

“Data Visualization on the Tweet Data using Geo Location on the Canadian Map”

Deep Patel
Department of Computer Science
Dalhousie University
deep.patel@dal.ca

Abstract

In today's digitalized world of technology, people commute on social media more than they commute in person as it reaches more people. Many of them like to share their thoughts to support or deny the ongoing events, share their location, share their happiness, and many more important details on social media platforms. That data is available and can be used to analyze and reflected as a visualization. That data can be represented to understand their thoughts and emotions through online public commutes such as tweets [1].

Introduction

In this project, the main focus is on representing human emotions over the map to get the knowledge of overall reactions from the area. Certain filtering options can be used to generate the map on certain data such as province-wise, city-wise, date-wise polarity, and also with keywords. Using all those data there will be a chance to visualize the data as per requirements to understand human emotions in a better way. Considering one scenario about the election can help determine which candidate has a better chance to win the election based on a sentimental analysis of tweets. Visualization of time series can be also used to represent the popularity of candidates to determine the peak period of a particular candidate [1].

Implementation

The implementation task for this project has been carried out in three steps as given below;

1. Data Gathering

First and foremost, tweet gathering was a challenging task to provide analysis on this. After getting a Twitter developer account and using an access token, access secret token, consumer key, and consumer secret token I extract tweets using python code (Tweepy library) for applying query queries on data that contains 3 words such as “COVID-19”, “Work From Home”, “Ukraine”. Inside each tweet's data section, we can find text written in a tweet which will help in determining the polarity of the text to understand the emotion of the person. Here is the sample extracted data from python code for tweets as shown in figure (1). The link for python code that can extract tweets is provided as a GitLab project (tweets extraction.py [function get_tweets]) with a link to the extracted data.

```
[Status(api=tweepy.api.API object at 0x808081B7F8D15D68), _json={'created_at': 'Sun Apr 03 19:23:31 +0800 2022', 'id': 1510699537832333322, 'id_str': '1510699537832333322', 'text': 'Are nyc homeless becoming more unhinged by the day or was it just from wfh for so long that I forgot?', 'truncated': False, 'entities': {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': []}, 'metadata': {'iso_language_code': 'en', 'result_type': 'recent'}, 'source': '<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>', 'in_reply_to_status_id': None, 'in_reply_to_status_id_str': None, 'in_reply_to_user_id': None, 'in_reply_to_user_id_str': None, 'in_reply_to_screen_name': None, 'user': {'id': 1501788769116114945, 'id_str': '1501788769116114945', 'name': 'Seth Bregman', 'screen_name': 'sethbregman', 'location': '', 'description': 'lmm consumer data room constructor; plug for 0 value add family office term sheets; closed count 2-28', 'url': None, 'entities': {'description': {'urls': []}}, 'protected': False, 'followers_count': 23, 'friends_count': 114, 'listed_count': 0, 'created_at': 'Thu Mar 10 05:15:36 +0800 2022', 'favourites_count': 411, 'utc_offset': None, 'time_zone': None, 'geo_enabled': False, 'verified': False, 'statuses_count': 103, 'lang': None, 'contributors_enabled': False, 'is_translator': False, 'is_translation_enabled': False, 'profile_background_color': 'F5F8FA', 'profile_background_image_url': None, 'profile_background_image_url_https': None, 'profile_background_tile': False, 'profile_image_url': 'http://pbs.twimg.com/profile_images/1501791275414081536/oq7b62n_normal.jpg', 'profile_image_url_https': 'https://pbs.twimg.com/profile_images/1501791275414081536/oq7b62n_normal.jpg', 'profile_banner_url': 'https://pbs.twimg.com/profile_banners/1501788769116114945/1644889791', 'profile_link_color': '1DA1F2', 'profile_sidebar_border_color': 'C0DEED', 'profile_sidebar_fill_color': 'DDEEF6', 'profile_text_color': '333333',
```

Figure 1: Tweet Extraction Example

2. Data Processing

Data processing has been achieved by the following steps;

Step-1: Data Cleaning

Before using the data for visualization data cleaning is important which involves RemoveURL: URL starts with HTML, www, Remove Noise Words: To remove emojis, hashtags, tagged names, images, Remove Numbers. To efficiently gather the geographical location some data manipulation is performed for deciding the location from which the tweet is present.

Step-2: Polarity Calculation

To showcase the emotions based on geographical location, I used the Text Blob python library which provided the polarity of each tweet.

- 1) Polarity > 0: Positive tweet
- 2) Polarity < 0: Negative tweet
- 3) Polarity = 0: Neutral tweet

I studied the internal functionality of the calculation of polarity and found there is a default bag of positive words and negative words are defined. Tweets are being split into words and then there is a calculation of several positive and negative words in the sentences and based on the maximum number the polarity is being decided. For the currently ongoing event on “Ukraine”, I filtered tweets with “Ukraine” and found below mentioned polarity shown in figure (2) for tweets in Canada.

```
Positive tweets percentage: 31.586849315868493 %  
Negative tweets percentage: 13.698630136986301 %  
Neutral tweets percentage: 54.794528547945284 %  
  
Positive tweets:  
RT @Lokims94: L'ukrain qui demande l'aide des soldats Sénégalais .. c'est plus gros manque de respect que j'ai pu voir en 2022  
@rightwingnutrs Before state of union NBC had muh here is ackshual history of Ukrain as proud independent nation Pu... https://t.co/84naDBaVM5  
ukrain ready to stop war but russia not ready  
this war result world war
```

Figure 2: Polarity for tweets

Step-3:

1. Tweets Sentiments based on Province (Pie Chart):

After finding sentiments of tweets, province-wise analysis is being provided where I calculated the total number of tweets for a particular province and the percentage distribution of sentimental tweets. Below mentioned figure (3) showcases the JSON data I used to create the chart.

```
[{"Percentage": 0.2, "location": "Ontario", "Type": "negative"},  
{"Percentage": 0.3, "location": "Ontario", "Type": "positive"},  
{"Percentage": 0.5, "location": "Ontario", "Type": "neutral"}]
```

Figure 3: Pie Chart Data

2. Words analysis based on polarity and frequency (Word Cloud):

Another visualization includes a province-wise word cloud where I calculated the volume of a particular word is being used and the occurrence of the words in a tweet that contains positive, negative, and neutral sentiment. Based on that I calculated the sentimental score. Below mentioned figure (4) showcases the JSON data I used to create the chart.

```
{
  "label": "Mask",
  "volume": 100,
  "sentiment": {
    "positive": 65,
    "neutral": 15,
    "negative": 20
  },
  "sentimentScore": 65,
  "location": "Ontario",
}
```

Figure 4: Word Cloud Data

3. Time Series Data

To collect the data for analysis of Canadian people's sentiments for each day during these events I have collected tweets for each date and stored sentiment value for each day. Below mentioned figure (5) showcases the JSON data I used to create the chart.

```
date,value
2020-03-02,-89
2020-03-03,0
2020-03-04,0
2020-03-05,-100
2020-03-06,79
```

Figure 5: Time Series Data

4. Donut Data

To collect the data for analysis of Canadian people's sentiments for each day during these events I have collected tweets for each date and stored sentiment value for each day. Below mentioned figure (5) showcases the JSON data I used to create the chart.

3. Data Visualization

The Twitter data has been fetched for either of 3 events – “Ukraine”, “Work from Home”, and “Covid-19” from the data gathering and data processing section. The final aim is to provide options to the user for choosing events to see emotions by geographic location from Tweeter Data Analysis and eventually based on the selection of province the visualization will be provided. Also, for the current date, the province-wise donut chart will be displayed based on the total happiness index.

When the user loads the website; the user will see the Canada map and a dropdown menu, in which 3 options are provided: Canada data, Ukraine data, and Work from Home data. Based on a selection below visualizations are provided:

[1] Canada Map

The complete map of Canada has been generated using d3 version 7 using the coordinated locations of each province. The entire map has been generated using NodeJS and the

functionality of d3 with SVG and displayed on the chrome browser with a view box of a particular size.

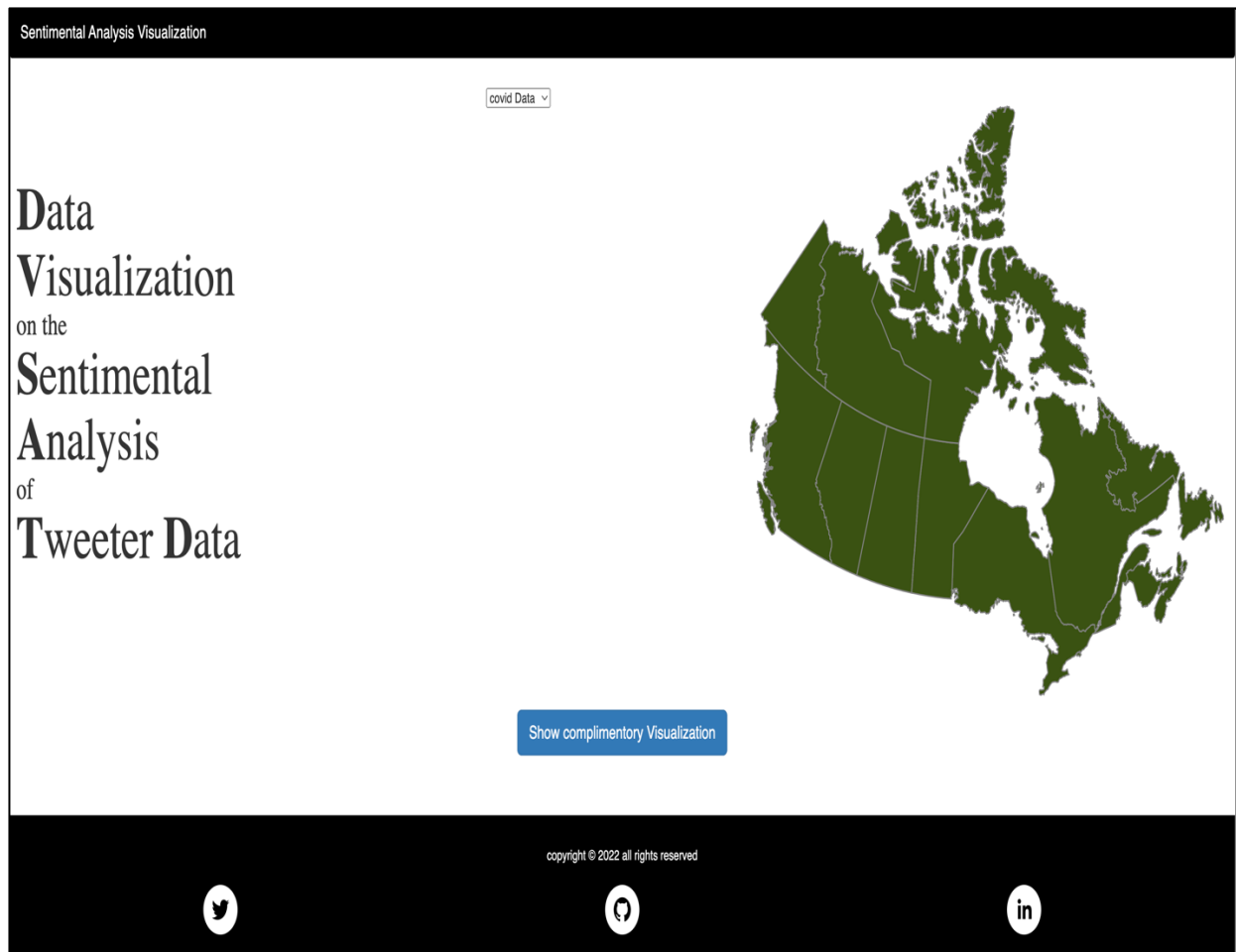


Figure 3: Home Page

1. Provided mouse hover effect

The mouse hover effect has been provided with a simple functionality of fetching the id of that particular block and changing its color. The final plan is to show details of the province and processed values from tweet data on the tooltip.



Figure 4: Canadian Map Visualization Mouse Over

2. Provided mouse click effect

The on-click event has been provided with the normal functionality of route to another page to showcase the province-wise sentimental tweet analysis.

[2] Donut Chart

As a complementary chart, a donut chart is generated that represents current province-wise covid-19 cases in percentage in Canada on the current date. To gather this data, I have used Covid-19 Current province-wise case API of government cite.

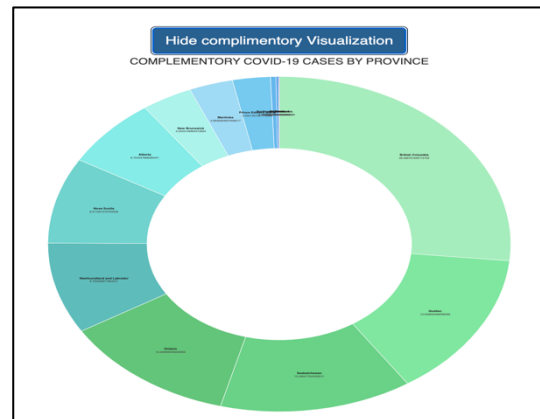


Figure 5: Complimentary Visualization for Covid Data

[3] Pie Chart

The pie chart will reflect the positive, negative, and neutral effects of some events based on the tweeter data. Currently, data has been extracted from the tweeter data using python code and the Node page has been created for the selected province.

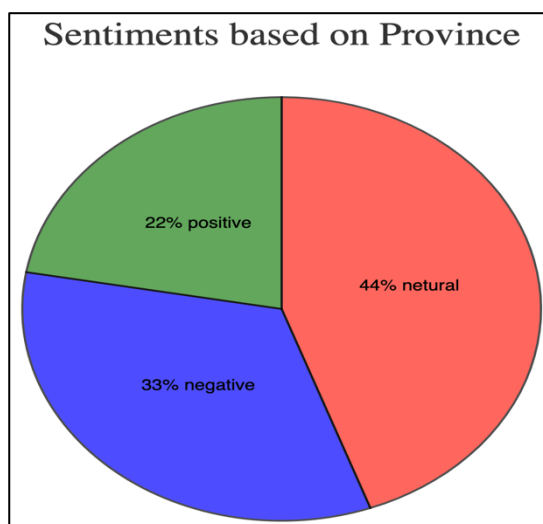


Figure 6: Sentiment Based Pie chart Visualization

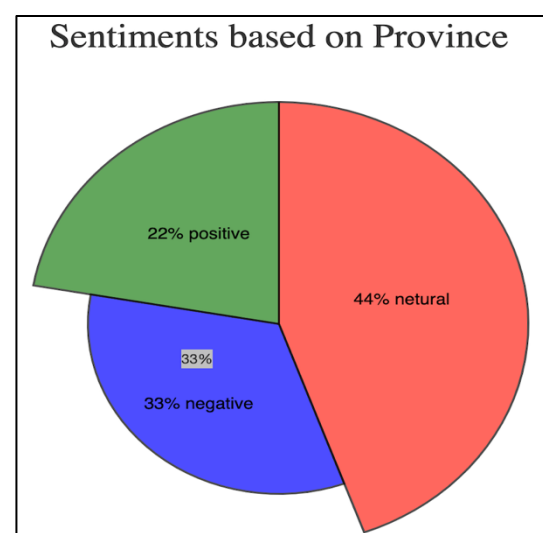


Figure 7: Sentiment Based Pie chart Visualization on Mouse Hover Effect

Some visualization effects are also applied to the Pie Chart as shown in figure (7):

1. Provided mouse hover effect: On mouse hover the radius of the pie chart increases to provide the idea of the current mouse movement on the chart.
2. Provided tooltip effect: On mouse hover the percentage distribution of the data is displayed as a tooltip.

[4] Word Cloud

The word cloud visualization is provided to analyze the sentiments based on words. Currently, data has been extracted from the tweeter data using python code the and Node page has been created for the selected province.

WordCloud based on polarity and frequency

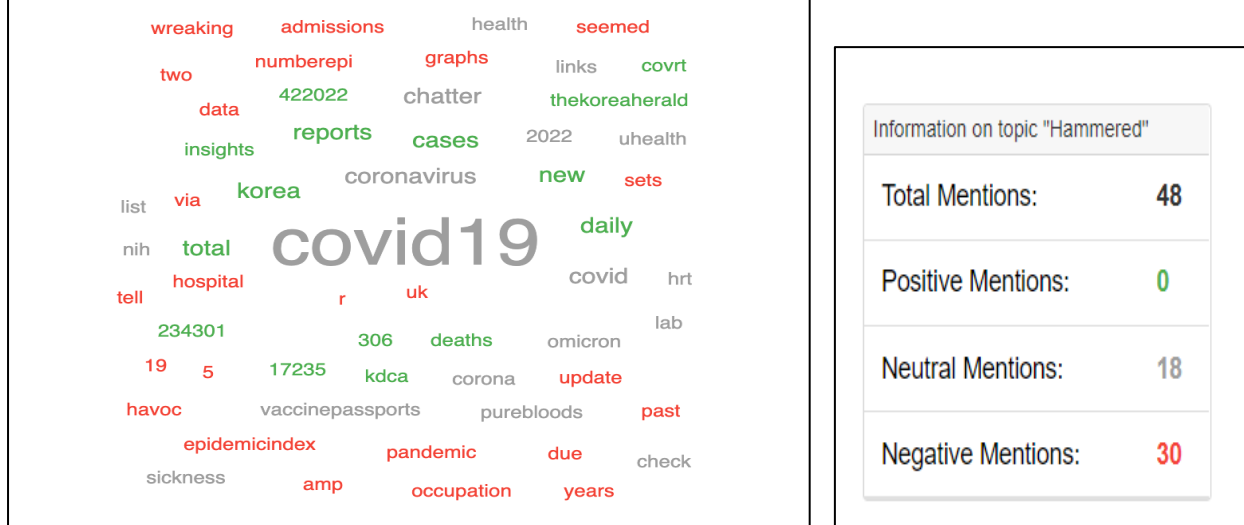


Figure 8: Word Cloud Data Visualization based on Polarity Frequency

Some visualization effects are also applied to the Word Cloud as shown in figure (8):

1. Provided Word Color based on sentiment: The word cloud contains
Word Color Red: Word used for majorly have Negative Sentiment in a tweet
Word Color Green: Word used for majorly have Positive Sentiment in a tweet
Word Color Grey: Word used for majorly have Neutral Sentiment in a tweet
2. Provided Word Size based on volume: Once the tweets are extracted frequency of words are being calculated which showcases volume in word cloud data based on the volume of the word size is being changed.
3. Provided sentiment details inbox view on click: If the user selects any word this view box provides details as shown in figure (8). The details of total mentions of the provided word in the tweets are being extracted for the event will be displayed. Adding to that, positive, negative, and neutral mentions are provided that give idea on how many times the word is being mentioned when the polarity of the tweet was positive/negative/neutral.

[5] Time Series Chart

The time-series visualization is provided to analyze the sentiment average score over time. Currently, data has been extracted from the tweeter data using python code and the Node page has been created for the selected province.

1. Provided zoom in effect: To provide details on date-wise sentiment score zoom effect is provide time-series graph as shown in figure (9).
2. Provided zoom-out effect: zoom out is available on double click.

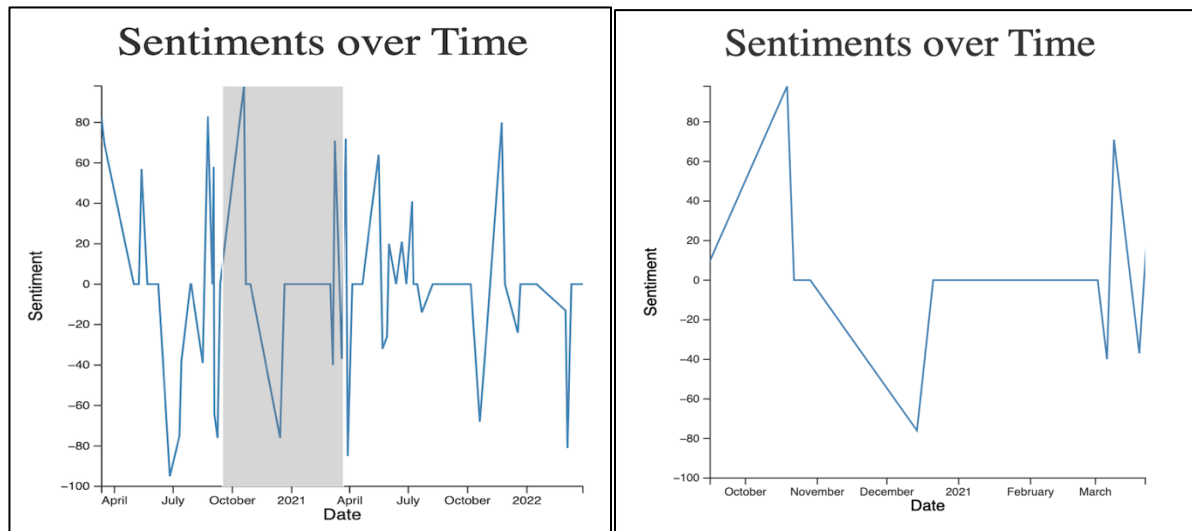


Figure 9: Sentiment Over Time Zoom in Effect

Limitations

In this study, I have used text blob library which uses bag of words, removing stop words, and hashtags that might not provide NLP techniques. Human language is very complex and it is evolving. There are many things like grammatical nuances, slang, cultural variations, trending word usages, etc. which need to be taken into consideration while evaluating complex moods like skepticism, sarcasm, anxiety, excitement, etc. To implement this visualization, Limited tweets were being considered as I am handling the data location due to additional expenses of Amazon Web Services (AWS). Also, having a huge dataset gives some lags while loading the visualizations. One of the limitations is that I have considered 3 events which is the limited sentimental analysis of provided options of the events.

Conclusion

In this study, I have explored the feasibility of using React Application for performing a sentiment analysis using Twitter data python (tweepy, text blob). Map visualizations done in d3 allows for accessibility of the results across any web browser. Current events are being considered while designing the visualizations for a topic like work from home, covid-19, and Ukraine. This study helps the user to analyze province-wise sentiments and also analyze sentiments throughout the time. Accessibility to sentiment analysis and opinion mining can be used for decision-making purposes to help better people's lives. This information can help the government to take appropriate measures in the areas like the Welfare Department, Health and Human services, or while implementing new policies. Other commercial businesses leverage the power of opinion mining to get feedback on their products and services to implement profitable changes [1].

Future Work

To make the system more generic one text area can be provided to the user rather than having 3 options in the system. Canada Map can be more attractive and effective if emoji, bubbles, and colour variation can be used to display based on geolocation and counts of sentiments. To

extract the live data Amazon web service can be used which can also handle a huge amount of tweet data. Advanced Machine Learning algorithms for natural language processing can be applied to this implementation for getting the exact polarity score for sentences [1].

Project Links

1. Git: https://git.cs.dal.ca/dppatel/csci6406_project.git
2. Live Link: <https://sentiment-visualization.herokuapp.com>

References

- [1]. A. -L. Alten, G. Gadre, S. Kulkarni and C. -S. Wu, "Analyzing Happiness Index Based on Geographical Locations," 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD), 2019, pp. 45-51, doi: 10.1109/ICAIBD.2019.8837010.
- [2]. I. Costa, R. Lima, C. G. R. dos Santos, B. S. Meiguins, A. G. M. Soares and R. Y. da SilvaFranco, "EmojiText: An Information Visualization Technique for Analyzing Phrases and Sentiments," 2021 25th International Conference Information Visualisation (IV), 2021, pp. 114-119, doi: 10.1109/IV53921.2021.00027.
- [3]. Twitter Sentiment Analysis using Python - GeeksforGeeks. (2017). Retrieved 5 March 2022, from <https://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/>
- [4]. Holtz, Y. (2022). Donut chart with group label in d3.js. Retrieved 5 March 2022, from https://www.d3-graph-gallery.com/graph/donut_label.html
- [5]. canada-map-d3 - CodeSandbox. (2020). Retrieved 5 March 2022, from <https://codesandbox.io/s/q87yj>
- [6]. GitHub - P7h/StormTweetsSentimentD3Viz: Computes and visualizes the sentiment analysis of tweets of US States in real-time using Storm. (2022). Retrieved 5 March 2022, from <https://github.com/P7h/StormTweetsSentimentD3Viz>
- [7]. G. Kavitha, B. Saveen and N. Imtiaz, "Discovering Public Opinions by Performing Sentimental Analysis on Real-Time Twitter Data," 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), 2018, pp. 1-4, doi: 10.1109/ICCSDET.2018.8821105.
- [8]. S. Raheja and A. Asthana, "Sentimental Analysis of Twitter Comments on Covid-19," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 704-708, doi: 10.1109/Confluence51648.2021.9377048.
- [9]. F. Heimerl, S. Lohmann, S. Lange and T. Ertl, "Word Cloud Explorer: Text Analytics Based on Word Clouds," 2014 47th Hawaii International Conference on System Sciences, 2014, pp. 1833-1842, doi: 10.1109/HICSS.2014.231.
- [10]. F. Neri, C. Aliprandi, F. Capecci, M. Cuadros and T. By, "Sentiment Analysis on Social Media," 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2012, pp. 919-926, doi: 10.1109/ASONAM.2012.164.