MENTAL ANXIETY AND DEPRESSION DETECTION DURING PANDEMIC USING MACHINE LEARNING

BY

Ryana Quadir
ID: 201-25-866

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer Science and Engineering

Supervised By

Dr. Sheak Rashed Haider Noori
Associate Professor and Associate Head
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY DHAKA, BANGLADESH

DECEMBER 2020

APPROVAL

This Project titled "**Mental Anxiety and Depression Detection during Pandemic using Machine Learning**", submitted by Ryana Quadir to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of M.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 20 January, 2021.

BOARD OF EXAMINERS

Dr. Touhid Bhuiyan **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

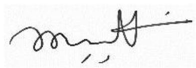
Dr. Sheak Rashed Haider Noori

Internal Examiner

Associate Professor and Associate Head

Dr. Md. Ismail Jabiullah **Internal Examiner**
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Dr. Mohammad Shorif Uddin **External Examiner**
Professor

Department of Computer Science and Engineering Jahangirnagar University

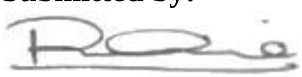
DECLARATION

I hereby declare that this project has been done by us under the supervision of **Dr. Sheak Rashed Haider Noori, Associate Professor and Associate Head, Department of CSE** Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Dr. Sheak Rashed Haider Noori
Associate Professor and Associate Head
Department of CSE
Daffodil International University

Submitted by:



Ryana Quadir
ID: 201-25-866
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

At first, we would like to thank Almighty God for His divine blessing and for giving us enough strength to complete this project. As we have been able to complete the thesis successfully, we would like to express our gratitude to the ones whose endless efforts has made this thesis successful.

I am grateful for the profound indebtedness of **Dr. Sheak Rashed Haider Noori, Associate Professor and Associate Head,** Department of CSE Daffodil International University, Dhaka. His depth of knowledge and keen interest made him the right person for conducting my research on Machine Learning and NLP (Natural Language Processing). His endless patience, scholarly guidance, continual encouragement, constant supervision, constructive criticism, valuable advices have made it possible to complete this project.

I would like to express my heartiest gratefulness to Professor Dr. Touhid Bhuiyan Professor and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty members and staff of CSE department of Daffodil International University.

Besides, I would like to express my sincere gratitude to all the people who completed the questionnaire I provided for creating my dataset. Without their time and support, it would not have been possible for me to get the authentic dataset.

Finally, I would like to thank my parents who gave their endless support during the time of project and mentally helped us. My other family members also deserve special credit for giving us adequate hope and praise to prepare this project with perfection.

ABSTRACT

As the novel coronavirus pandemic sweeps the globe and people take to their homes to avoid getting and spreading the contagion, it makes proper sense that much of the conversation about this is taking place online. With the rise of Social Media usage, web surfing and a long period of uncertainty during this Pandemic time, there is a sheer concern about the mental health and anxiety disorders among people. People are now using the internet to share information, air their anxieties, and spend time while in quarantine. This increate rate of online Social Media Use (SMU) opened the possibility to identify some common traits among people with various mental disorders and anxiety by the large dataset provided. The moments when those online conversations light up also tell us a lot about how our feelings around the pandemic are evolving. In recent years, this research area has started to evolve, but it would be extremely valuable during this crisis period. Although it is a complex task to perform as mental illness patterns are very complicated, it showed the light of hope in the past. Previously, adoptive supervised machine learning, such as deep neural network approaches were used to predict the pattern and level of mental illness; but they failed due to lack of annotated training data. In this research, we are proposing an effective machine learning architecture, based on Cluster analysis, Natural Language Processing (NLP) technique in the analysis of unstructured data extraction from Social media platforms and couple of psychological screener to classify mental condition of people.

TABLE OF CONTENTS

CONTENTS **PAGE**

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

With the outburst of the coronavirus disease Covid-19 early in 2020, the time spent at home has increased sharply for people around the world. But it was after the upsurge of the pandemic, somewhere in between February and March, people started to experience strict lockdown situations to help flatten the pandemic's deadly curve. To be exact, mentions of "Coronavirus" across social platforms and news media started to take off in late February and spiked in March 2020.

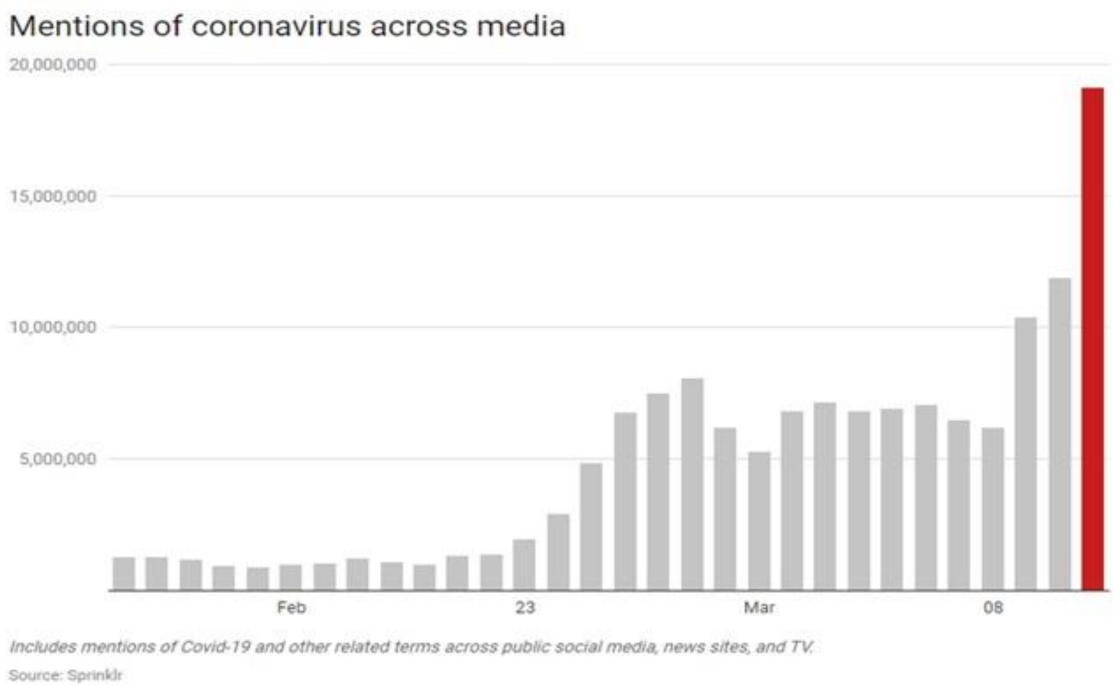


Fig 1: Coronavirus mentions across Social media

During a long stay at home, they tend to spend more time on Social Media platforms, providing an unexpected boost to engagement on these platforms. At the same time, as a byproduct of the coronavirus disease, people start to feel a strange kind of stress, fear, and anxiety about being put in such an area of uncertainty. Of course, there is grief about the sufferings and deaths that people are witnessing everyday; but even so, the helplessness is even stronger when people feel they are losing things in terms of the ability to move around freely. In the wide spectrum of mental health disorders, Anxiety and Depression are the most common ones and major health problems worldwide. There are multidimensional effects of anxiety-depression disorders and these are responsible for multiple somatic symptoms, namely gastritis, acid reflux, palpitation, tremor, insomnia or hypersomnia, significant weight gain or loss etc. There are also some psycho-social indications like suicidal ideation or attempt, depressed mood, social withdrawal, decreased productivity in the workplace, lack of concentration and many more that can be observed easily. These behavior and negative feelings are reflected directly on people's daily activities and consequently exposed in Social Media. Social media analytic companies found that the sentiment surrounding coronavirus posts is, unsurprisingly, mostly negative. The most prevalent emotion has been disgust, with many of those mentions centering around handwashing and the second-most-common emotion in these posts was fear.

8433	Recently fallen off waking up and going for a ...
8549	What! Your #depression is back? Here are some ...
8562	#selegiline dosage for depression https://goo...
8638	Mom's depression tied to kids' emotional, inte...
8685	@hamrick_krista @itsAshleyBB Well Geeez Ashley...
8748	i also have a lot to say about asian immigrant...
8750	I want glittery, fun, sexy, crazy party music ...
8775	Mom's depression tied to kids' emotional, inte...
8795	I was supposed to revise business but I forgot...
9007	@ethomasson is a senior correspondent @Reuters...
9062	@LaurakBuzz sadly not as deep as i thought. Bu...
9111	Worst thing about depression is forgetting to ...
9150	Exercising regularly cuts risk of depression b...
9256	But I know depression and drug addiction don't...
9375	@gabrielflorin01 Sis what happened? Depression...
9377	so glad winter is finally over. Goodbye season...
9381	@BethFratesMD That's what I've been trying to ...
9415	My depression won't let me work out and be hea...
9622	Why your doctor doesn't provide #IVKetamine as...
9643	How exercising can slash the risk of depressio...
9732	Going back and istening to an older album of a...
9750	@WhatTheFFacts I don't like to say that I'm go...
9763	Study finds exercise may lower risk of depress...
9779	Poor recognition of #depression and #anxiety i...
9812	Depression is real

Fig 2: Segment of Depressive Tweets

It is not always in terms of sharing the stress, frustration or fear, it is sometimes only the connection and impact of the overall Social Media. In other words, the association between SMU variables and mental health may be indicative of a user's experience and attitude rather than the volume of Social Media Consumption. In this regard, SMU would only be used as a measure to detect the users' mental stability and psychiatric disorder (if there is any); but not to demonstrate the negative effect of Social Media usage among the users.

In reality, there are multifactorial causes associated with Anxiety and Depression including biological, socio-economic state, environmental reasons and sometimes cultural background. When a psychiatrist or psychologist diagnose a patient, it is usually based on Statistical Manual of Mental Disorder (DSM-5)[19] or the International Classification of Diseases (ICD)[20] 10. But most of the time, the illness stays undiagnosed since patients are not aware of the mental conditions. And if it remains unidentified, the suffering continues, and it can get more severe day by day.

During this odd time, it is a must to arrange some kind of psychological screenings in a monthly or bimonthly basis in organizations. And it would be extremely valuable to develop an early warning system that can identify people with high risk. Those people can quickly be referred to psychological counselling and treatment center. It would be beneficial for creating a mentally healthy workforce for industry and would enable them to lead socially and economically productive lives.

CHAPTER 2 RELATED WORK

There were many approaches made for detecting mental disorder and depression of the social media users in the past. Many of them dealt with the volume of usage as a measure of detecting potential SNMD (Social Network Mental Disorder). That means, they had a different goal to prove than ours. Nonetheless, these models helped us identify the feature extraction well and to formulate a better classification for clusters.

2.1: Affected Content Analysis of Online Depression Community by Thin Nguyen

Thin Nguyen, Dinh Phung, Bo Dao, Svetha Venkateshy, and Michael Berk et al.[10] used affective content analysis method of Online Depression Community to design the algorithm for detecting depression stigma. They used machine learning and different statistical methods to discriminate online messages between depression and control communities using mood, psycholinguistic processes and content topics extracted from the posts generated by members of online communities. The dataset was collected from 24 mental health communities called the CLINICAL group. In this process, first, the linguistic styles have been found as indicative feature of depression. Then the topics were used to characterize text documents to enable understanding of what (topics) people are interested in and how (language styles) they discuss their interests in social media. The posts were then classified according to the depression and mental disorder symptoms.

The data-modeling performance was improved by implementing classification algorithm like, Logistic Regression model called Lasso, SVM (Support Vector Machine), Latent Dirichlet Allocation (LDA) as a Bayesian probabilistic modelling framework, RF (Random Forest) etc.

(a) Distribution of moods (in cloud visualization) tagged by the CLINICAL communities.

(b) Distribution of moods tagged by the CONTROL communities.

Fig 3: Mood Usage by two different communities

This approach only worked with users' social media posts and their syntactic analysis and classification; but there could be some unexpressed feelings indicating depression stigma that were not exposed in the posts. This area was not dealt in this research. Moreover, the affective aspect, in particular, the role of sentiment and mood, has not been thoroughly studied in this research.

2.2: Depressive Symptoms Analysis Using Active and Passive SMU

Social Media Usage was classified as Active and Passive to analyze depression[1] symptoms by César G. Escobar-Viera, Ariel Shensa, Jaime E Sidani, Brian A. Primack, Nicholas David Bowman, Jennifer Knight and Everette James. Their classification was mostly based on the way social media is used, how much time people are spending on it and how many platforms are they visiting. The Active users are found out to be of improved well-being whereas the Passive users are likely to have social anxiety and of decreased well-being. These opposing effects had varied association with depression symptoms. They used Ordered logistic regression to assess associations between both passive and active

SMU. They examined their data-distribution of 7 SMU variables using Shapiro-Wilk test of Normality and histogram, Q-Q plots graphical methods. The sample was assessed as factorable by the Kaiser–Meyer– Olkin (KMO) test of sampling adequacy, which indicated that the sample was factorable. They initiated the process from the hypothesis that passive users would have positive association with mental illness, depression whereas no significant association would be found between active SMU and depression.

FACTOR STRUCTURE AND SCALE DEVELOPMENT OF PASSIVE AND ACTIVE SOCIAL MEDIA USE

Complete item[a]	*Factor loading I "Active use"*	*Factor loading II "Passive use"*
Read discussions	0.11	0.80
Read comments/reviews	0.14	0.86
Watch videos or view pictures	0.15	0.70
Share others' content (e.g., retweet, share posts or status updates)	0.81	0.07
Like/favorite/voting	0.57	0.47
Comment on, or respond to someone else's content	0.76	0.28
Post your own content (e.g., tweet, status update)	0.86	0.07
Cronbach's α[b]	0.80	0.72
Factor variance[c]	0.34	0.30

Fig 4: Factor Structure and Scale development of Active and Passive SMU

This straightforward association between usage pattern and depression turned out to be true that depression was not found in the active users, which proved their initial hypothesis. But active use of social media can sometimes involve problematic SMU and addiction, which is a potential indicator of mental illness. That extreme case should have been considered in constructing the model.

2.3: Convolution Neural Networks Applied in feature learning of Social Media posts

Preserving users' privacy with the help of separating agents, Sridharan et al. [18] proposed the depression detection diagnostics using Convolution Neural Networks (CNN). As the dataset, he used the data posted by different users in social networks. Another filtering mechanism was used to filter out redundant data to increase the learning of CNN.

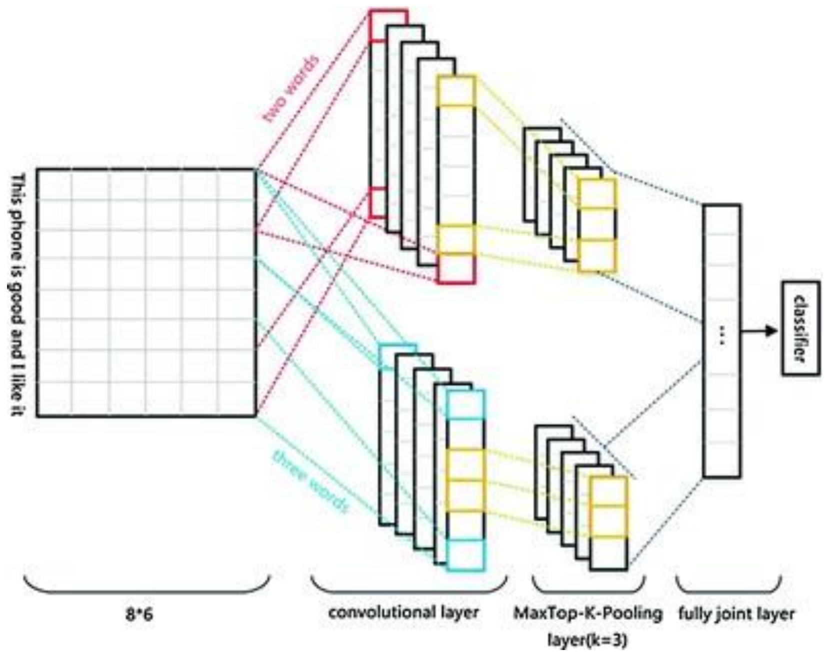


Fig 5: Text-post Feature Extraction and Classification Based on CNN

This approach made an attempt to automatically identify potential online users with Social Network online disorder detection-framework. The collected dataset was analyzed using syntax and semantic analysis, which gave the sense of depression stigma among the social media posts by different age groups. The posts where keywords matched but didn't qualify for stigma-set, belong to the Non-stigma group. It managed to explore various features from various data-logs and this is where a new tensor technique was applied for deriving latent features from multiple Social Networks.

Again, this method worked only with the posts from users and tried to identify the latent features from there. But the intensity of social media usage or emotional attachment pattern were not dealt in finding the disorder or depression symptoms.

CHAPTER 3 PROPOSED METHOD

As stated before, it is a complicated task to determine the mental illness as the disorder patterns are very unpredictable and complicated, we developed patterns of different userclusters with extreme vulnerability. For example, one cluster might have users with addictive or problematic levels of SMU, which may be associated with increased anxiety and mental stress. This association might be placeable to the increased likelihood of individuals who experience long term anxiety specially during pandemic and subsequently develop addictive behaviors. On the other hand, there is a group of sensitive users who do moderate level of SMU but has immensely strong emotional connection with social media. So, we would like to define these clusters first and then start to ascertain the susceptibility to severe anxiety and alarming level of depression.

Besides, mental disorder is perceived not by observing someone feeling down or depressed or fearful just for a moment; it would persist for longer time instances that affects general cognitive function of the brain. And as a consequence of witnessing the death toll around and the inability to interact socially, the mental illness became a common phenomenon worldwide.

The 4-phase model that we came up with can be called 'LMHA-model' since it determines the usage-pattern and characteristics of 4 groups of users.

As we did not specify how many clusters there would be, we chose the Divisive Hierarchical Clustering in this regard. This clustering technique allowed us to identify optimal number of clusters with distinguishing usage-pattern. After feeding data against all 3 parameters, our IDE: RStudio unpacked the packages `factoMineR` and `factoExtra` and with the help of package `ggplot2` and `graphics`, we concluded with a Dendrogram-plot with 4 distinctive clusters.

While choosing social media platforms, we selected 10 most widely used sites, namely:

Facebook, YouTube, Instagram, Twitter, Tiktok, Weibo, Snapchat, Pinterest, QQ, and Qzone. Now, considering the time spent on these social media, frequency of usage, intensity of use and level of abusing the SMU, these are the 4 clusters:

1. Low Usage (LU): Not affecting personal life
2. Moderate Usage (MU): Strong emotional connection
3. High Usage (HU): High but not problematic
4. Addictive Usage (AU): Problematic and addictive social media uses

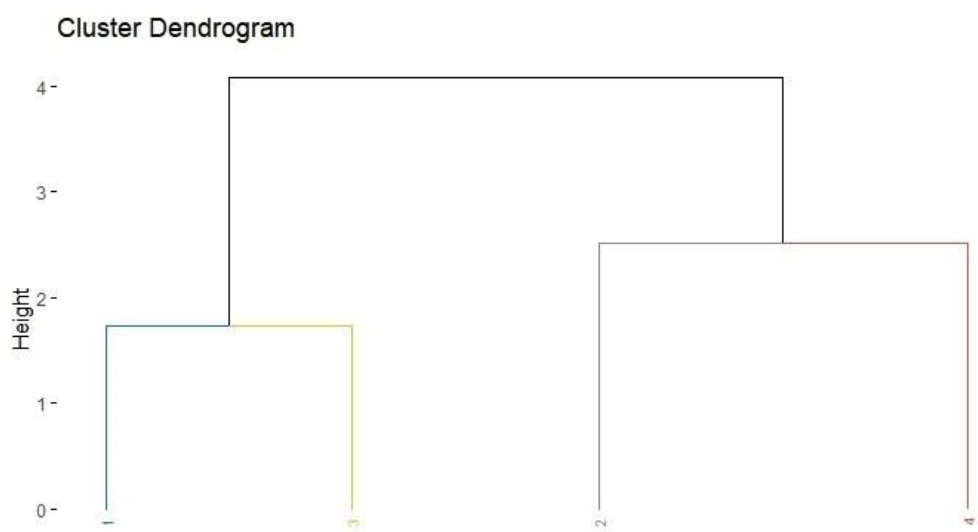


Fig 6: Cluster Dendrogram by Divisive Hierarchical Clustering

The measurable parameters that are used to define these clusters are:
i. Time spent (per day) ii. Frequency of use (per day) iii. Number of SM platforms

3.1: Phase 1: Initial Clustering

First, an initial classification was drawn for 4 clusters based on these parameters above. Here, we used the data from our social network and online statistics from datareportal.com[3], businessinsider.com[4], statista.com[5], smartInsights.com[6], globalWebIndex.com, jmir.org[7] tables.

Total 200 respondents from our network provided their statistics about average spending time daily on social media, number of times these platforms visited per day and number of social media active accounts they have and use during COVID-19 period. About 100 of them also participated in filling up tables for their emotional statistics. The ranges for each parameter were specified after analyzing the collected data by the Cluster analysis. After having the product of cluster range-value and their corresponding weightage, and by normalizing the data, the first table looked like this:

Table 1:

Parameters	**Cluster1 LU**	**Cluster2 MU**	**Cluster3 HU**	**Cluster4 AU**
Time spent	0.02835 (<1)	0.3942 (1-2)	0.4695 (2-3)	0.4521 (>3)
Frequency	0.04136 (1-3)	0.15963 (4-6)	0.17301 (7-9)	0.28943 (>9)
No. of platforms	0.1626 (1-4)	0.0858 (4-7)	0.0248 (7-9)	0.00211 (>9)

Table1: Cluster degree of memberships by social media usage parameters

After getting the initial clusters from the 3 parameters, the graph looked like this:

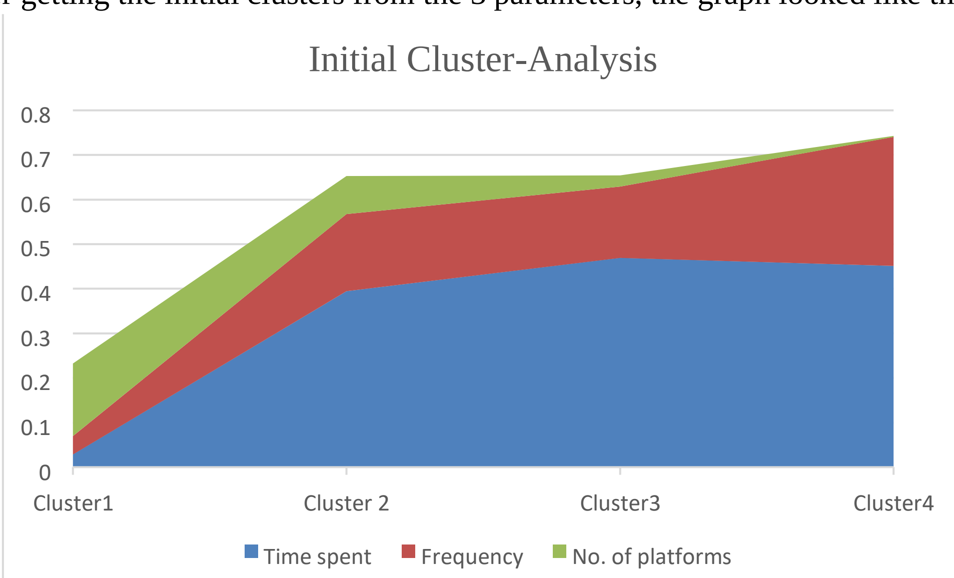


Fig 7: Initial Cluster-Analysis from parametric values

3.2: Phase 2: Cluster Optimization by keyword extraction

Secondly, we gathered posts and popular hashtags of these individuals and started to extract the keywords using NLP (Natural Language Processing) from these posts for syntactic analysis. Again, we relied on R and package was mostly `Udpipe` and dependencies were: `library(dplyr)` `library(igraph)` `library(tm)` `library(NLP)` `library(openNLP)` `library(openNLPdata)` `library(coreNLP)` `library(stringr)` `library(koRpus)` `library(koRpus.lang.en)` `library(TreeTagger)` and again `ggplot2`. Udpipe Package provides pretrained language models that we used. Here, we collected mostly Facebook, Twitter and Weibo posts and used NLP to identify the most focused words and examined the effective keywords by tuning them up. After getting the keywords, the association was made from keyword-based analysis regarding individual's mental condition.

So, from 2nd phase, we got our refined clusters with criteria-related validity and 3 emotional indices from Kaggle Andbrain Emotions Sensor Dataset. This Emotions Sensor Dataset contain top 23730 English words classified statistically using Naive Bayes Algorithm into 7 basic emotions, namely: disgust, surprise, fear, anger, depressed, happy and anxious. We took 3 emotional indices from these 7 emotions, which were relevant with this research. The Naive Bayes Algorithm was used to calculate the probabilities of existence of these words belonging to each emotional index. A part of the analysis looked like this:

Table 2:

SL	**Posts and hashtags**	**Keywords**	**Tuned up keyword**	**Depressionindex**	**Anxiety - index**	**Fear-index**
1	97,000 kids tested Covid19 positive in the US within 2 weeks of School opening. Bangladesh is planning to open schools by September.	Tested, Covid- 19, positive, planning, open	Covid -19, positive	0.2114	0.5168	0.358

2	Empty airports and planes at the height of the #COVID-19 pandemic	Empty airport s, height, pande mic, planes	Empt y, pand emic	0.41	0.35	0.112
3	Encouraging people to get #Coronavirus in the name of Virtue	virtue, people	vi rt ue , pe op le	0.516	0.131	0.082
4	This implies there will be no 'after' of it and if there's no vaccine, this virus will remain with us	virus, vaccine	vir us, vac cin e	0.423	0.315	0.259
5	Trying not to be angry about this #coronavirus is hard, talking listening and even seeing the effects of it	angry, effects	an gr y, ef fe ct s	0.457	0.241	0.238

Table 2: Keyword extraction from posts and analyzed with emotion -index

The emotion sensor indices are then put into a Multivariant correlation matrix, where the 3 variables are: Depression, Anxiety and fear. From that the correlation between Depression and Anxiety was taken. The Multivariant correlation matrix was calculated this way:

5	**Depression-index**	**Anxiety-index**	**Fear-index**
Depressionindex	1		
Anxietyindex	0.110137	1	
Fear-index	0.108766	0.057358	1

Table 6: Multivariate correlation matrix for a post

The 2nd phase was about optimizing the initial cluster and it is visible from the flowchart of the entire process:

Flowchart:

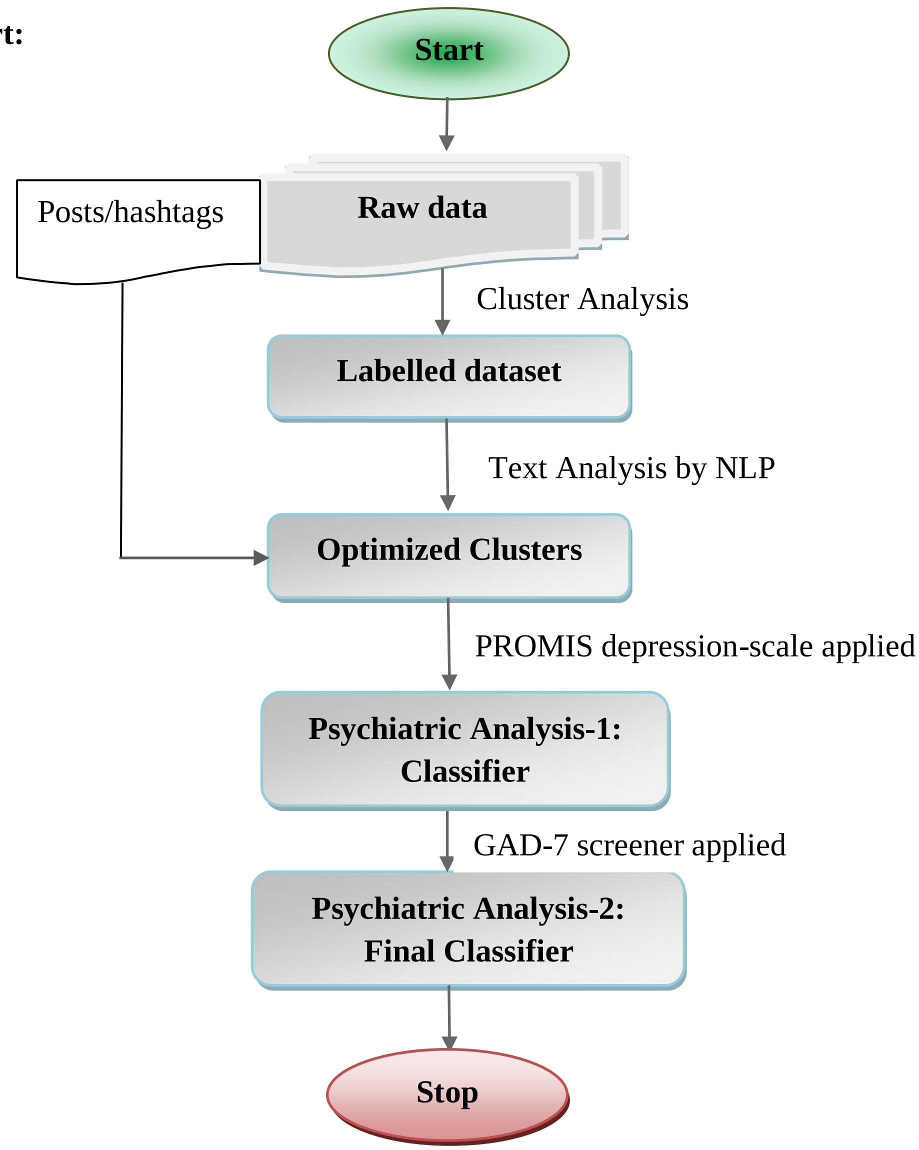


Fig 8: Flowchart of the entire process

3.3: Phase 3: Applying PROMIS scale

Thirdly, we tried to get the users' emotional statistics and for that we asked 100 participants from all 4 clusters of the same sample to complete the table about their feelings over the last 7 days. The PROMIS (Patient-Reported Outcomes Measurement Information System) depression-scale is a well stablished way of obtaining the emotional state of people. This is a 4-item scale of measuring users' feelings and items are scored using a 5-point Likert-type scale. It was applied to all the clusters including Cluster1 (Low Usage). This step is for emphasizing on the emotional attachment with social media even when users are not exposing themselves enough to the online world.

The raw scores from this table would have a range from 4 to 20 and greater scores indicate increased severity of mental issue, specially depression symptoms. The probabilities were calculated from individuals' score from this table:

Table 3:

	Emotions	**Never (1)**	**Rarely (2)**	**Sometimes (3)**	**Often (4)**	**Always (5)**
1	Hopeless			√		
2	Worthless		√			
3	Helpless				√	
4	Depressed					√

Table 3: An example of how a participant completed the 4-item form

3.4: Phase 4: Applying GAD-7 scale

In the final phase, there were again 100 respondents who agreed on completing the 7item Generalized Anxiety Disorders Scale (GAD-7) to have a clear identification of the anxiety vulnerable users among the clusters. GAD-7 is a screener for generalized anxiety disorder in primary care settings. Those respondents were asked to state the frequency of the 7 items that they have been bothered over the last 4 weeks. Again, the probability of having Anxiety-symptom was calculated for individual's responses.

The GAD-7 looked like this:

Table 4:

	Emotions	**Never (0)**	**Several days (1)**	**Over half the days (2)**	**Nearly every day (3)**
1	Feeling nervous, anxious, or on edge		√		
2	Not being able to stop worrying		√		

3	Worrying too much about different things				√
4	Trouble relaxing			√	
5	Being so restless that it's hard to sit still	√			
6	Becoming easily annoyed or irritable				√
7	Feeling afraid as if something awful might happen			√	

Table 4: An example of how a participant completed the GAD-7 form

The Anxiety measurement:

The probability of having anxiety would be: (1+1+3+2+0+3+2)/21

= 12/21 = 0.5714

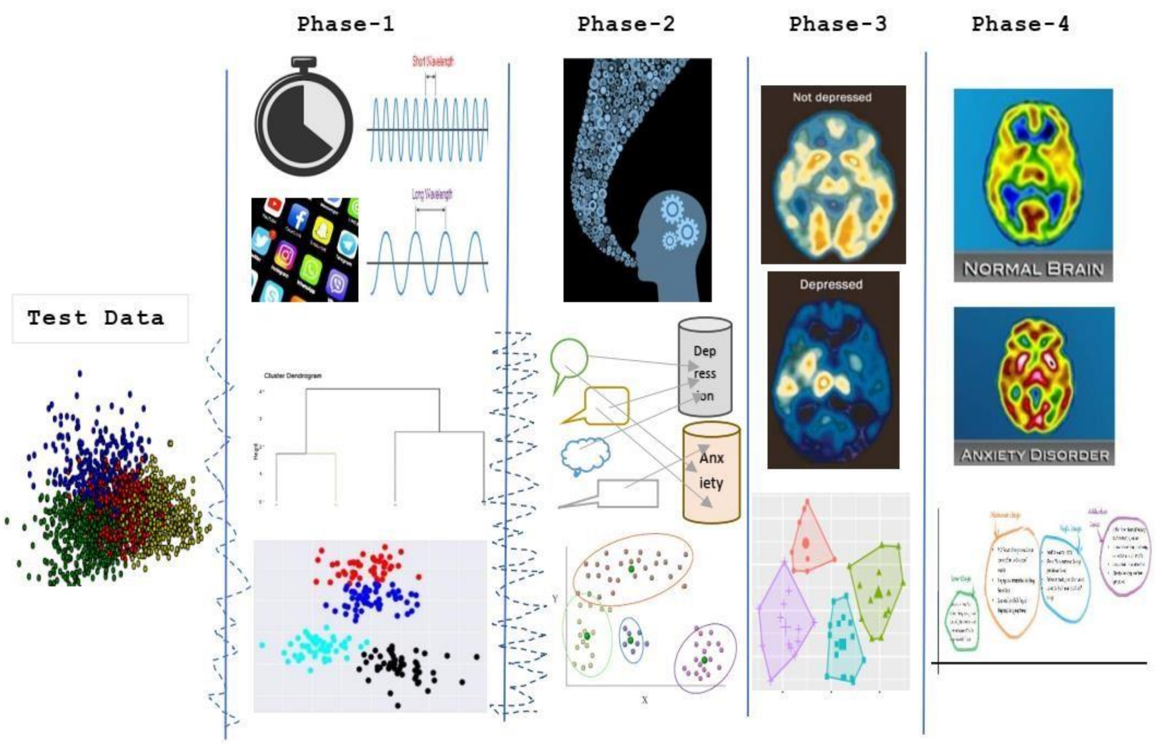


Fig 9: The 4-phase of LMHA-model

CHAPTER 4 DATA COLLECTION AND ANALYSIS

4.1: Data Collection

The task of data collection was conducted in 2 parts. First, we started collecting statistics from online sites and portals like datareportal.com, businessinsider.com, statista.com, smartInsights.com, globalWebIndex.com to get detail-oriented data. These data were labelled by different parameters, but we chose the 3 variables for our research- time spending on different social media platforms, number of times visited per day and number of active accounts. Also, a big number of tweets, hashtags and posts were collected from online platforms which were public. But as this research is also focused on users' emotional statistics, it needs to have user-specific data. As a result, we relied mostly on our own network for collecting most of the raw data. This was done by providing questionnaire to the participants who are basically social media users. In this regard, we made sure to preserve the privacy of the users by filling up the form anonymously. The participants were asked to provide 14 answers like the following:

Basic Parameters	1.1 Daily time spent on any Social Media 1.2 Number of times per day visiting social media 1.3 Number of 'active' social media accounts use	MCQ type
Text Analysis from user posts	Please write one of your Social media posts/status about Covid-19 (if any) in the last 6 months	Open-ended answer
PROMIS test-data	During lockdown at home, how frequently you felt— Hopeless, Worthless, Helpless, Depressed	5-point Likert-type selection
GAD-7 test-data	During lockdown period, how often in 4 weeks (~1 month) have you been bothered by the following problems?	4-point Likert-type selection

Table 5: Questions in the questionnaire provided

As this research was based on Social media data, we did not reach to those persons without social media endorsement, meaning, who did not have or use social media at all. But we did consider those users with less involvement in Social networks but can potentially contribute to the usage-pattern survey. We did not categorize the participants by any demographic or ethnic group as it would not add any new value to our research.

The participants were found to be fall into our 4-clusters like the following way:

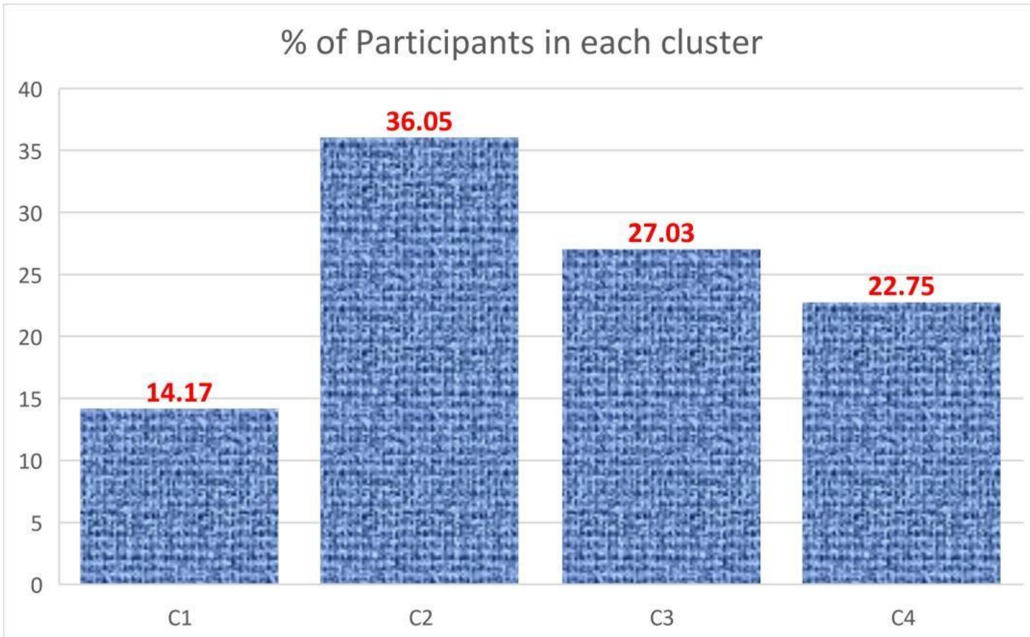


Fig 10: The percentage of participants in four-clusters

4.2: Data Analysis

The analysis of the raw data started by labelling them and identifying number of clusters needed to fit those data. For that, the Divisive Hierarchical Clustering was used to get the optimal number of clusters and label the collected data.

4.2.1: We used the software RStudio Version 1.3.1093, which is the most updated one. And the reason of choosing R as the language is it gives us more options for Statistical testing and more appropriate library and packages which were needed for this kind of research. From Table 1, we can see the cluster-membership was classified by users approving some levels of each SMU characteristic. Also, not any one cluster is exclusively giving high values for all parameters. It is rather 3 of them- MU, HU and AU which were seemingly found as vulnerable in terms of higher value. The collinearity between these predictor variables was assessed and Cluster1 (LU) fell below 0.01, that justifies the initial classification. But we are not going to omit this cluster for the next step and parsing of text of the posts would be done on posts of all 4 cluster users

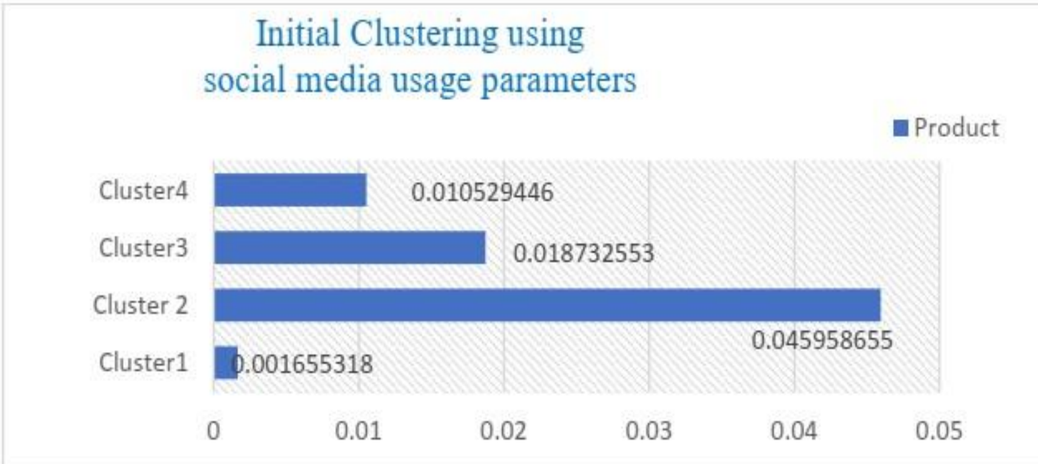


Fig 11: Collinearity between Cluster-variables

4.2.2: Next, the keywords were obtained from the social media posts which were used as positive-labelled training data for an NLP classifier. Here, we took just one post from single user, although many users had multiple posts. After selecting the keywords from the posts, these keywords were projected on emotions sensor dataset containing top English words classified statistically using Naive Bayes Algorithm into 7 basic emotions for the emotion analysis. We picked up indices of three emotions that are relevant with our goal: fear, anxiety, and depression. These indices are then examined for multicollinearity by the multivariate correlation matrix. The correlation between DepressionAnxiety thus becomes the emotion-sensor for each of the posts. This helped evaluating the criterion-related validity of the cluster solution.

At the end of this step, we would have a more concise version of the correlation and more optimized clusters.

4.2.3: After step 2, we got more optimized clusters and now, we would put two more classifier which would lead the data to a psychological screening vision. We would identify whose individuals having higher index-values for anxiety and depression and their belonging clusters. We would put on this classifier on that data by analyzing the PROMIS-score for users' depression measurement. For that we would examine the probability of each table filled up by users. Then the probabilities would be ranked on Depressive symptoms like:

Symptom level	Calculated Probability	Cluster
None to slight	P < 0.4	LU
Mild	0.4< P<=0.5	HU
Moderate	0.5< P =0.8	MU
Severe	0.8< P = 1	AU

Table 7: Depressive Symptom level from PROMIS

4.2.4: Lastly, we would put the final classifier- the GAD-7 test on the data and focus on individual's anxiety and fear measurement during pandemic. The probability of anxiety level was calculated from the GAD-7 tables. It is another Likert-type table where the probability of users' anxiety was measured.

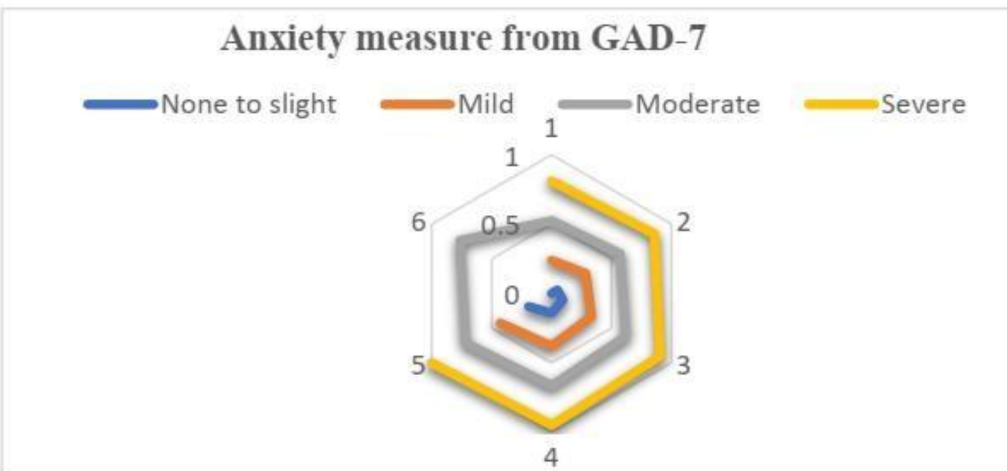


Fig 12: Anxiety level from GAD-7 table

From this final classifier, four different categories of anxiety level were observed. This was then put into our data from phase-3 of data-analysis. The final output was found after applying this GAD-7 classifier.

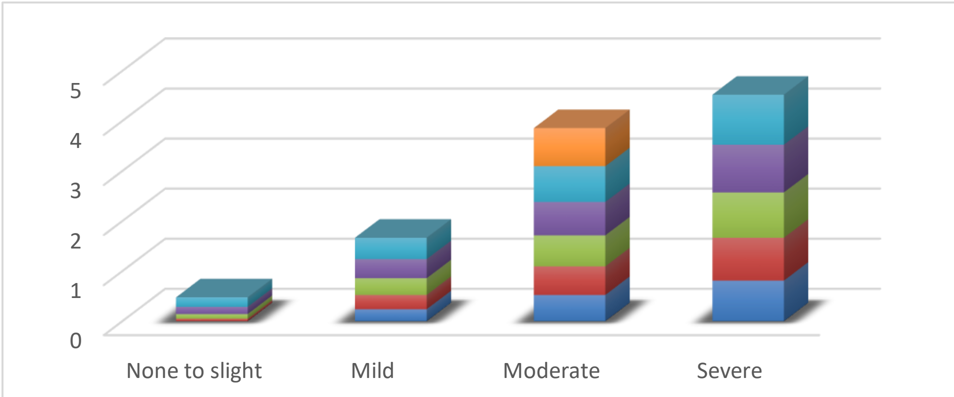


Fig 13: Different categories of anxiety level classifier

CHAPTER 5 RESULT AND DISCUSSION

5.1: Result:

We started the entire procedure with 120 participants' statistics for initial cluster analysis and analyzed 150 posts from social media platforms of 100 users. In the third phase, we managed to get 80 individuals who completed the PROMIS form and we ended up having 71 people filling up the final GAD-7 table. Although the number of participants along with the percentages varied in different phases, it yielded 4 distinct patterns of their social media use and their relevant mental condition.

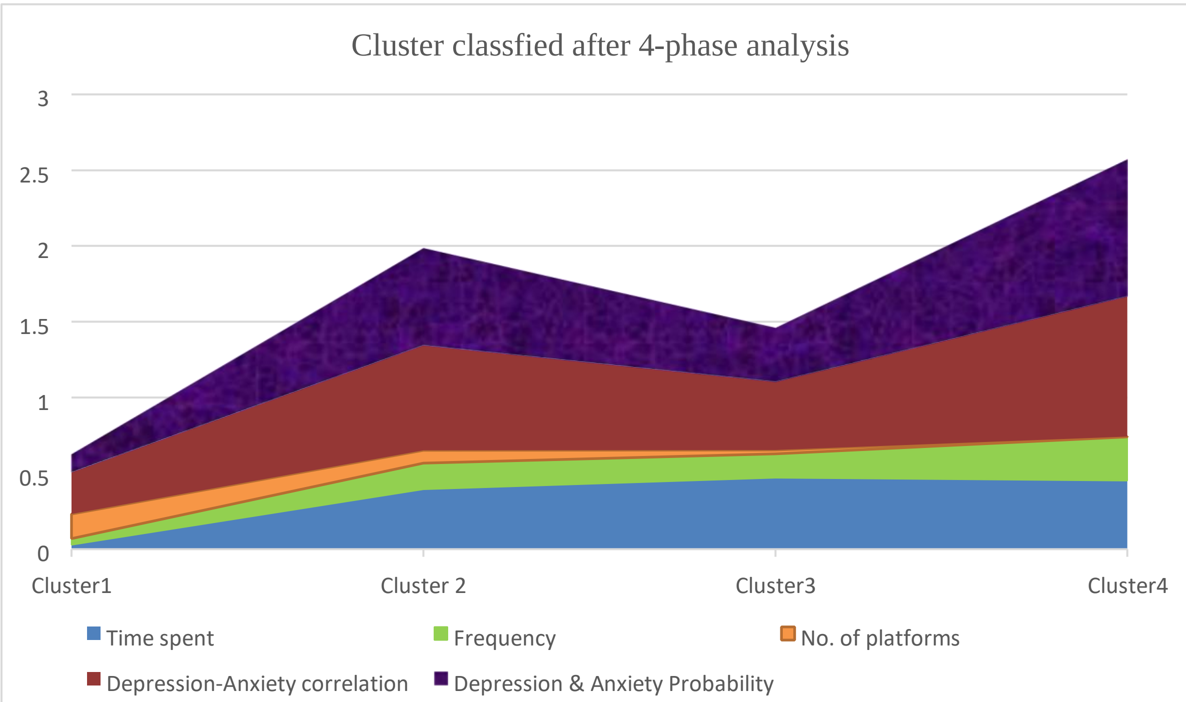


Fig 14: Final Cluster result from 4-phase analysis

5.2: Discussion:

During the initial cluster analysis, we found that the LU-cluster (Low Usage) had lowest score for all the 3 variables- time, frequency and no. of platforms use. As a result, they had fewer participation in posting feelings in social media. For applying the 2 psychological disorder screeners, we did not take out the LU-cluster even though they failed to look vulnerable in terms of any mental ailment. From the keyword extraction phase, this LU-cluster began to fade out and ended up as the group that is ‘not associated’ with an any mental illness.

For the MU-cluster, the time spent was a moderate 1-2 hours and they had 4-7 active SM accounts, but the frequency is in bit higher range (7-9 times per day). This group indicates those people who spent not a huge amount of time but have strong emotional connection with social media and engage in attention seeking behaviors like frequent status-updates and subsequent checking for ‘Like’ in online world. This obsession may lead to depression if the individuals do not receive the desired feedback from his or her social media audience . It might get elevated during lockdown phase which was reflected from the PROMIS - scale result as they showed moderate (0.5-0.8) probability of Depressive symptoms. No wonder, they demonstrated maximum participation and hence had more posts for keyword extraction.

We found interesting analysis for HU-cluster, who spent 2-3 hours every day having 7-9 active SM accounts. They tend to stay connected to their online world and showed less frequency in checking their own account for feedback. In both psychological disorder screeners, they exhibited probabilities below 0.5, signifying their positive mental state. They tend to share life experience even during pandemic-time and respond quickly and frequently to other users which leads to their improved well-being.

Finally, we defined Cluster-4 as ‘Addictive’ since all 3 variables associated with them gave us the highest values in doing initial cluster-analysis. The ‘Fear Of Missing Out’ (FOMO) felling made them stay continually connected in social media losing real-life interactions triggered the elevated anxiety and fear symptoms. In case of keyword extraction, their posts indicated high index-values of not just depression and anxiety, they were also linked with fear. Additionally, they unveiled probabilities in the range 08-0.95 in both psychological tests which yields extremely high risk of mental syndromes.

Scores	0-4	5-10	11-16	17-21
Probability	0 -0.1904	0.238 – 0.47619	0.5238- 0.7679	0.809- 0.9523
Cluster	LU	HU	MU	AU

Table 8: Results from GAD-7 scale

Hence from this 4-phase analysis, it was clear that people from Cluster-2 (MU) and Cluster-4 (AU) are closely associated with mental ailment. Some border -line values of Cluster-3 (HU) individuals also need to have a close look at the mental health for their well-being.

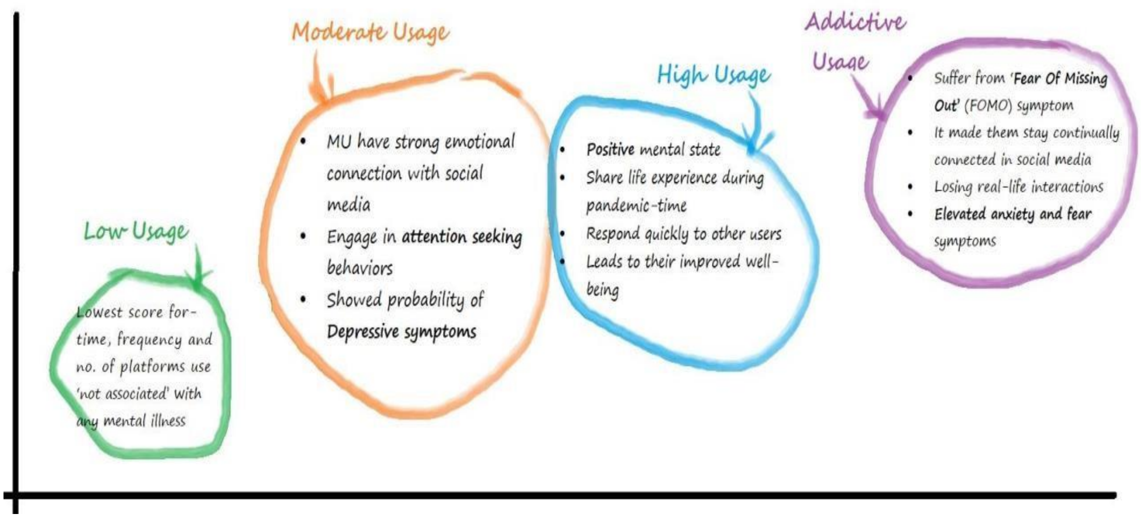


Fig 15: Visualization of 4 clusters as 4 usage patterns

CHAPTER 6 LIMITATION AND CONCLUSION

6.1: Limitation:

The dataset we constructed was primarily from our network and a portion was from the public posts and information from online portals. Thus, we were able to get the values for the two psychological screeners of users. Without accessing our own network, these sensitive values were not possible to collect. Moreover, responses from the screener-tests and posts and other information must be from the same sample or respondents since we tried to identify the pattern of social media usage and thus track the mental health of them. As a result, the primary data collected was based on our own network. On the contrary, these responses could not get diverse in terms of different Ethnicity and race and social positions. Yet, we tried to have variations in selecting social media users concerning their age, gender, occupation, and religion. But we did not want to label the sample by different religion, ethnicity, gender or age, as it would not add any value in our research. Only the Social media consumption and the pattern were the labelling features for this research.

Social stigma played a role in terms of treating mental illness during the data collection. As the goal of the research was to predict the mental health of people and apparently the risk of having disorders, there was a hurdle between people and their screener-test responses. So even if we started with 200 respondents, for the PROMIS and GAD-7 screener tests, we manage to involve 105 people exposing fact about their mental health.

Besides, we did not include any users with history of previous mental health issues or history of drug addiction, alcohol abuse. The participants were seemingly and fit general people having social media presence.

We could have done an analysis of Phase-1 and phase-2 from online datasets, but the mental-health screening would not be conducted in that case. Even though this dataset is seemingly small, the data collected were authentic and legitimate and the information were taken carefully preserving their privacy.

6.2: Conclusion:

The deterioration of people's mental health during pandemic period can result in severe mental disorder. For different people, the combination of emotions is different - for most of them, it's anxiety for the family, relatives and his/herself. It is a tremendous grief for losing our normal day-to-day social interactions which leads to severe depression. Sometimes it is not just general anxiety or depression, it is the fear about what is going to happen to the society, to the economy and for how long it is going to last. If the uncertainty lasts for more than a year and people suffer from that during the entire time or longer, it will become daunting. And it needs to be identified and treated as early as possible. The four-phase LMHA-model analysis using cluster analysis, NLP and two psychological screener that we proposed in this paper and the associations between the parameters are quite relevant and obtainable. Among the 4 distinct clusters, the 2 specific patterns- MU and AU with elevated symptom levels of depression and anxiety are undoubtedly at high risk. This realistic diagnosis of mental disorder should be considered seriously in clinical interventions and therefore should be taken care at the early stage for the individuals at highest risk. The implication of this research is hence a demonstration of the potential of social media as a new means of mental health screening and monitoring, which can provide a foundation for Online Early Warning System for mental disorder.

REFERENCES

[1] Ce´sar G. Escobar-Viera, Ariel Shensa, Nicholas D. Bowman, Jaime E. Sidani, Jennifer Knight, Everette James and Brian A. Primack. 2018. Passive and Active Social Media Use and Depressive Symptoms Among United States Adults. In CYBERPSYCHOLOGY, BEHAVIOR, AND SOCIAL NETWORKING Volume 21, Number 7.

[2] Devakunchari Ramalingam, Vaibhav Sharma and Priyanka Zar. May 2019. Study of Depression Analysis using Machine Learning Techniques. In International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075,Volume-8, Issue-7C2.

[3] Simon Kemp. January 30, 2020. Digital 2020: Global Digital Overview. Retrieved from DataReportal: https://datareportal.com/reports/digital-2020-global-digital-overview

[4] Alexandra Samet. June 9, 2020, 10:58 PM. 2020 US SOCIAL MEDIA

USAGE: How the Coronavirus is Changing Consumer Behavior. Retrieved from Business Insider URL: https://www.businessinsider.com/2020-us-social-media-usagereport?r=US&IR=T

[5] J. Clement. June 19, 2020. Estimated U.S. social media usage increase due to coronavirus home isolation 2020. In reports from Statista URL: https://www.statista.com/statistics/1106343/social-usage-increase-due-tocoronavirus-home-usa/

[6] Dave Chaffey. August 3, 2020. Global social media research summary July 2020. From smartInsights URL: https://www.smartinsights.com/socialmedia-marketing/socialmediastrategy/new-global-social-mediaresearch/

[7] Ahmad AR, Murad HR. May 19, 2020. The Impact of Social Media on Panic During the COVID-19 Pandemic in Iraqi Kurdistan: Online Questionnaire Study. Vol 22, No 5 (2020). Retrieved from: https://www.jmir.org/2020/5/e19556/

[8] Ahmed Husseini Orabi, Prasadith Buddhitha, Mahmoud Husseini Orabi and Diana Inkpen. June 5, 2018. Deep Learning for Depression Detection of Twitter Users. In Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, pages 88–97 New Orleans, Louisiana

[9] Bethany McCord, Thomas L.Rodebaugh, Cheri A. Levinson. 2014. Research Report named Facebook: Social uses and anxiety. Computers in Human Behavior 34 (2014) 23–27. Retrieved from www.elsevier.com/locate/comphumbeh

[10] Nguyen T., Phung D., Dao B. and Venkatesh S., Article named Affective and Content Analysis of Online Depression Communities in IEEE Transactions on Affective Computing · July 2014 DOI: 10.1109/TAFFC.2014.2315623.

[11] Margaret Mitchell, Kristy Hollingshead and Glen Coppersmith. 13 December 2016. Quantifying the Language of Schizophrenia in Social Media. Retrieved from https://www.researchgate.net/publication/291373346

[12] Kirsten Weir. March 16, 2020. Seven crucial research findings that can help people deal with COVID-19. In American Psychological Association. Retrieved from URL: https://www.apa.org/news/apa/2020/03/covid-19-research-findings

[13] Samantha K Brooks, Rebecca K Webster, Louise E Smith, Lisa Woodland, Prof Simon Wessely, Prof Neil Greenberg, et al. March 14, 2020. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Volume 395, ISSUE 10227, P912-920

[14] Katie Keane and Elise Ruland. In R 2 integrated blog. Retrieved from: https://www.r2integrated.com/r2insights/covid19s-impact-on-social-media-usageand-strategies

[15] Emily Chen, Kristina Lerman and Emilio Ferrara. Published online May 29, 2020. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. Retrieved from URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7265654/

[16] C. O. Alm, D. Roth, and R. Sproat. 2005. Emotions from text: Machine learning for text-based emotion prediction. In Proceedings of HLT/EMNLP, pages 579–586

[17] Sverre Urnes Johnson, Pål Gunnar Ulvenes, Tuva Øktedalen and Asle Hoffart. Published online Aug 6, 2019. Psychometric Properties of the General Anxiety Disorder 7-Item (GAD-7) Scale in a Heterogeneous Psychiatric Sample. In Front Psychol. 2019; 10: 1713. Retrieved from URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6691128/#:~:text=The%20seven%20items%20assess%20(1,if%20something%20awful%20might%20happen.

[18] S.Sridharan, AkilaBanu, M. Bakkiyalakshmi, A. Buvana P; "Detection and Diagnosis on online Social network Mental Disorders using conventional Neural Networks"; In International Journal of Engineering Science and Computing; 2018.

[19] American Psychiatric Association. Force DSMT. Diagnostic and statistical manual of mental disorders: DSM-5 (2013). American Psychiatric Association; 2016.

[20] World Health Organization. The ICD-10 classification of mental and behavioral disorders: diagnostic criteria for research. 1993.

Appendix A

(Research Reflection: Questionnaire & Posts) Questionnaire sent to Users:

Research on Mental health during COVID-19

- Social Media basic usage type
- Generalized Anxiety Disorder 7-item (GAD-7) scale
- Patient-Reported Outcomes Measurement Information System (PROMIS) 4-item scale
- User post/status

* Required

Name (Optional)

Your answer

Occupation (Optional)

Your answer

Please help me complete my research providing the following information:

1. Average spending time daily on social media
2. Number of times per day these platforms visited
3. Number of social media active accounts you have and use

1.1 Daily time spent on any Social Media * 1 point

- ○ Less than 1 hr
- ○ 1 - 2 hrs
- ○ 2 - 3 hrs
- ○ More than 3 hrs

1.2 Number of times per day visiting social media * 1 point

- ○ 1 - 3 times
- ○ 4 - 6 times
- ○ 7 - 9 times
- ○ More than 9 times

1.3. Number of 'active' social media accounts have and use * 1 point

- ○ 1 - 3
- ○ 4 - 7
- ○ 7 - 9
- ○ More than 9

GAD-7:: During lockdown period, how often in 4 weeks (~1 month) have you been bothered by the following problems?

GAD-7:: One of the 2 Major scales for getting emotional statistics:

During lockdown period, how often in 4 weeks have you been bothered by the following problems? *

1 = Not at all sure 2 = Several days 3 = Over half the days 4 = Nearly Every day

	Not at all sure (1)	Several days (2)	Over half the days ...	Nearly Every day (4)
2.1 Feeling nervous...	○	○	○	○
2.2 Not being able t...	○	○	○	○
2.3 Worrying too m...	○	○	○	○
2.4 Trouble relaxing	○	○	○	○
2.5 Being so restles...	○	○	○	○
2.6 Becoming easil...	○	○	○	○
2.7 Feeling afraid a...	○	○	○	○

PROMIS-scale:: During lockdown at home, how frequently in 7-days period, you felt--

PROMIS-scale: Well established depression-scale

During lockdown at home, how frequently in 7-days period, you felt-- *

	Never (1)	Rarely (2)	Sometimes (3)	Often (4)	Always (5)
3.1 Hopeless	○	○	○	○	○
3.2 Worthless	○	○	○	○	○
3.3 Helpless	○	○	○	○	○
3.4 Depressed	○	○	○	○	○

Please write one of your facebook/twitter posts/status about Covid-19 (if any) in the last 6 months.

Your answer

Submit

Text from posts collected from Users:

Date	Post
Oct 14, 2020 08:29pm	If you wanna see social distancing, lend a person some money
30-Aug-20	56 hotel staffs in Cox's Bazar are already infected with COVID-19 for the surge of tourists after 17th August, 2020. Lacking a clear SOP and awareness, the situation may become more unstable in future. Source: Dr. Md. Shahjahan Nazir, Cox Bazar Sadar Hospital
28-Aug-20	4 months ago when you were commenting on the stupidity of the people from lower income and lower middle income group vacationing in Cox's Bazar or going back to their villages, did you think that you would be doing the same thing 4 months later when the virus has spread mercilessly across the country, which is undoubtedly a lot worse compared to the situation we had a couple of months ago?
13-Aug-20	97,000 kids tested Covid-19 positive in US within 2 weeks of School opening. Bangladesh is planning to open schools by September.
25-Jan-20	Coronavirus is man-made and was already there in the 1950's. Nowadays only modified. Again, used upor the people. Mainly to spread fear and used for the vaccination-campaigns through mainstream media. #coronavirus #manmade #fear #antivax #wakeUp
12-Jul-20	AB+ Plasma required for Corona patient in Labaid Hospital. If you know anyone who can help. Please let me know.
11-Jul-20	My Nana Shoshur has passed away few minutes back. Inna lillahi Wa Inna Iliaihe Rajiun. He was diagnosed with Covid-19, Pneumonia. He was in life support since last evening. My Nani Shashuri passed away last month with similar symptoms. May Allah grant them Jannatul Ferdous. Please keep them in your prayers.
21-Aug-20	Transparency. Contained. Sure, it is. #coronavirus #SoAngry
18-Jun-20	Coronavirus made us realize that the only solution to any problem is to avoid people

19-Oct-20	Protesting against grandeur and celebration of Durga Puja in West Bengal during this pandemic.
20-Oct-20	Our NGO helped needy people in this situation
2-Jun-20	Corona Day 13: No Smell No Taste Dry Cough Headache Chest Pain Weakness/ Dizziness When does it end
22-Sep-20	Here as well, are people who wear masks under their noses 🙆 So, today I saw a random lady going to two such people inside the tram and asking whether it is too hard to cover their noses as well. Those people really didn't care much but what that lady did, seemed quite bold.
14-Aug-20	No one has clearly understood Covid-19 till date. Not WHO, any Government, Politician, Medical Agency. Trial & Error is the only tool at hand to face the Challenging situation, which is further aggravated due to blunderbuss ignorance all around. Unfortunately, due to this Pandemic, all pre-existing diseases and ailments or medical emergencies are relegated to insignificance by the Authorities. At a most damaging impact over Humanity, globally.
28-Mar-20	This generosity at a critical time in the #coronavirus outbreak is so good to see. Thank you @billgates @melindagates @gatesfoundation
25-Oct-20	Waiting for the morning when I'll got to know there is no corona virus case ..
16 May, 2020	2020 is a Curse...
25 Aug, 2020	We are the proud family who won against covid. Hope everything will be over one day and we can breathe again without any fear and get back the time spending with our near and dear ones.

23-Jan-20	#Chinavirus #coronavirus How SARS and CORONAVIRUS are bioweapons4 "Vaccine Big Business. Read before they censor this!!
21-Oct-20	Empty airports and planes at the height of the #COVID-19 pandemic
28-Apr-20	Encouraging people to get #Coronavirus in the name of Virtue
6-May-20	This implies there will be no 'after' of it and if there's no vaccine, this virus will remain with us
16-Apr-20	Trying not to be angry about this #coronavirus is hard, talking listening and even seeing the effects of it

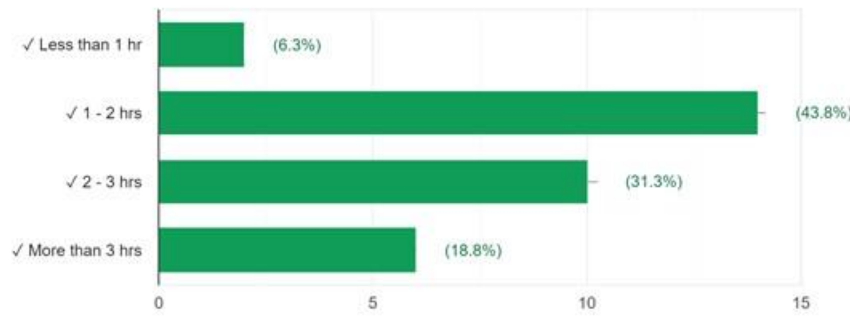
Time	UserName	Tweet_text
2020.05.13 23:59	CarePageAU	By gathering regular feedback #agedcare providers can be responsive to the changing mood
2020.05.13 23:59	NewsMedical	Should school closures related to COVID-19 be continued long-term https://t.co/C8Tc8NdWI
2020.05.13 23:59	lexi_kenney	In the past 3 months weâ€™ve laughed together mourned drank danced and debated To me
2020.05.13 23:59	ByrdAdatto	Episode 43 ByrdAdatto discusses updates to advertising rules and regulations for businesses
2020.05.13 23:59	ActonInstitute	. on the Dan Proft Show COVID-19 lockdown orders are the state-mandated â€marginalizatio
2020.05.13 23:59	AprilGarbusjuk	Empathy is a critical component in communications and relationship-building both now and
2020.05.13 23:59	oliviadeng1	Fuck anyone who says Charlie Baker is handling the COVID-19 pandemic well It took him a m
2020.05.13 23:59	bettycjung	US formally accuses China of hacking US entities working on COVID-19 research https://t.co/
2020.05.13 23:59	guy123_g	Blocked for thinking all teens are ignorant of COVID-19.
2020.05.13 23:59	MairiMMcInnes	_UK's new analysis on the â€return to work strategy and what a phased approach to lifting se
2020.05.13 23:59	JasonClark829	If someone has a heart condition and wasnâ€™t going to die but contracts covid and dies yes
2020.05.13 23:59	Coffee_2222	#CCPvirus #Wuhan #Coronavirus China Removes All Traces of â€Thousand Talents Program C
2020.05.13 23:59	LightTherapy_	Wow this is a Success for Wisconsin Stay at home order is Unconstitutional from Court đź‡±đ:
2020.05.13 23:59	q_abolitionist	Yo â¦ donâ€™t get any ideas This wonâ€™t go the way you think it will https://t.co/yqsXKwF
2020.05.13 23:59	Sucksie_Sucks	Almonds are high in Zinc which is Trump's covid's home remedy I will get cuz Jaz to ban them
2020.05.13 23:59	TUDigitalMedia	With many concerned with how the pandemicâ€™s trend line will impact future media mark
2020.05.13 23:59	TheChandlerDude	@__ListenUp_ Very few jurisdictions have been reporting COVID-19 mortality properly "Exce
2020.05.13 23:59	cannonhillpark	A Wings and Scrubs Angel in Lightwoods Park on Hagley Rd by sculptor Luke Perry shines a li
2020.05.13 23:59	michele_kay_	The end of an erađź™Źđźź Fuck you Covid for pillaging the remainder of my youth
2020.05.13 23:59	903KAZU	On our COVID-19 blog today: -No restrictions on who can get tested in Monterey County.. ev
2020.05.13 23:59	StephenNee2	Which figures More than 30,000 excess deaths so far this year in UK v any previous average F
2020.05.13 23:59	compass_housing	Compass Professor Adamson is a leading advocate for #HousingForAll and recently participa
2020.05.13 23:59	AamirKhan9feb97	Day 208 19th&20th Ramadan Mubarak Allah ap &amp apki family ko Salaamat rakhe Aar
2020.05.13 23:59	weact2	Vitamin D determines severity in COVID-19 so government advice needs to change https://t
2020.05.13 23:59	Feminismandfre1	What if it isnâ€™t covid and is Kawasaki Serious question.
2020.05.13 23:59	oranglaut	'This virus may never go away, WHO says https://t.co/C4cfHwc7YM
2020.05.13 23:59	frankkimmel4	With hospitals incented by Medicare payments to report all deaths as CoVid-related regardl
2020.05.13 23:59	jeffjameslee	John Ivison Ottawa's COVID-19 debt binge runs very real risk of ruining the next generation I
2020.05.13 23:59	Degeniusmedia	BREAKING COVID-19 Cases In Nigeria Near 5000 As Lagos Record Fresh 51 Cases #NoLagosLoc
2020.05.13 23:59	shortfamly	There are now reports showing that children from 1 to 18 are developing different effects fr
2020.05.13 23:59	Akamai49	One more thing since you are personally responsible for the evening news content how do j
2020.05.13 23:59	dyreerawrs	Yeah.. So this is why the US is never going to be rid of covid. CNN Wisconsin Supreme Court :
2020.05.13 23:59	azfamily	Maricopa County Public Health hires trains more contact tracers to track COVID-19 to help de
2020.05.13 23:59	USEmbassySeoul	The gathers #COVID19 data and makes it accessible to the public so that practitioners arounc

Time	UserName	Tweet_text
2020.04.15 23:59	John_R_Amelia	CenturyLink's adaptive networks flourish during COVID-19 crisis https://t.co/iy9zDQcxrz https://t.co/da8amB4nOR
2020.04.15 23:59	TheRecipe6	If not for covid 19 you were just enjoying your rich celebrity life Don't act like you care now please Lol But just for the record the answer is "Nothing
2020.04.15 23:59	BitcoinBiology	_salad It literally says "COVID-19 originated in a Wuhan laboratory not as a bioweapon" Are you assuming every biosafety lab leak is a bioweapon ?
2020.04.15 23:59	HayleySnider2	Not COVID but Iâ€™ve literally had people tell me if I donâ€™t believe in god my CF will be much worse and that I can be saved if I pray or whateve
2020.04.15 23:59	meganbushway15	_Boom _mckraken _tony _SecondFleet Then there's this: https://t.co/wTg4lbeuul
2020.04.15 23:59	theladralph	#EXCLUSIVE: Fox News is reporting that Covid-19 originated in a Wuhan laboratory (China not as a bioweapon but as part of China's effort to demon
2020.04.15 23:59	RadWagon	It's good to see some politicians making sure they let everyone know that they are in this as well Isn't it I'm mean isn't it #Tories I mean isn't it ? htt
2020.04.15 23:59	imleftcoast	"People should know the facts. Blames WHO &amp claims Covid-19 is the 19th strain of the virus #TrumpPlague
2020.04.15 23:59	jimgoldstein	UW team illustrates the adverse impact of visiting â€¦just one friend during COVID-19 lockdown UW News https://t.co/eoqid0nSLz
2020.04.15 23:59	ArrighiOrosz	But they do seem to have the best resistance to the Covid virus Maybe that's the key eat like a toddler đź¤Łđź¤Łđź¤Ł
2020.04.15 23:59	VestaviaVoice	Rachel Brockwell has been selling handmade scrunchies at The Clotheshorse since last fall but when the COVID-19 pandemic hit she created a patt
2020.04.15 23:59	stockx	Hey there our team just followed up through email Thank you for your patience at this time we appreciate it! https://t.co/Xkqzr9lmL8
2020.04.15 23:59	DavidACohen_MD	Inpatient thoughts: I canâ€™t count how many hundreds of times Iâ€™ve told a patient they have cancer. But telling them they have COVID seems
2020.04.15 23:59	KaitMarieox	One day after CNN ran Chinese propaganda its being confirmed that Covid-19 was created in a Chinese laboratory &amp their government covered
2020.04.15 23:59	daniialonzo	COVID-19 Sonora https://t.co/UvGWa51HbW
2020.04.15 23:59	deluded_jim	Ok so those deaths are COVID-19 but ERs and ICUs that are suddenly so full Trump brings a hospital ship up to New York is probably because of an u
2020.04.15 23:59	phifedogg1	over 65,000 people died from the flu in 2019 a lot of those were children no child in the US has died from covid-19 under the age of 17 The joke is yo
2020.04.15 23:59	Khelrauko	The UK is currently contributing approximately 9 of the total deaths from COVID-19 in the world A country with a world population share of 0.87% I'
2020.04.15 23:59	Nanjimo	Maybe itâ€™s a new symptom of COVID-19 (For Kellyanne thatâ€™s 2019)
2020.04.15 23:59	bluetreepoint	With the spread and impact of COVID-19 have we now entered a Bearish Market period #bearish #trading #market #investing #markets #covid19 ht

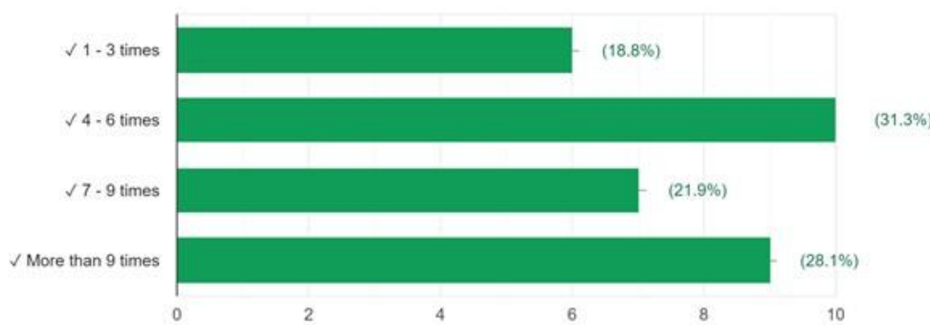
Time	UserName	Tweet_text
2020.04.14 23:59	SCOTTYSIMM	White House Attacks Voice Of America Over China Coronavirus Coverage https://t.co/qEzZIgC0WW
2020.04.14 23:59	unlewis	Trump says he doesn't think Michigan Dem will vote for 'Sleepy Joe after surviving coronavirus Fox News h
2020.04.14 23:59	sdakshin	Word of the year https://t.co/mi0pzuITyb via
2020.04.14 23:59	PeepsRide	Elderly people who have suspended in-home services during the coronavirus crisis will soon receive welfa
2020.04.14 23:59	SeattleMet	Dig around a little You probably already have something mask-worthy at home https://t.co/NOpObfEwIs
2020.04.14 23:59	rep4bettergovt	Coronavirus Live Updates Trump Halts U.S Funding of World Health Organization Perhaps funding of the Re
2020.04.14 23:59	Milesandmiles30	Why have so many people in this thread stopped reading after coronavirus You know that's not the end of
2020.04.14 23:59	muhammad_fall	Haunting photos of empty airports and planes at the height of the #COVID-19 pandemic show the airline in
2020.04.14 23:59	AllOnMedicare	US for-profit healthcare sector cuts thousands of jobs as pandemic rages https://t.co/eGL63ujnlR
2020.04.14 23:59	Jsurfer0730	Trump campaign still bragging about 'lowest unemployment rate in years https://t.co/CHBHDBfD3A
2020.04.14 23:59	kbaseballump	One Arrested as Raleigh NC Police Suspend First Amendment Declare Coronavirus Lockdown Protest "Non-
2020.04.14 23:59	brisuelou	He thinks that the WHO mismanaged this? Are you fucking kidding me? https://t.co/IzU2YG6ssx
2020.04.14 23:59	Verda_777	_camera _defender Here is the chronological order provided by the Hudson Institute China and WHO are di
2020.04.14 23:59	oracleofliberty	wow out of control for federal funds. #CoronavirusPandemic #coronavirus #newyorkcoronavirus #AndrewC
2020.04.14 23:59	hocott_david	. I urge you to #cancelstudentdebt in the next #coronavirus package A #StudentDebtStimulus will help the
2020.04.14 23:59	aprilhenrybooks	The Art (and Awkwardness of a Virtual Haircut https://t.co/WIQn3MHN0b
2020.04.14 23:59	WarWraith	This implies there will be an "after" If there's no effective vaccine and this coronavirus remains with us the
2020.04.14 23:59	AdvocaatBroeke	U.S will halt funding to World Health Organization while it investigates groupâ€™s response to Coronavirus
2020.04.14 23:59	FulviaCalcagni	[... there is little reason to expect the coronavirus crisis to accelerate climate-friendly decoupling UNLESS p
2020.04.14 23:59	nisar__adil	The study found that the genetic sequence of a coronavirus discovered in lung samples of pangolins was hi
2020.04.14 23:59	manatelugumovie	Corona In India 644 In Telangana 52 Just From Today! The number of Coronavirus in https://t.co/acLVNgeAM
2020.04.14 23:59	ChiTownLionPSU	Data from https://t.co/MywM5iAcVh wrt Coronavirus/Covid19. Comparing the United States to some of ou
2020.04.14 23:59	MickSPrice	#WhereIsNana Crime Family Pelosi extorting taxpayer funds for decades I see her dad was a corrupt politic
2020.04.14 23:59	jjbigcity	TRUMP KNEW ABOUT CHINA'S OUTBREAKS OF CORONAVIRUS IN NOVEMBER 2019 IN FACT OF THIS PRESIDEI
2020.04.14 23:59	RonRonkmfa	There are those among us who are chomping at the bit to congregate and mingle without regard for the eff
2020.04.14 23:59	RichinWriterss	Trump Coronavirus Task Force hold press briefing at White House 4/14/20 https://t.co/dwvPGr9DaK
2020.04.14 23:59	diimsa	The Word Challenge High School (7:00 PM 7:05 PM CST). #DIIMSAVBOARD #DIIMSACHALLENGE #DIIMSA #Cli
2020.04.14 23:59	Reuters	As virus tears through reservation Navajos give lifeline to elders and families https://t.co/g72TcIelZE https:
2020.04.14 23:59	steviekim222	What Does Coronavirus Mean For Brands on Social Media /feed/what-does-coronavirus-mean-for-brands-
2020.04.14 23:59	BagalueSunab	https://t.co/tfezR73pbt Coronavirus mutation could threaten the race to develop vaccine A strain found in
2020.04.14 23:59	mharisman	(4) Dog of a Covid-19 patient in Hong Kong has tested â€śweak-positive for Covid19 but officials say there i:
2020.04.14 23:59	Ronnie_Zen	Together we can stop the spread of Coronavirus Disease (COVID-19) #AyoBersamaLawanKorona #DiRumah.
2020.04.14 23:59	The_Real_Fly	Boeing Co on Tuesday reported another 75 cancellations for its 737 MAX jetliner in March as the coronaviru
2020.04.14 23:59	DonDoncarey4	Student files class-action lawsuit against Liberty University over coronavirus response TheHill https://t.co/
2020.04.14 23:59	ParhamMansor	Chinaâ€™s â€Donation Diplomacy Raises Tensions With U.S https://t.co/7rgzDduAQr
2020.04.14 23:59	khw_lker	Trying not to be angry about this coronavirus is hard Talking listening and even seeing the effects of it is tir
2020.04.14 23:59	HausSante	encouraging people to get Coronavirus In the name of Virtue https://t.co/vqI8uIkhVf

Graphical view of collected data:

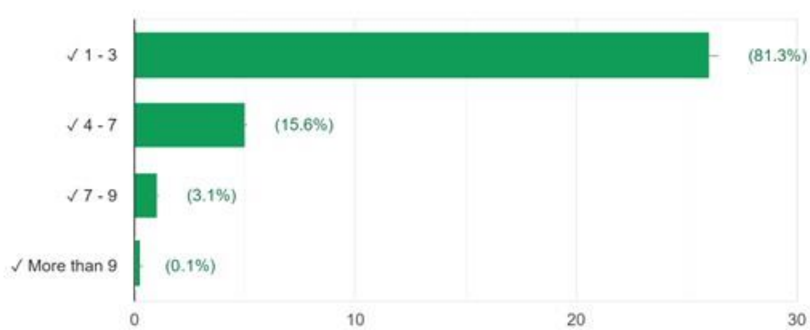
1.1 Daily time spent on any Social Media



1.2 Number of times per day visiting social media



1.3. Number of 'active' social media accounts have and use



Appendix B

(Research Source Code)

R Codes for divisive hierarchical Clustering:

```
install.packages("FactoMineR") library(FactoMineR)

 install.packages("factoextra") library(factoextra)

library(class) fviz_dend(res.diana, cex=0.5, k=4,
```

```
palette= "jco") print() library(foreign)

library(compiler) library(Matrix)

fviz_cluster(res.diana, cex = 0.5, k = 4)

fviz_cluster(res.diana) fviz_cluster()

fviz_dend(res.diana, cex = 0.5, k = 4, palette = "jco")

library("tidytext") install.packages("tidytext")

library(tidytext)


CRAN_mirrors <- read_csv("R/R-4.0.2/doc/CRAN_mirrors.csv")

View(CRAN_mirrors) library(readr) smNumber <-

read_csv("R/SM dataset.csv")

View(smNumber) plot(smNumber) library(factoextra)

library(cluster) res.diana <- diana(smNumber, stand =

TRUE) library(factoextra) fviz_dend(res.diana, cex =

0.5, k = 4, palette = "jco") fviz_dend(res.diana, cex =

0.8, k = 3, palette = "jco")
```

R Codes for POS-tagging:

```
 install.packages("udpipe")

library("udpipe")

POStag <- function(object = object){

require("stringr")   require("NLP")

require("openNLP")

require("openNLPmodels.en")

corpus.tmp <- object
```

```
    corpus.tmp       <-       lapply(corpus.tmp,
function(x){    x <- paste(x, collapse = " ")   }
)    corpus.tmp <- lapply(corpus.tmp, function(x)
{    x <- enc2utf8(x)   }  )    corpus.tmp <- gsub("
{2,}",   "   ",   corpus.tmp)       corpus.tmp   <-
str_trim(corpus.tmp,      side      =      "both")
sent_token_annotator                             <-
Maxent_Sent_Token_Annotator()
word_token_annotator                             <-
Maxent_Word_Token_Annotator()
  pos_tag_annotator <- Maxent_POS_Tag_Annotator(language =
"en", probs = FALSE,
    model = " C:\Users\ryq\Documents\R\R-
4.0.3\library\openNLPdata\models\\en-pos-maxent.bin")
text <- readLines("D:\Ry_docu\Courses\Project
Thesis\Dataset\SM dataset.xlsx") textpost_en
  Corpus <- lapply(corpus.tmp, function(x){
x  <-  as.String(x)    }    )     lapply(Corpus,
function(x){    lapply(object, function(x){
    y1 <- annotate(x, list(sent_token_annotator,
word_token_annotator))    y2<- annotate(x,
pos_tag_annotator, y1)
   y3 <- annotate(x, Maxent_POS_Tag_Annotator(probs =
TRUE), y1)    y2w <- subset(y2, type == "word")    tags
<-   sapply(y2w$features,  '[[',  "POS")        r1   <-
```

```
sprintf("%s/%s",  x[y2w],  tags)       r2  <-  paste(r1,
collapse = " ")     return(r2)  }  )
  }
```

POS Tagging with TreeTagger:

```
library(koRpus) library(koRpus.lang.en)

set.kRp.env(TT.cmd="C:\\TreeTagger\\bin\\tag-english.bat", lang="en")

postagged <- treetag("D:\Ry_docu\Courses\Project
Thesis\Dataset\\testcorpus/linguistics07.txt")

datatable(postagged@tokens, rownames = FALSE, options =
list(pageLength = 5, scrollX=T), filter = "none")
```

POS Tagging with coreNLP:

```
library(coreNLP) initCoreNLP()

annotation <- annotateString(text)

text <- readLines("D:\Ry_docu\Courses\Project Thesis\Dataset\SM
dataset.txt")

# convert character to string s

<- as.String(text)

# define sentence and word token annotator sent_token_annotator <-

Maxent_Sent_Token_Annotator() word_token_annotator <-

Maxent_Word_Token_Annotator()

a2 <- NLP::annotate(s, list(sent_token_annotator,

word_token_annotator)) parse_annotator <-

Parse_Annotator() p <- parse_annotator(s, a2)

ptexts <- sapply(p$features, '[[', "parse")
```

```
ptexts ptrees <- lapply(ptexts, Tree_parse)

ptrees[[1]]
```