronavirus approached at high levels in the U.S. of America [4]. Hassan et al. SentiCircles, a dictionary put together approach for sentiment knowing with respect to social media content like twitter tweets. The approach considered the knowing of the things that they themselves thought at both entity level and tweet-level [7].

Xiaolong et al. introduced the hashtag- level sentimental mining which consequently produces the general sentimental analysis for a given hashtag in a specific timespan, which particularly varies from the regular sentence-level and report level feeling investigation [8]. Rameshwer et al. presented the significance of notion investigation in forecasts [9]. As now a days, every human being in the world is working under COVID-19 pandemic condition. So, to evaluate people's notion over this condition is useful for maintain positive surroundings [10]. So, taking in the thought of this in mind, we are doing sentiment recognition for the same.

III. Sentimental Analysis

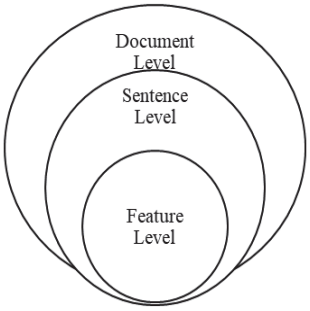
The key point while performing sentimental analysis at as a rule is to recognize either sentence is subjective or

objective. If the given line is delegated as objective, no other major undertakings are necessary, whereas if the line is delegated as subjective, its polarity (positive, negative, nonpartisan) must be known [11]. As illustrated in fig. 1, sentimental analysis does 2 types of separation on a sentence: subjectivity and polarity.



Sentence Classification

Client created content on Internet can be inspected principally in three extraordinary granularity levels namely documentation level, sentence level and feature level [12] as shown in fig 2.



Different Granularity Levels of Sentimental Analysis

- Document Level: It is used to classify the entire document into notions like positive, negative, or unbiased.
- Sentence Level: It divides the document into multiple sentences and performs the classification at sentence level into positive, negative, or neutral [12].
- Feature Level: It uses one feature and provides the results with respect to selected feature. For instance, the sentence "The iPhone is awesome, however they despite everything need to chip away at battery life and security issues" assesses 3- perspectives: “iPhone” (positive), “battery life” (negative), and” security” (neutral) [13].

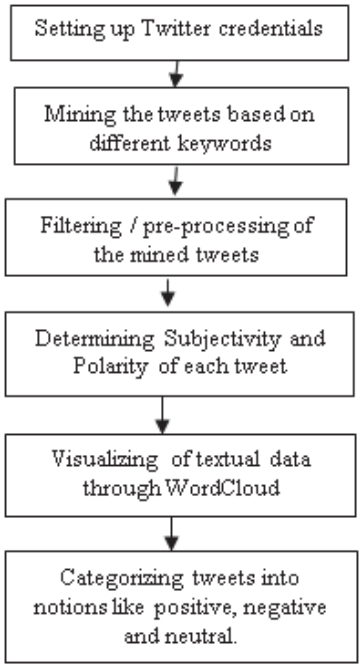
A. Handling of Emojis

The quantity of positive and negative emojis in the tweet are checked and the accompanying standards are applied [14]:

- If a tweet consists of minimum of one positive emojis and no negative emojis, it will be termed as positive.
- If a tweet consists of minimum of one negative emojis and no positive emojis, it will be termed so as negative.
- If not even one of the other principles above can be applied, the tweet is marked as obscure [14].

IV. METHODOLOGY USED

Figure 3 Displays the steps of proposed work. This part shows the first three steps one by one.



Flow of Proposed Work

B. Setting Twitter Credentials and Mining Tweets

There is a need of authenticity from Twitter website for accessing the information. So, first step is to set the twitter credentials to access the twitter comments. The key focus of the Sentimental Analysis is on textual data which requires more text processing [15]. In the present work, we prepared testing data with respect to three search keywords “COVID”, “Coronavirus” and “Covid-19”. These three keywords are commonly and mostly used by users on social media. So, this work tried all these keywords as search string.

C. Preprocessing of Data Set

Before performing classification on any data set, pre-processing is required which means cleaning of data. In our work, pre-processing is to get rid of the material or content that is not at all required in sentimental analysis like punctuations, images, hyperlinks [16,15] etc. Figure 4. illustrates the snippet for cleaning the data.

```
def cleanTxt(text):
  text=re.sub(r'@[A-Za-z0-9]+','',text)#removed mentions
  text=re.sub(r'#', '',text)#removed #
  text=re.sub(r'RT[\s]+', '',text)#removing RT
  text=re.sub(r'https?:\/\/\s+', '',text)#remove the hyper link
  text=re.sub(r'_:+"' , '',text)
  return text

df['Tweets']=df['Tweets'].apply(cleanTxt)
df
```

Fig. 4 Cleaning of mined data

V. Results and Analysis

This part of the paper shows the outcomes and analysis of sentimental analysis on twitter comments [1,6,7] for three keywords "COVID", "CORONAVIRUS" and "COVID-19".

A. Sentimental Analysis for Keyword "COVID"

To represent how many times the word had come "COVID" in a twitter dataset, we have made a WordCloud as shown in fig. So, through the WordCloud, we can see the frequency of the keyword "COVID" [4,17]. After that, we have computed the subjectivity vs. polarity for the same and illustrated through the scatter plot in fig. 5. It uses the cartesian coordinates to display the values for 2 constants in the data set [18]. Subjectivity checks that the particular word is subjective or objective. However, polarity tells us about the positive and negative behavior of human being for respective keyword. in the polarity column, there is a point "zero" [18,19]. So, on left of zero all the points on the graph signifies non positive comments and right to zero positive comments.

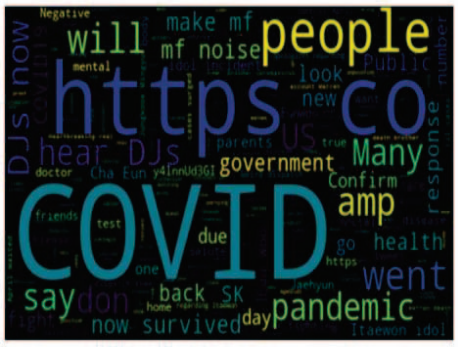


Fig. 5 WordCloud for Keyword "COVID"

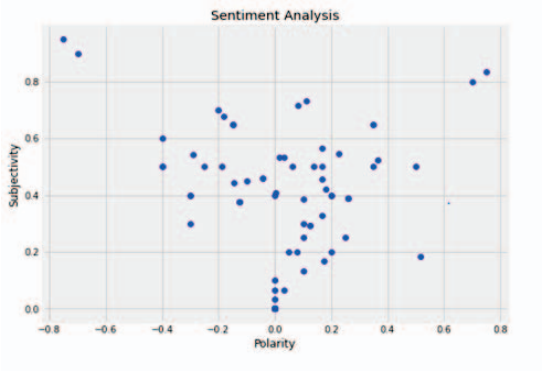


Fig. 6 Polarity vs. Subjectivity for keyword "COVID"

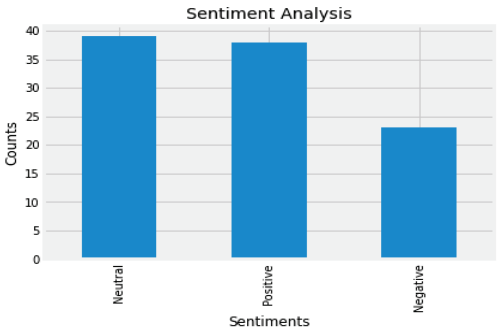


Fig. 7 Sentiment Analysis for keyword "COVID"

Figure 7 shows the sentiment analysis [7] in terms of percentage of positive and negative tweets in the respective data set. Here, there are 38% positive, 23% negative and 39% neutral tweets for the keyword "COVID". We can clearly see that the number of neutral tweets is higher than the total number of positive tweets, and the number of positive tweets is higher than the negative tweets. Similarly, we have performed the sentimental analysis for the two words: "CORONAVIRUS" and "COVID - 19". Fig.8 represents the wordcloud [4,17] and fig. 9 represents the subjectivity and polarity [11] for keyword "Corona virus".

Fig. 8 Word Cloud for "corona virus"

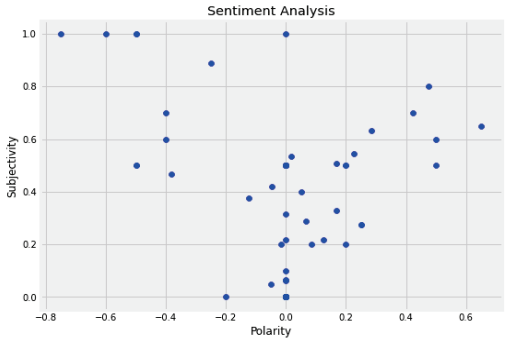


Fig. 9 Subjectivity vs Polarity for “corona virus”

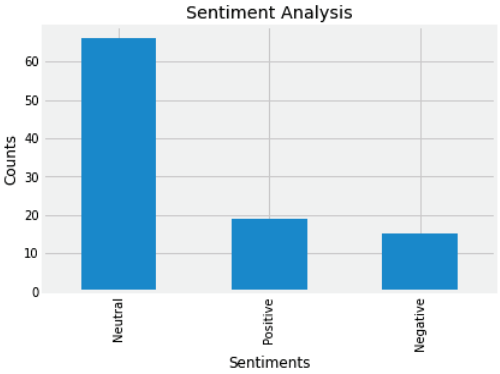


Fig. 10 Sentiment Analysis for keyword “corona virus”

Figure 10 shows the percentage of positive and negative tweets in the respective data set. Here, there are 19% positive and 15% negative tweets. By this we can concludingly say that there are 66% neutral tweets.

Similar steps we have performed for the keyword “COVID-19”. Fig. 11 and fig. 12 shows the wordcloud and the subjectivity and polarity for the same search string.

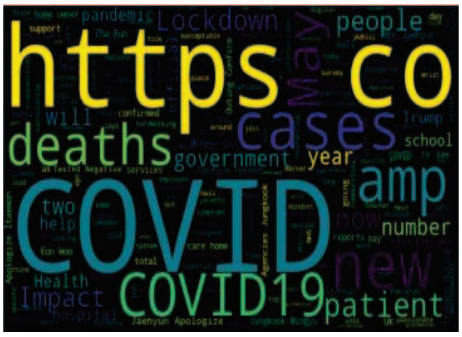


Fig. 11 Word Cloud for “COVID19”

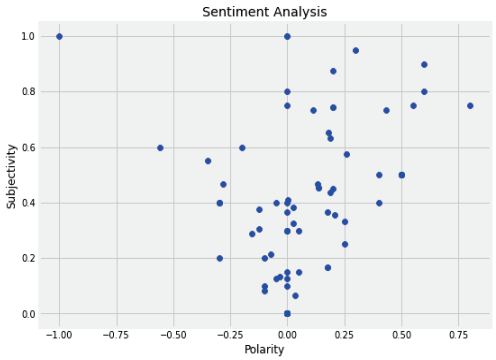


Fig. 12 Subjectivity vs Polarity for “COVID19”

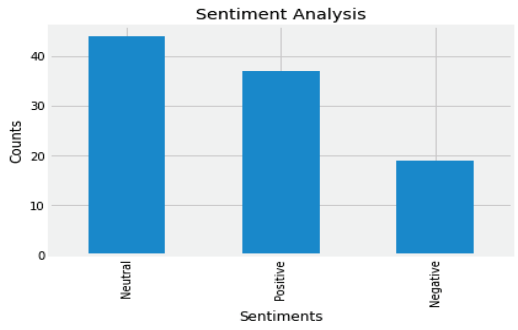


Fig. 13 Sentiment Analysis for keyword “COVID19”

Figure 13 illustrates the percentage of positive and negative tweets in the respective data set. In this particular case, there are 37% positive, 19% negative and 44% neutral tweets. Here the number of neutral tweets is higher than the number of positive tweets, and the number of positive tweets is much higher than the negative tweets. From the sentimental analysis for all three keywords, we have seen that the percentage of neural tweets is higher as compare to positive and negative tweets. However, the percentage of positive tweets is more as compare to negative tweets which signifies that even in this chaotic infection spread out conditions [2,3] , people are maintaining positive as well as neutral attitude.

VI. Conclusion

The paper talks about the criticalness of social network analysis. Twitter is one of the important media which is used by people to share their opinions. This paper performed the sentimental analysis for three important keywords (COVID, CORONA VIRUS, COVID – 19) of COVID-19 pandemic condition on more than 370 tweets from twitter. Based on these keywords there are around 31% of positive tweets and 19 % of negative tweets. 50% of neutral tweets that is half of the tweets on twitter based of these respective keywords are neutral in sentiments. The polarity results concluded that the neutral sentiments are high over positive and non-positive sentiments for COVID situation.

REFERENCES

[1] S. Aliza, C. Nadam, and S. Basr,. "Twitter sentiment analysis." In Proceedings of the 6th International conference on Information Technology and Multimedia, pp. 212-216. IEEE, 2014.

[2] COVID-19 pandemic in India. Wikipedia. Retrieved June 22, 2020, from https://en.wikipedia.org/wiki/COVID-19_pandemic_in_India.

[3] J. Samuel, G. G. Ali, Md Rahman, Ek Esawi, and Yana Samuel, "Covid-19 public sentiment insights and machine learning for tweets classification.", Information, vol. 11, no. 6, pp. 314, 2020.

[4] https://en.wikipedia.org/wiki/COVID-19_pandemic_in_India

[5] Samuel, J., Ali, G. G., Rahman, M., Esawi, E., & Samuel, Y. (2020). Covid-19 public sentiment insights and machine learning for tweets classification. *Information*, *11*(6), 314.

[6] Patel, A. P., Patel, A. V., Butani, S. G., & Sawant, P. B. (2017). Literature Survey on Sentiment Analysis of Twitter Data using Machine Learning Approaches. *IJIRST-International journal for Innovative Reasearch in Scinece & Technology, Volume3*, (10).

[7] Bagheri, H., & Islam, M. J. (2017). Sentiment analysis of twitter data. *arXiv preprint arXiv:1711.10377*.

[8] Saif, H., He, Y., Fernandez, M., & Alani, H. (2016). Contextual semantics for sentiment analysis of Twitter. *Information Processing & Management*, *52*(1), 5-19.

[9] Wang, X., Wei, F., Liu, X., Zhou, M., & Zhang, M. (2011, October). 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* (pp. 1031-1040).

[10] Singh, R., Singh, R., & Bhatia, A. (2018). Sentiment analysis using Machine Learning technique to predict outbreaks and epidemics. *Int. J. Adv. Sci. Res*, *3*(2), 19-24.

[11] Hamzah, F. B., Lau, C., Nazri, H., Ligot, D. V., Lee, G., Tan, C. L., & Shaib, M. K. B. M. (2020). CoronaTracker: worldwide COVID-19 outbreak data analysis and prediction. *Bull World Health Organ*, *1*(32).

[12] Liu, B. (2010). Sentiment analysis and subjectivity. *Handbook of natural language processing*, *2*(2010), 627-666.

[13] Jagtap, V. S., & Pawar, K. (2013). Analysis of different approaches to sentence-level sentiment classification. *International Journal of Scientific Engineering and Technology*, *2*(3), 164-170.

[14] Yaqub, U., Sharma, N., Pabreja, R., Chun, S. A., Atluri, V., & Vaidya, J. (2018, May). Analysis and visualization of subjectivity and polarity of Twitter location data. In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age* (pp. 1-10).

[15] Shiha, M., & Ayvaz, S. (2017). The effects of emoji in sentiment analysis. *Int. J. Comput. Electr. Eng.(IJCEE.)*, *9*(1), 360-369.

[16] Meduru, M., Mahimkar, A., Subramanian, K., Padiya, P. Y., & Gunjgur, P. N. (2017). Opinion mining using twitter feeds for political analysis. *Int. J. Comput.(IJC)*, *25*(1), 116-123.

[17] Angiani, G., Ferrari, L., Fontanini, T., Fornacciari, P., Iotti, E., Magliani, F., & Manicardi, S. (2016, September). A Comparison between Preprocessing Techniques for Sentiment Analysis in Twitter. In *KDWeb*.

[18] Ramsden, A., & Bate, A. (2008). Using word clouds in teaching and learning. *Retrieved on September*, *1*, 2013.

[19] Touchette, P. E., MacDonald, R. F., & Langer, S. N. (1985). A scatter plot for identifying stimulus control of problem behavior. *Journal of applied behavior analysis*, *18*(4), 343-351.

[20] Moon, K. R., van Dijk, D., Wang, Z., Gigante, S., Burkhardt, D. B., Chen, W. S., ... & Ivanova, N. B. (2019). Visualizing structure and transitions in high-dimensional biological data. *Nature biotechnology*, *37*(12), 1482-1492.