1-1-2020

UNDERSTANDING DEPRESSION DURING THE COVID-19 PANDEMIC THROUGH SOCIAL MEDIA DATA

Nusrat Armin
University of Mississippi

UNDERSTANDING DEPRESSION DURING THE COVID-19 PANDEMIC THROUGH SOCIAL MEDIA DATA

A Thesis
presented in partial fulfillment of requirements
for the degree of Master of Science
in the Computer and Information Science
The University of Mississippi

by

Nusrat Armin

May 2021

ABSTRACT

The COVID-19 pandemic has dramatically affected peoples' daily lives all over the world - physically, economically, and emotionally. Due to the virus, many people have died, and many hospitalized. A record number of people have lost their job, and many businesses have closed. The global economy is at risk. People are facing new realities of their lives. Studies have shown that the level of depression is three times higher than before this pandemic. Previous studies have shown that people use social media to express their emotions and feelings. The purpose of this study is to understand the depression during this COVID-19 pandemic using social media data. To study the depression-related tweets, I have used a COVID-19 related dataset that is made available by IEEE. Regarding methods, I have employed unsupervised techniques such as clustering and topic modeling. Besides, I have used sentiment analysis to understand the emotions and subjectivity in the clusters. The result from the analysis shows the change of depression related discussions during this pandemic over a five and a half months period of time. It also shows the characteristics of the overall discussion around depression. The findings may be useful for future depression studies.

DEDICATION

To Ammu, Abbu,

Naeemul (Husband),

Our two sons Abdullah and Ahmad (Unborn)

And the miscarried baby.

ACKNOWLEDGEMENTS

I would like to acknowledge everyone who has played a role during this journey of pursuing my master's degree. At first, I would like to thank my advisor, and supervisor Dr. Dawn Wilkins for her advice, encouragement, motivation, and patience. Without her direction and motivation, this journey won't be that smooth. I also want to thank her for guiding me in every step of my thesis. I would also like to thank my committee members, Dr. Yixin Chen and Dr. Philip Rhodes, for their time, comments during the defense, and also for making corrections to my thesis draft.

I am grateful to the faculties who have taught me during the last four years. Special thanks to Mr. Joseph Carlisle, Mrs. Hui Xiong, Dr. Kristin Davidson, and Dr. Conrad Cunningham for the language classes, Dr. Yixin Chen for the algorithm classes, Dr. Feng Wang for the operating system classes, and Dr. Byunghyun Jang for the Computer Architecture class. I am thankful to Dr. Adam Jones for the Human Centric Computing (HCC) class, the class was great to motivate one in research and thinking process. I am thankful to Dr. Philip Rhodes for the Cloud Computing class. In this class, I got a chance to work with big data in the Cloud. Again, I am thankful to Dr. Dawn Wilkins for the Advanced Natural Language class. There is no doubt that she is an excellent teacher and presents every difficult topic easy for the students. This class has encouraged me to choose my thesis topic and encouraged me to work with Natural Language Processing (NLP).

I am thankful to the University of Mississippi and the Computer and In-

formation Science Department for allowing me to pursue my degree and also for the assistantship. The environment of the University and the beautiful campus has encouraged me to concentrate on my study.

I am also thankful to the staff of the department for administrative and technical support. I would like to thank some of my friends - Carla, Silu, and Tina for the encouragement, and for helping me several times.

I am so much grateful to all my family members. I want to show my gratitude to my parents for their unconditional love, encouragement, support, caring, and prayers. Most of all, I am grateful to them for their sacrifices for bringing me up and preparing me for a bright future. I am also thankful to my husband (Naeemul) for his love, care, support, encouragement. Without his motivation, encouragement, understanding, sponsoring, and support, pursuing this degree would be very difficult for me.

Finally, I want to praise and thank my Lord Allah. Every time I have faced difficulties, He has shown me a path and made it easy for me. Without His favor, nothing would be possible.

December 07, 2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The Coronavirus disease 2019 (COVID-19) is a viral disease first identified in Wuhan, China in December 2019. After that, it has continued to spread nearly in all countries in the world. By March 11, 2020, more than 118,000 cases were found, 4,291 people had lost their lives, and thousands of people were hospitalized due to COVID-19 in 114 countries. The World Health Organization (WHO) became concerned about the spread and severity of the outbreak and declared it as a pandemic [WHO (March 11, 2020)]. The countries had started to take steps to control the spread of the virus with their own rule, including lockdowns, curfews, quarantines, stay-at-home orders, shelter-in-place orders, and other restrictions that were enforced or recommended (total or partially) in many countries. Due to the restrictions, the overall movements of people are restricted, international and in some places, domestic travel bans, schools, universities, religious places are closed, business centers such as markets, restaurants are closed, social or cultural gatherings are also restricted. This pandemic not only affected the global economy but also the day to day life of general people. People are facing new realities in their life. Many people have lost their job, some are working from home, many are temporarily unemployed. In some places, children are doing homeschooling, avoiding face-to-face interactions, wearing masks, taking extra hygiene procedures, and so on. All of these consequences have affected the mental health of people [WHO (2020)]. Depression, anxiety, and stress are common mental health problems due to the pandemic. According to the experts, this pandemic became a traumatic event on a large scale [Drillinger (September 10, 2020)]. Some studies are saying that depression symptoms are three times higher during the COVID-19 lockdown than usual [Drillinger (September 10, 2020)]. President Trump said, "There's depression, alcohol, drugs at a level nobody's ever

seen before. The cure cannot be worse than the problem itself." [FOXNEWS (October 23, 2020)]

Social media is a place where people share their feelings and emotions about daily happenings. It can act as a resource to study the discussions of the general people who are depressed or posting depression-related tweets. In this study, I have chosen Twitter as a source of the data to be studied. Here, I have studied the discussions related to the COVID-19 pandemic and depression over Twitter data to understand depression during this pandemic.

In this study, I have used the dataset of IEEE that is related to COVID-19. I have collected about 600 million tweets and filtered all the "depress" related tweets from them. I have found a total of 890,838 tweets that are related to depression. But this data set is full of noise that makes it necessary to preprocess before analyzing them. So, I have cleaned the data after doing the preprocessing. After cleaning, there are 216,701 tweets in the filtered dataset. In the analysis part, I have studied the depression trend over time. Then, I have clustered the tweets and applied topic modeling to understand depression-related topics. Here, I have applied "k-means clustering" and "biterm topic modeling" and have identified seven distinct clusters. Later, I have further analyzed those clusters using sentiment analysis.

In the following chapters, first, I am going to describe the literature review of this study related to the COVID-19 pandemic and mental health. Then, I will provide a detailed description of the data source, how I have collected those data, a description of the dataset, and how I have preprocessed them. Then, I will explain the analysis that I have done to understand depression on Twitter data. Finally, I will describe the future works, the limitations of my study, and the conclusion of my study.

CHAPTER 2

LITERATURE REVIEW

Since the novel COVID-19 pandemic has started, researchers from multiple domains are working continuously to understand the pandemic, reduce its impact, and analyze the data. This chapter has addressed some of the works that are relevant to Mental Health, COVID-19, and Social Media.

2.1 Depression studies based on Social Media

In one work, authors [Balani and De Choudhury (2015)] have developed a self-disclosure detection algorithm which measured the levels of self-disclosure in social media. They have focused on different mental health forums on Reddit. They have built a classifier to identify the levels of self-disclosure. The classifier can characterize a Reddit post to be of high, low, or no self-disclosure with 78% accuracy. They have found prominent differences in the ways individuals socially engage in this forum. They are hoping that their self-disclosure detection algorithm will help in psychological therapy via social media. They have suggested future research on how to design intervention strategies without compromising mental health patient's identities.

2.2 Studying COVID-19 and Depression using Social Media

In one research from the public health domain, authors [Gao et al. (2020)] have studied the prevalence of mental health problem associated with social media exposure (SME). They focused on anxiety and depression. The authors did online surveys and measured the level of depression and anxiety using the WHO-Five well-being index. Then, they did a cross-sectional study on their frequency of exposing to social media. Their findings showed that

there is a high prevalence of mental health problems, which is positively associated with the frequency of SME during the COVID-19 outbreak. As it is a cross-sectional study, it difficult to accurately elucidate causal relationships between SME and mental health. They have suggested additional longitudinal studies such as cohort studies and nested case-control studies for the future. The result they found could not cover all age groups evenly since few senior citizens' participation may have affected the results. The authors have realized from their study that social media may lead to disinformation, which turns mental health problems. So, it is necessary to take action against disinformation by monitoring and filtering false information and promoting accurate information through cross-section collaboration.

In another work, authors [Li et al. (2020)] have applied natural language processing (NLP) techniques to classify the tweets in terms of mental health. They have collected about 8,148,202 tweets related to COVID-19 and mental health in different languages from March 24 to 26, 2020. Then they built an EmoCT (Emotion-Covid19-Tweet) dataset for the training purpose. The dataset consists of randomly chosen 1,000 English tweets that are labeled manually as one, two, or three emotion labels. They have used the EmoCT to classify the tweets into eight emotions: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. They have proposed two models of classifications: single-label and multi-label. They have applied a pre-trained multilingual version of the Bidirectional Encoder Representations from Transformers (BERT) model that can predict up to 104 languages. They have trained 1,181,342 unlabeled randomly selected tweets. They got higher accuracy from the single-label model than the multi-label. So, they have taken the single-label model as their primary model for further analysis. They have kept the tweets labeled two emotions - sad, and fear, as these two emotions are more related to severe negative sentiments such as depression. To understand why people are feeling fears and sad, they have analyzed words and phrases that highly correlated with both emotions. They have chosen the top 500 keywords and phrases based on the frequency and found some informative keywords and phrases. For future work, they want to do some detailed analysis after grouping the tweets based on languages and

locations. They also want to track the Twitter data for a longer time to see how people rebuild their trust and joy from sadness and fear.

Authors in [Zhang et al. (2020)] have identified 2,575 distinct Twitter users who could be depressed. They have used regular expressions to find out the tweets of depressed users. Then, they have randomly selected another 2,575 distinct Twitter users who could not be depressed. After that, they have trained these datasets using three transformer-based depression classification models. Finally, they used the models to analyze community-level depression on Twitter. However, the dataset was small, and the data did not manually annotate.

CHAPTER 3

DATA COLLECTION

3.1 Source

Many data sources can be found on websites that are related to the COVID-19 pandemic. In this study, I have used the dataset "COV19 Tweets Dataset" that was published in IEEE [Lamsal (November 20, 2020)]. They have collected the tweets using 90+ keywords and hashtags that are related to the COVID-19 pandemic such as "corona", "covid", "covid19", "ncov", "pandemic", "quarantine", "lockdown", "social distancing", "work from home", "wearamask", and so on. They are publishing the data from March 19, 2020, 01:37 AM, and updating it on a daily basis. This dataset has considered globally published English tweets. Twitter content redistribution policy only allows to share the tweet IDs and/or user IDs. They do not allow to share the actual tweets [Lamsal (November 20, 2020)]. So, the dataset published by IEEE has only the tweet IDs and the related sentiment scores to that ID in the form of CSV files.

In my study, I have used the data from March 19, 2020, to September 4th, 2020. I have collected about 600 million tweet IDs from this period.

3.2 Collection Procedure

To download the CSV files, I had to create an account in IEEE. As the downloaded CSV files contain both the IDs and the sentiments, I had to separate the IDs in different CSV files before starting the hydration[1]. Rather than hydrating each file separately, I have merged 7 files before starting the hydration. I have noticed the files have some duplicate IDs.

[1]Hydration is the method to get the details (tweets, creation time, URLs, hashtags, user name, etc.) of a collection of Tweet IDs.

So, while merging, I have removed the duplicates. To hydrate the tweet IDs, I have used both Hydrator (an application) and Twarc (a python library) at a time using two different tweet IDs. Each merged files were 62MB to 535MB. It took more than one month to collect more than five months of data. After hydration, I have found approximately 10%-30% of those tweets are deleted and some tweets had an error. Table 3.1 shows the description of the data found from IEEE files.

Attributes	Summary
First day of tweet	March 19, 2020
Last day of tweet	September 04, 2020
Total number of days	170
Number of tweets from IEEE Dataset (COVID-19 related tweets)	Aboout 600 million
Attributes of the hydrated data	"coordinates", "created_at', "hashtags", "media", "urls", "favorite_count", "id", "in_reply_to_screen_name", "in_reply_to_status_id", "in_reply_to_user_id", "lang", "place", "possibly_sensitive", "retweet_count", "retweet_id", "retweet_screen_name", "source", "text", "tweet_url", "user_created_at", "user_screen_name", "user_default_profile_image", "user_description", "user_favourites_count", "user_followers_count", "user_friends_count", "user_listed_count", "user_location", "user_name", "user_screen_name.1" , "user_statuses_count", 'user_time_zone", "user_urls", "user_verified"

Table 3.1. Data description before preprocessing

CHAPTER 4

DATA PREPROCESSING

4.1 Data Filtration

4.1.1 Filter depression related tweets

To work with the depression-related tweets, at the very beginning, I have filtered the tweets that contain the term "depress" in it from the COVID data set. While filtering, I have ignored the case to capture the multiple variations of the term "depress". For example, "Depress", "Depression", "DepreSSion", "DEPRESSING", "depressed" and so on. After filtration, I have created a depression corpus with all those tweets. This dataset consists of 890,838 tweets.

4.1.2 Remove Retweets

From the depression corpus, I have removed all the retweets. To do that, I have identified the rows of data-frame whose "retweet_id" is "NaN". Then, I have used only those tweets for further analysis. After removing retweets, I have found 265,862 tweets.

4.1.3 Remove Duplicate IDs

I have also removed the duplicate IDs from the depression corpus. After removing duplicate IDs, I have found 252,214 tweets.

4.1.4 Removing spammer or extreme outlier users

I have distributed users to understand the number of tweets posted by users. Figure 4.3 shows the user distribution. Here, I have found that more than 86% of the users have posted just one tweet, and about 14% have posted more than once. I have checked some of

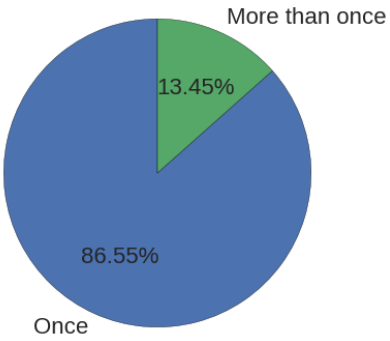

Figure 4.1. Percentages of the users who have twitted once and more than once

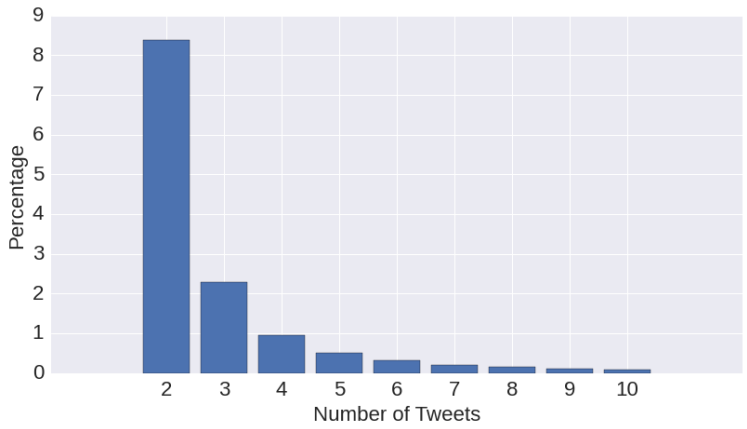

Figure 4.2. Distribution of the users who have twitted more than once

Figure 4.3. User distribution

the top outliers and found that many of them are posting ads using different hashtags that are related to COVID-19 and depression. To avoid the spammers, and noise in the collections, I have decided to consider only the tweets of the users who have posted a maximum of 10 tweets during these five months as they have covered more than 97% of the tweets.

Figure 4.4 shows the sizes of the dataset before and after filtering the data.

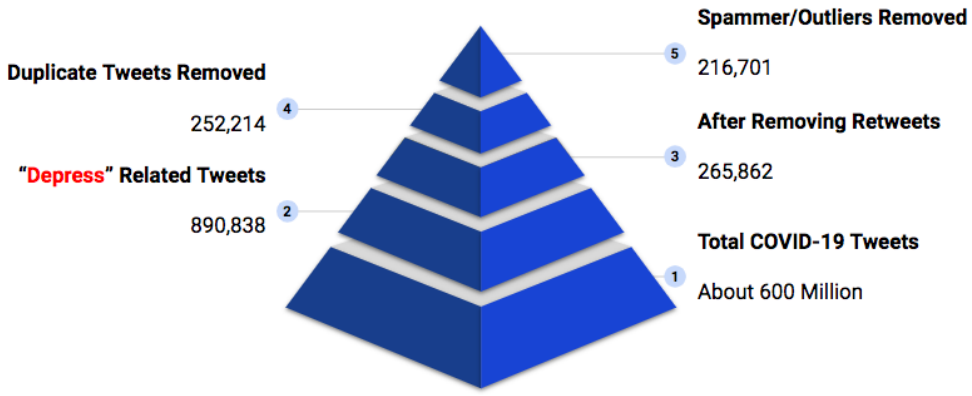

Figure 4.4. Sizes of the collected data before and after preprocessing

4.2 Data Cleaning

After filtering all of the data, the dataset now consists of only one CSV file with 216,701 tweets. The tweets are raw and unstructured. It is hard to work with such raw data with unnecessary information in it. So, it is necessary to clean the data before starting the analysis. Here are the steps I have followed:

4.2.1 Clean URL, MENTION, remove stop words and common words

After removing the retweets and the tweets of the outliers, I have cleaned the remaining data using the tweet-preprocessor package in Python. Using it, I have cleaned the URL, MENTION over the Twitter account, and HASHTAGS. Then I have converted the cleaned tweets to lower case. I did not remove any word from it for sentiment analysis purposes. Because if remove stopwords, and common words, it may not give an accurate score. I have named it as partially cleaned text.

For further cleaning of data, I have used NLTK (Natural Language Toolkit), which is one of the best libraries for preprocessing text data. At first, I have used NLTK word_tokenize package to tokenize the texts. Then, I have removed the STOP words and some common words related to the term "depress", "corona", "covid" and the word "amp" (for &) from the partially cleaned text as they are not useful. To remove the stop words, I have used the "Stop Word Dictionary" of NLTK. Finally, I have taken only the texts which contain only the alphabetic characters in them to avoid punctuation or numerical values. I also have lemmatized[1] the cleaned texts. To do that, I have used NLTK "WordNetLemmatizer" package. I have replaced all the empty texts with the empty string " " to avoid errors.

After all the cleaning, I have found 216,701 Tweets. The rest of the thesis focuses on this cleaned data.

[1]Lemmatization is the text preprocessing technique of converting the different forms of words to their lemma (root words). For example, the lemma of "walked", "walking" is "walk".

CHAPTER 5

METHODOLOGY

Various data analysis methods have been applied to the tweets that I have collected and cleaned. This analysis can be helpful to understand the discussion of depression in social media during the COVID-19 pandemic. Further data analysis can be done to produce more useful insights. Below are the methods that I have followed to produce necessary information regarding the analysis that will be discussed.

5.1 Clustering

I have a large amount of unlabeled and unclassified data to be analyzed, and there is no easy way to label them. So, I found unsupervised learning as an appropriate option to analyze a large amount of data. Unsupervised learning is a type of machine learning that does not require any labeled data or pre-trained data to identify patterns and relationships in the data set with minimum human guidance [Education (September 21, 2020)]. Clustering is a common unsupervised machine learning approach that can use to group the unlabelled data on the basis of the similarity or dissimilarity among the given information [Education (September 21, 2020)]. There are many algorithms to cluster the unlabelled data. Among those many types, I have used K-means because of its simplicity and efficiency.

5.1.1 K-means

K-means is an exclusive type of unsupervised clustering method. Like other clustering methods, K-means clustering groups the data based on their similarities. In this method, n observations are assigned into K groups or clusters. The partition between the clusters is done in such a way so that each observation is closest to its centroid. Here, the size of the

clusters depends on the value of K. The bigger the values of K, the smaller the size of the cluster, and vice versa [Education (September 21, 2020)].

5.1.1.1 Decide the Number of Clusters

In the K-means algorithm, the value of K or the number of clusters needs to set before applying the algorithm. In this study, I have used the elbow method to determine the number of clusters (K). In this method, it is necessary to calculate the sum of squared error(SSE) for some value of K. SSE can be defined as the sum of the squared distances between the centroid and each observation or member of the cluster [SSE (2020)]. Here is the methodology to calculate SSE :

- At first, I have selected the range of K 3 to 20.
- For each value of k, I have to initialize k-means.
- After that, I have calculated the SSE of each k.
- Finally, I have plotted the K against the SSE graph.

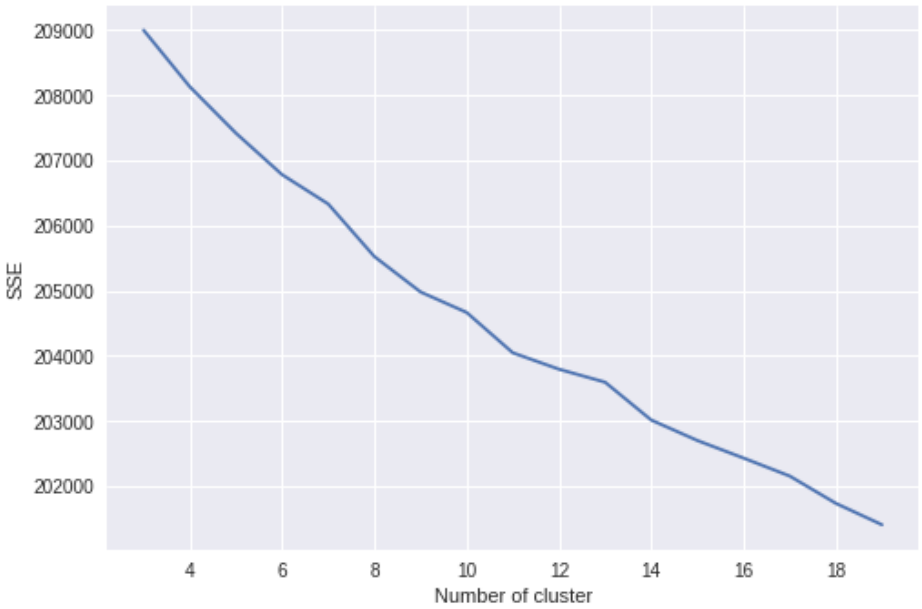

Figure 5.1. SSE vs. Number of Clusters

In figure 5.1 the plot of SSE vs. K has been shown. Unfortunately, in this plot, the elbow (the point at which the value of SSE seems to be decreasing for an increase in k) is hard to figure out as the line is mostly diagonal with many breaks within it. I found the minimum breakpoint on 6 and 7. So, it seems that k = 6 is a good place to start. To become more careful, I have considered the range of k between 4-8 as my target to get the minimum number of clusters to get less granular clusters.

Here is the methodology that I have followed to cluster the depression-related tweets for K(4-8) clusters :

- Used the K-means from scikit learn in order to apply the K-means algorithm over the depression dataset.
- To tokenize the texts, I have applied TfidfVectorizer over the cleaned lemmatized texts of the depression corpus. Here, 2000 maximum features were taken.
- Applied K-means algorithm for k clusters over the tokenized texts.
- Followed the same methodology for different values of k (4,5,6,7,8). For further analysis chosen k = 7 clusters, as it has given a good balance between the quality and quantity of the clusters.

5.2 Topic Modeling

In order to understand, explore, visualize the relevant topics in the collection topic modeling can be very useful. Topic modeling is a statistical model in Machine Learning or Natural Language Processing that can be used to get the themes of a collection of documents. There are many algorithms that can be used to get the topics of the collection such as Latent Dirichlet allocation (LDA), probabilistic latent semantic analysis (pLSA), biterm topic model (BTM), etc. While the traditional algorithms like LDA or pLSA usually perform well for documents with long texts, they may not perform well for short texts like tweets whereas BTM was designed to work basically with short texts. Studies showed that when applied

BTM and W2V-GMM, and modifying LDA over Twitter data compared, BTM performed better compared to all other models when working with short documents [Jónsson and Stolee (2015)]. So, in my study, I have used biterm topic model or (BTM) in order to extract and identify the topics in the depression corpus.

5.2.1 Biterm Topic Modeling (BTM)

Here is the methodology that I have followed to get the topics in the depression corpus using BTM :

- Randomly took 150,000 tweets from the cleaned depression corpus.
- Converted the sample tweets to count-vectors using CountVectorizer.
- Removed tweets having less than two words. BTM Topic Modeling requires this. After removing 145860 tweets were there to work with.
- Applied BTM Topic modeling using 20 iterations for 4,5,6,7 topics. Get better results for 7 topics.
- Used pyLDAvis package and t-Distributed Stochastic Neighbor Embedding (t-SNE) methods to visualize the topics.

5.3 Sentiment Analysis

To quickly understand the reactions and emotions of a large amount of tweet collection, I have performed sentiment analysis. Sentiment analysis is an easy way to view the polarity of the text, whether it is positive, negative, or neutral. This technique applies both Natural language processing and Machine learning to get a weighted sentiment score of the text. Different types of sentiment analysis techniques can use to get the sentiment scores of the text. Among all those techniques, Valence Aware Dictionary and Entiment Reasoner (VADER) performs outstanding for social media analysis [Gilbert and Hutto (2014)]. It does

not require any training data and can easily apply to any general dataset. So, in my study, I have used VADER to get the polarity of the data.

5.3.1 VADER sentiment

Valence Aware Dictionary and Entiment Reasoner (VADER) is a lexicon and rule-based technique that can use to get the sentiment value of the texts. The VADER lexicon is consists of sentiment related words and icons with their sentiment scores [Lexicon (2017)]. This lexicon is very rich with a large number of words, has validated by human, and its quality has considered as "Gold Standard" [Gilbert and Hutto (2014)]. When Vader is applied to a text, it matches the words with the words in its lexicon and gives a value to each word (positive, negative, or neutral). Finally, it calculates all the values and gives a compound score.

Here is the methodology that I have followed to get the sentiment of the tweets using VADER sentiment :

- I have used NLTK.sentiment.Vader package to get the sentiment of the tweets. It is a powerful package and easy to use. The package return sentiment score between -1.0 to 1.0.

- I have applied the Vader over the partially cleaned corpus of the dataset. In the partially cleaned tweets, I have cleaned only the URL, mentions, and hashtags as it may change the sentiment of the text.

- I have grouped the sentiments according to their date and took their average to see the overall change in the sentiments. I have also grouped them according to the month and calculated their average to get a more clear view of the change.

- I also applied Vader to get the sentiment scores for each cluster.

CHAPTER 6

ANALYSIS AND RESULTS

6.1 Depression Trend Analysis

To get a quick overview of the depression tweets, I have plotted the frequency of depression tweets per day. Figure 6.1 is showing the timeline of depression tweets. Here, the black line is indicating the trend before data preprocessing, and the blue line is indicating the trend after preprocessing the data. So, the blue trend line contains only the tweets of those users who have tweeted at most 10.

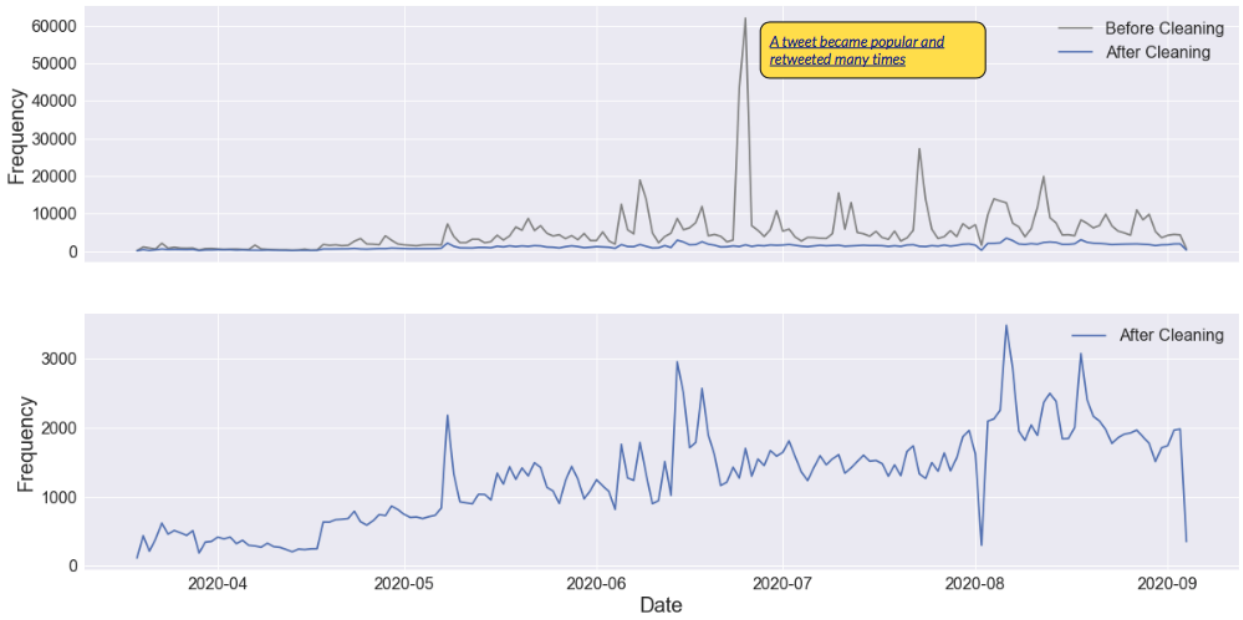

Figure 6.1. Frequency of Depression tweets per day

If we observe the trend before preprocessing (the black trend line), the trend is upward over time with many ups and downs. In this figure some spikes are clear and the spike on June 25 is eye-catching than the others. What happened on that day that has made it so

dominant. This trend line is filled with a lot of noises and lacks a plausible explanation, making it unable to reflect the actual depression trends of the population. Figure 6.2 shows the word cloud of the outliers where the most frequent word is "subscribe". So, it was necessary to plot another graph to search for the query. The blue trend line represents the graph after preprocessing. If we compare both lines most of the peaks in the black trend are missing in the blue trend line. After comparing both lines I found the reason behind the spike on June 25. The tweet in figure 6.3 was so popular on that day. The actual creation of that was on 23 June. On June 25, that tweet was retweeted 148.6k times.

Figure 6.2. Word Cloud of the Outliers

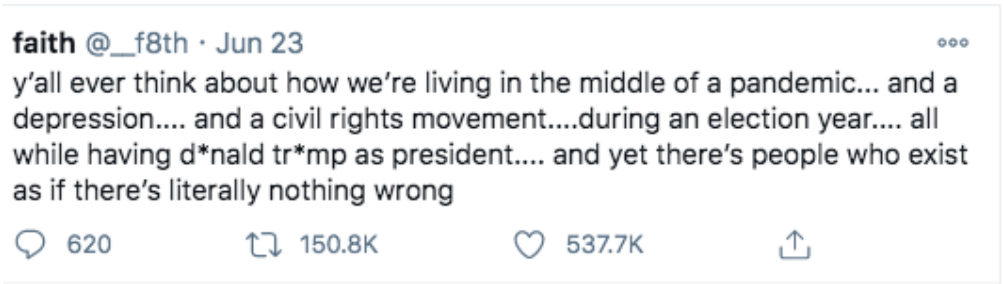

Figure 6.3. Popular Retweet

However, if we observe the figure 6.4, the timeline after data preprocessing is indicating the increase of depression tweets over time with some spikes. I have tried to figure out

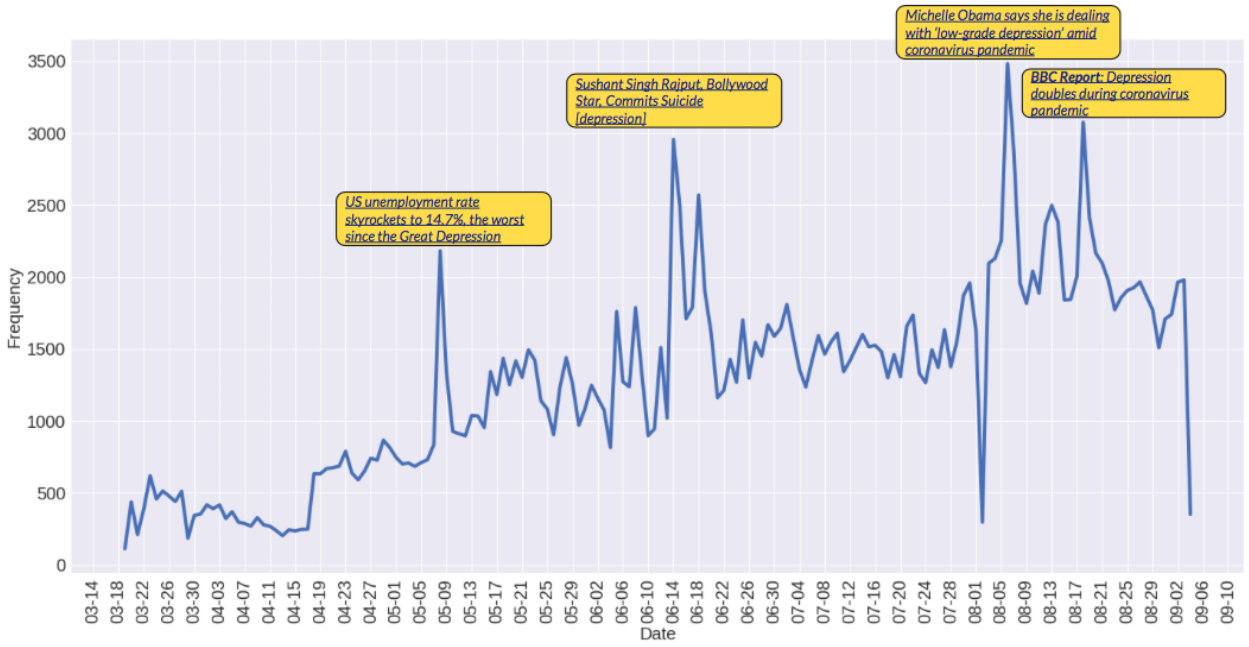

Figure 6.4. Frequency of Depression tweets per day after data preprocessing

the reasons for the top four spikes after reading some sample tweets. The first spike I have found on May 08, 2020. The reason for that spike was crossing of the unemployment rate than the rate during the great depression. "The U.S. economy lost a staggering 20.5 million jobs in April, pushing the unemployment rate to 14.7% according to data released Friday by the U.S. Bureau of Labor Statistics" [ABCNEWS (May 08, 2020)]. The second spike I have found on June 14, 2020. On that day, Bollywood filmstar "Sushant Singh Rajput" committed suicide due to depression [NYTIMES (June 14, 2020), TIMESOFINDIA (June 15, 2020)]. The news on [NYTIMES (August 6, 2020)], "Michelle Obama says she is dealing with 'low-grade depression' amid coronavirus pandemic, racial injustice in the US" was the reason for the peak on August 06, 2020. The last peak I have found in the timeline was on August 18, 2020. The reason for that peak was the news on [BBCNEWS (August 18, 2020)], which reports that during the COVID-19 pandemic, the depression symptom among the adults in Britain found twice compared to the previous year.

6.2 Sentiment Analysis

To understand the change of the emotions and reactions of people over time, I have plotted the average sentiments for each day and month. In the beginning, it may seem that the word "depression" itself is a negative word, so it may not be necessary to analyze their sentiments. But the analysis is telling some other stories.

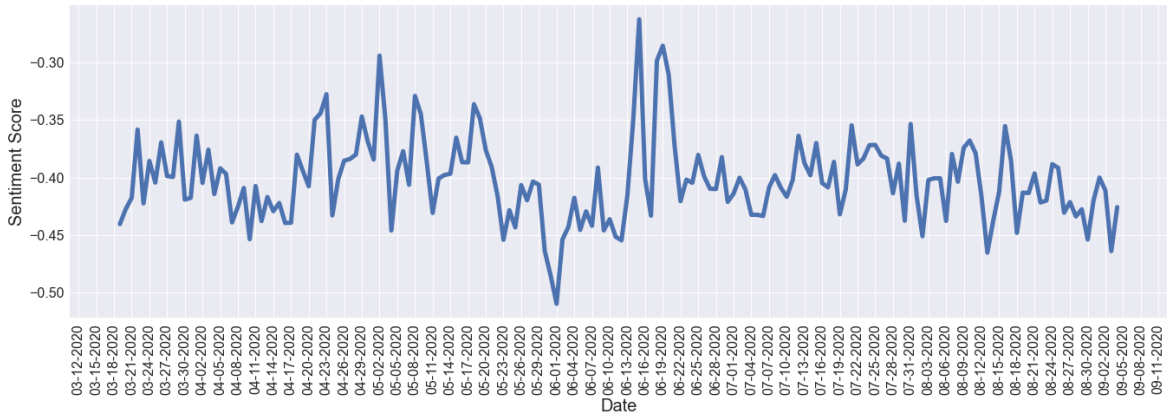

Figure 6.5. Average Sentiment Scores Per Day

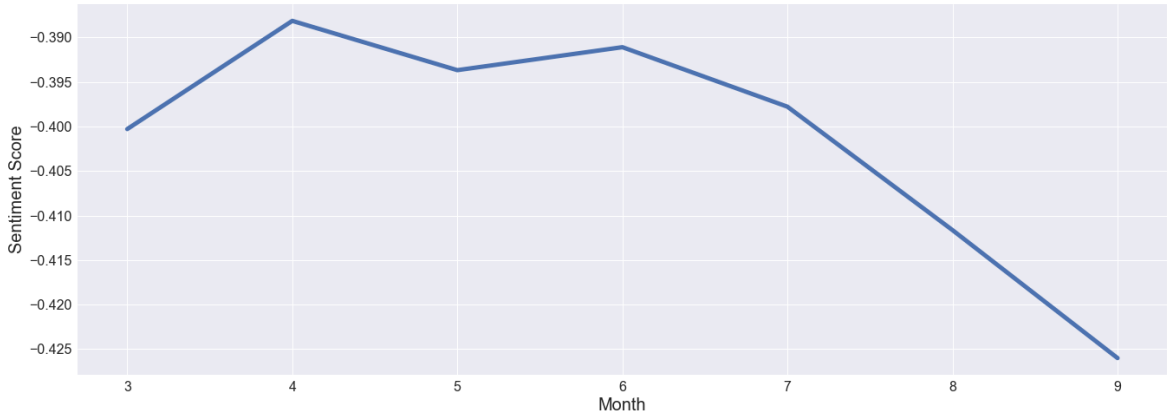

Figure 6.6. Average Sentiment Scores Per Month

Figure 6.7. Sentiment Analysis

Figure 6.5 visualizes the average sentiment values of the depression tweets per day. The trend line is always negative with many ups and downs of the sentiments. The lowest

average sentiment score (less than -0.5) I found on June 1st, 2020, and the highest average sentiment score (more than -0.3) on June 15, 2020, which is closer to the positive. I have tried to figure out the reasons behind these extreme changes but did not any reason.

Figure 6.6 is visualizing the average sentiment values of the depression tweets per month. In this visualization, the trend line is pretty clear. The negative sentiment is consistently higher than the positive sentiment with little ups and downs at the beginning months. But after June, it is consistently downward towards the negative.

6.3 Cluster Analysis

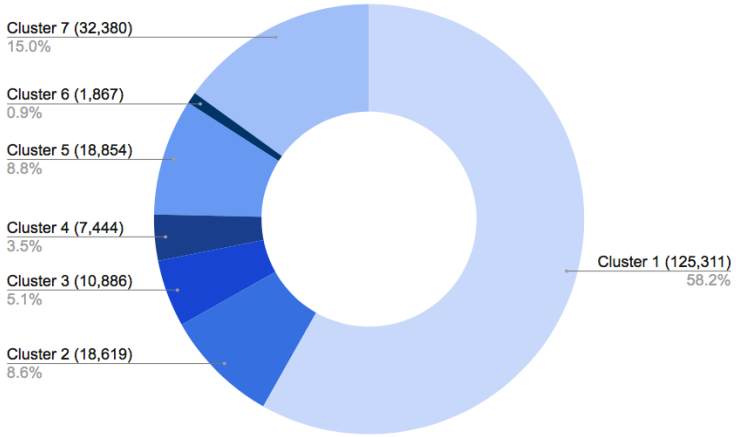

Figure 6.8. Cluster Size

After applying the K-means algorithm I have found 7 distinct clusters. Each cluster represents the tweets that have a similarity between them. All clusters are not the same in size. Some clusters have many more tweets than others. Figure 6.8 shows the percentages of each cluster. To visualize the top frequent bigrams of the clusters I have used Word Clouds. Figure 6.9 to 6.22 visualizes the top bigrams and hashtags of the clusters. Here, each Word Cloud represents one cluster. Figure 6.23 to 6.29 shows the sample tweets of each cluster. K-means labels the clusters with numbers. So, after reading the sample tweets of each cluster and observing the WordClouds, I have labeled them according to their themes. Figure 6.30 to 6.36 visualizes the sentiment graphs of each cluster. This visualization is indicating the

polarity of each cluster. The timeline of depression tweets for each cluster can be visualized in figure 6.37 to 6.43.

Below is the summary of the clusters are described according to their size, themes, sentiments and timeline:

- **Cluster 1 (Personal Feelings):** This cluster is the largest cluster I have found. It contains 125311 tweets, which covered 58.2% of the total tweets.

  The bigrams representing the clusters are "social distancing", "wear mask", "feel like", "working home", "lost job", "every day" and so on. After looking at the bigrams it seems this cluster is representing the tweets that are discussing people's personal life related to the pandemic. In the sample tweets figure 6.23, people are expressing their personal feelings and opinions regarding the social distancing and the pandemic. So, the overall theme of this cluster suggests the label as "Personal Feelings". The average sentiment of this cluster is -0.39. The sentiment graph in figure 6.30 indicates that most of the tweets in this cluster are highly negative, many of them are neutral and few of them are positive. So, this cluster can be considered as a negative cluster.

  The timeline of this cluster in 6.37 is pretty similar to the timeline that is visualizing the overall change in the depression tweets. The overall trend of this cluster is increasing with the time, though there are some ups and downs.

- **Cluster 2 (Economy/Unemployment):** This cluster contains 18619 tweets which covered 8.6% of the total tweets.

  This cluster is representing the economic problems due to the pandemic. The bigrams "since great", "unemployment rate", "highest unemployment", "unemployment science", "worst economic" are representing the tweets that are related to the economy. Figure 6.24 shows the sample tweets for this cluster, where people are discussing the economic problem specially the unemployment problem due to COVID-19. So, I have labeled this cluster as "Economy/Unemployment".

The sentiment of this cluster is pretty similar to cluster 1. The average sentiment of this cluster is -0.32. The sentiment graph in figure 6.31 indicates that most of the tweets in this cluster are highly negative, many of them are neutral and few of them are positive. So, this cluster can be considered as a negative cluster.

Figure 6.38 visualizes the timeline of depression tweets for cluster 2. The overall trend of this cluster is also increasing with time with some ups and downs. The highest peak for this cluster on May 08, 2020. This peak has also triggered the peak of the overall depression timeline. We found earlier the reason for this peak, which is exceeding the rate of unemployment from the rate during the great depression.

- **Cluster 3 (Trump's Policy/ Politics):** This cluster contains 10886 tweets which covered 5.1% of the total tweets.

  The top bigrams for this clusters are "donald trump", "trump administration", "great trump", "michelle obama", "pandemic trump", "trump america", "race riot" and so on. The top hashtags for this clusters are "trumpion", "trump", "trumphasnoplan", "biden2020" and so on. Here, the hashtag "trumpion" is for "trumpdepression". The depress word is missing here because of the cleaning I did before the analysis. These bigrams and hashtags are related to Trump's policy and other political issues in the US due to the pandemic. Also in the sample tweets in figure 6.25, people are discussing US political issues and especially most of them are blaming Trump's policy related to the pandemic. So, I have labeled this cluster as "Trump's Policy/ Politics".

  The average sentiment score for this cluster is -0.52. In figure 6.32 the sentiment graph indicates that it is a highly negative cluster. Most of the tweets in this cluster are negative and very few of them are neutral and positive.

  Figure 6.39 visualizes the timeline of depression tweets for cluster 3. The reasons for the two spikes on May 08(High Unemployment rate) and August 06 (The news of Michelle Obama's depression) have been explained earlier.

- **Cluster 4 (Student's concern regarding exams):** This cluster contains 3.5% of the total tweets which is 7444.

  The top bigrams I have found here are "sir please", "please sir", "sir give", "justice listen", "exam student", "situation student" and so on. The hashtags I have found here "studentsinscforjustice", "cancelfinalyearexam", "studentslifesmatter", "speakupforstudents", "cancelcompartment" and some others. The discussions in the sample tweets in figure 6.26, indicates the concern of the students regarding their education, especially in India. So, I have labeled this cluster as "Student's concern regarding exams".

  The average sentiment score for this cluster is -0.29. Like the previous clusters, this cluster can also be considered as a negative cluster. In figure 6.33 the sentiment graph indicates that most of the tweets in this cluster are negative, but many of them are positive and few of them are neutral.

  The timeline for this cluster in figure 6.40 indicates the increase of depression tweets after the month of May.

- **Cluster 5 (Mental health):** This cluster is consists of 18854 tweets. It covered 8.8% of the total tweets.

  The most frequent bigrams for this cluster are "mental health", "health issue", "anxiety pandemic", "stress anxiety", and others. Some of the top hashtags for this cluster are "mentalhealth", "mentalhealthmatters", "mentalhealth anxiety", "health" and so on. It seems that the tweets within this cluster are related to mental health-related issues. In the sample tweets in figure 6.27, people are discussing their mental health problems, for example, sleeping problems, anxiety, and depression. They are also advising those who are suffering from depression. So, I have simply labeled this cluster as "Mental Health".

  The average sentiment score for this cluster is -0.53. This is the most negative cluster

among all the clusters. In figure 6.34 the sentiment graph indicates that most of the tweets in this cluster are negative and very few of them are neutral and positive.

The timeline for this cluster in figure 6.41 indicates the increase of depression tweets over time. So, the overall mental health-related issues have increased with time.

- **Cluster 6 (Showing Empathy):** This is the smallest cluster I have found. This cluster consists of 1867 tweets. It covered %.09 of the total depression tweets.

  The top frequent bigrams for this clusters are "repost lockdown", "could two", "two twitter", "copy repost", "repost lockdown" and so on. The sample tweets in figure 6.28 suggests that people are sharing helpline numbers for those who are suffering from depression. So, I have labeled this cluster as "Support for depression".

  The average sentiment score for this cluster is 0.033. This cluster is a neutral cluster and totally different from the other clusters in respect to their sentiment. In figure 6.35 the sentiment graph indicates that most of the tweets in this cluster is neutral and very few of them are positive or negative.

  The timeline for this cluster in figure 6.42 shows that this cluster has started in May and ended in August. The only peak I found on June 16.

- **Cluster 7 (General Discussions related to Pandemic)**: This is the second-largest cluster that contains 15.0% of the tweets which is 32380 tweets.

  Some of the most frequent top words for this clusters are "global pandemic", "pandemic economic", "middle pandemic", "double pandemic", "race riot" and so on. The top hashtags for this clusters are "mental health", "trump", "pandemic", "anxiety", "blacklivesmatter" and others. Though most of the top hashtags for this cluster overlap with the other clusters, the top bigrams indicate it as a global issue related cluster. If we observe the sample tweets in figure 6.29, people are discussing different issues related to the pandemic. So, I have labeled this cluster as "General Discussions related to Pandemic".

The average sentiment score for this cluster is -0.44. Like most of the cluster this cluster is negative. In figure 6.36 the sentiment graph indicates that most of the tweets in this cluster are negative and few of them are neutral or negative.

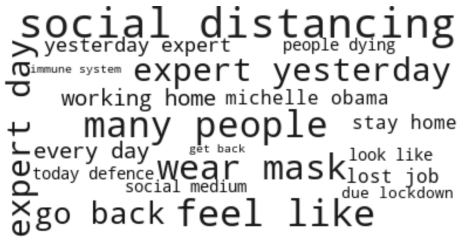

Figure 6.9. Word Cloud (Cluster 1)

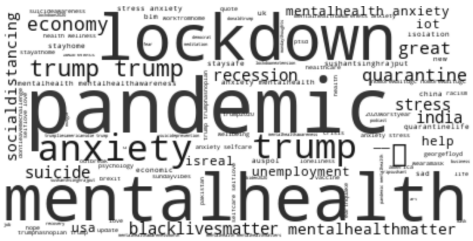

Figure 6.10. Hash Cloud (Cluster 1)

Figure 6.11. Word Cloud (Cluster 2)

Figure 6.12. Hash Cloud (Cluster 2)

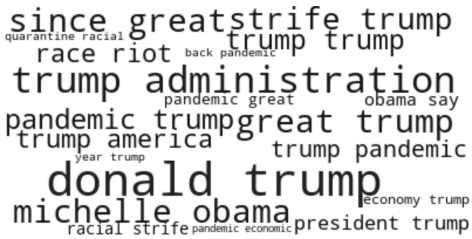

Figure 6.13. Word Cloud (Cluster 3)

Figure 6.14. Hash Cloud (Cluster 3)

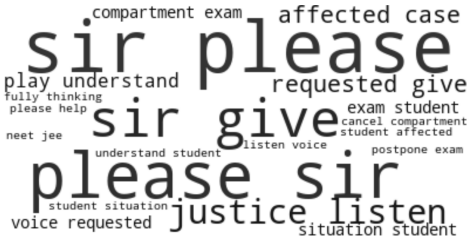

Figure 6.15. Word Cloud (Cluster 4)

Figure 6.16. Hash Cloud (Cluster 4)

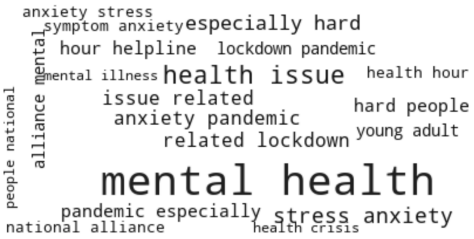

Figure 6.17. Word Cloud (Cluster 5)

Figure 6.18. Hash Cloud (Cluster 5)

Figure 6.19. Word Cloud (Cluster 6)

Figure 6.20. Hash Cloud (Cluster 6)

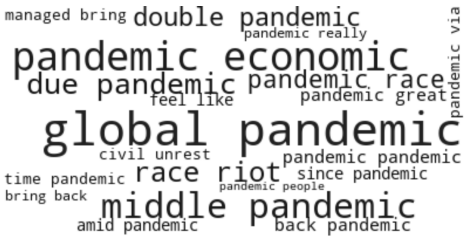

Figure 6.21. Word Cloud (Cluster 7)

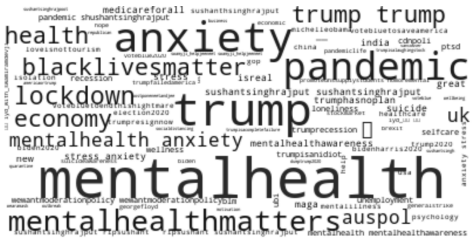

Figure 6.22. Hash Cloud (Cluster 7)

Sample Tweets	Interpretation of the theme	Labeling
i agree 💯 %. i can't see my children or my grandson because of this lockdown. luckily i'm not in care but i'm getting depressed. i'm sick but i can't see my loved ones. □. it isn't fair. this government is heartless. it's drunk on power! 😡	Social distancing	Personal feeling and mixed, can be considered as general themes
depression is that which needs to be killed before corona.	Personal feeling	
talk about how we can stay happy in quarantine because i'm about to be depressed i'm literally so sad i miss my school i miss my friends	Social distancing	
is the lockdown making you depressed, or are you just bored?	General discussion/ personal feeling	
these small countries should be in a union. too wee, too poor, too stupid, surely? how can they cope with covid, the banking crisis, the depression, and still have stuff to trade and still pay pensions far in excess than is paid in the uk?	Global discussion	

Figure 6.23. Sample tweets of Cluster 1

Sample Tweets	Interpretation of the theme	Labeling
"the u.s. unemployment rate jumped to 14.7 percent in april, the highest level since the great depression, as most businesses shut down or severely curtailed operations to fight the deadly coronavirus."	Unemplyment	Unemployment/ Economy
everything you do benefits russia and hurts america and americans. proof? 170,000+ dead americans as pandemic out of control. economy ruined, unemployment numbers worst since depression, no foreign policy. you hate us and it shows.	Unemployment/ Economy	
unemployment surged to 14.7% in april, highest since great depression, as coronavirus triggered 20.5 million job losses	Unemployment	
how long we can expect the u.s. economic catastrophe to endure: the coronavirus pandemic leveled the u.s. economy in the second quarter of the year, leading to the worst collapse since the great depression. gross…	US economy	
let us not forget that trump was negligent in his handling of the coronavirus,105,000+american people have died,the unemployment rate is lower than it has been since the great depression,and the economy is tanked, so what was he saying about making america great again???????!!	Unemployment/ economy	

Figure 6.24. Sample tweets of Cluster 2

Sample Tweets	Interpretation of the theme	Labeling
michelle says she's dealing with 'low-grade depression' because of quarantine, racial strife and trump	Politics	US Politics and Trumps policy
russia interfering in our election, kids in cages, totally corrupt administration, pandemic totally out of control, economic depression, peaceful protesters beaten, hate crime to no end, making enemies out of allies. this is trump's version of what makes america great again!	Politics/ Blaming trumps policy	
thanks to you, seniors living in nursing homes have been isolated from each other for months. many have died in 2020 from depression and loneliness. even if they don't die of covid, covid (and trump) killed them. 8/4/2020 8:45 am est 4,863,077 cases 158,975 dead	Blamming trump	
trump wants you to think joe biden would make things worse than he has made them. us is first in the world in covid infections and deaths, worst economy since the depression, and race relations have never been worse than now. foh!	Politics/Trump	
socially and morally bankruptcy of trump is destroying america from the inside to out. footnote of history: trump's buffoonish presidency nearly bankrupted america, caused national unrest and caused the second great depression because he failed to control a raging coronavirus.	Blaming trumps policy	

Figure 6.25. Sample tweets of Cluster 3

Sample Tweets	Interpretation of the theme	Labeling
australian government: give australian students the choice to continue learning from home - please sign the petition!	Students in Australia want to study from home	Imapct of COVID-19 on Students specially in INDIA
sir do something and cancel the exams sir we are depressed 😠 about the exams and corona cases are increased 33 lakhs in india how can we write exams sir 😭😭 cancel ugc exams sir please🙏🙏	Students are depressed about exams in India	
sir please ,we students are mentally depressed in these pandemic times.98 percent of the students want to postpone exams . please address us and tell your opinion sir. and 13 students already committed suicide because of this issue. please help	Request to postpone exams	
we international students assure you we will follow each and every rule about quarantine. it's a request let us in we have visa before march 18th.	Students are conserns about the visa issue	
every student's life is important for development of nation. student community is currently suffering a lot from mental anguish & depression.conducting exams in this pandemic will be dangerous. please & demand to postpone exams until normalcy is restored.	Situation of students due to pandemic	

Figure 6.26. Sample tweets of Cluster 4

Sample Tweets	Interpretation of the theme	Labeling
yeah anxiety depression has been kicked up even worse during pandemic. have had to see a doctor for extra help. most of the time i get 4-5 hours of sleep a night max now. rough.	Sleep problem due to anxiety and depression	Mental Health
reach out to family and friends this weekend! staying connected is a good way to reduce the anxiety, depression and loneliness social distancing can bring on.	Advice how to cope with the depression	
the mental health issues related to lockdown and the pandemic are especially hard for people with depression. the wonderful charity mind have a 24 hour helpline: 03001233393. ❤️❤️	Helpline for depression	
the percentage of people facing anxiety and depression is no doubt rising during the covid-19 pandemic. if you know of someone facing these symptoms try to form connection with them. connection right now is so key to help us feel that we are not alone.	Help people who are depressed	
parental depression, anxiety during covid-19 will affect kids too	Kids are in anxiety/depression	
this sad story is around the world now among medical workers. before covid-19 health care workers were already vulnerable to depression and suicide mental health experts now fear even more will be prone to trauma-related disorders. can not take out braina	Mental health	

Figure 6.27. Sample tweets of Cluster 5

Sample Tweets	Interpretation of the theme	Labeling
please could any 2 of my twitter followers please copy and repost? this lockdown period is especially hard for people with depression. samaritans uk tel no. 116 123. just two. any two. say done.	Copy and repost the helpline number	Helpline for Depression
the mental health issues related to lockdown and the pandemic are especially hard for people with depression. the wonderful charity mind have a 24 hour helpline: 03001233393 please could any two of my twitter friends just copy and repost to share the helpline far and wide.	Copy and repost the helpline number	
this lockdown period is especially hard for people with depression 24hr helpline: 0800 456 789	Helpline number	

Figure 6.28. Sample tweets of Cluster 6

Sample Tweets	Interpretation of the theme	Labeling
its very heart ruin news .i feel many young people like sushant singh daily face lot of depressed things due to pandemic and no body can understand their feelings and they are finishing their lives and everything going under suspense eventhough govt.'s activities	General discussion about depression due to pandemic	General and global discussion related to pandemic
pandemic to worsen japan's suicide rate? distress calls have increased after japan announced its lockdown and state of emergency. people are now fighting two wars-pandemic and depression. tells you more	Pandemic in Japan	
pandemic depression global economy will never be the same-world bank estimates as many as 60 million people globally will be pushed into extreme poverty as a result of the pandemic. un recently warned the world is facing the worst food crisis in 50 years.	General discussion about global pandemic	
the way discussion about what to do about the pandemic, as individuals and organizations, has really become about who be the most tediously moralistic rather than about what is actually relevant for public health and it's just depressing.	General discussion about pandemic	
we search for solutions. however, solutions are hard to come by when we don't know the truth about the pandemic. black workers are suffering under the distress of a depression era unemployment rate of 26%. may's national jobless rate was 21.2%, not 13.3%.	General discussion about pandemic	

Figure 6.29. Sample tweets of Cluster 7

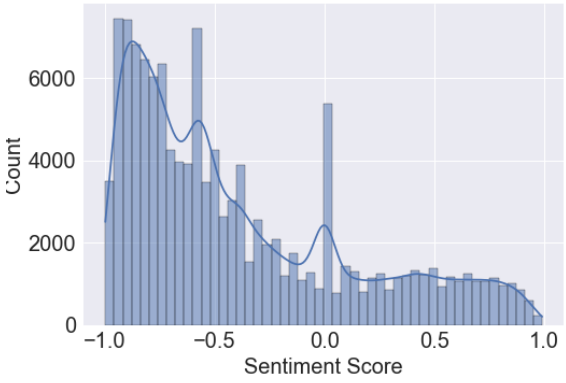

Figure 6.30. Sentiments (Cluster 1)

Figure 6.31. Sentiments (Cluster 2)

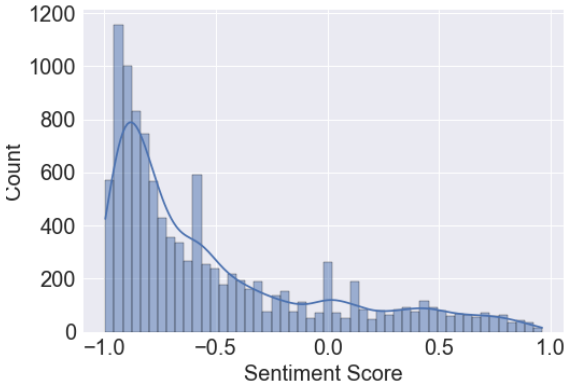

Figure 6.32. Sentiments (Cluster 3)

Figure 6.33. Sentiments (Cluster 4)

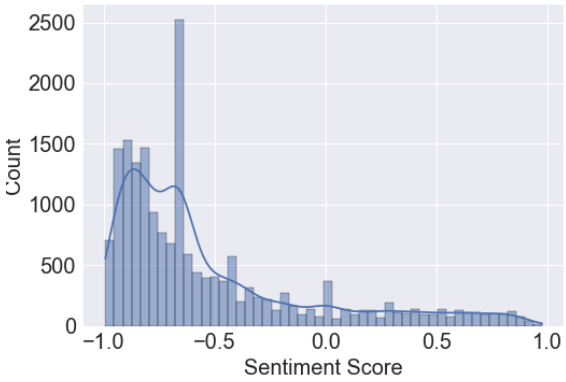

Figure 6.34. Sentiments (Cluster 5)

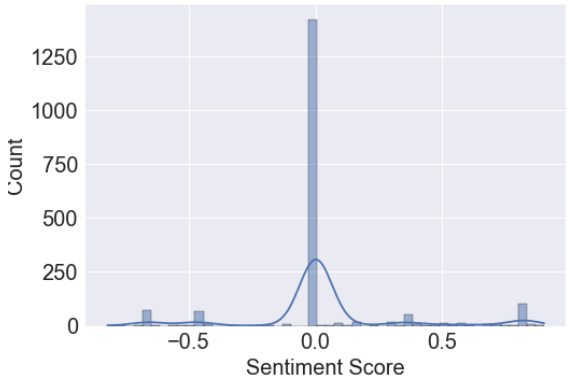

Figure 6.35. Sentiments (Cluster 6)

Figure 6.36. Sentiments (Cluster 7)

Cluster ID	Mean Sentiment	Std
1	-0.387131	0.513961
2	-0.31642	0.538817
3	-0.516995	0.477779
4	-0.287206	0.530624
5	-0.527737	0.452534
6	0.033234	0.289603
7	-0.435072	0.498558

Table 6.1. Average sentiment scores for each cluster

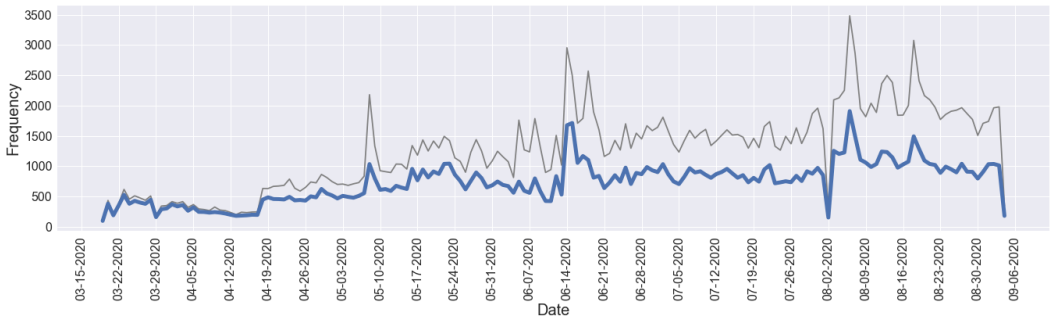

Figure 6.37. Cluster 1 Timeline

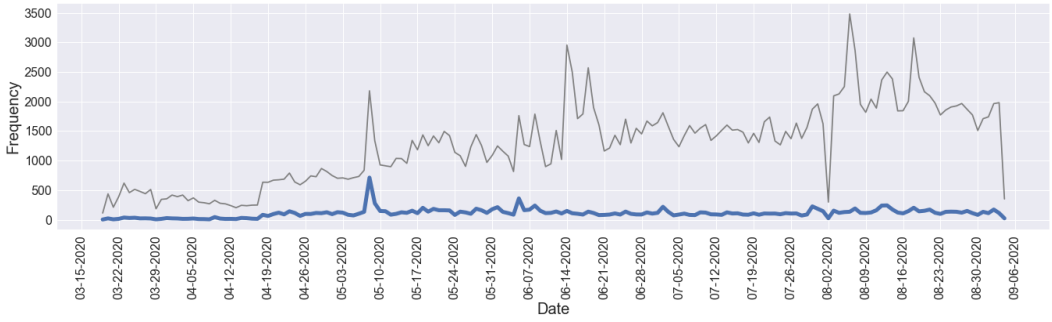

Figure 6.38. Cluster 2 Timeline

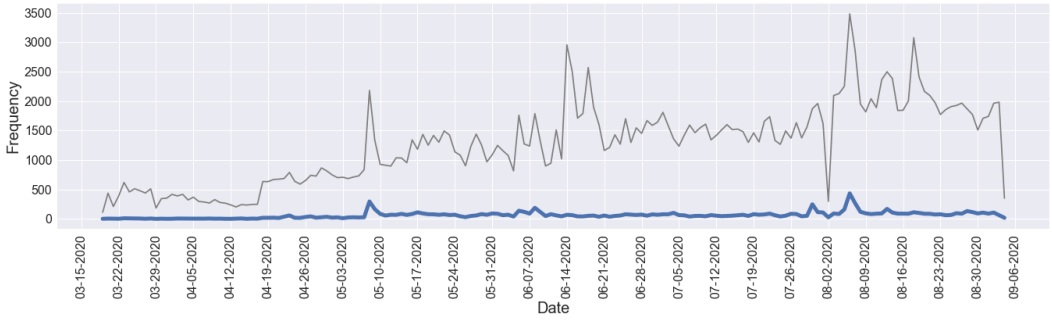

Figure 6.39. Cluster 3 Timeline

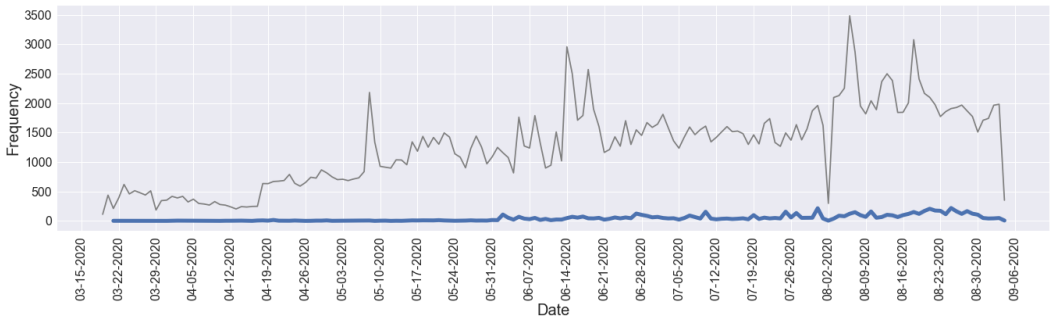

Figure 6.40. Cluster 4 Timeline

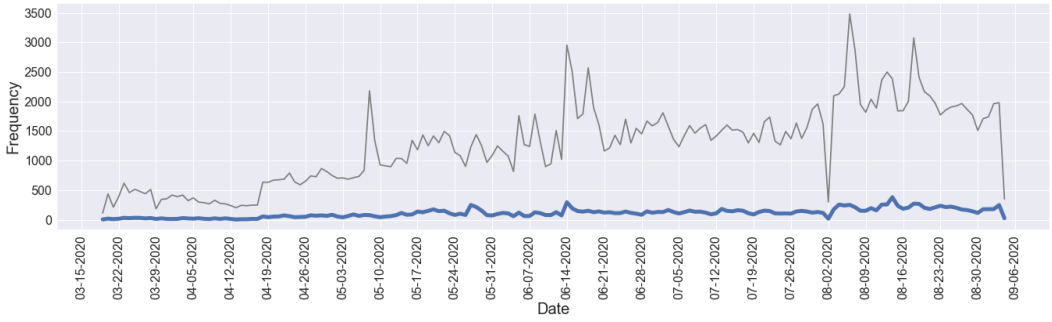

Figure 6.41. Cluster 5 Timeline

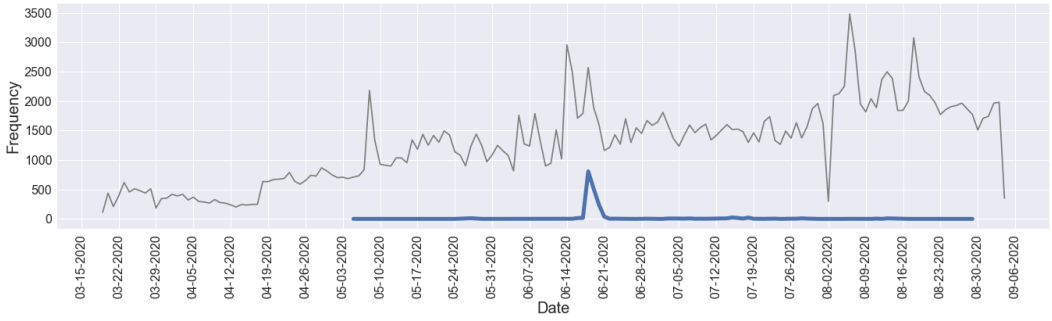

Figure 6.42. Cluster 6 Timeline

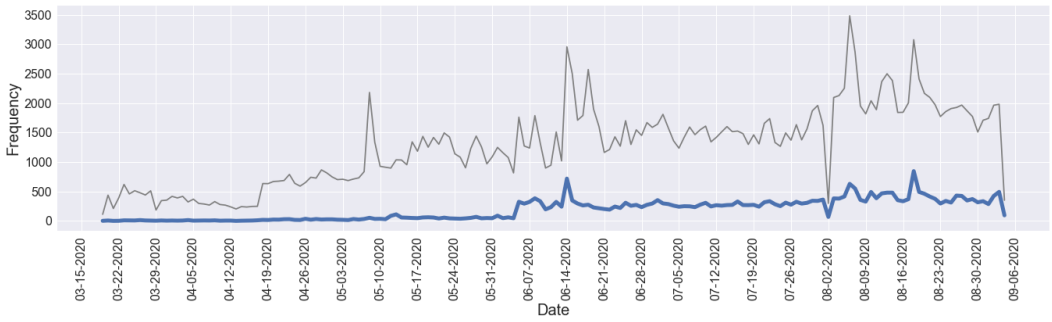

Figure 6.43. Cluster 7 Timeline

6.4 Topic Analysis

After applying biterm topic modeling for multiple times, found 7 topics that seems to be covered by the all co-ordinates without overlapping any of them. I also tried to keep the 7 as their themes seems to be matched with the 7 clusters I found earlier. The 7 topics represent 7 themes about the depression-related tweets. Figure 6.44 represents the positions of the topics in the inter topic distance map. Here, each circle represents each topic and the distances between them are presented in a 2D plane. Figure 6.45 shows the 30 most important terms in the whole corpus and figure 6.46 to 6.52 shows the the 30 most important terms for each topic.

The biterm algorithm does not provide the labels for the topics. So, I have labelled the topics by myself. I have tried to use the same labels for the topics that I have used for the clusters. The exception I have found, Cluster 7 has discussed global and general issues related to pandemic and cluster 2 has both economy and unemployment related discussions, whereas topic 2 indicates the "Economy" and topic 6 indicates "Unemployment" related terms. But the overall themes are same. Below top ten words for each topic and their themes are presented:

- **Topic 1 (Personal Feelings):** like, get, feel, really, home, go, thing, know, time, day
- **Topic 2 (Mental Health):** anxiety, suicide, people, health, lockdown, mental, social, pandemic, isolation, many
- **Topic 3 (Impact on Education):** student, anxiety, exam, sir, mental, pandemic, stress, health, due, help
- **Topic 4 (Economy):** economic, people, economy, pandemic, great, death, business, job, country, would
- **Topic 5 (Politics/ Trump):** pandemic, trump, obama, riot, racial, race, great, say, michelle, president

Figure 6.44. Intertopic Distance Map

- **Topic 6 (Unemployment):** great, unemployment, since, worst, american, pandemic, economy, trump, highest, economic
- **Topic 7 (Helpline):** especially, hard, please, two, twitter, friend, period, helpline, lockdown, could

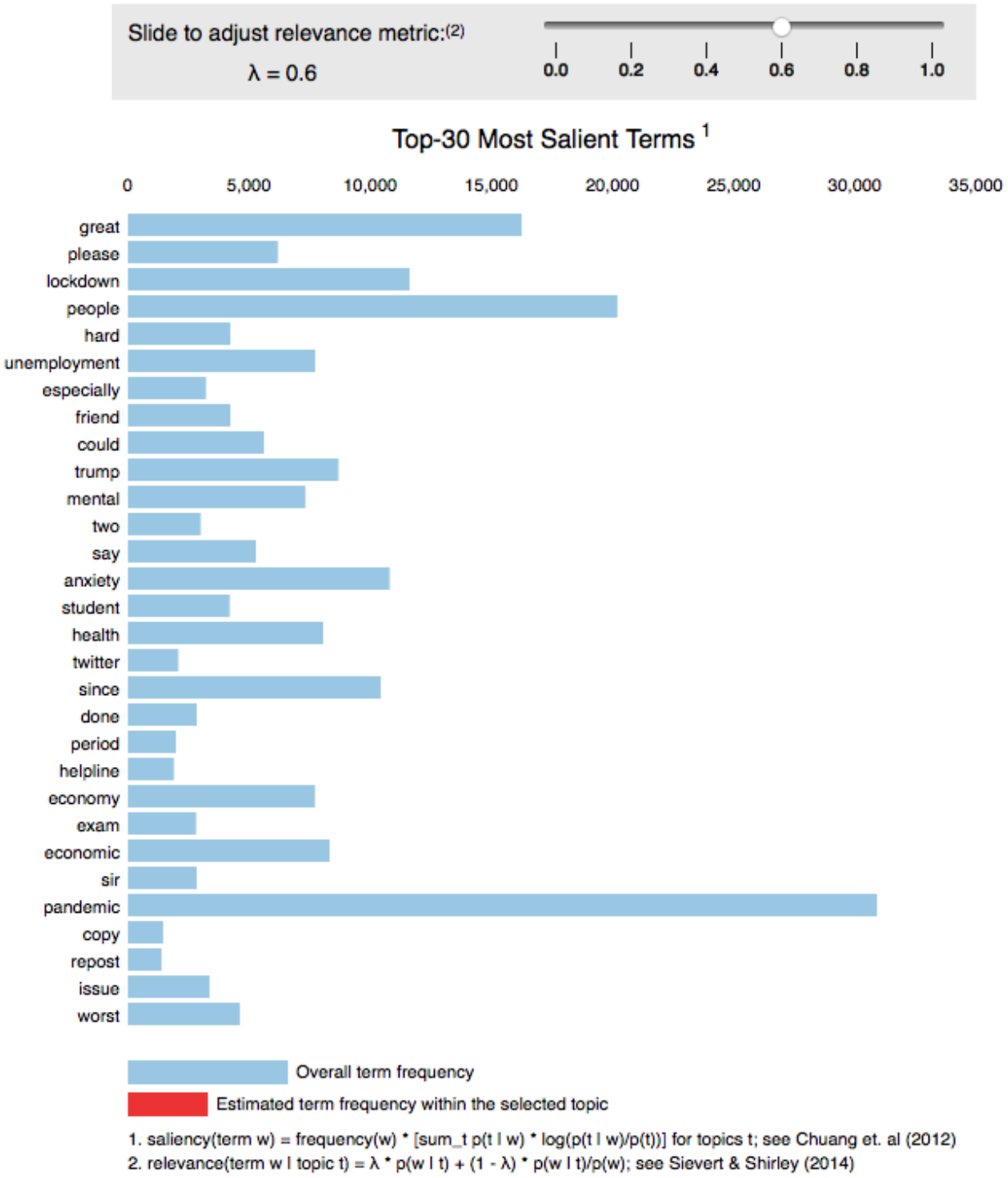

Figure 6.45. 30 Most Salient Terms

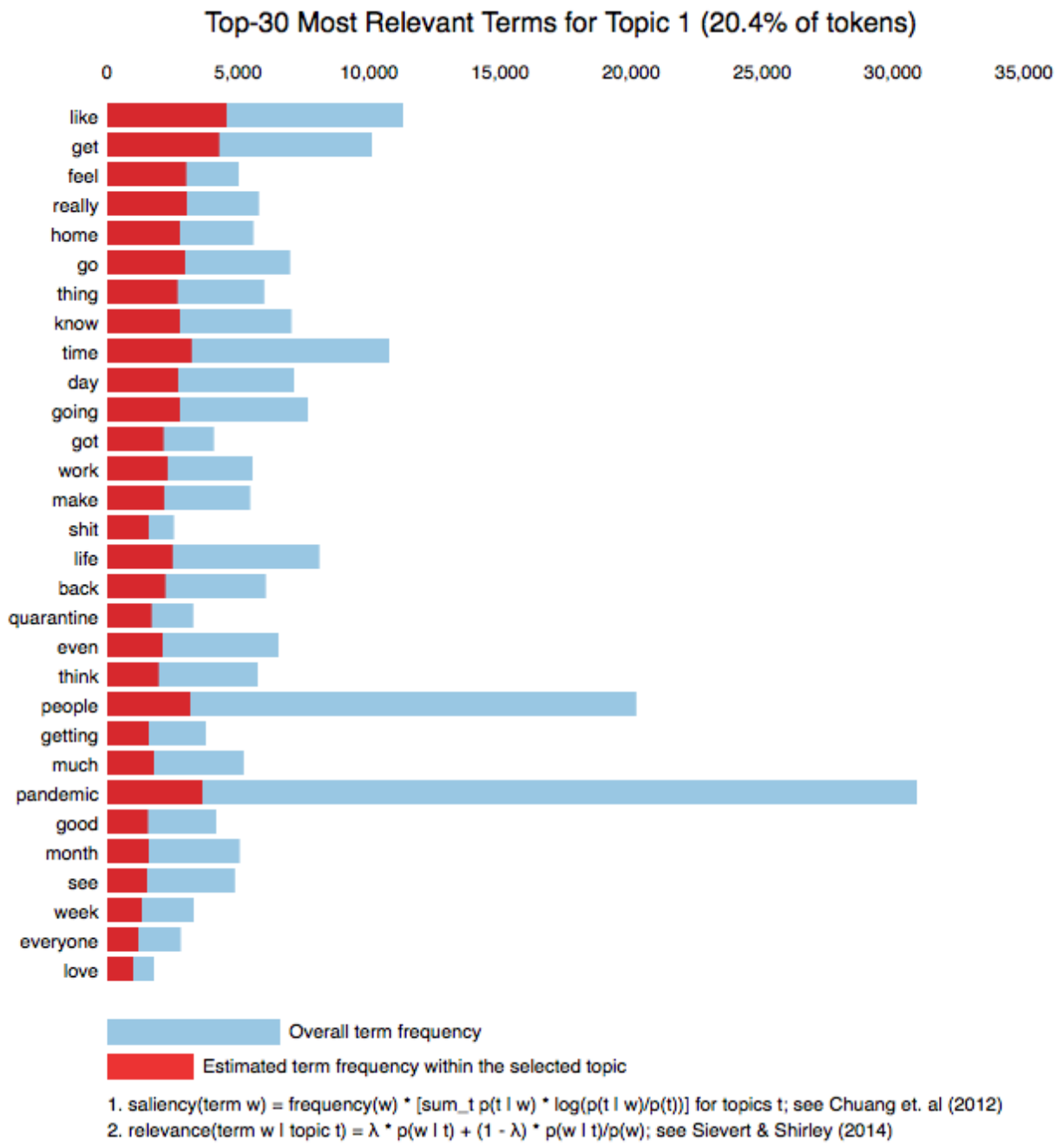

Figure 6.46. Topic 1

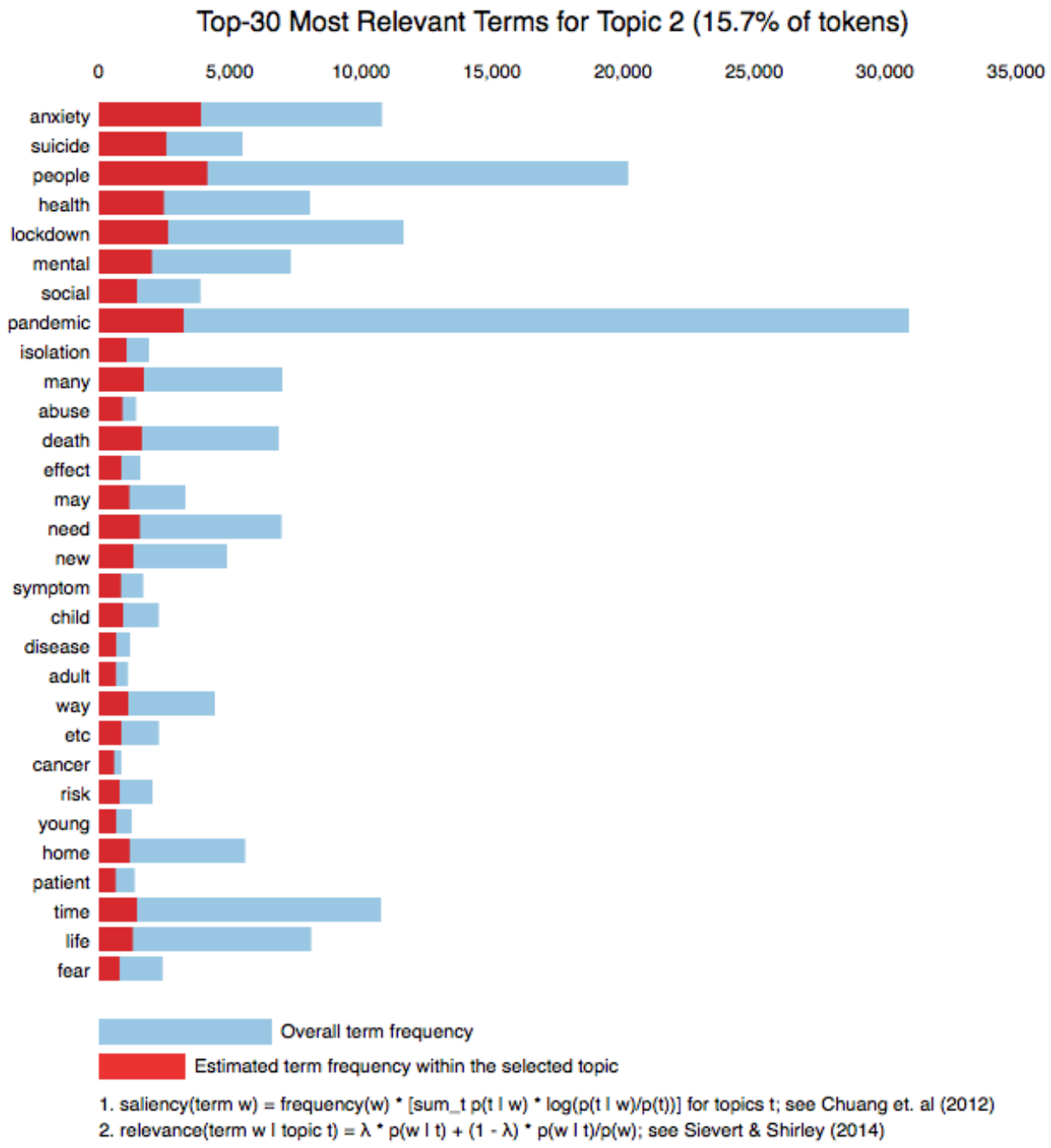

Figure 6.47. Topic 2

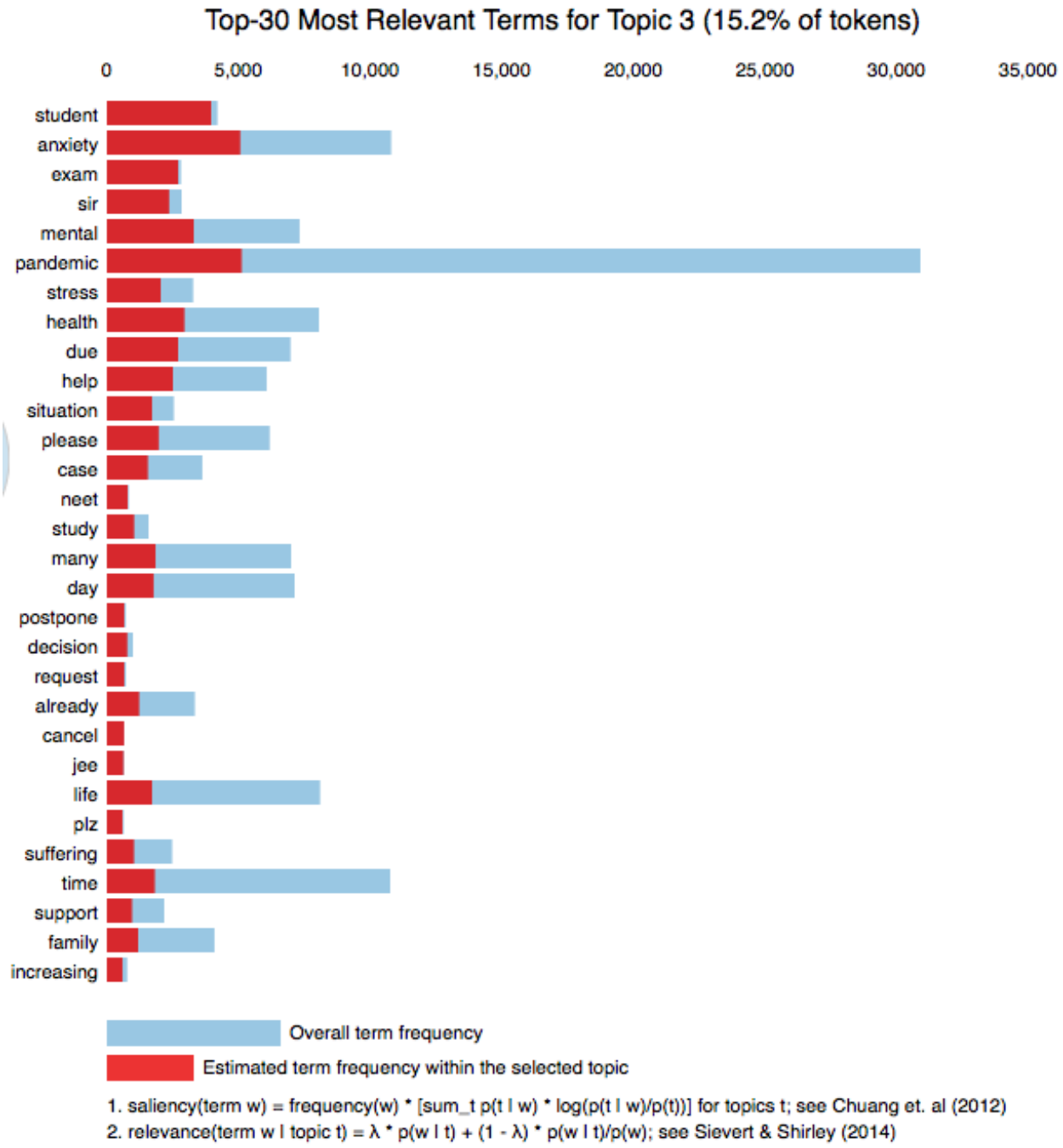

Figure 6.48. Topic 3

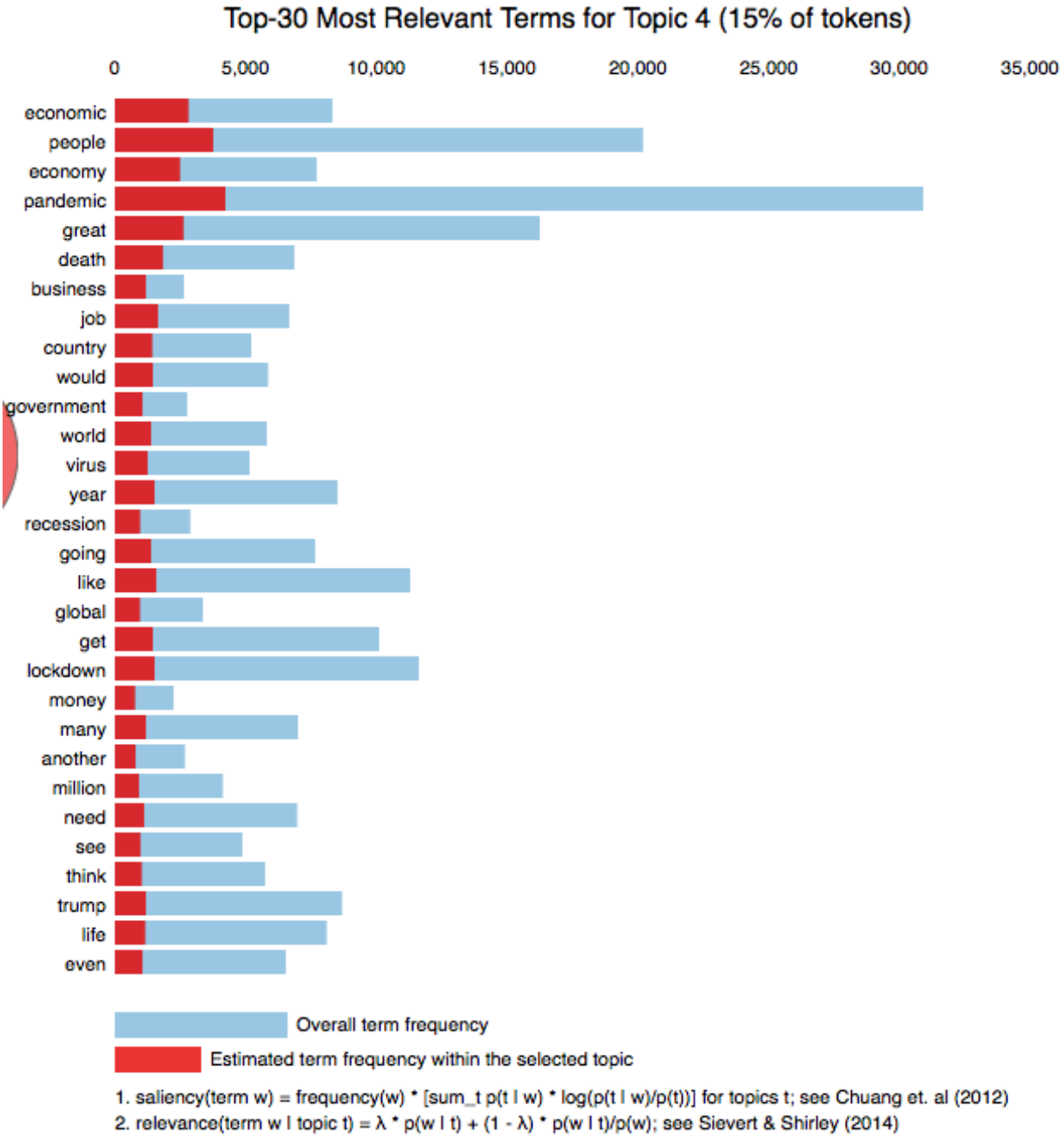

Figure 6.49. Topic 4

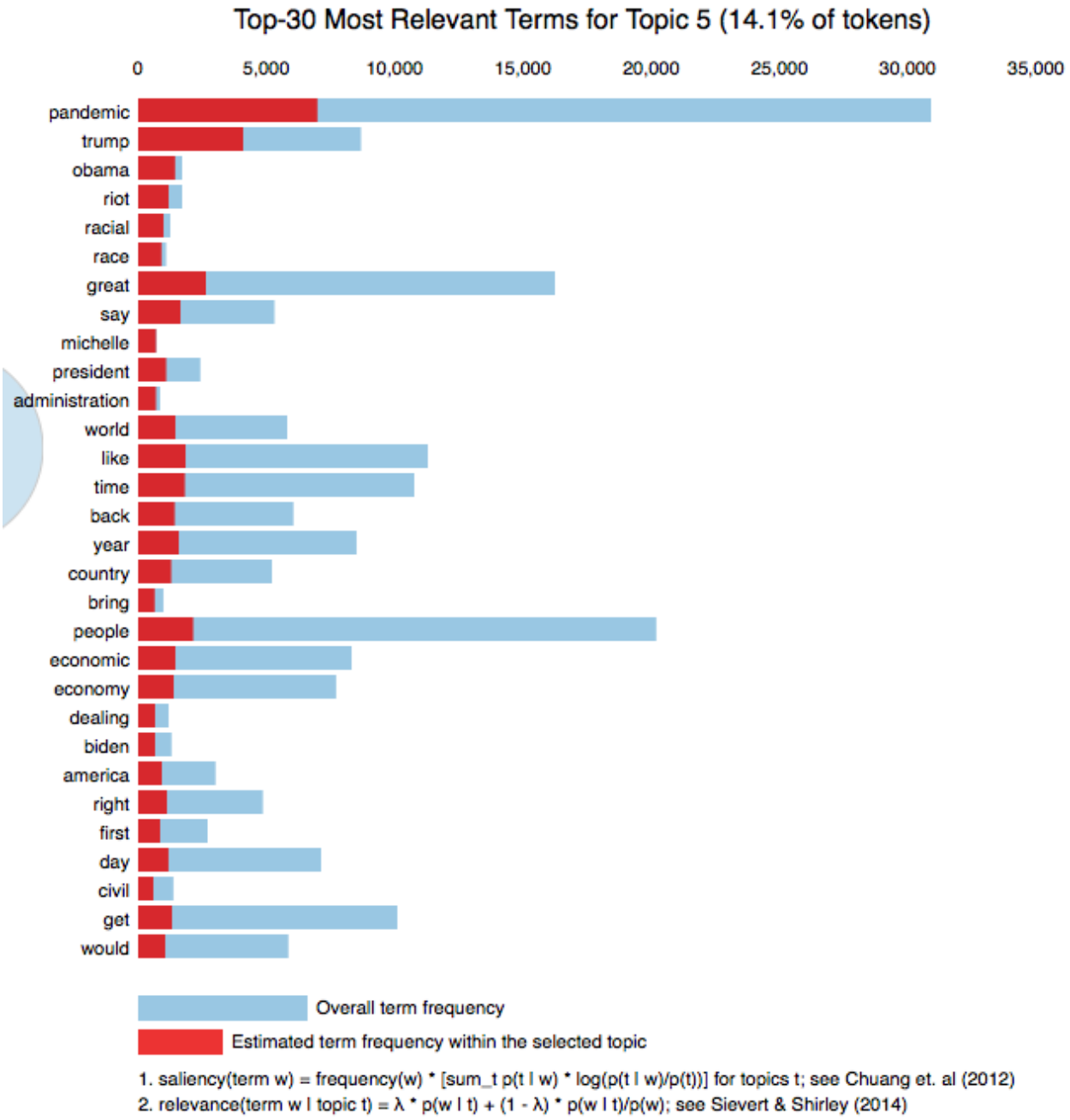

Figure 6.50. Topic 5

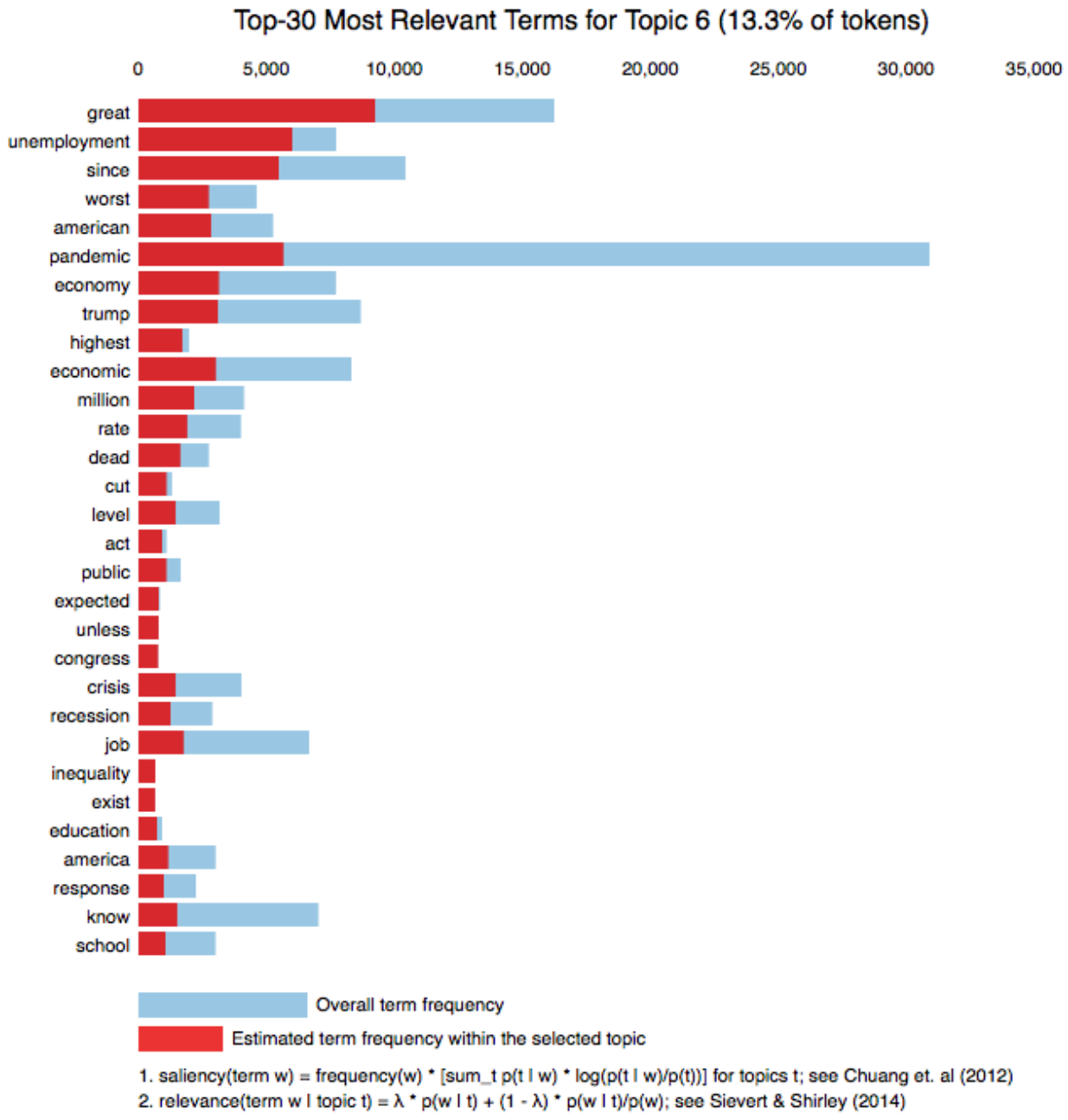

Figure 6.51. Topic 6

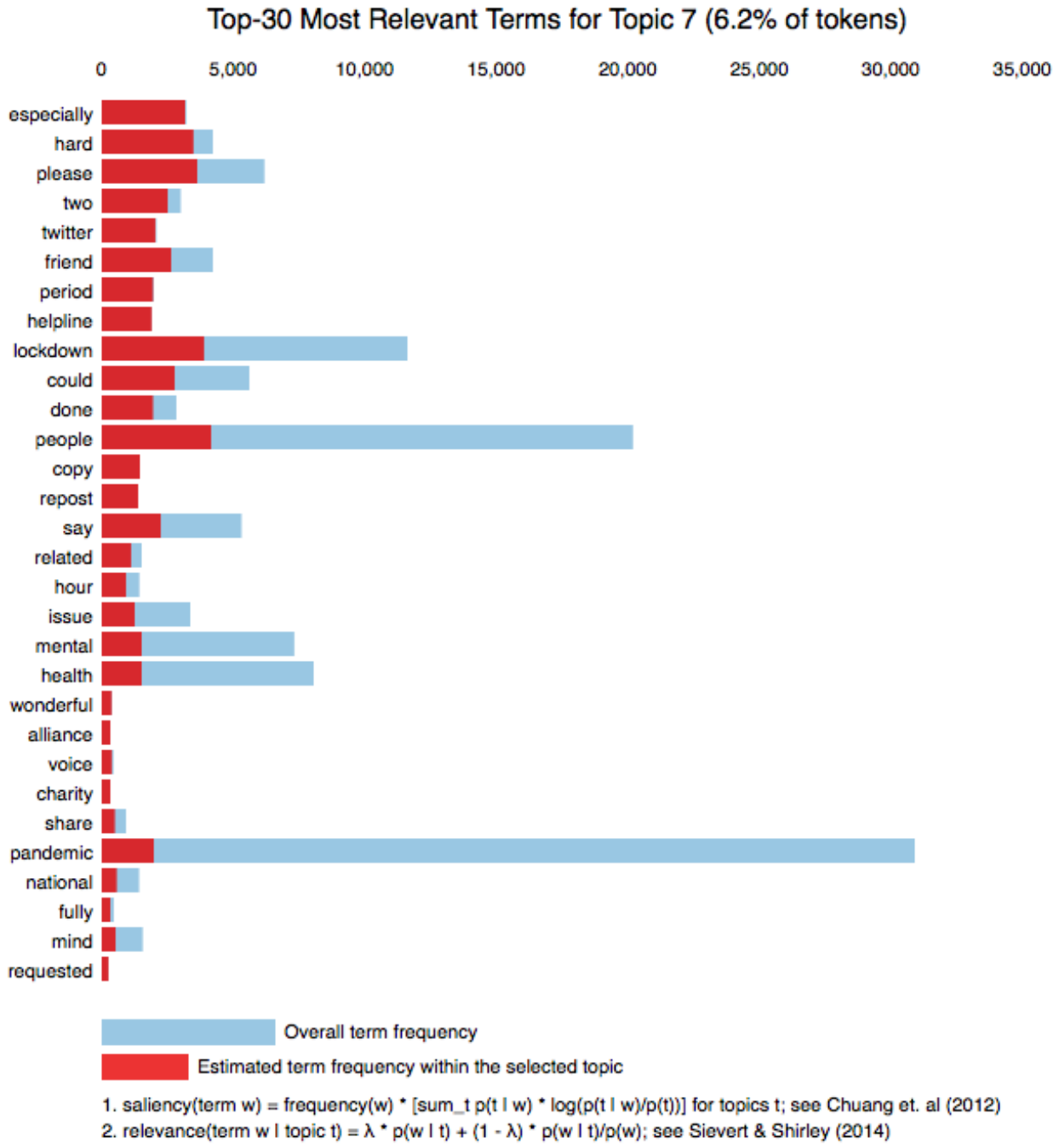

Figure 6.52. Topic 7

CHAPTER 7

LIMITATIONS AND FUTURE WORKS

7.1 Limitations

There are a few limitations in this study. First, the dataset I have used is in English. It does not cover the other languages. So, the overall number of depression tweets during this time period is missing here and the results are limited to a particular group. Secondly, I have studied only the collection of tweets that is published in IEEE dataset. They have added many hashtags and terms later to collect the data. Some data are deleted by the users. So, this dataset does not cover the all depression tweets that are in English.

7.2 Future Works

The clusters that I have found can be used to label the data so that supervised learning can be applied over future data. Each of the clusters can be used to do further analysis, for example the economy cluster can be used to get the detail analysis related to COVID-19, depression, and economy.

The analysis can continue by tracking the data for a longer period of time and track the changes over time. Besides, geo-location based analysis can be done to see the emotions of people in different regions and how they change over time.

By collecting depression related tweets before and after the pandemic a comparison can be done between the depression related tweets before, during and after the COVID-19 pandemic to see how the depression related issues evolve.

CHAPTER 8

CONCLUSION

The purpose of this study is to understand the discussions related to depression about the COVID-19 pandemic on Twitter. To analyze those discussions, I have performed a trend analysis of the collected data to see the overall change in the depression. In this analysis, I have found some days when the depression level was higher than the average. I also have performed the sentiment analysis over those data and found the overall sentiment always negative and over time the average sentiment is downward toward the negative. To get a better understanding of the depression, I have performed a cluster analysis and found 7 distinct clusters. The clusters are representing the overall picture of the depression tweets. To understand the clusters, I have performed top bigram analysis and labeled them according to their themes. I also did a sentiment analysis of those clusters and found six of them are mostly negative and one neutral cluster. I have analyzed the trends for each cluster over time and found cluster 1 similar to the pattern of the overall change in depression tweets. This is the biggest cluster and that indicating the reason for dominating the overall trend. For further analysis, I have performed a topic analysis and found 7 distinct topics. I have found similarities between the themes of the clusters and the themes of the topics. The cross-checking has performed well.

Though there are limitations of this work, it has given an overview of depression-related discussion during a pandemic. This study can be the basis for many works in the future. This study can be helpful for researchers who are working with depression in the pandemic.

BIBLIOGRAPHY

BIBLIOGRAPHY

ABCNEWS (May 08, 2020), Us unemployment rate skyrockets to 14.7%, the worst since the great depression.

Balani, S., and M. De Choudhury (2015), Detecting and characterizing mental health related self-disclosure in social media, in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 1373–1378.

BBCNEWS (August 18, 2020), Depression doubles during coronavirus pandemic.

Drillinger, M. (September 10, 2020), Depression symptoms 3 times higher during covid-19 lockdown.

Education, I. C. (September 21, 2020), Unsupervised learning.

FOXNEWS (October 23, 2020), Trump addresses addiction, depression due to covid-19 lockdowns.

Gao, J., P. Zheng, Y. Jia, H. Chen, Y. Mao, S. Chen, Y. Wang, H. Fu, and J. Dai (2020), Mental health problems and social media exposure during covid-19 outbreak, *Plos one*, 15(4), e0231,924.

Gilbert, C., and E. Hutto (2014), Vader: A parsimonious rule-based model for sentiment analysis of social media text, in *Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf*, vol. 81, p. 82.

Jónsson, E., and J. Stolee (2015), An evaluation of topic modelling techniques for twitter.

Lamsal, R. (November 20, 2020), Coronavirus (covid-19) tweets dataset.

Lexicon, V. (2017), Lexicon use for the vader sentiment algorithm.

Li, I., Y. Li, T. Li, S. Alvarez-Napagao, and D. Garcia (2020), What are we depressed about when we talk about covid19: Mental health analysis on tweets using natural language processing, *arXiv preprint arXiv:2004.10899*.

NYTIMES (August 6, 2020), Michelle obama says she is dealing with 'low-grade depression' amid coronavirus pandemic, racial injustice in us.

NYTIMES (June 14, 2020), Sushant singh rajput, bollywood star, dies at 34.

SSE (2020), Error sum of squares (sse).

TIMESOFINDIA (June 15, 2020), Sushant singh rajput commits suicide: Here are some warning signs of depression you should look out for.

WHO (2020), Mental health & covid-19.

WHO (March 11, 2020), Who director-general's opening remarks at the media briefing on covid-19 - 11 march 2020.

Zhang, Y., H. Lyu, Y. Liu, X. Zhang, Y. Wang, and J. Luo (2020), Monitoring depression trend on twitter during the covid-19 pandemic, *arXiv preprint arXiv:2007.00228*.

VITA

Contact Information

Name: Nusrat Armin

Email: narmin@go.olemiss.edu

Education

- **University of Mississippi**,

  Graduate Student (January 2017 - Present),

  Department of Computer and Information Science

- **Bangladesh University of Engineering and Technology**,

  Bachelor in Architecture (February 2011)

Experience

Graduate Teaching Assistant (Fall 2018 - Spring 2019)

Technical Skills

- Programming Language: C, JAVA, Python, Haskell, Scala
- Modeling Tools: SketchUp, AutoCAD, CorelDRAW

Research Interests

- Data Analysis
- Natural Language Processing (NLP)
- Machine Learning
- Human Computer Interaction (HCI)