MSCI-623 Final Project

Course Instructor - Prof. Lukasz Golab

Analysis of Mental Health Impact due to COVID-19

Submitted by

Anh Truc Quan Nguyen

Student ID: 20818546
Department of Management Sciences
University of Waterloo

Jeswin Maria Cinthia Palthasar

Student ID: 20837700
Department of Management Sciences
University of Waterloo

Abstract

The impact on mental health due to COVID-19 is a rising global concern. Several measures have been taken by the government to reduce the spread and it has profoundly affected day to day life of people. Our objective is to analyse the emotional wellness of people with respect to change in different aspects of regular lifestyle. This analysis aims to apply appropriate machine learning algorithms to diagnose mental state of the survey respondents and to identify the level of assistance they require to ensure mental stability. The analysis is done in three stages. The initial stage explores the influence of the mental health due to different variables in the environment. The next stage of analysis is predictive analysis which involves prediction of the mental stability according to a set of predictor variables. The third stage of analysis deals with the finding of association between different factors that can impact the mental wellness.

1 Introduction

The coronavirus disease (COVID-19) has been escalating in the United States since beginning in March of 2020. With the rapid changes in societies including university closures, remoting work, loss of work, people are facing some mental health challenges. This report sought to identify potential factors that affect emotional health of adults during the pandemic. A survey conducted by NORC at the University of Chicago addressed the overall public perception and conditions [1]. Using this dataset, we want to predict the mental health condition of participants using 27 preprocessed explanatory variables. K-means clustering algorithm operates by updating the means of the cluster. Even though the categorical data can be integer coded, the distance between these variables are not appropriate to measure the mental wellness. We later used association rule mining to explain the rules related with mental well-being.

2 Related work

The impact of COVID-19 on the society and individual, especially mental health, has been addressed by several publications. Holmes (2020) stated that an emergency action for mental health science for the COVID-19 pandemic should be taken by researchers. This is an urgent need for them to identify the drives of mental health and psychological effects by using the quality data across the population and vulnerable groups [2]. The article gains strong interest from researchers, it was later cited by 353 different publications. Another earlier survey in 190 Chinese cities conducted by Wang (2020) with 1738 respondents shows that there is a significant reduction in mental health after 4 weeks of COVID outbreak. Other highlights are female gender has higher association with psychological impact and respondents who wear masks and use hand sanitizer have reduction in mental impact [3]. Increased feelings of anxiety and depression has also been addressed in several reports [4] [5] [6]. When the COVID-19 hit the United States, older adult age is the vulnerable group due to the higher fatality risk. However, a nationally representative address-based sample of 6,666 U.S. adults suggests that this group was also associated with seeing lower risks of getting COVID-19, less economic effects, less depression and less anxiety [7]. There are no age differences in depression and anxiety based on the report. Still, it is suggested that promoting good mental health is vital for older population during quarantine [8]. Another group that needs mental care assistance is youth. The mass school closures increase the mental health problems in this group [9] [10] [11]. First line workers including health workers who must contact close with the virus also experiencing stress and mentally unwell [11]. There is strong evidence that the epidemic is having a profound effect on all society.

3 Data

3.1 Data collection

The data is published by Associated Press which is based on a survey conducted in the United States after one month of the coronavirus hit the country [1]. This survey gives the information about physical health, mental health, changes in lifestyle and economic conditions of the habitants in 10 states (California, Colorado, Florida, Louisiana, Minnesota, Missouri, Montana, New York, Oregon and Texas) and eight metropolitan areas (Atlanta, Baltimore, Birmingham, Chicago, Cleveland, Columbus, Phoenix and Pittsburgh)[19]. The dataset is reported in April 2020 by NORC at the University of Chicago which focuses on three criteria:

- Physical Health: Symptoms related to COVID-19, relevant existing conditions and health insurance coverage.
- Economic and Financial Health: Employment, food security, and government cash assistance.
- Social and Mental Health: Communication with friends and family, neighbours, anxiety and volunteerism.

3.2 Data Preprocessing

Since the data is based on the Questionnaire from the COVID-19 survey, it needed exhaustive preprocessing. The original dataset has 178 columns with respect to 36 questions. The following columns from the original dataset has been modified and customized to analyse the impact in Mental health due to COVID-19.

SOC 1 – Trustworthiness of Neighbourhood

SOC 1 column is based on how much people in the neighbourhood are trustworthy. People were given four response options All, Most, Some, None. This variable did not require any processing and kept as such.

SOC 2A – Contact with neighbours in the past month (during pandemic)

The variable is based on the frequency of communication during the pandemic. Six response options were given such as Basically everyday, A few times a week, A few times a month, Once a month, Not at all and Not sure. This variable did not require any processing and kept as such.

SOC 2B – Contact with neighbours prior to pandemic

The variable is based on the frequency of communication prior to March – 1, 2020. Six response options were given such as Basically everyday, A few times a week, A few times a month, Once a month, Not at all and Not sure. This variable did not require any processing and kept as such.

CONTACT_NEIGHBOURS_DIFF variable is created to observe the difference in communication before and after pandemic

SOC3A - Contact with family and friends in the past month (during pandemic)

The variable is based on the frequency of communication during the pandemic. Six response options were given such as Basically everyday, A few times a week, A few times a month, Once a month, Not at all and Not sure. This variable did not require any processing and kept as such.

SOC 3B– Contact with family and friends prior to pandemic

The variable is based on the frequency of communication prior to March – 1, 2020. Six response options were given such as Basically everyday, A few times a week, A few times a month, Once a month, Not at all and Not sure. This variable did not require any processing and kept as such.

CONTACT_FAMILY/FRIENDS_DIFF variable is created to observe the difference in communication before and after pandemic

SOC 4A – Volunteered for any organization or association during pandemic.

The variable is in binary form. The response yes indicated that the person involved in volunteering activities during pandemic and the response no indicated that the person was not involved in volunteering during pandemic.

SOC 4B – Volunteered for any organization or association before pandemic.

The variable is in binary form. The response yes indicated that the person involved in volunteering activities before March-1,2020 and the response no indicated that the person was not involved in volunteering during pandemic.

VOLUNTEERING_DIFF variable is created to record the difference in involvement in volunteering activities before and after pandemic

PHYS8 – Physical health

The respondents were asked how they feel about their health at that time. They were given five options such as Excellent, Very good, Good, Fair, Poor.

PHYS1 – Symptoms of COVID -19

The respondents were asked if they were experiencing symptoms. The were given 17 options like Fever, cough, change in appetite. These symptoms were categorized into three categories

- MOST_COMMON_SYMPTOMS - Fever, Cough, Tiredness or Fatigue
- SERIOUS_SYMPTOMS - Chest Congestion, Shortness of breath

- LESS_COMMON_SYMPTOMS - Chills, Runny or Stuffy nose, Skin rash, sore throat, sneezing, muscles aches, headaches, Abdominal discomfort, Nausea, Vomiting, Diarrhea, change or lost sense of taste or smell and lost of appetite

The categorization was done with reference to self-assessment details provided by Government[0]. The count of symptoms with respect to categories and recorded for all the respondents.

PHYS2 - FOLLOWED_OFFICIAL_GUIDANCES

The respondents were asked if they are following any measures in response to COVID-19. Out of the given options we have selected the measures advised by the government. The options are as follows.

- Canceled a doctor appointment [12]
- Worn a face mask [13]
- Avoided public or crowded places [13]
- Washed or sanitized hands [13]
- Kept six feet distance from those outside my household your household [13]
- Wiped packages entering my your home [13]
- Canceled or postponed work activities [14]
- Canceled or postponed school activities [14]
- Canceled or postponed dentist or other appointment [14]
- Worked from home [14]
- Studied at home [14]
- Canceled or postponed pleasure, social, or recreational activities [14]
- Avoided contact with high-risk people [14]
- Stayed home because I felt unwell you felt unwell [15]

The number of activities has been counted and they have been grouped into following categories.

- 0 - 4 : low adherence
- 5 - 9: moderate adherence
- 10 - 14: strong adherence

The count indicated the adherence to the rules and the preventive measures suggested by the government. The rest of the response options were not included because it indicated self disciplined activities. It is assumed that they would not have a much impact on the Mental health.

ECON8 - Impact on personal plans.

The respondents were asked if their personal plans have been affected due to new rules imposed. This question is further categorized to four categories such as

COMMUNITY_LOCATIONS_AND_EVENTS

- Closure of bars
- Closure of restaurants
- Closure of other businesses

- Canceled sport events
- Closure of place of worship
- Closure of work
- Work from home requirements

CHILDCARE_EDUCATION
- K-12 school closure
- Pre-K or child care closure
- College or training closure

PUBLICS_SERVICE_AND_INFRASTRUCTURE
- Reduced public transportation
- Other reduced public services
- International travel restrictions or bans
- Domestic travel restrictions or bans

PERSONAL_AND_SOCIAL_ACTIVITIES

- Ban on gatherings of 250 people or more
- Ban on gatherings of 50 people or more
- Ban on gatherings of 10 people or more
- Closure of gyms or fitness facilities
- Quarantine requirements or stay-at-home orders

These categories are binary variables. The response 'Yes' denotes their personal plans have been affected and vice versa.

ECON8Q – Compulsory Quarantine requirements

This variable is a binary variable. The response 'yes' indicates that the person should be isolated and vice versa.

ECON 6 – Government and Community assistance

The respondents were asked if they received, applied or not applied for income assistance. This question is categorized into two parts as Government assistance and community assistance. If the respondent has received any of the assistance the option 'yes' is recorded and vice versa

Government assistance:

- Unemployment Insurance
- SNAP (Supplemental Nutrition Assistance Program) called Supplemental Nutrition Assistance Program or Food Stamps
- TANF (Temporary Assistance for Needy Families) [CATI] called Temporary Assistance for Needy Families
- Social Security

- Supplemental Social Security
- Any kind of government health insurance or health coverage plan including Medicaid, Medical Assistance or Medicare
- Other aid from the government

Community assistance:

- Assistance from a church or religious organization
- Assistance from another community organization
- A food pantry
- Other assistance

ECON 5A – Food/Money Security
The respondents were asked if they were worried about running out of groceries and did not have enough money to buy them. The response options often true, sometimes true and never true were recorded.

PHYS9 - Covered by insurance
The respondents were asked if they were covered by health insurance or health plan coverage. It is a binary column. The response option 'Yes; indicates they are covered by health insurance and vice versa.

PHYS3 – NO_OF_CONDITION_AT_RISK

We based on CDC list of underlying medical conditions that might be at an increased risk for severe illness from COVID-19 [16]. The conditions are as follows.

A. Diabetes
B. High blood pressure or hypertension
C. Heart disease, heart attack or stroke
D. Asthma
E. Chronic lung disease and COPD
F. Bronchitis and emphysema
G. Allergies
H. A mental health condition
I. Cystic fibrosis
J. Liver disease or end stage liver disease
K. Cancer
L. A compromised immune system
M. Overweight or obesity

The number of activities has been counted and they have been grouped into following categories.
- Conditions at risk: C, M, A, D, B, L, E, F, I, J, K
- Mental condition: H

CDC does not state patients with cancer and mental problem have higher risk. However, based on the understanding that some cancer treatments such as chemotherapy can weaken immune system leading to immunocompromised [17]. We decide to classify cancer as conditions as risk. With mental conditions, we suspect it would impact the mental emotional health (SOC5) under COVID-19. By putting the condition to separated column with binary value, it could help us in better visualization of how mental health is influenced by other variables.

PHYS4 – Tested positive
The respondents were asked if they are tested positive for Corona. This is a binary variable. The response 'Yes' indicates that they are tested positive and vice versa

PHYS5 – contact with COVID-19 positive
The respondents were asked if they are in contact or came in contact with COBID positive person. This is a binary variable. The response 'Yes' indicates that they have contacted COVID positive person and vice versa

AGEGROUP
The respondents were asked to select the age group they belong. The age group is as follows

- 18-29
- 30-44
- 45-59
- 60+

HHsize – House Hold Size
The respondents were asked to select the no of persons living in the house.

HouseHold Income
The respondents were asked to select the income of the house. The response is as follows.

- Under 10K
- 10K-20K
- 20K-30K
- 30K-40K
- 40K-50K
- 50K-75K
- 75K-100K
- 100-150K
- 150K+

SOC5 - EMOTIONAL HEALTH

The respondents were asked whether they are having the following reactions such as nervous, anxious, or on edge, depressed, lonely, hopeless about the future and had physical reactions such as sweating, trouble breathing, nausea or a pounding heart when thinking about your experience with the coronavirus pandemic. Based on the number of days they had the above reactions, their emotional health is classified as ***Stable, Need Assistance and Need Immediate Assistance***. If the individual experience the symptoms for 1- 4 day(s) they have been categorized as need assistance. If the individual experience the above symptoms for more than 5 days, they have been categorized as Need immediate Assistance and the individuals who did not experience any symptoms have been categorized as stable.

4 Exploratory Data Analysis(EDA)

4.1 Integer Coding of data

Most of the data collected is categorical data. In order to feed them into inputs to a classifier and EDA they have to be encoded. The categorical data has been converted into numerical data and listed in Table 1. The respondents were given an option to skip the question. For the skipped questions a default value of 0 is assigned.

VARIABLES	CATOGORIES	ENCODINGS
TRUSTWORTHINESS_OF_NEIGHBOURHOOD	None	0
	Most	1
	Some	2
	All	3
CONTACT_WITH_NEIGHBOURS_DIFFERENCE	No	0
	Yes	1
CONTACT_WITH_NEIGHBOURS_DIFFERENCE	No	0
	Yes	1
VOLUNTEERING_DIFF	No	0
	Yes	1
PHYSICAL_HEALTH	Poor	0
	Fair	1
	Good	2
	Very Good	3
	Excellent	4
MOST_COMMON_SYMPTOMS	(Numeric)	
SERIOUS SYMPTOMS	(Numeric)	
LESS COMMON SYMPTOMS	(Numeric)	
FOLOWED_OFFICIAL_GUIDELINES	Low adherence	0
	Mild adherence	1
	Strong adherence	2
COMPULSORY QUARANTINED(ECON8Q)	No	0
	Yes	1
COMMUNITY_LOCATIONS_AND_EVENTS	No	0
	Yes	1
CHILDCARE_EDUCATION	No	0
	Yes	1
PUBLICS_SERVICE_AND_INFRASTRUCTURE	No	0

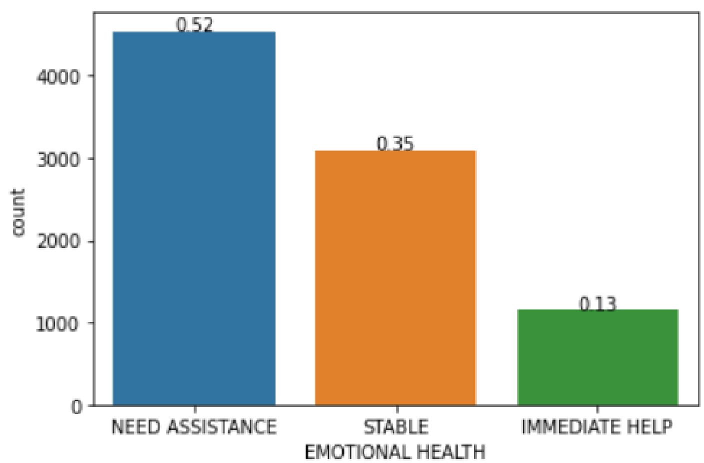
	Yes	1
PERSONAL_AND_SOCIAL_ACTIVITIES	No	0
	Yes	1
FOOD/MONEY_SECURITY	Skipped/Refused	0
	Don't Know	1
	Often true	2
	Sometimes true	3
	Never True	4
GOVERNMENT ASSISTANCE		One hot Encodings
COMMUNITY ASSISTANCE		One hot Encodings
COVERED BY INSURNANCE	No	0
	Yes	1
COVID POSITIVE	No	0
	Yes	1
CONTACT WITH COVID POSITIVE	No	0
	Yes	1
DEATH OF FAMILY MEMBER FROM COVID/ILLNESS	No	0
	Yes	1
AGE_GROUP	18-29	0
	30-44	1
	45-59	2
	60+	3
GENDER	Female	0
	Male	1
EDUCATION		One hot Encodings
HHSIZE	One person	1
	Two person	2
	Three persons	3
	Four Persons	4
	Five Persons	5
	Six persons or more	6
HHINCOME	Under 10K	0
	10K-20K	1
	20K-30K	2
	30K-40K	3
	40K-50K	4
	50K-75K	5
	75K-100K	6
	100-150K	7
	150K+	8
EMOTIONAL HEALTH	STABLE	2

	NEED ASSISTANCE	1
	NEED IMMEDIATE HELP	0

Table 1: Integer Encoding of Variables

4.2 Data Visualization

After cleaning and preprocessing of data, the next step is to explore the data to get the insights from it. The variable Mental Health is chosen to be an outcome variable and rest of the variables are taken as explanatory variables. From the bar graph plotted, it is observed that 35% of the respondents are stable, the majority of the respondents (52%) needs some assistance to deal with anxiety/depression and 13% needs immediate attention. This figure denotes that majority of the people's mental health is impacted by COVID-19. Since the current situation is new for many people and it is impacting the regular lifestyle, it is difficult to adapt.



.

Fig 1: Distribution of classes in Mental Health

Out of 13% of people needing immediate attention, 6% of the people have been diagnosed with mental illness. On further analysis, it is found that 1% of the respondents have been tested positive for COVID-19 and 4% of the COVID positive patients have encountered death of family member due to COVID or respiratory illness. The statistics also shows that above 30% of the respondents have received Government or community assistance and 93% of the respondents are covered by insurance.

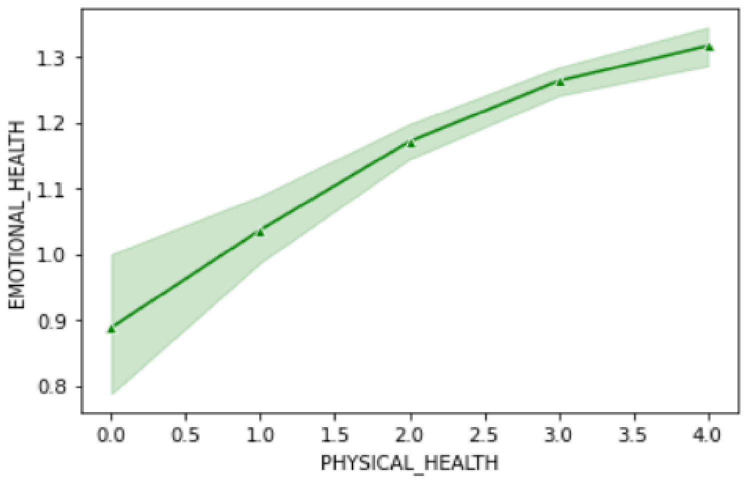


Fig 2: Physical health Vs Mental Health

The next step is to investigate the impact of mental health with physical health. The line plot showed in fig 2 indicates that as the physical heath improves, mental health also improves. It is found that poor physical health has profound negative impact on mental health. This deterioration of mental health may be because, individuals with chronic illness are at increased risk to COVID-19.

On further examination of impact in mental health due the number of chronic illness or increased conditions at risk, it is observed that the mental health declines as the number of conditions at risk increases. The mental health also fluctuates when the number of conditions at risk goes beyond 7. This trend in Fig. 3 is clearly an evidence that COVID-19 has created worries amongst the individual with chronic illness.

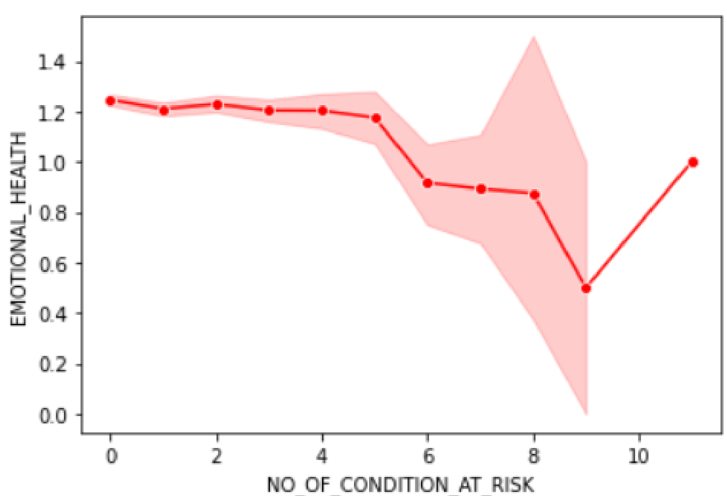


Fig 3: Emotional Health Vs No of Conditions at risk

The impact of mental health is also examined with experience of symptoms of COVID-19. A line plot is plotted between Emotional health and number of the most common symptoms, less common symptoms and serious symptoms. In each scenario it is observed that as the number of symptoms experienced increases, the negative impact on mental health increases.

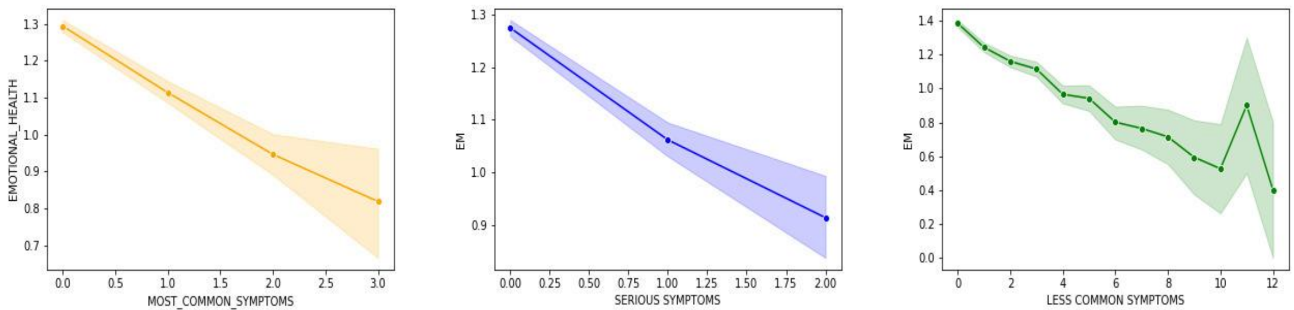


Fig 4: Emotional Health Vs Symptoms

When the impact of mental health is analysed with adherence of individual following official guidelines, it is surprising to see that people thoroughly following best practices are more prone to anxiety and depression. The adaption to the change in lifestyle has made the individual more anxious. The reason is that this kind of rules is new to the people and we have not been exposed to such pandemic in our lifetime.

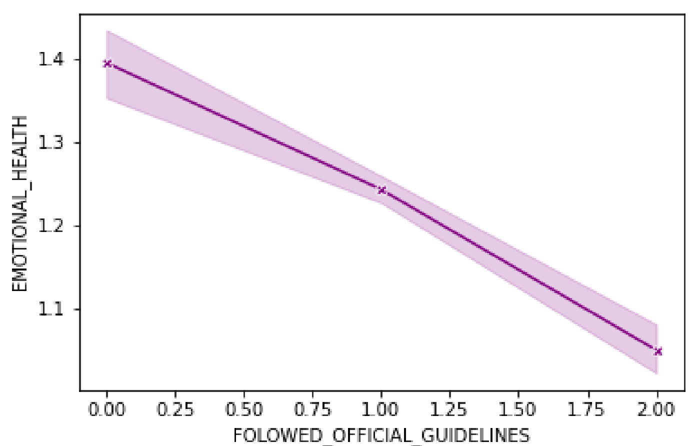


Fig 5: Emotional Health Vs Followed_Official_Guidelines

It is observed from the figure that the mental health declines as the people were trying to follow rules and regulations. Due to these new rules' enforcement, communication with families and friends have been impacted. The recorded response shows that 38% of the responded have encountered decreased contact with neighbours and 29% of the respondents have reduced contact with the family and friends. Furthermore, 22% of the respondents have discontinued their volunteering activities due to COVID-19.

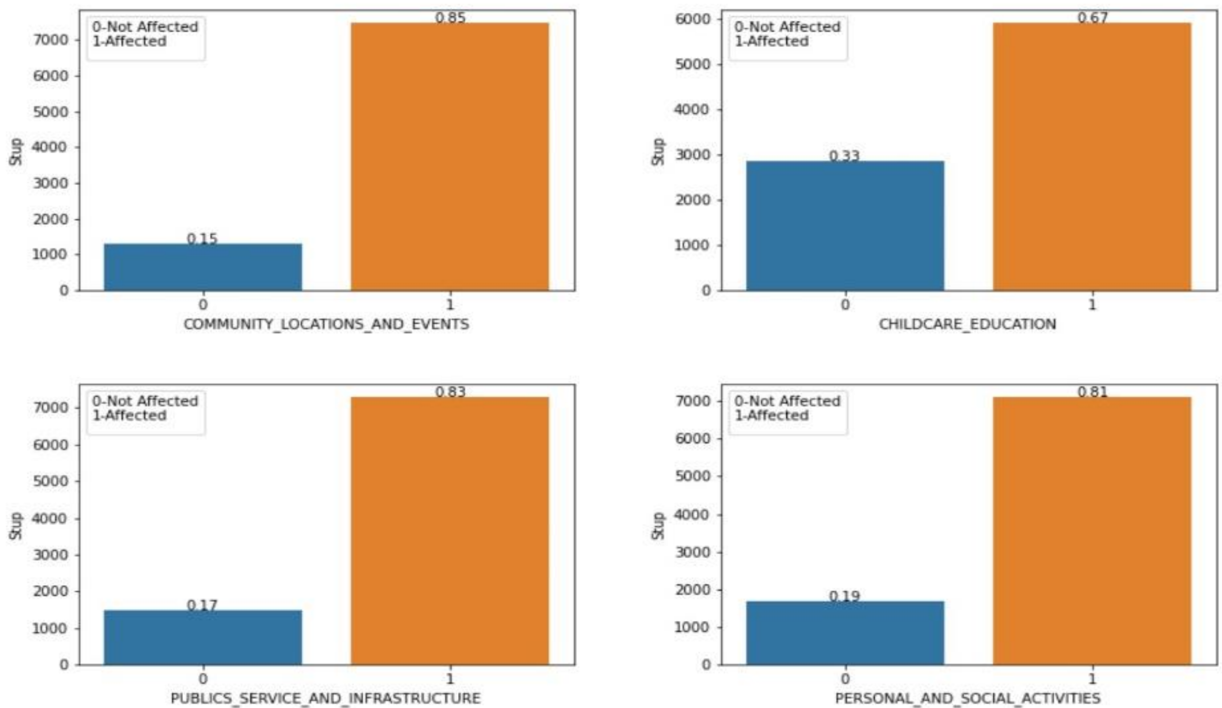


Fig 6: % of Personal Activities affected by COVID-19

From Fig 6 it is noted that more than 80% of Community events, Personal activities and public services have been affected and 67% of childcare service and educations have been closed.

The next step is to examine the mental health impact in various age groups. It is observed that young individuals are more vulnerable to anxiety and depression at this pandemic. From the graph it is noted that as the age increases the negative impact on mental health decreases.

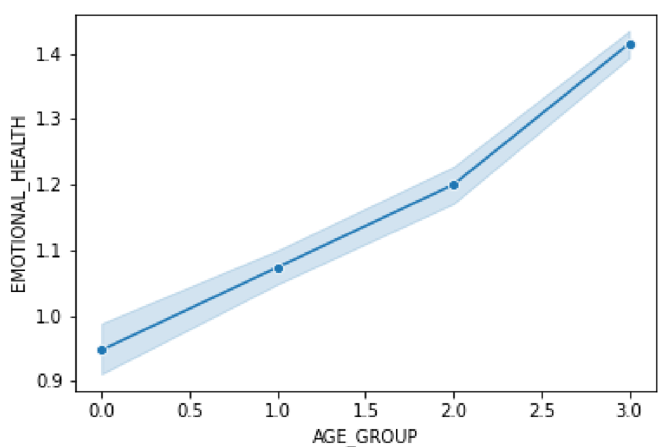


Fig 7: Emotional health Vs Age group

When the impact of mental health is studied with respect to number persons living in a household, it is observed that when the household size is 2, there is not much impact on mental health. For rest of the categories, the stability of the mental health fluctuates. For households which have more than 4 members, the variability becomes larger compared with the groups have less than 4.

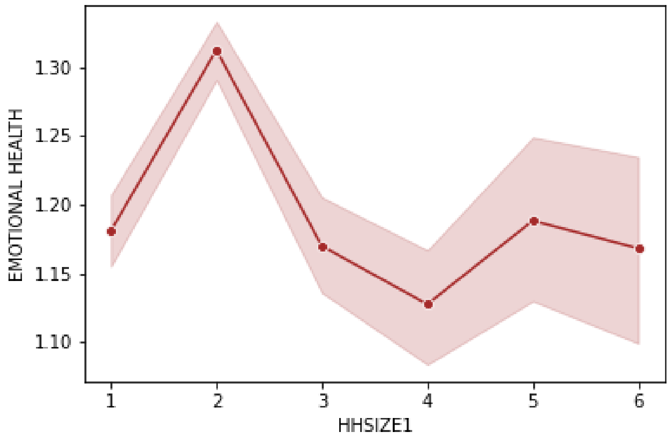


Fig 8: Emotional Health Vs HHsize

Thus, the data visualization concludes that many factors account for the impact of mental health on individuals. It is evident that people are struggling to adjust to the new lifestyle and has created a huge pressure to save oneself from COVID-19.

5 Supervised learning Algorithms

The mental health of the respondents is predicted with help of Sklearn classifiers.

5.1 Modeling

The following four models have been used for multi label classification.

- Logistic Regression
- SVC
- Gradient boost
- Naïve Bayes

The processed data is fed as input to the classifiers. The data is split in the ratio of 80:20 for training and testing. Each model is trained on train data and the cross-validation score is recorded. It is observed that gradient boosting classifier, SVC and logistic regression has given the best accuracy of 58%. The Naïve Bayes has given less accuracy when compared to other algorithms. This proves that Naïve Bayes is bad estimator with assumption of independent predictors. The cross validated accuracy for the models are given below

Models	Accuracy
Logistic Regression	58.9
Gradient Boost	58.7
SVC	58.16
Naïve Bayes	54.69

Table 2 : Supervised learning models and validation Accuracy

After recording the validation accuracy, the test data is fed into the classifier and the accuracy is recorded. The test accuracy for the classifiers is given below. It is observed from the results that logistic regression has given the least accuracy. The logistic regression model has overfitted on the training data and it performed worst on the test data. The reason is that the linear regression tends to over fit if the features are strongly corelated to each other. The other classifiers have given approximately the same accuracy as validation accuracy. Gradient boost classifier and SVC has worked well with data with giving the best accuracy. Collinearity is not an issue in Gradient boost classifier. If the features are strongly corelated, one of the features is arbitrarily chosen and split is applied. Hence it does not affect the model and works well with data. SVC works well by minimizing squared hinge loss and penalizing the size of bias.

Models	Accuracy
Logistic Regression	45.9
Gradient Boost	56.9
SVC	56.5
Naïve Bayes	54.3

Table 3 : Supervised learning models and test Accuracy

5.2 Variable Importance

The feature score of the Gradient boost classifier and logistic regression is computed. From the obtained output from gradient boost classifier, it is noticed that the variables Mental_condition(0.19), Age_group(0.19) and less_common_symptoms(0.127) have greater feature score when compared to other variables. These numbers denote that the current mental condition has a major impact in determining the mental wellness of the respondents. In addition to that age group and the fear of getting COVID-19 due to experiencing fewer common symptoms also has significant impact on mental health. The feature score obtained from logistic regression also implies that previous mental condition(1.044) is an important factor for determining mental wellness. The Government assistance and community assistance with score of 0.25 has an essential contribution towards the improvement of mental health. The score for Covid-19 positive is (-0.34) which indicates it is an important feature to assess the mental stability of the respondents.

5.3 Hyperparameter Tuning

The baseline was executed without any parameters. The classifiers are tuned with GridSearch CV to get the best score and best parameters. The accuracy score of the classifiers with 10-fold cross validation is recorded with best parameters are given in the table. A slight improvement in accuracy is observed after parameter tuning.

Classifier	Parameters	Value	Accuracy
Logistic Regression	C	10.0	58.88
	penalty	L2	
Gradient Boost	max_depth	3	59.47
	n_estimators	100	

	subsample	0.7	
SVC	C	100	58.3
	gamma	0.001	
	kernel	rbf	
Naïve Bayes	alpha	0.421	54.72

Table 4: Tuned parameters for Supervised learning models

6 Unsupervised learning algorithms

Apriori association is used in this report. With 33 columns in current dataset, the algorithms could generate thousands of association rules and many of which might no interesting. Hence, we divide the dataset into two set of criteria for evaluation by correlation and subjective arguments [18].

6.1 Association patterns based on correlation

We want to establish a set of objective interestingness measure to improve the quality of association rules. Based on the correlation graph which were produced in exploratory data analysis, we select below variables as our measurements:

- PHYSICAL_HEALTH
- MOST_COMMON_SYMPTOMS_EXPERIENCED
- SERIOUS_SYMPTOMS_EXPERIENCED
- LESS_COMMON_SYMPTOMS_EXPERIENCED
- EMOTIONAL HEALTH
- FOLOWED_OFFICIAL_GUIDELINES
- COMMUNITY_LOCATIONS_AND_EVENTS
- CHILDCARE_EDUCATION
- PUBLICS_SERVICE_AND_INFRASTRUCTURE
- PERSONAL_AND_SOCIAL_ACTIVITIES
- AGE4
- HHSIZE1
- NO_OF_CONDITION_AT_RISK

We later transform the dataset to a list of lists and check the Apriori rules with at least 70% confidence, 30% support. There are 590 association rules for this dataset with 63 frequent itemset that related to emotional health. Below table shows the maximal frequent item set and its rules related to mental health. With more than 1.1 lift value for listed rules, it is suggested that the antecedents and consequents are slightly positive correlation in these rules.

Table 5: Association rules of COVID-19 survey with 13 variables

Rule	Association Rule	Support	Confidence	Lift
367	{CHILDCARE_EDUCATION = YES, EMOTIONAL HEALTH = NEED ASSISTANCE} -> {PERSONAL_AND_SOCIAL_ACTIVITIES = YES, COMMUNITY_LOCATIONS_AND_EVENTS = YES}	31.3%	86.5%	1.2
145	{CHILDCARE_EDUCATION = YES, EMOTIONAL HEALTH = NEED ASSISTANCE} -> {PERSONAL_AND_SOCIAL_ACTIVITIES = YES}	32.7%	90.6%	1.1
148	{PERSONAL_AND_SOCIAL_ACTIVITIES = YES, EMOTIONAL HEALTH = NEED ASSISTANCE} -> {COMMUNITY_LOCATIONS_AND_EVENTS = YES}	40.3%	92.9%	1.1
361	{PUBLICS_SERVICE_AND_INFRASTRUCTURE = YES, SERIOUS_SYMPTOMS_EXPERIENCED = NO, EMOTIONAL HEALTH = NEED ASSISTANCE} -> {PERSONAL_AND_SOCIAL_ACTIVITIES = YES}	30.2%	89.6%	1.1

Above rules suggest that with 80 % minimum confidence 80% and 30% minimum support, the first rule suggests that most of the participants who experience mental health symptoms in 1 – 4 consecutive day(s) and have restriction in accessing school or childcare service are also have community events, personal and social activities affected. Additionally, the second rule suggests there is stronger confidence in association between limited access to educational service, participant's emotional health and how their personal and social plans affected. The last two rules suggest people who have no serious symptom of COVID-19 also need mental assistance under the changes of governor services.

In addition to this, below are three rules associated with mental health with highest support value in our dataset. We observe that theses rules have lower confidence and lift value compared with the rules in Table 5 for the same pairs of items. With lift values equal to 1.0, the antecedents and consequents are independent. By comparing between the rules in two tables, it is suggested that by having more than one emergency restrictions, people are more likely to develop mental issues.

Table 6: Association rules of COVID-19 survey with 13 variables

Rule	Association Rule	Support	Confidence	Lift
17	{EMOTIONAL HEALTH = NEED ASSISTANCE} -> {PERSONAL_AND_SOCIAL_ACTIVITIES = YES}	43.4%	84.0%	1.0
16	{EMOTIONAL HEALTH = NEED ASSISTANCE} -> {COMMUNITY_LOCATIONS_AND_EVENTS = YES}	45.4%	87.8%	1.0
18	{EMOTIONAL HEALTH = NEED ASSISTANCE} -> {PUBLICS_SERVICE_AND_INFRASTRUCTURE = YES}	44.2%	85.5%	1.0

6.2 Association patterns based on subjective arguments

In the rules based on correlation, the relationship represented may seem rather obvious. Hence, we eliminate 12 variables (except emotional health) in above rule and create a dataset with 20 remaining variables to find the unexpected association related to mental health. With at least 70% confidence, 30% support, Apriori algorithms generates 35745 association rules from this dataset with 1687 rules in relation to mental health.

Table 7: Association rules of COVID-19 survey with 21 variables

Rule	Association Rule	Support	Confidence	Lift
6303	{EMOTIONAL HEALTH = STABLE} -> {MENTAL _CONDITION = NO, COVERED BY INSURNANCE = YES, CONTACT WITH COVID POSITIVE = NO}	30.1%	85.6%	1.2
6197	{EMOTIONAL HEALTH = STABLE} -> {MENTAL _CONDITION = NO, COVERED BY INSURNANCE = YES, COVID POSITIVE = NO}	30.4%	86.5%	1.2
5383	{EMOTIONAL HEALTH = STABLE, CONTACT WITH COVID POSITIVE = NO} -> {FOOD/MONEY_SECURITY = NEVER TRUE, COVID POSITIVE = NO}	30.9%	89.8%	1.1

Above rules suggest that with 80 % minimum confidence 80% and 30% minimum support, the first two suggest that participants with no mental health problem, being covered by an insurance plan, neither have COVID-19 nor live with infected patients are more likely to cope with the sudden changes due to the coronavirus pandemic. The last rule suggests that participants without food insecurity and do not have COVID-19 are also have good mental health and most of the cases do not live in same household with infected case.

7 Conclusion

The analysis done in this project denotes that there is a significant impact in mental health due to COVID-19. The impact is due to many factors such as adaption to a new lifestyle, following new rules and regulations, changes in communication with family and friends and so on.

It is observed from the supervised learning methods that COVID-19 positive people are not only affected physically but also mentally. People with mental concerns are affected and require assistance. It is also found that younger generations are more mentally affected and people experiencing less common symptoms of COVID-19 are more anxious. They have a constant fear of testing positive, even though they are not experiencing serious symptoms of COVID-19. On the other hand, assistance provided by government and other communities have helped the people to stay optimistic, thereby improving the emotional wellness of the people.

Using association rule, we find that public restrictions could lead to mental illness. When limiting the variables to 21, the rule suggests that participants will not experience mental problem when they are covered by insurance plan and do not have mental health disorders. Food security also show stronger associated with good mental health along with the condition of living without infected people. We would like stress the importance of the mental health interventions to help those who are in risk. Increasing mental communications and supports for the vulnerable groups may be needed to mitigate the risks of mental health problems.

Our analysis has potential limitations as this was a survey conducted on April 2020, which was one month after the COVID-19 entered the United States. As the number of infected cases increases, some variables become less pronounced as it is in the report. Another limitation we want to address is the individuals who are ill and vulnerable are less likely to respond to the survey.

References

[1] The Associated Press (2020). COVID Impact Survey – Public Data. https://data.world/associatedpress/covid-impact-survey-public-data.

[2] Holmes, Emily A., et al. "Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science." The Lancet Psychiatry (2020).

[3] Wang, Cuiyan, et al. "A longitudinal study on the mental health of general population during the COVID-19 epidemic in China." Brain, behavior, and immunity (2020).

[4] Yao, H., Chen, J. H., & Xu, Y. F. (2020). Patients with mental health disorders in the COVID-19 epidemic. The Lancet Psychiatry, 7(4), e21.

[5] Brooks, S. K., Webster, R. K., Smith, L. E., Woodland, L., Wessely, S., Greenberg, N., & Rubin, G. J. (2020). The psychological impact of quarantine and how to reduce it: rapid review of the evidence. The Lancet.

[6] Gao, J., Zheng, P., Jia, Y., Chen, H., Mao, Y., Chen, S. & Dai, J. (2020). Mental health problems and social media exposure during COVID-19 outbreak. Plos one, 15(4), e0231924.

[7] Bruine de Bruin, Wändi. "Age differences in COVID-19 risk perceptions and mental health: Evidence from a national US survey conducted in March 2020." The Journals of Gerontology: Series B (2020).

[8]Armitage, R., & Nellums, L. B. (2020). COVID-19 and the consequences of isolating the elderly. The Lancet Public Health, 5(5), e256.

[9] Ford, T., Vizard, T., Sadler, K., McManus, S., Goodman, A., Merad, S. & Collinson, D. (2020). Data resource profile: the mental health of children and young people surveys (MHCYP). International Journal of Epidemiology.

[10] Collishaw, S. (2015). Annual research review: secular trends in child and adolescent mental health. Journal of Child Psychology and Psychiatry, 56(3), 370-393.

[10] Sellers, R., Warne, N., Pickles, A., Maughan, B., Thapar, A., & Collishaw, S. (2019). Cross-cohort change in adolescent outcomes for children with mental health problems. Journal of Child Psychology and Psychiatry, 60(7), 813-821.

[11] Li, Z., Ge, J., Yang, M., Feng, J., Qiao, M., Jiang, R., ... & Zhou, Q. (2020). Vicarious traumatization in the general public, members, and non-members of medical teams aiding in COVID-19 control. Brain, behavior, and immunity.

[12] CDC (2020). Doctor Visits and Getting Medicines. https://www.cdc.gov/coronavirus/2019-ncov/daily-life-coping/doctor-visits-medicine.html.

[13] CDC (2020). How to Protect Yourself & Others. https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html.

[14] CDC (2020). Communities, Schools, Workplaces, and Events. https://www.cdc.gov/coronavirus/2019-ncov/community.

[15] CDC (2020). What to Do If You Are Sick. https://www.cdc.gov/coronavirus/2019-ncov/if-you-are-sick/steps-when-sick.html.

[16] CDC (2020). Evidence used to update the list of underlying medical conditions that increase a person's risk of severe illness from COVID-19. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/evidence-table.html.

[17] NIH (2020). Coronavirus: What People with Cancer Should Know. https://www.cancer.gov/about-cancer/coronavirus/coronavirus-cancer-patient-information.

[18] Tan, P. N., Steinbach, M., & Kumar, V. (2016). Introduction to data mining. Pearson Education India.

[19] NORC (2020). COVID Impact Survey. https://www.norc.org/Research/Projects/Pages/covid-impact-survey.aspx.