Sentiment analysis for COVID-19 pandemic on Twitter Data

Trisha Chakraborty
Computer Science and Engineering
Mississippi State University
Mississippi State, MS
tc2006@msstate.edu

Spurthi Buchireddy
Computer Science and Engineering
Mississippi State University
Mississippi State, MS
sb3581@msstate.edu

Md Rayhan Amin
Computer Science and Engineering
Mississippi State University
Mississippi State, MS
ma1962@msstate.edu

***Abstract*—The COVID-19 pandemic has caused ruin over the globe. Bounty lost their lives. Economy is smashing and regular folks losing their positions, MNCs failing. These extraordinary circumstances are causing blended responses among people groups. Individuals utilizing web-based media stages as their first retreat to communicate their feeling seeing their nearby also worldwide COVID circumstances. Twitter is one of the most well known online media plat-structures where individuals could impart their considerations and insights with different clients. Physically mining this conclusions isn't a choice among bunch measure of tweets accessible and afterward filer them out to examine and handle the suppositions of humankind. Our task supposition examination intends to deplete such assets like Twitter's tweets to mine the assessment and feelings on a circumstance and study the good, negative and nonpartisan effect of a circumstance. Besides, we utilized various ways to deal with perform conclusion examination and thought about the instruments, investigated the information utilizing WordCloud library and pictured the information bunches utilizing T-SNE Visualization.**

***Index Terms*—sentiment analysis, twitter analysis, opinion mining, COVID-19, coronavirus**

I. Introduction

The human race is enduring an unprecedented event due the rise of the COVID-19 pandemic. Everyday numerous people are getting affected as well as numerous people are dying due to this virus. As a result, our life is being affected like we never thought before. People are losing their jobs, companies are getting bankrupt. There is chaos everywhere. People are getting concerned and expressing their feelings about the situation through social media platforms among which Facebook and Twitter are notable. These social media platforms can be considered as a significant source of information that can get us acquainted with the thoughts, concerns, fears among peoples mind regarding this potentially life changing event. They can also help us understand the current social trends and contribute in making key decisions to improve the situations.

The COVID-19, also called as the coronavirus is caused by the SARS-CoV-2 virus. The virus can be deadly and life threatening specially to older adults as well as people with heart or lung problems. Up until now the virus affected 55.6 million people worldwide and claimed 1.34 million precious life and it still has not stopped [1]. The virus spreads through close contacts from person to person. Moreover, every affected person might not show any symptoms, which makes it more deadly as affected people with no symptoms can spread the virus among other people even without knowing. As a result, staying apart from people is considered the most effective way to keep healthy.

As our world is changing due to the pandemic, it has become quite unpredictable to understand the challenges lying ahead of us. People all over the world are facing the same crisis though the perception of the situation is not same. People are divided based on race, gender, country and religion etc. These divisions created a differences in the way we think and perceive. These perceptions can be very helpful in understanding the situations in different places and help the respected authority to take necessary steps to limit casualties. Moreover, the COVID-19 has been impacting people psychologically as well [2]. That is why Thakur et.al have labelled the pandemic as a psychological one as well.

As the situations are not improving, people are getting very anxious and expressing their feelings using a number social media platforms. These social media platforms can be considered as a significant source which can help us gather information regarding the psychological and mental status of its wide range of users around the world. As a result, it might help people in need of psychological help to overcome the situation. Also, these information can help us understand the collective feelings of a society. But going through these information manually is not only cumbersome but also time consuming. Thankfully, the development of technologies like data mining and machine learning has made these tasks easy and less time consuming. Now, using different APIs(Application Programming Interfaces) we can collect bulk data and analyze them using different machine learning techniques. In this paper, will discuss a method of collecting information from Twitter, which is social networking service and micro blogging service. After collecting the data from Twitter, we will analyze them to find out how citizens of world-wide counties are dealing with the current pandemic situation using sentiment analysis techniques.

Sentiment analysis can be defined as a process of analyzing textual data in an automated way and sorting them into sentiments namely positive, negative or neutral [3]. Sentiment analysis can be a useful tool for analyzing Twitter data to understand peoples feelings regarding the situation. Different methods of sentiment analysis are widely available and they are very easy to use. WordCloud is another interesting technology which is a way of visualizing the most used words from bulk data. They can help us identify the current topic of discussion in given period of time if we can collect relevant data of that specific time.

In this paper, we proposed a method of collecting Twitter data related to COVID-19 and analyze the polarity of the data using sentiment analysis. We also performed month-wise visualization of the collected data using WordCLoud. The rest of paper is organized in the following manner. Section II will discusse the related works regarding Twitter data as well as the use of sentiment analysis and WordCloud technology. Section III will discuss the proposed methodology. Section IV will focus on the experimental procedure as well as explain the results from the experiments. Finally, section V will conclude the paper by analyzing the results from the previous section.

II. Background and Related Work

The specific field of study that focuses on analyzing people's sentiments, opinions, evaluations, emotions and attitudes is called sentiment analysis and opinion mining [4]. In the study of data mining and other relevant topics, sentiment analysis is a very active research area. Numerous researches have been done and still going that uses sentiment analysis tools to analyze emotions from textual data as well address different aspects related to sentiment analysis.

Nasukawa et.al proposed a method of extracting sentiments for specific subject from a document which are associated with either positive or negative polarity [5]. The sentiment analysis accuracy can be greatly enhanced by properly identifying the semantic relationship between the sentiment expression and the relevant subject. The proposed method performed semantic analysis along with a syntactic parser that was complemented with a sentiment lexicon. The resultant method achieved high precision in finding sentiments within news articles as well web pages. It is also possible to model rich lexical meaning by using unsupervised vector-based approach to semantics. But they seem to fail while capturing sentiment information because the sentiment is related to the many word in the documents. To address this issue, Mass et.al presented a model capable of learning word vectors while capturing semantic term from documents by using a mix of supervised and unsupervised techniques [6]. Their proposed model could leverage continuous as well as multi-dimensional sentiment information.

A novel framework called the joint sentiment/topic model or JST was proposed by Lin et.al which was capable of detecting sentiments and topics simultaneously from textual data [7]. The framework was based on Latent Dirichlet Allocation also know as LDA and did not require any label as it was a fully unsupervised framework. The framework produced promising results and one of the highly regarded techniques in sentiment analysis. Another approach of sentiment analysis at phrase level was introduced by Wilson et.al where they first detected whether an expression is neutral or polar and then based on the result, they disambiguated the polarity of that given expression [8]. The approach was capable of identifying the contextual polarity automatically from a large subset of sentiment expression while resulting a performance better than the baseline performance.

Analyzing the sentiment of data collected from Twitter has always been an active research area. Bagheri et.al presented the application of sentiment analysis on Twitter data where they performed a number of queries related to different issues such as politics, humanity etc. and analyzed the results [9]. Their results suggested that most of the tweets are of neutral polarity.

Another application of sentiment analysis comes in the form of stock market prediction. Based on a number of research, it was found that stock market movements are related to public opinions and sentiments. And Twitter is big source of public opinion. Pagolu et.al proposed a method predicting the stock market movements though anlyzing the sentiment of Twitter data [10]. The proposed method used two popular textual representation methods namely Word2Vec and N-gram for analyzing the sentiments.

Another interesting research on Twitter data was performed by El Rahman et.al which was based on sentiment analysis as well. Here, the authors used a combinations of supervised and unsupervised machine learning techniques to find out the popularity between two of the biggest fast food chain in the US namely the McDonald's and KFC [11]. They collected tweets related to these two fast food chains and performed sentiment analysis after cleaning and structuring the data. Agarwal et.al performed another sentiment analysis on Twitter data based on POS-specific prior polarity features while also exploring the usage of a tree kernel and their method performed better than state-of-the art baseline [12].

As the coronavirus rages on, researchers have been using different methods to assess the situation. Rajput et.al performed two types of empirical analysis to assess the current trends regarding coronavirus [13]. They used word frequency and sentiment analysis on twitter messages to understand the current trends. Tweets from general users as well as WHO were part of the corresponding corpus. Their result showed that most of the tweets were of positive polarity. Another research performed by Xue et.al tried to address the psychological issues during the pandemic by first collecting twitter data and then using machine learning techniques to classify the tweets into 10 categories [14]. Their results showed no significant amount of tweets regarding the treatments and symptoms. But the number of tweets regarding people's fear due the unknown nature of the virus was overwhelming.

Wearing mask during this pandemic is one of the most effective way protect ourselves from the virus. But this direction of wearing masks has been received differently by different communities. Starting from January to May in the year of 2020, Sanders et.al analyzed over one million tweets to understand the public opinion regarding masks with the help of sentiment analysis [15]. Their results showed that tweets regarding masks during the pandemic have increased quite significantly. Another research performed by Kuhn et.al employed a neural network based sentiment analysis technique capable of analyzing multilingual twitter data collected during the first month of 2020 to understand people's mood based on their country of origin [16].

All these researches related to COVID-19 and sentiment analysis motivated us to understand how these techniques work and employ some of them to uncover how things are evolving in the first six month of 2020 regarding the pandemic. We hope, our research would help us understand the current trend regarding the pandemic as well as understand how the pandemic is impacting the suicidal rates across the world. Our results would help the respected authority to understand the situation and take the necessary steps to improve the situation.

III. Proposed Method

1) Dataset: The dataset used in this project are tweets collected for 6 months from January till June using tweet IDs [17].The tweets are related to over 75 keywords and the total tweet count is 736,297,019 tweets. We have created user authentication keys from Twitter Developer account to fetch the tweets. From the tweet ids collected, the tweet texts are fetched using Twitter API v2 by making use of Tweepy library: GET statuses/show/:tweet_id.

Sentiment analysis techniques can be classified into three types: Rule-based, automatic and hybrid. Rule based approach uses a bunch of rules that are created manually, whereas automatic approach rely on machine learning techniques and hybrid approach is a combining both rule based and automatic approaches.

A. Rule-Based Approach

In this method, we define a set of rules using techniques in Natural Language Processing like tokenization, stemming, and parsing. These processed data is fed to a model of TextBlob Library which uses Naïve Bayes classifier to calculate the sentiment polarity of tweets. The flow of processes in sentiment analysis using TextBlob is shown in the figure 1.

1) Data Preprocessing: For the tweets collected using Tweepy, the special characters like @, #, and URLs. Retweet tags like RT and numbers have been removed. NLTK stopwords like a, an, the, etc. have been removed from the tweets and the words with sequence of characters have been corrected such as 'Coooool' to 'Cool'. All the text in the tweets is then converted to lowercase, and at the end the emoticons have been removed. The figure 2 shows the original tweet and the tweet after preprocessing.

2) Feature Extraction: The twitter text that is preprocessed is now fed to the python sentiment analysis library TextBlob which is used for processing textual data. TextBlob offers simple API and features like parts-of-speech tagging, noun phrase extraction, tokenization, etc. TextBlob uses

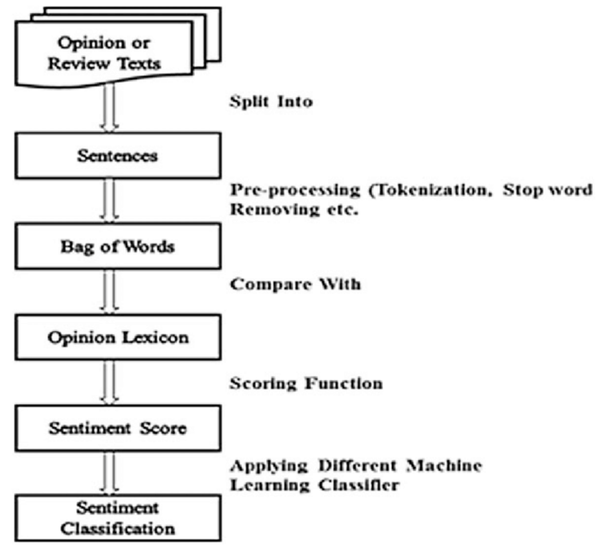
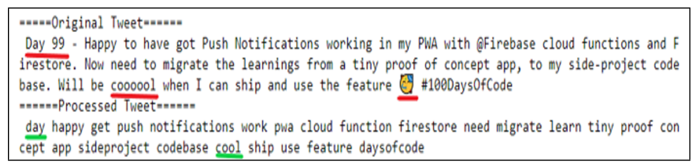


Fig. 1: Process flow of Sentiment analysis using TextBlob [18]



=====Original Tweet======
Day 99 - Happy to have got Push Notifications working in my PWA with @Firebase cloud functions and Firestore. Now need to migrate the learnings from a tiny proof of concept app, to my side-project code base. Will be cooooool when I can ship and use the feature #100DaysOfCode
======Processed Tweet======
day happy get push notifications work pwa cloud function firestore need migrate learn tiny proof concept app sideproject codebase cool ship use feature daysofcode

Fig. 2: Original tweet and preprocessed tweet

textblob.sentiments module and exploits Naïve Bayes analyzer to calculate the sentiment score.

Naïve Bayes is a group of probabilistic algorithms that utilizes Bayes' Theorem to anticipate the category of a text. The Naive Bayes Analyzer classifier is pre-trained on Stanford NLTK movie review corpus. The figure 3 shows equation of Bayes theorem that is used to predict the sentiment probability.

$$P(label|features) = \frac{P(label) * P(features|label)}{P(features)}$$

Fig. 3: Bayes equation to predict the sentiment probability

The sentiment like "Positive", "Negative" and "Neutral" are calculated by TextBlob using Naïve Bayes classifier and the polarity score is determined by sentiment score. If the

Sentiment score $<0 \rightarrow Negativepolarity$
Sentiment score $>0 \rightarrow Positivepolarity$
Sentiment score = 0 $\rightarrow Neutralpolarity$

B. Automatic Approach

Automatic approach for sentiment analysis rely on Machine learning techniques such as classification but does not depend on manually created rules. In this approach, the text is fed to the classifier which returns the category of sentiment i.e., positive sentiment, negative sentiment, and neutral sentiment.

In this, there are two stages: training and prediction. During training, sample data is used to train the sentiment analysis model which tags the text as positive, negative, or neutral. In prediction, unseen data are fed to the model to predict sentiments.

C. Hybrid Approach

Hybrid approach is a combination of both the methods which are: rule-based approach and automatic approach. This approach delivers more accurate results compared to the other two approaches. Stanford CoreNLP is a hybrid approach that develops a representation of entire sentences dependent on the sentence structure. It computes the sentiment based on how words make the importance out of longer expressions. The process flow of Stanford CoreNLP is shown in the figure 4.

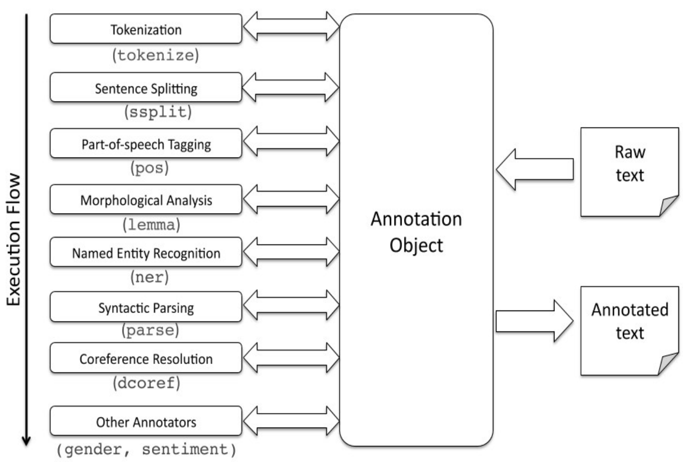


Fig. 4: Process flow of Stanford CoreNLP [19]

1) Data Preprocessing: CoreNLP is based on Java and it is required to install Java on the device. Preprocessing the raw data in CoreNLP can be done in two ways: using py-corenlp or using NLTK. The main difference between the two is that the output of py-corenlp is a raw JSON file that can be used to extract features or sentiment while NLTK provides the flexibility to write our own functions for data preprocessing.

2) Feature Extraction: Stanford's CoreNLP is an important tool to use for: sentiment analysis, named entity recognition, parts of speech tagging. Named Entity Recognition is useful in identifying the key components in a text, such as names of people, places, brands, etc. The main entities in a text can be extracted to help sort the unstructured data and to detect the important information, which is critical while dealing with large datasets. Unlike TextBlob, sentiment score in CoreNLP ranges from 0 - 4.

0 represents Very Negative sentiment,

1 represents Negative sentiment,

2 represents Neutral sentiment,

3 represents Positive sentiment,

4 represents Very Positive sentiment.

IV. Experiments and Discussion

In this section, we present report the experiment and performance of the two models we used for sentiment analysis. Our code is publicly available at https://github.com/trishac97/Twitter-Sentiment-Analysis.git.
This experiment reports month-wise sentiment analysis on how the World has been reacting in the early 6 months of COVID - spread. The experiment was implemented on 1-2 lakh tweets each month. Figure 3 reports the overall positive, negative and neutral tweet in % for Texblob API, our first method for sentiment evaluation.

Month	Positive Tweets (%)	Negative Tweets (%)	Neutral Tweets (%)
January	47.87	21.59	30.53
February	39.6	25.8	34.5
March	40.2	24.4	35.3
April	42.66	23.52	33.80
May	43.83	22.22	33.94
June	43.23	32.01	24.75

Any observation from sentiment analysis and making an point from tweets is complex and abstract. From the reported results, the % of neutral tweets over the 6 months did not fluctuated, and there is a drop of positive tweets moving from January to February and a rise of negative tweets in May to June. Any other fluctuations are not significant enough to make any observation. In motivation of getting more transparent results, we explored our second method, Stanford's CoreNLP. Fig 4 reports the overall positive, negative and neutral tweet for % for Stanford's CoreNLP.

A. Analysing behavior of two approach

Clearly, Standford's CoreNLP produced fine-grained scale consisting of 5 sentiment classes: Very Positive, Positive, Very Negative, Negative, Neutral. This characteristic helps to make an observation concrete. With the continuous rising negative tweets and corresponding falling positive tweet percentage estimates the current scenario of the world.
The portrayed difference in positive, negative and neutral sentiment percentages can be conspiracy of this two reasons.

- The intuition is sketchy in terms of translating practical sentiment percentages into real-life sentimental values. No signicant change in negative tweet percentage can be justified by assuming that humanity is accepting to take challenges and becoming warmed up to this new way of life.
- An originally positive tweet as in Figure 5 is wrongly classified by TextBlob API to belong in Negative class of sentiments, but Standford's coreNLP rightly classifies it to Positive class of sentiments.

Building on the second point, we reported one such example of wrongly classifies tweets. This is one the main big reason of why the inflicting changes in the percentage change for two methods used. Here is an analysis on the classifier

Month	Very Positive Tweets (%)	Positive Tweets (%)	Very Negative Tweets (%)	Negative Tweets (%)	Neutral Tweets (%)
January	26.59	38.69	15.63	10.23	8.86
February	18.69	34.69	18.63	16.23	12.09
March	16.96	32.47	19.25	19.23	12.09
April	15.56	32.56	19.69	25.23	6.96
May	15.23	30.87	25.66	26.59	2.65
June	15.18	23.19	28.47	30.28	2.88

TABLE I: Table 2. Sentiment Analysis with CoreNLP API

```
=====Original Tweet======
 Also my dad was exposed to covid recently but has thankfully tested negative so maybe I'm catching a
break this time
======Processed Tweet======
 also dad expose covid recently thankfully test negative maybe im catch break time
======Polarity Score========
 -0.15
======Sentiment========
 negative
```

Fig. 5: Example of Tweet Sentiment Classification.

being used by TextBlob API. The reason behind how Naive Bayes sometimes fails to detect the correct sentiment can be described as follows: Let's say we have a sentence called

$$P(negative \mid I\ got\ covid\ positive) = \frac{P(negative) * P(I\ got\ covid\ positive \mid negative)}{P(I\ got\ covid\ positive)}$$

$$P(I\ got\ covid\ positive \mid negative) = P(I \mid negative) * P(got \mid negative) * P(covid \mid negative) * P(positive \mid negative)$$

Fig. 6: Example on Naive Bayes Analyzer.

"I got covid positive" and Naive Bayes is trying to find the probability of how much the sentence belongs to a negative class. To Find out the root of the inaccuracy, as we know, Naive Bayes Classifier is pre-trained on movie review corpus from which it predicts the probability of a label, given a feature. So, we are finding the probability of the sentence "I got covid positive" given negative class. This problem simply translates to Bayes theorem. The term "I got covid positive" given negative class can be found by finding the probability of how much each word in the sentence gives a negative sense, and then multiplying all together to get the final term. But the red circled term that is probability of the word "positive" given "negative" class would clearly produce a probability value almost 0 and hence the total term ends up being probability 0. Which is the root of the incorrect result in this case. Hence for any natural language processing task, it's important to consider the context and not just the lexical content of the sentence.

B. Twitter Data Visualization

1) WordCloud: Word Cloud is an information perception procedure utilized for speaking to message information in which the size of each word shows its recurrence or significance. It is a python library that can be utilized to locate the best 100 high recurrence words. The size of the content means the recurrence of that word. As should be obvious, the word Covid is the biggest content in both Jan and Feb wordcloud, which implies these are the most happening words in the content. Such wordclouds can be helpful in Analyzing client and representative input.

From the visualization tool, we made the notable observations that translates to the situation:

- We collected wordcloud for first 6 months. We observed the word social distancing became popular among tweeters from March.
- Another notable incident happened during June was the protest for George Floyd and the usage of Tear gas.

The figure 8 displays the month-wise visaulization from January till June using WordCloud.

C. T-SNE Visualization for Vaccine Sentiment

We attempted to record the reaction to the term Vaccine during Pandemic. Formed cluster of top 100 high frequency words around the keywords using TSNE visualization. We used keywords: Vaccine and Pandemic.

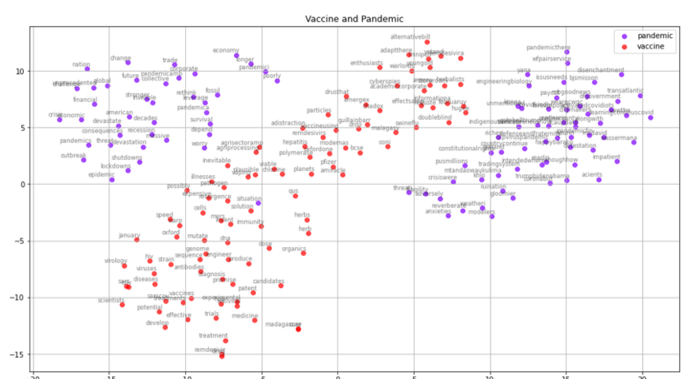
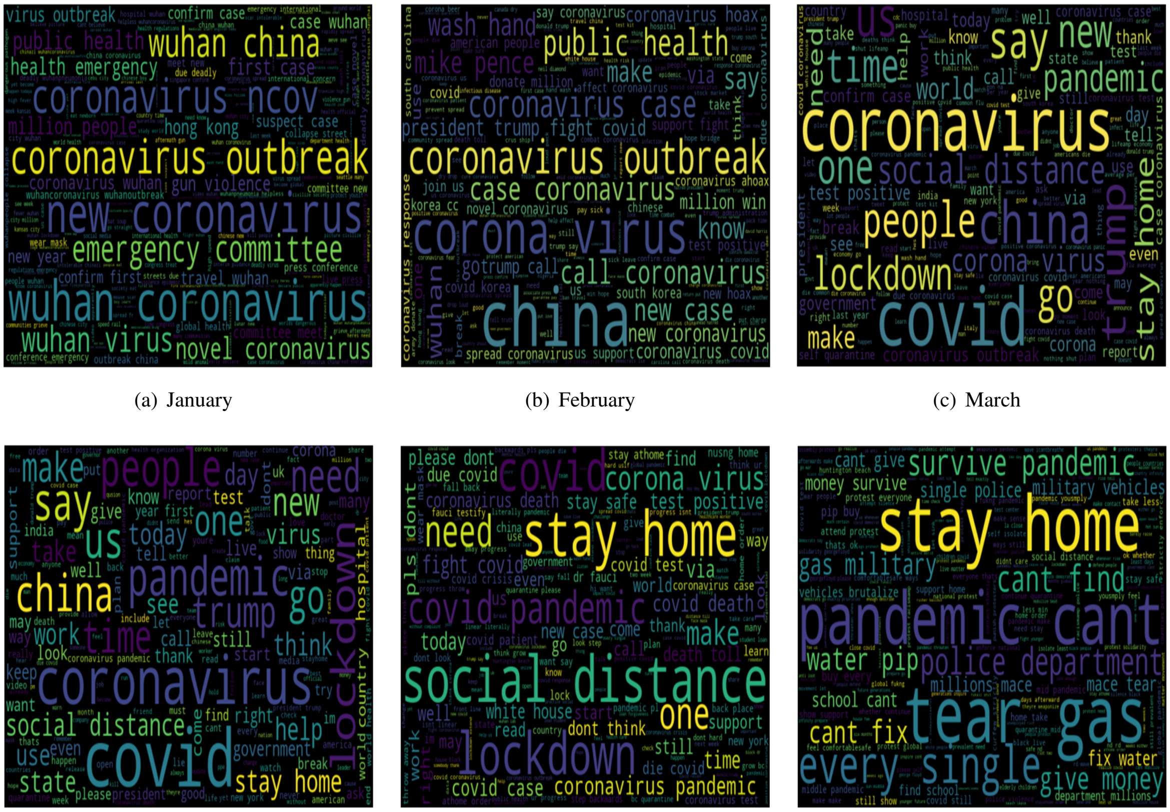


Fig. 7: T-SNE Visualization

In the figure 7, Red dots are words related to the term vaccine and purple for the word pandemic. The cluster for **vaccine** shows the occurrence of **Oxford** university which is currently getting all the **funds** and running **trials** and filing **patents** for **patents** vaccines. Furthermore, some other proposed **candidate** vaccines are **herb** based and **organic**. The idea that the initial vaccines are going to be **expensive** can also be inferred from this figure. Also, the words related to the work **pandemic** situation which has caused **change** in **economy** and **trade**. The mental health of a large population of humankind is **worrisome**,**anxityful** and **impatient**.



(a) January (b) February (c) March

(d) April (e) May (f) June

Fig. 8: Month-wise Visualization Using WordCloud

V. Conclusions

Our goal was to study the sentiment analysis during the first 6 months of COVID spread. Our contribution in the project revolves around exploring various Sentiment Analysis tools modeled for sentiment analysis classification task. A simple sentiment analysis task can lead to answering the simple question: If the emotion of the sentence is positive or negative? A more complex Sentiment Analysis task can lead to answering: What is the emotion of the sentence in a ranking scale of 1-5, 1 being very positive and 5 being negative. Such emotion scaling are called finely grained emotions. For effective sentiment classification: Negation is important. When dealing with bag of words and each word has same influence on deciding the sentiment of the sentence,correct manipulation of word negation is crucial. Also, in some cases, word occurrence may matter more than word frequency. Using all words (pure naïve bayes) works well for some tasks but not all. While the first method in the paper, Naive Bayes was more easier to setup, but CoreNLP used multinomial naïve bayes and found better fine grained classifications. There are many more NLP sentiment classification tools used by the researchers in recent times as discussed in the related section of the paper. Significant takeaway from the paper is particular on what investigation to be performed is imperative to evade superfluous calculations that may back you off when working with bigger informational collections.

VI. Contribution of Members

Our contribution to this project can be divided into three parts:

- **Project Slides and Presentation** Trisha Chakraborty and Spurthi Buchireddy drafted and formalized the slides content. On the presentation day, we presented the slide content in 50-50% respectively.
- **Coding** Trisha Chakraborty and Spurthi Buchireddy wrote the code for performing the experiments.
- **Report Writing** Trisha Chakraborty, Spurthi Buchireddy and Md Rayhan Amin drafted and formalized the report content. We spilitted 33.33% of the writing workload among us.

REFERENCES

[1] "Template:covid-19 pandemic data," Nov 2020.

[2] V. Thakur and A. Jain, "Covid 2019-suicides: A global psychological pandemic," *Brain, behavior, and immunity*, 2020.

[3] "Twitter sentiment analysis with machine learning," Jun 2019.

[4] B. Liu, "Sentiment analysis and opinion mining," *Synthesis lectures on human language technologies*, vol. 5, no. 1, pp. 1–167, 2012.

[5] T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," in *Proceedings of the 2nd international conference on Knowledge capture*, pp. 70–77, 2003.

[6] A. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies*, pp. 142–150, 2011.

[7] C. Lin and Y. He, "Joint sentiment/topic model for sentiment analysis," in *Proceedings of the 18th ACM conference on Information and knowledge management*, pp. 375–384, 2009.

[8] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," in *Proceedings of human language technology conference and conference on empirical methods in natural language processing*, pp. 347–354, 2005.

[9] H. Bagheri and M. J. Islam, "Sentiment analysis of twitter data," *arXiv preprint arXiv:1711.10377*, 2017.

[10] V. S. Pagolu, K. N. Reddy, G. Panda, and B. Majhi, "Sentiment analysis of twitter data for predicting stock market movements," in *2016 international conference on signal processing, communication, power and embedded system (SCOPES)*, pp. 1345–1350, IEEE, 2016.

[11] S. A. El Rahman, F. A. AlOtaibi, and W. A. AlShehri, "Sentiment analysis of twitter data," in *2019 International Conference on Computer and Information Sciences (ICCIS)*, pp. 1–4, IEEE, 2019.

[12] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. J. Passonneau, "Sentiment analysis of twitter data," in *Proceedings of the workshop on language in social media (LSM 2011)*, pp. 30–38, 2011.

[13] N. K. Rajput, B. A. Grover, and V. K. Rathi, "Word frequency and sentiment analysis of twitter messages during coronavirus pandemic," *arXiv preprint arXiv:2004.03925*, 2020.

[14] J. Xue, J. Chen, C. Chen, C. Zheng, S. Li, and T. Zhu, "Public discourse and sentiment during the covid 19 pandemic: Using latent dirichlet allocation for topic modeling on twitter," *PloS one*, vol. 15, no. 9, p. e0239441, 2020.

[15] A. Sanders, R. White, L. Severson, R. Ma, R. McQueen, H. C. A. Paulo, Y. Zhang, J. S. Erickson, and K. P. Bennett, "Unmasking the conversation on masks: Natural language processing for topical sentiment analysis of covid-19 twitter discourse," *medRxiv*, 2020.

[16] A. Kruspe, M. Häberle, I. Kuhn, and X. X. Zhu, "Cross-language sentiment analysis of european twitter messages duringthe covid-19 pandemic," *arXiv preprint arXiv:2008.12172*, 2020.

[17] "https://github.com/echen102/covid-19-tweetids."

[18] R. S. Jagdale, V. S. Shirsat, and S. N. Deshmukh, "Sentiment analysis on product reviews using machine learning techniques," in *Cognitive Informatics and Soft Computing*, pp. 639–647, Springer, 2019.

[19] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky, "The stanford corenlp natural language processing toolkit," in *ACL*, 2014.