Thesis
Depicting Mental Health problems through Social Media Analytics

Name: Bisma Arshad

Abstract

Mental health problems are considered as serious health issues all over the world because these problems can lead to severe outcomes like suicide, violence, abuse or murder. Majority of people with mental health problems are either untreated or undiagnosed. People who end up having schizophrenia, depression, bipolar or OCD for more than a year might undergo mental disability (Government Digital Service (GDS), 2014). There is a need to identify such problems before it gets worse. Fortunately, social media is as a platform where people express their feelings and problems openly. Data science can extract and analyze these statements to perform verity of functions for people suffering from mental health problems.

This dissertation contributes to the psychology field by providing a system which depicts the Twitter users who are potentially suffering/ suffered from depression, bipolar disorder and schizophrenia in US, UAE and Singapore. Different numbers of tweets are extracted from twitter using keywords and hashtags in these three countries. Snscraper was used to extract tweets from twitter. These tweets are then filtered out to people who convey signs of mental health problems only. Data is extracted from four timestamps within 2019-2022 to find the impact of Covid19 among twitter users. The dissertation includes all the methods, tools and techniques used in the project to achieve its aim and objectives.

Chapter 1

1.1 Introduction

We are living in the age of digital communication where everything is surrounded by technologies. Social media is emerging as a platform for digital community. One of the social media platforms called Twitter is the new way of online communication which uses micro blogging to share messages with the audience. The information that the user posts can be

spread within seconds because of the follower-following relationships. It provides user to share message within the limit of 280 characters which was extended from 140 characters. According to the world statistics by (Dean, 2022) Twitter has currently 396.5 million user profiles for which 206 million users use twitter on daily basses. Most of the users fall between the age group of 25-34 with a percentage of 38.5%, and 20.7% users are of age 35 to 49. It is ranked amongst the top social media, 8.85% of the world's population uses Twitter from which the average of men users are higher than women users. In year 2020 twitter was downloaded more than 6 million times in the world. It is used most in US with 73 million users. Along with this, people not only use twitter as a source of entertainment but also as a source of information. Almost 48% of twitter users in US use twitter as a source of news and for entertainment. Recently 2 top trending hashtags on twitter were #Covid19 and #Stayhome, these types of hashtages keeps the twitter community up to date on the trends and news. Hashtags are also used in other social media platforms like Facebook and Instagram.

Nowadays, technology is playing an increasing role in the medical field. In the past few years, machine learning and data mining emerged with new techniques for detecting, analyzing and visualizing data. In which, social media facilitate machine learning models to classify or predict mental health problems. Although mental health is a complex and comprehensive issue, but it's also revealed that people suffering from mental health problems tends to prefer sharing their health issues through social media to get help and concerns (Kim, Lee, & Park, 2021). Data for machine learning models can be pulled together by collecting and tracking activities of users who gave public access to their information. Data of private accounts cannot be accessed. The collected raw data is used to predict the type of mental health problem a person is suffering from. To do so, (Mostafa, Ahmed, & Junayed, 2021) used supervised machine learning classifiers to the data that includes SVM, logistic regression, KNN, Naïve Bayes and Neural Network Classifier. According to (Shybeko, 2021) predictive models are capable of improving diagnoses or providing better medical care to patients. These models can help in the fields of health insurance, mental health, medical Imaging, palliative care and pharmacy service.

According to (Mental Health Conditions, 2022) mental health problem is a state of mind in which a person's feelings, mood, behavior and thoughts are affected which can impact the person's living style and any other human beings related to that person. However, 75% of times, people can begin to face mental health problems by the age of 25 and 50% by age 14. Also, a mental illness is not a condition that occurs unexpectedly, psychologists suggest that it can be caused by environment, lifestyle of a person or genetics. Classification of a mental health problem requires specific diagnosis from a psychiatrist by undergoing few steps. It's hard to identify someone with the type of mental illness they are suffering from through online platforms. This was the reason why this dissertation is visualization rather than predictive analysis.

This dissertation intends to depict the prevalence of depression, schizophrenia and bipolar disorder among twitter users by analyzing text from social media (Twitter). The goal for this project is to explore and visualize the data that can provide statistical views of three mental health problems in United States, UAE and Singapore. These statistics may differ from actual rate of Schizophrenia, Depression or bipolar disorder in a country but this must be highlighted that the dissertation is focused on only those people who declared having schizophrenia, depression or bipolar disorder on Twitter.

Dissertation is focused on:

Social Media:	Twitter
Users:	Public users
Tweets:	Self diagnosed tweets
Mental health topics:	Schizophrenia, Depression & Bipolar disorder
Language:	English
Time:	3 months from each year (2019-22)

Expected results are to obtain useful statistical visualization of Depression, Bipolar disorder and Schizophrenia in three countries, US, UAE and Singapore. The visualization will present the analyzed tweets in a user-friendly format which will be useful to psychology

staff and students. This will be beneficial for psychologists and psychology students to gain relevant information regarding mental health problems, like how people online express their mental state and symptoms of any mental health problem. Furthermore, the final dissertation anticipates the impact of Covid19 on mental health of people. This is done by analyzing tweets within specific time frames, during Covid19 waves. For example, during covid19 pandemic, number of U.S Adults facing stress, depression, anxiety and other disorders were increased because they were experiencing sadness, loneliness and fear due to social isolation, financial pressure and altered daily routines. (Staff, 2021).

1.1.1 Aims

The aim of this project is to have a comparison of different mental health problems in three countries US, UAE and Singapore. Tweets will be extracted to perform sentiment analysis and other techniques for visualizing Schizophrenia, Depression and Bipolar disorder faced by people who express themselves on Twitter. In addition, its aim is to uncover the impact of Covid19 on people suffering from these three mental health problems. **1.1.2 Objectives**

Objectives of this project are:

Collecting data

- Extract more than 10,000 tweets in English language related to depression, schizophrenia and bipolar disorder taken only from US, UAE and Singapore within four time frames with three months of period. One before Covid19, and three during Covid19 peak waves. These peak are manually selected. This is done by using snscrape.

Pre-processing data

- Apply cleaning techniques to the dataset using NLTK preprocessor for removal of urls, usernames, special characters, numbers and to perform word tokenization, stemming and lemmatization .

Exploratory Data Analysis

- Analyze and visualize the raw data to gain knowledge on valuable information. WordCloud is used to display all the words in the dataset in different font sizes according to their significance.

Sentiment Analysis

- Perform sentiment analysis on the dataset to obtain the sentiment of each tweet, whether the tweet is negative, neutral or positive.

Visualization

- Compare, contrast and visualize Depression, schizophrenia and bipolar disorder in US, UAE and Singapore to give clearing understanding of the data collected from Twitter to psychologists and psychology students.

Key Words: Twitter, Tweets, Mental Health, Depression, Schizophrenia, Bipolar Disorder, Data Visualization.

1.1.3 Motivation

Psychology fitness of people is a trending topic these days. This is because in USA in year 2020, suicide was the 9th leading cause of death among people ages 10-64 years and 2nd leading cause of death among people ages 10-14 years and 25-34 years. Apart from this statistics, the intention of suicides were even higher than committed suicides. (USA.gov, 2022). A person going through Depression or anxiety for too long may cause intentions of suicide. Famous celebrities like Kate Spade, Robin Williams and Sushant Sing Rajput also committed suicide. Few individuals suffering from mental health problems don't talk about their feelings freely, it's important to identify such problems and spread awareness accordingly so that people going through mental health problems may find help without asking.

This dissertation is focusing only on three mental health problems Depression, Bipolar disorder and Schizophrenia, taken from three different types of groups. Very familiar problem called depression is included because this problem is commonly found in surroundings and mostly ignored by society. Depression is common because most of the mental health problem either have depression as symptom or they all end up as depression.

For example, at some point, people suffering with stress and anxiety ends up having depression as well. Other two mental health illness are Schizophrenia and bipolar disorder. The purpose for choosing these two problems was that they are the leading cause for mental disabilities and are very common in people especially in US (Avenaim, 2021). People have a lesser amount of knowledge about Schizophrenia and bipolar disorder compared to depression. Also, very few machine learning models are created for Schizophrenia and Bipolar disorder.

Chapter 2

2.1 Literature Review

This chapter explains the background and main ideas of the project with respect to some related work to support the research.

2.1.1 Background

Social media is a popular platform for people to exchange thoughts, it made life of individuals easy in a sense that they can talk or express themselves freely. One of the best advantages of social media is that it has given people an opportunity to openly discuss about their mental health. In return, people do get support but not necessary enough to solve a mental health problem. Looking back few years ago, mental health problems was rarely a topic for discussion. Although mental illnesses were common in people that time but because of limited data and lack of information, hardly anyone took it seriously. In fact back in illiteracy time in late 1600s, people suffering from mental health were considered possessed by demons and were forced to perform exorcisms, even after that they were either imprisoned or executed (BATON ROUGE BEHAVIORAL HOSPITAL , 2022). Eventually by the 18th century, people get to know mental health a little and mental asylums were built but patients there received no treatments. People though that mental illness is a permanent mental disability. Today, mental health is taken as a serious health problem which is treated by psychiatrists in hospitals and clinics. (OpenStax, 2022)

There are several machine learning models with different techniques and classifiers which extracts user's online social media posts regarding various illnesses like PTSD, stress, anxiety, or even disorders like eating disorder, sleep deprivation or Seasonal Affective Disorder. Recently many machine learning models are developed using data related to covid19. There are models that can also predict person's mood on social media (Coppersmith, Dredze, & Harman, 2014).

This dissertation is based on a project that performs an analysis on tweets, extracted from twitter that will provide relevant information about number of people suffering/ suffered from Depression, Schizophrenia and Bipolar disorder. The tweets will be extracted on four specific time periods in year 2019 to 2022 where Covid-19 was on peak; the time period will be used to analyze the impact of Covid19 on people. It will also identify the number of people who are aware of these three mental health problems and number of people who actually said they have / had these problems. These statistics can contribute in preventing serious outcomes like suicide or mental disability which can be done with the help of government in spreading awareness and free facilities in countries where rate of such mental health problems are high. The statistics can be used by psychologist for educational purpose as it provides an insight of social media tweets where expressed feelings, words and symptoms of people having schizophrenia, depression and bipolar disorders are portrayed. It can also be used in giving better medical services or better diagnoses. These steps are beyond the scope of the project and can be added as future work.

2.1.2 Depicting Mental Health on Social Media

Several studies and researches are found related to the topic "mental health problems" with help of social media. Following section highlights the work of few researchers on relation of mental health and social media. One of the common social media sites is twitter. People expressed that Twitter is a platform where they can socialize and connect with everyone. People are interested in Twitter because they feel it's a platform where every user express themselves freely and they can interact with any other user with shared understandings.

(Berry, Lobban, Belousov, Emsley, Nenadic, & al., 2017) Have explored some reasons of why people use Twitter to convey their mental health problems. They analyzed that people speak out loud about their health or other problems without restraints, some users support them by expressing that they are going through equivalent pain and other users support them by spreading hope and peace. Few public tweets from their finding are listed:

"If it helps even one person recognize they aren't alone in their pain" ***Twitter***

"Suicide might be complicated but extending a hand to someone is simple and it may save their life." ***Twitter***

"Perhaps, my tweets and experiences may help others, even if it's only one person." ***Twitter***

They noted that due to open resources and information provided on Twitter, users were able to find materials related to their problems which worked as self directed psychology education. This not only helped in getting clinical aid but also helped them to feel more relaxed and calm.

Additionally, talking about social media posts, people use informal ways to communicate or to express their feelings. The way of writing personal posts is different than normal communication. People use different phrases, abbreviations and emojis to convey their message to others. The process of converting slangs or abbreviations into proper input data for models is enlightened by (Kumar, 2020) to classify user's perspective by means of tweets. According to him, this can be done by using few cleaning techniques to make the data in a readable form for the model.

Tweets contains special characters, numbers, hashtags, URLs and repeating words which are incomprehensible by the machine learning models. Few techniques and steps for cleaning the data according to (Shah, 2020) are

- Using preprocessor for removal of URLs, @mentions, reserved words and emojis.
- Using NLTK for ○ Removal of numbers, punctuations and converting text into lower case ○ Tokenization, built in TweetTokenizer() can be used
    - ○ Converting words to its stem or root words by Stemming and Lemmatization ○ Removal of stop words by pre-defined English stop words list

Comparison on these researches is shown below in a table where each author is compared with others according to the social media platform they used, is the research was based on a mental health and was the research done on a specific illness.

Authors	**Research about**	**Social media used**	**Based on mental health**	**Specific mental health problem**
(Berry, Lobban,	Proves why people use Twitter	Twitter	Somehow	No
Belousov, Emsley, Nenadic, & al., 2017)	to express themselves			
(Kumar, 2020)	Describes how to use Twitter data for modeling	Twitter	No	No
(Weerasinghe, Kediel, & Rachel, 2019)	Early detection of mental health problems	Twitter	Yes	No
(Caroline, Seren, Jiawen, Bilal, Marc, & Felicitas, 2021)	Predicts depression rate in Germany during COVID 19	Twitter	Yes	Yes
(Steven Gangloff, 2021)	Detect which platform had more information about Epilepsy	Twitter & Instagram	Yes	Yes

Table 1: List of authors and their researches

2.1.3 Significance

We are living in the era where inhabitants are fighting to be acceptable by the society; this competition is draining their brains to a point where it results in malfunctioning of mind. The thing with mental health illness is that it needs to be trenched till root cause whereas physical illnesses are usually easily diagnosed via scans, X-rays, ultrasounds etc. Detecting a metal health problem before it gets worse can prevent from grater loss.

If it is possible to analyze the approximate number of people with mental health problems in a country, it will be easier to perform or conduct events according to the severalty of problems faced by people. In most countries, people feel shame to consult a psychiatrist and in majority cases people don't even know they have such type of illness.

(Tiller, 2013) Stated that almost every third patient with depression, anxiety or stress disorder do not look for treatment. And less than half patients are proposed with beneficial treatment. He also suggests that we should raise more public awareness and educate people with such mental disorders; this may result in better outcomes.

By considering the importance of mental health problems, three mental illnesses are taken into account, named as Depression, Schizophrenia and Bipolar disorder. This dissertation is a comparative study which mainly focused on these three types of mental health problems.

2.1.4 Depression

Depression is a mental health disorder that has severe effects on a person's life. Considering ICD-11 (Maj, 2020), a person having at least 5 out of 10 symptoms of depression that occurs nearly every day for at least 2 consecutive weeks is considered depression. These 10 symptoms are depressive mood, poor concentration and attention, hopelessness for their future, low interest or pleasure in daily activities, low self-worth or inappropriate guilt, thoughts of death or suicide or attempt of suicide, disturbed or excessive sleep, change in appetite or weight, fatigue or reduced energy and psychomotor agitation or retardation. ICD-11 classifies depression as mild, moderate or severe. This depends on the intensity of symptoms and degree of functional impairment.

Depression can result in suicides. (Smith, 2014) Stated that one of the dangerous disease listed according to death and disability, depression is ranked 9th in the list after Heart disease, HIV and stroke. Not every mental health problem is deadly but almost every problem leads to Depression. According to (Organization, 2022)Depression is a term use for mental disorder. A person suffering from depression goes through sadness and lack of interest. It can also cause loss of appetite and sleep which can result in poor concentration and tiredness. It is also stated that the average number of women having depression is higher than men. Problems like anxiety and stress can lead to depression as well. A study proves that 39% people that are going through anxiety are likely to have symptoms for depression also. Depression has stages according to which one can be cured. A person's last

stage of depression can result in suicide or mental disability. According to (Organization, 2022)it is recorded that 75% of people suffering from depression are untreated in low and middle income countries or in under developing countries. In an article by (Tiller, 2013) one of the best treatment suggested for depression disorder is cognitive behavior therapy (CBT).

2.1.5 Schizophrenia

The second mental health disorder taken into account is called Schizophrenia. Schizophrenia is considered a serious mental health problem that effects on person's behaviors, their thinking and how they feel. The cause of this disease is still unknown but (WHO, 2022)states that environmental factors or interaction between genes may cause schizophrenia. According to DSM-5 and ICD-11 a person is diagnosed with schizophrenia when at least 2 of 7 symptoms (at least 1 core symptom) occur most of the time for minimum 1 month. Core symptoms include hallucinations, passivity or control, delusion, experience of influence and disorganized thinking. (Gaebel, 2020)

Symptoms of schizophrenia can be categorizes as negative and positive. Hallucination and delusion comes under positive symptoms which is build up commonly in age group 16 to 30. Negative symptoms can be lack of speech or emotion, these symptoms can be sometimes misdiagnosed as depression. Like depression, schizophrenia is also found in a number of societies at similar rates. Although the risks of this disorder is no less than any other disease still we can find a lack of discussion on this topic (Stanford University, 2015). By look at the statistics according to (Stanford University, 2015) it is observe that people suffering from schizophrenia may have higher risk of suicides, about 6% of people are incarceration and other 6% of people are experiencing homelessness.

2.1.6 Bipolar disorder

Bipolar disorder is the third mental health problem which is chosen. This is a disorder which effects on a person's mood, or ability to function. A normal person can experience mood swings which last only few hours but people with bipolar disorder experience intense emotional mood swings which have a period of days to weeks. People suffering from bipolar disorder may also have anxiety disorder or depression. (Association, 2022)

The ICD-11 reflects bipolar disorder as mood disorder. It considers a person with 1 or more manic or mixed episodes as bipolar I disorder. ICD-11 requires duration of minimum 1 week for mania. Bipolar disorder is classified as bipolar type I, II, III and IV. Each person with a bipolar disorder is categorized according to their mood episodes. The four types of mood episodes are mania, hypomania, depression and mixed episodes. (Michael Berk, 2020)

As twitter allows user to communicate with each other, it is also used as an emerging tool for mental health research. Despite of this advanced feature extraction method, the conditions like bipolar or seasonal affective disorder (SAD) are rarely considered because diagnoses are not done on basis of texts only. Bipolar disorder can be treated by medication, psychotherapy or talk therapy depending on the type of bipolar disorder.

2.1.7 Methods

Every problem is different in identification and in handling. Even their diagnoses, stages and treatments are dissimilar to each other. It's tough to pull out data from any social media in a single way. Few mental health problems can be diagnosed by looking at the patterns or signs in text. On the other hand, some problems are hard to categorize or notice. There are number of techniques that can be used to detect mental health problems like

- **Time of posting**: Use to predict problems similar to Seasonal Affective Disorder.
- **Emotions**: Use to predict mental health problems like bipolar or anxiety disorder
- **Feelings**: Help in predicting illnesses along with emotions. Eating disorder or sleep deprivation can be detected with expressed feelings or moods.

Methods used in this project are:

- **Medical words**: Use of words related to medical or symptoms of a problem like stroke, stress, depression, suicide, clinical, medicine, etc are used to extract most of the tweets related to specific problems. Words are used to classify mental health problems. In this project, tweets are extracted using three words Depression, Schizophrenia and Bipolar. This extracted only those tweets which contain these words only.

- **Hashtags** : Users use hashtags in their posts to describe a specific topic. This is used in spreading awareness about certain topics around the world. In the project, user's who self diagnose a problem uses hashtags to state their problems are extracted.

  Along with filtering medical words, three hashtags Depression, Schizophrenia and Bipolar were also added to extract tweets.

- **Self diagnosis**: People declare and confirm a diagnosed problem by themselves using the problem's name. This helps out in extracting mental problems for visualization rather than predicting. In the project, after extracting data, all tweets were further filtered out using first person pronoun and sentences like 'I had bipolar', 'I am suffering from Schizophrenia or 'I am going through depression'.

These methods are used with Snscraper to extract tweets from 2019 to 2022. Tweepy can also be used to collect tweets but Twitter API has limited access to only 1 week old tweets. After extracting tweets using words and hashtags, further preprocessing is done for cleaning purpose. Once all numeric values, special characters, urls, hashtags and emojis are removed, file is checked for null values and duplicates. All null or duplicates are dropped for advance filtering process. This new dataset is filtered out using first person pronouns to check self diagnosed tweets only, details of why we use first person pronouns are discussed in section 2.1.8 Psychology patterns. The filtered dataset is tokenized using stemming and lemmatization process. Each tweet goes through textblob and vader for text polarity. The tokens are converted to vectors to feed the sentiment analysis model. Count, tfidf and hashing vectorizer are used to convert tokens into vectors; these vectors are used for sentiment analysis. The sentiment analysis provides the results of how negative, neutral or positive the tweet is. These tweets along with their polarity are used for visualization.

2.1.8 Psychology patterns

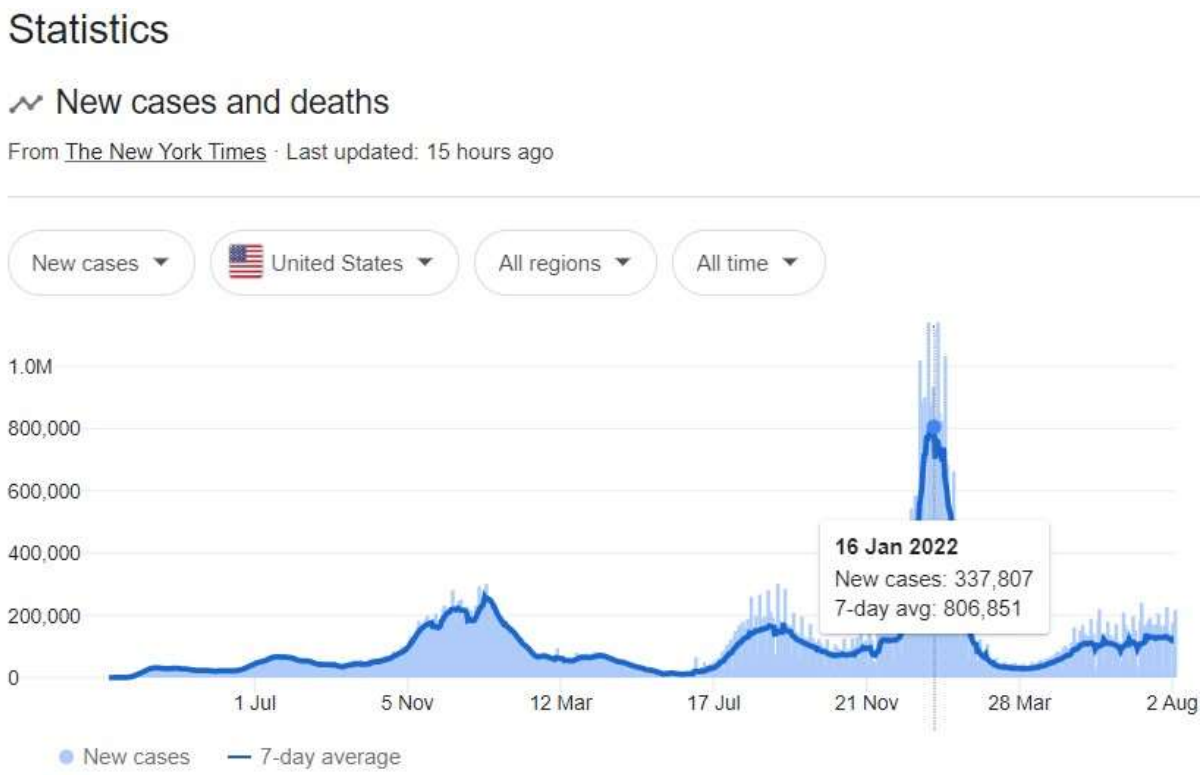
Social media is a source of information by people all around the world in real time. People use these platforms to express their inner feelings. This made it easy to spot out different disorders by looking and analyzing person's writing style. According to (KALE,

2015), people suffering from mental health issues tend to use clinical words in their posts which are uncommon and unfamiliar to normal person. According to his research, first person singular pronouns like “me”,”my”, “I” are linked to person’s level of mental illness. It is also mentioned that users who use first person pronouns in their posts are more likely to be suffering from mental illness than users who use plural pronouns like “we” or ”us”.

A function is created to go through each tweet and filter tweets related to first person pronouns. This step will factorize all the tweets that are linked with the user itself. All the tweets that are related to third person will be eliminated. For e.g. “she was diagnosed with depression” will be eliminated. To run this on the dataset, first person pronouns or self related words list are considered which contains words like I’m, I’ve, I’ll, I’d, me, my, myself, mine etc.

2.1.9 Data visualization

Data extracted on the basis of keywords, self-diagnosed words or psychological patterns are not enough for better visualization, work is required on comparing and creating relations in data. According to (shanthababu, 2021) Data visualization is a process of displaying complex data with graphical representation which allows viewers to indentify and understand the insights of the dataset. Data visualization techniques help to attain the main aim and objectives of this project as it assist in aspects like story telling of the dataset which explains the evolution of mental health problems in people during Covid19. Tweets are extracted in 3 months time where Covid-19 was on peak. These peaks were taken manually from global Covid-19 chart on Google presented by New York Times.



From four years, four highest peaks were taken into account. Along with Covid-19 cases in US, UAE and Singapore, peak of deaths due to Covid-19 were also taken for visualization. Details of dataset in is section 3.2.1 Datasets. The patterns in the dataset are identified by using different graphs and charts. All the visualization is done on the bases of self diagnosed tweets where people declared having/had depression, schizophrenia or bipolar disorder. List of graphical representation techniques used in the project are

- Line chart using matplotlib.pyplot for displaying number of people suffering/ed from mental health problems is US, UAE and Singapore.
- Pie chart using matplotlib.pyplot for comparing total tweets extracted related to 3 mental health problems with filtered tweets of people actually having/had mental health problem.
- Bar chart using matplotlib.pyplot for comparing results of textblob and vader. Also, for comparing negative, neutral and positive polarity of each datasets.
- Area plot using matplotlib.pyplot for displaying Covid-19 peaks. Also, for comparing three mental health problems with Covid-19 deaths.

- Scatter plot using matplotlib.pyplot to display number of filtered tweets of each country in 4 years.
- WordCloud using matplotlib.pyplot to display frequent words in dataset of each mental health problem in each country. Wordcloud is also used to display frequent words of tweets of specific users.
- Donut chart using matplotlib.pyplot to compare each mental health problem in a country.
- Word frequency plot using gensim model to display a descriptive word frequency count of each dataset. This is used to display word count of tweets of specific users to get better understanding of symptoms of users having/had any mental health problem.

2.1.10 Evaluation

The evaluation of the project is done by getting a feedback. The feedback is collected from psychiatrist, psychology students and few data science students from Herriot-Watt University Dubai. All the final results of comparisons and visualization were displayed in a pdf file and a survey was created related to the results of the project for evaluation, to avoid PLES issues, no dataset or any other file was shared. The survey consists of 11 small questions regarding the project's findings where each person will rank according to the worth of information provided. The answers are rated on the bases of how useful the results are for psychology department, or how valuable the visualized information is.

2.1.10.1 Results of Evaluation

The online survey was collected from a psychiatrist, two data science students and few psychology students. Results of the survey are shown below. The survey is presented in the appendix of the report.

Question 1:

Question 2:

Question 3:

Question 4:

Question 5:

Question 6:

Question 7:

Question 8:

Question 9:

Question 10:

Question 11:

2.1.11 Limitations

Few limitations of the project are

- Only English language tweets are considered, apart from that no other language will be used, this will limit the selection of the countries.
- False positive and false negative will be high as it's not a face to face conversation but just some text data. To overcome this, few techniques will be used but still hardly any sarcasm will be detected.
- Many people who are suffering from a mental health don't post anything. The research will be done on those people only who posted about their mental health problems or anything related to it on Twitter.
- People in underdeveloped countries which are facing similar issues will be neglected because there will be not enough data to analyze.
- Limited time is allocated for the project in which only few countries and few mental health problems were taken into account.

Chapter 3

3.1 Requirements Analysis

Data Science projects are a bit short of requirement analysis. In this chapter all the mandatory and optional requirements are displayed along with functional and non functional requirements. This chapter also highlights the methodology of the project.

3.1.1MoSCoW

- MoSCoW explains what must be done, should be done, could be done and would not be done in the project.

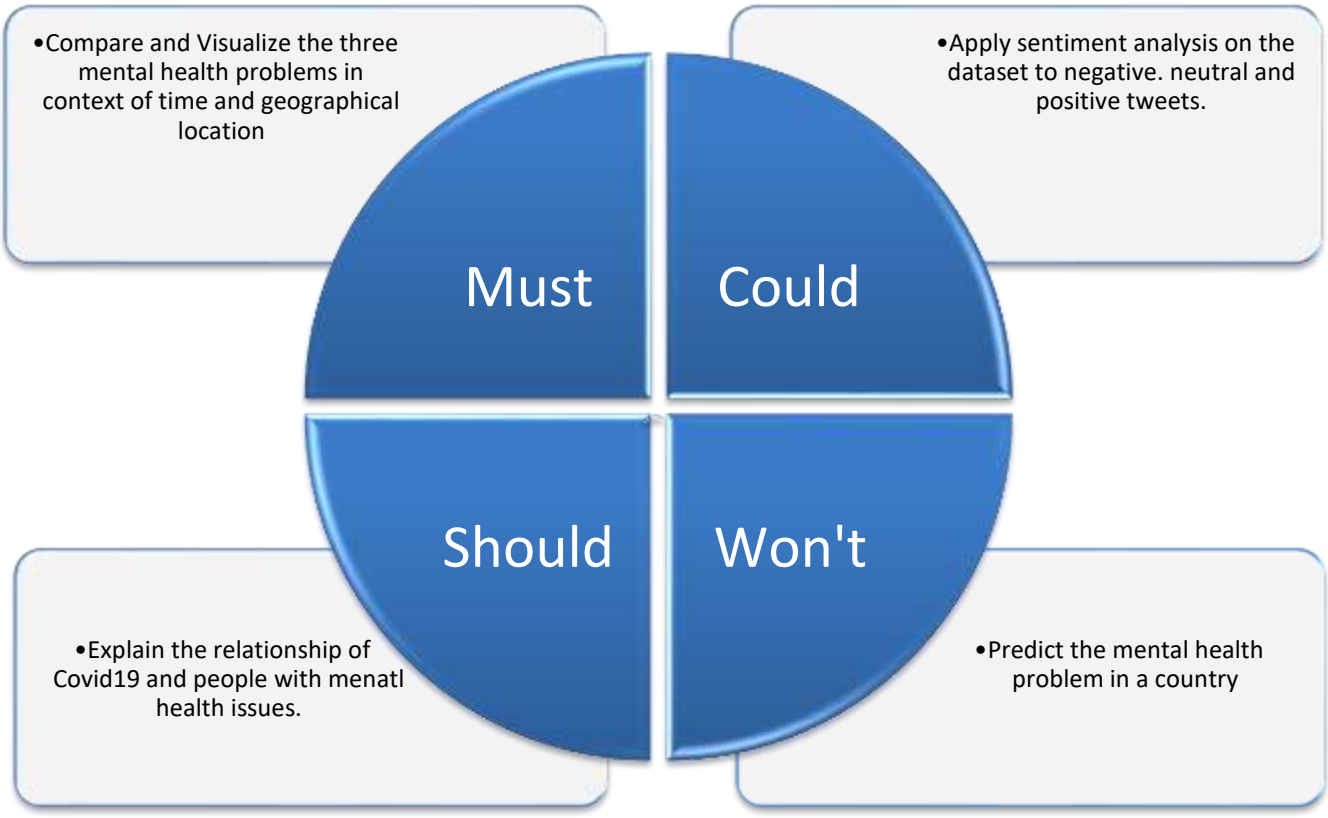


Figure 1: MoSCoW

3.1.2 Functional requirements

Requirements	Description	methods
Usability	Data visualized should be understandable by the users	2 files are created, one for tweet extraction and one for visualization to avoid jumbled code and ease to the user.
Data Visualization	Compare and visualize the three mental health problems in a user friendly way so that essentials of all datasets are conveyed properly.	Different types of charts and graphs are used to display information of datasets. To see more details on data visualization, see section 2.1.9 Data visualization
Flaunt tweets	Present symptoms or words of specific users who declared having/had any mental health problem by extracting almost 500 of their tweets.	This is done by using wordcloud which display most used words in a cloud and pyLDAvis to display detailed graph of word count of each dataset.

Table 2: Functional Requirements

3.1.3 Non Functional requirements

Requirements	Description	Methods
Data Security	Data should be protected and not shared with anyone	Snapshots of results will be shared for evaluation.
Performance	High system performance while running	Different files for visualization and tweets extraction to reduce running load.
Data integrity	Completeness, accuracy and consistency of data	Restricted access for editing any project file.

Table 3: Non Functional Requirements

3.2 Methodology

The project is designed on Juypter notebook using python v3.8 programming language. A library called snscrape is used to extract tweets from twitter. Tweepy was first planned to use for tweet extraction. But after getting approval and API from twitter, the access of tweepy was limited to extract one week old tweets only.

```
# Authenticate with twitter portal
auth = tw.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tw.API(auth, wait_on_rate_limit=True)
```

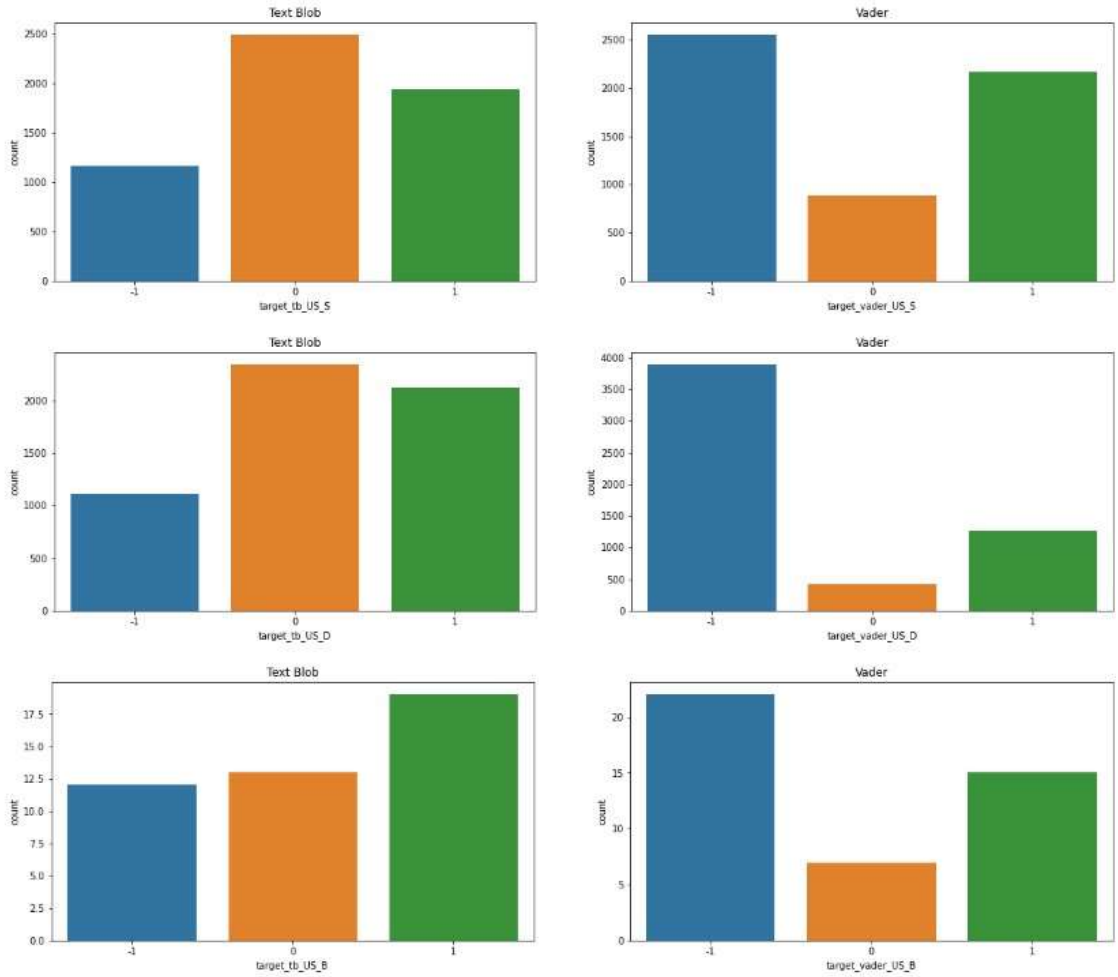
Screenshot 1: Access Twitter with Twitter API

This was the reason to use Snscraper which allows extracting tweets from any year.

```
    myfile = writer(file)
for i, tweet in enumerate(sntwitter.TwitterSearchScraper(search_words + date).get_items()):
    if i > 10000:
        break
```

Screenshot 2: Tweet extraction using Snscraper

Preprocessor is used to perform data cleaning which removes all the hashtags, urls, numbers, emojis and special characters from the dataset, it also converts all the words to lowercase. Preprocessed data is manually filtered for tweets of people having/had mental health problem only. TextBlob and Vader are used to give polarity to the tweets. The comparison of both polarity results are shown in a bar chart.



This shows that Vader gives better polarity to text as number of neutral tweets is less than TextBlob results. StopWords from NLTK are used for Tokenization and stemming. Tokens of each file is created and saved for visualization in the form of wordcloud. Also these tokens are converted into vectors for sentiment analysis. Three types of vectors are used count vectorizer, tfidf vectorizer and hashing vectorizer. Each vectorizer is fed to three classifiers Naïve Bayes, KNN, Decision tree and Random forest along with vader polarity. Data visualizations are done on countries, Covid19 cases, Covid19 deaths, mental health problems and posts of specific users. More details on visualization part are in section 2.1.9 Data visualization.

3.2.1 Datasets

No pre defined dataset is used in the project; the dataset is gathered by extracting tweets from twitter for three mental health problems named as Schizophrenia, depression and bipolar disorder. One way to extract tweets is by using twitter API. To get Twitter API you have to apply for a request. Once done, Tweepy is used in Python to extract all desired tweets. Twitter API allows to access data from only past 1 week, in order to get tweets from months or years ago paid premium twitter API can be used. An alternate way to access old tweets is to use a library called snscrape. It requires Python version 3.8 or higher to run. Snscrape can be used directly from CLI or else Python's OS library can be used in Jupyter notebook to execute CLI commands (Beck, 2020). Snscraper was used to extract all the tweets in the project. The mental health topics for the project depression, schizophrenia and bipolar disorder are taken from US, UAE and Singapore. Tweets were extracted using keywords and hashtags of depression, schizophrenia and bipolar from 2019 to 2022.

Four time frames of three month period are considered to extract the data; 1st time frame is March-April-May of 2019 where there was no Covid-19, 2nd and 3rd time frames are taken from January-February-March of 2020 and 2021 where Covid19 wave was on peak and 4th time frame is taken from January-February-March of 2022 where hard times started to decline. The dataset contains id number, tweets, time, location, usernames and labels. All other information like age, reposts, likes or account name is not extracted. Data related to these three mental health problems are pulled out using words and hashtags. This dataset is further downsized by eliminating tweets that does not follow or fall under the psychology patters. Psychologists have found patters in people's tweets who suffer from mental health problems. See 2.1.8 Psychology patterns under literature review for more details. The table below gives the summary of tweets extraction from every country.

Country	Mental Health problem	Total tweets extracted	Total filtered tweets
US	Depression	13,394	5,579
	Schizophrenia	19,002	5,593

	Bipolar disorder	13,008	4,921
UAE	Depression	147	57
	Schizophrenia	71	18
	Bipolar disorder	53	20
Singapore	Depression	84	33
	Schizophrenia	88	24
	Bipolar disorder	68	19

Table 4: Tweets extracted

Another dataset was created using tweets of 6 users from each US dataset of depression, schizophrenia and bipolar who declared having/had mental health problems. This was done to identify the symptoms of these users by collecting maximum of 500 tweets each. These new tweets went through preprocessing and tokenization process for visualization.

Mental health problem	No of users	Tweets per user	Tweets extracted
Schizophrenia	6	500	1,666
Depression	6	500	2,523
Bipolar	6	500	2,453

3.2.2 MS Excel

There are total of 48 excel files in which each edited dataframe is stored separately. Firstly 9 original excel files were created to store extracted tweets for three mental health topic in three countries; files were concatenated afterwards according to visualization need. The files are then cleaned and saved to different excel files. Filtered files are also saved in separate excel files. Apart from this, all lists, tweets of specific users and dataset with urls are saved in csv formats. These files are not shared or accessed by anyone else except the researcher.

3.2.3 Preprocessing and analyzing data

Few steps, techniques and algorithms done before visualizing the mental health problems used in this project are:

- Use dependencies numpy, pandas, nltk, regex, seaborn, wordcloud, pyLDAvis, gensim and matplotlib.
- Steps like removing stopwords, emojis, numeric values or special characters are performed next to converting sentence to lowercase. Along with this, all the URL's and repeating words are removed which do not have much significant value. All null values and duplicates are dropped. Once all this is done, further Tokenization is done by Stemming and Lemmatization process for reducing the words to their stem or root words (for e.g. "hated" to "hate" and "loving" to "love"), this was done by using PorterStemmer.
- Most common words in a dataset can be viewed with WordCloud, this is done to observe the significance of the words. Frequency of word count is displayed using pyLDAvis.
- Tokenized text is transformed into vectors with TF-IDF vectorizer, hashing vectorizer and count vectorizer word. Sometimes reduction of dimensionality data is used to visualize the enhanced data.
- For sentiment analysis, 4 classifiers are trained Naïve Bayes, KNN, Decision tree and Random Forest classifiers, one by one with output of three vectorizor. This is done to obtain negative, neutral and positive sentiment of tweets.
- Data visualization is done by using charts and graphs like bar charts, line charts, pie charts, area plot, scatter plot, wordcloud and word count graph. Details on data visualization are presented in section 2.1.9 Data visualization

Chapter 4

4.1 Professional issues

- Every step of the project is documented clearly
- All terms and conditions of Twitter are accepted
- Data is extracted according to Twitter's policy
- Anything that is used in the project have proper licensing and documentation

- Information for this research that is taken from journals, articles and websites are cited properly in Harvard style.

4.2 Legal issues

There is no legal issue faced in the project.

4.3 Ethical issues

- Twitter data is not used for personal means and is not shared with anyone.
- Only snapshots of visualization are shown in pdf file for evaluation. Raw data is not displayed or redistributed to anyone.
- Data is secured throughout the project using a restricted folder for all the files and will be removed after project is done.
- No sensitive or personal information of user is exploited while working on the project. All the Twitter data is publically available. Data from private accounts are not collected.
- User's information like Twitter username, URL, followers, post likes and age is explicitly eliminated.

4.4 Social issues

- The research or Twitter data will not be published anywhere.
- No survey of twitter users will be conducted. No contact will be made to any user of the post.

Chapter 5

5.1 Project Plan

The key to start a project is project plan. This chapter consists of Gantt chart which displays the tasks, their allocated time and project division. Risk analysis explains each risk that may be faced during the project.

5.1.1 Flow chart

A flow chart describes the flow of the project. Visualizing all the steps and sequences helps better in understanding the project. Below is the flow chart of the final project which displays all the processes in rectangle, decisions in diamond and preprocess in hexagon.

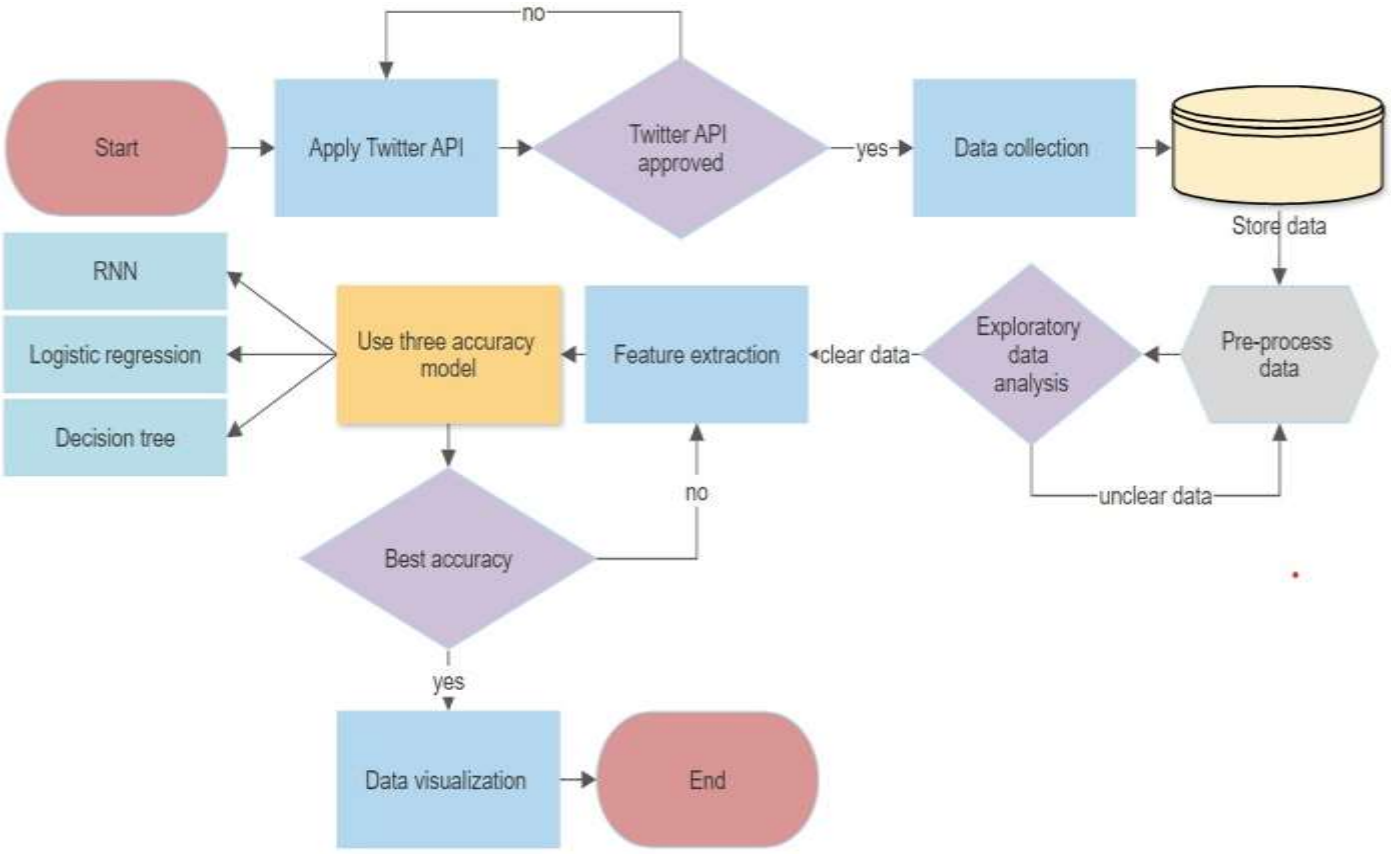
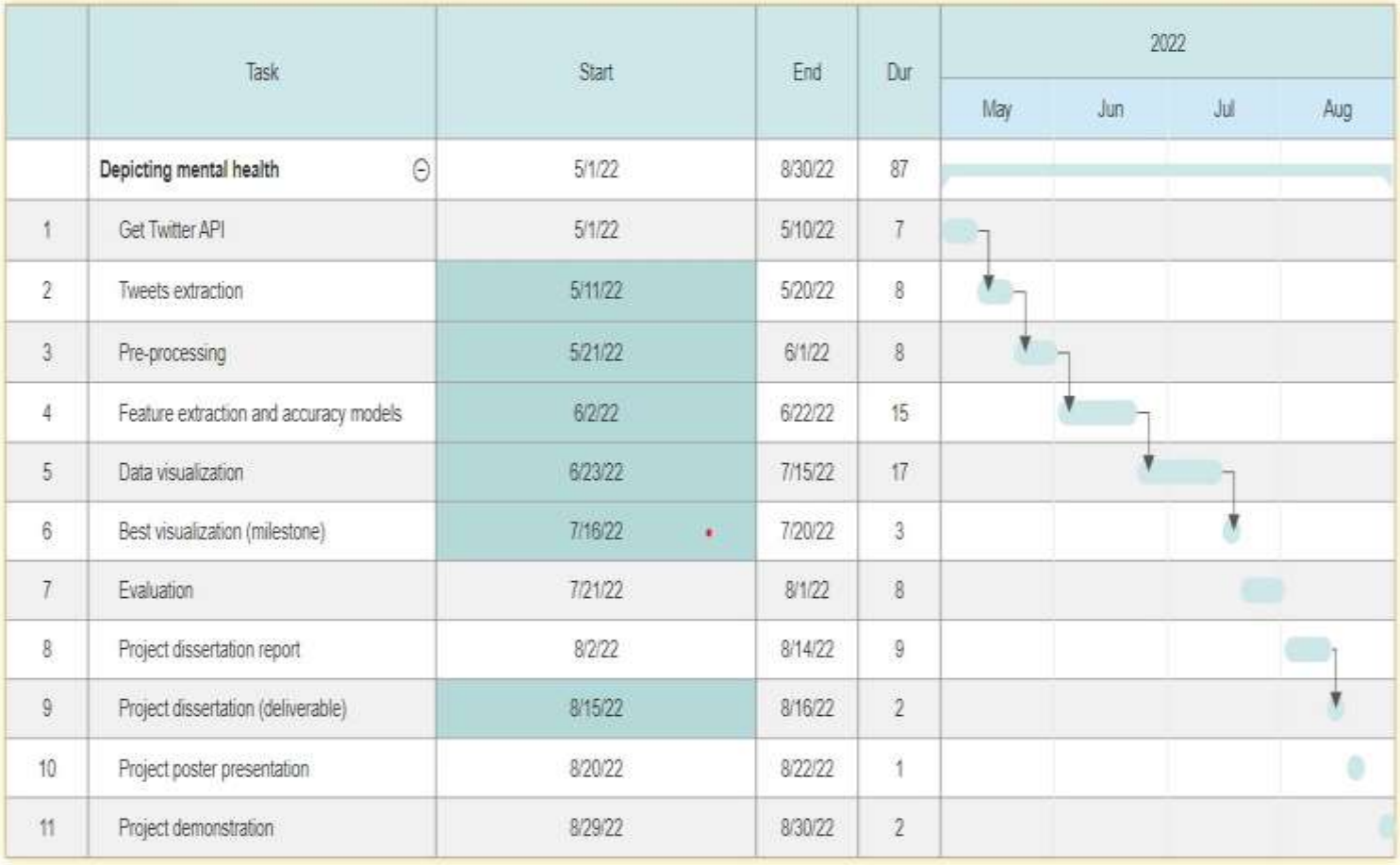


Figure 2: Flow chart

The first step for the project is to get Twitter APIs. Twitter APIs are used to access and extract tweets for the project. The data extracted is stored in a password protected excel file. All the pre-processing steps are applied to the data stored for EDA. If data is not properly cleaned, more cleaning techniques will be performed until data is cleaned and clear for feature extraction step. This is done to get the best accuracy for the dataset. Different combinations of features are used to train different models to get the best accuracy. Last step is to visualize all the mental health data of twitter users according to time and geographic location in a user friendly way so that it can present valuable information to psychologists and psychology students.

5.1.2 Gantt chart



	Task	Start	End	Dur
	Depicting mental health	5/1/22	8/30/22	87
1	Get Twitter API	5/1/22	5/10/22	7
2	Tweets extraction	5/11/22	5/20/22	8
3	Pre-processing	5/21/22	6/1/22	8
4	Feature extraction and accuracy models	6/2/22	6/22/22	15
5	Data visualization	6/23/22	7/15/22	17
6	Best visualization (milestone)	7/16/22	7/20/22	3
7	Evaluation	7/21/22	8/1/22	8
8	Project dissertation report	8/2/22	8/14/22	9
9	Project dissertation (deliverable)	8/15/22	8/16/22	2
10	Project poster presentation	8/20/22	8/22/22	1
11	Project demonstration	8/29/22	8/30/22	2

Figure 3: Gantt chart

This Gantt chart represents the tasks, milestones and deliverables of the project. A buffer of 4 days is allocated around the milestone as extra time period to cover any uncompleted work.

5.1.4 Risk Analysis

It's better to conduct risk analysis on a project. This will help to identify all the risks that may appear while doing a project. It will assist in categorizing the risks as high and low priority.

Moreover a risk register consists of information about risk effects and how to mitigate these risks. Risks may change during the project, either the risk is eliminated by its own or new risks appears. The priority of risks also changes through the project. All the risks that appeared in the project are highlighted.

Risk register

NO.	Risk category	Risk name	Risk description	Priority	Probability	Effects	Response

1.	Technology	Twitter Access	Twitter Access is unauthorized or not granted	High	High	The whole project depends on the twitter datasets. Project will fail if this happens	Apply for Twitter API's before starting the project. If API's is not granted, then look for other existing datasets.
2.	Technology	Data Analysis	Problems may occur with the datasets like less features, imbalanced classes or random characters in text	High	High	Problem in analyzing data can cause the model to visualize wrong data	More techniques should be applied on the dataset for cleaning and preprocessing before working with the dataset to mitigate these risks.
3.	Technology	loss of Data	Hardware or Software problem which causes loss of data	High	High	This happens usually when performing analysis on larger datasets. If data is loss, project reaches back to its starting position.	Files were saved on 3 different places, Laptop, email and Github. 2 separate files are used 1 for tweet extraction and other for data visualization to decrease the processing load.

4.	External risk	Psychology	This risk may appear in evaluation part. The model may not have enough valuable information	Medium	Low	This will affect the grades of the project.	Evaluation should not be left till the end. Continuous information trading should be done with the supervisor or psychiatrist
5.	People	Author's skills	The person conduction the research and building the model may have lack of expertise and skills	Medium	Low	Scope of the project may decrease	Practice makes this better. Asked for guidelines from someone with good experience in this field.
6.	Requirements	Requirements changes	Requirements of the project may change due to time limit or lack of experience	Medium	Moderate	Requirements changing in the start of the project are fine but changes at the end of the project would create a problem in the documentation and research part.	Clear out the requirements with the supervisor before starting the project. Document every changed requirement.

7.	Design risk	Data visualization	less data displayed for visualization	Low	Low	It will affect the objectives of the project and the evaluation part	Visualize as much information as possible.
8.	People	Time management	Poor time management	Low	Moderate	Project plan will be disturbed and can affect the quality of the project.	Move according to the project plan or have a gap between to overcome previous steps.
9.	People	Supervisor	Supervisor may have personal problems or reallocation of supervisor is done	Low	Low	Supervisors plays important role. This risk may take the project on a wrong path	Have regular meetings with the supervisor. Use online tools or meet in person.

Table 5: Risk register

New risks appeared in the project

N0.	Risk category	Risk name	Risk description	Priority	Probability	Effects	Response
10.	Technology	Tweet extraction	Problem in extracting tweets of specific users or specific time period	High	High	Dataset is the most important part of the project, without dataset project may fail	Tweepy didn't work. An alternative tool was used called Snscraper to scrape tweets.
11.	Requirements	Data Visualization	Less amount of tweets for visualization	Medium	Low	Accurate comparison will not be done which may not	Display data separately to avoid unbalanced number of tweets. Comparison was done side by side instead of

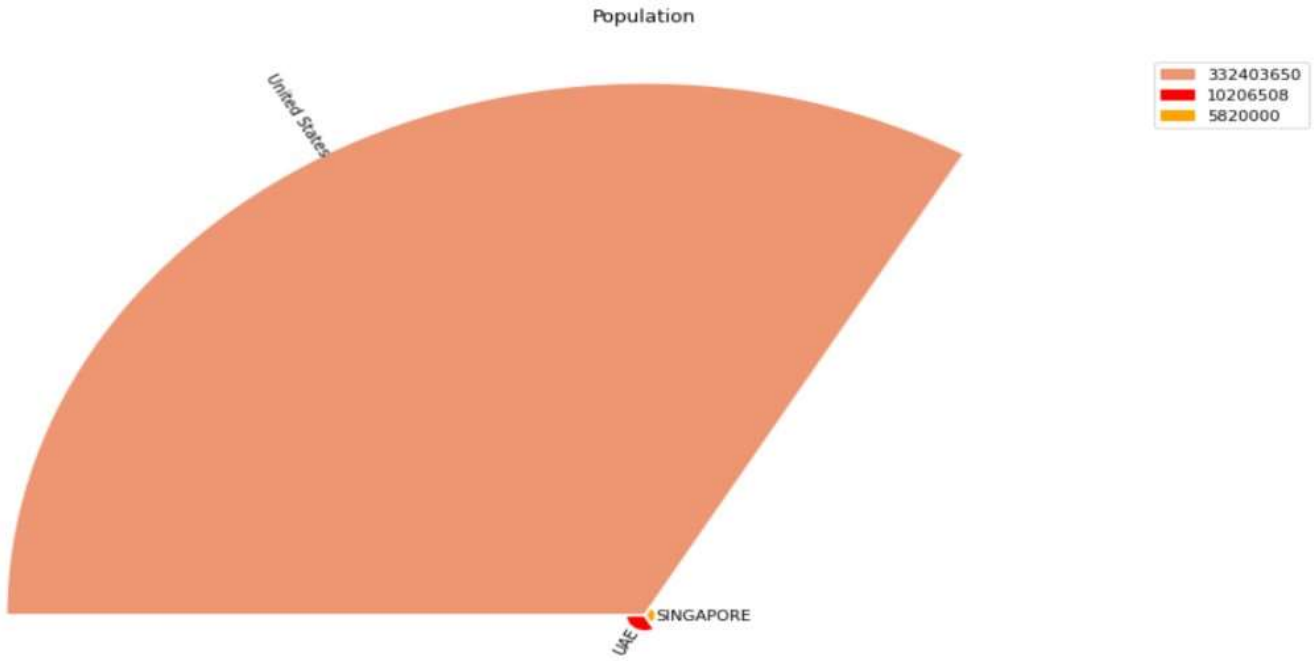
						complete project requirements.	comparing on same plot.

Table 6: New risk register

Results

The following results are based on visualizing three mental health problems (Schizophrenia, Depression and Bipolar disorder) by analyzing Twitter text from users of three countries (US, UAE, Singapore) in 4 years of time (2019,2020,2021 and 2022). Tweets are extracted within a 3 month time period each year. These months were chosen according to the Covid_19 peaks.

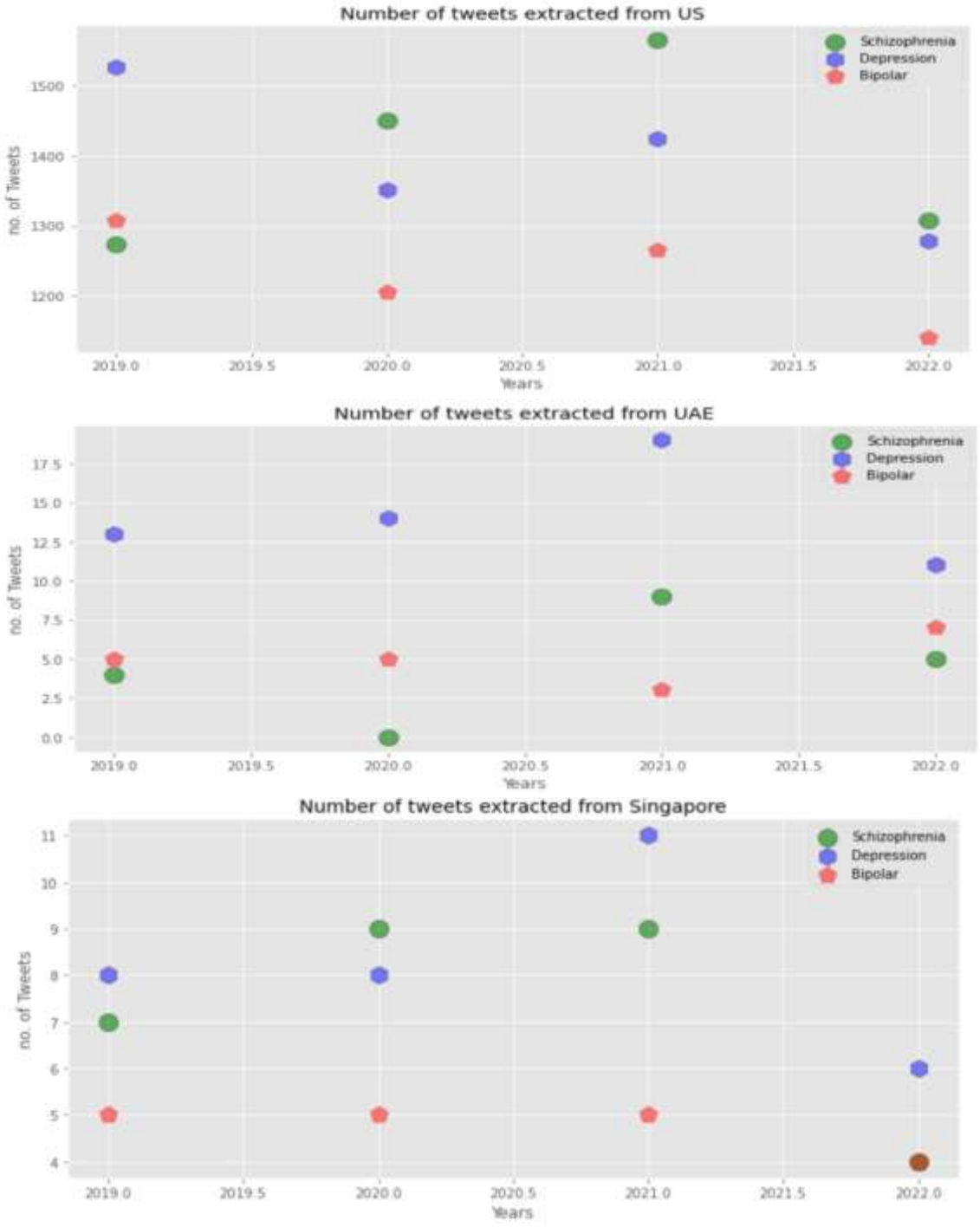
Total Population of US, UAE and Singapore



Result 1: Population

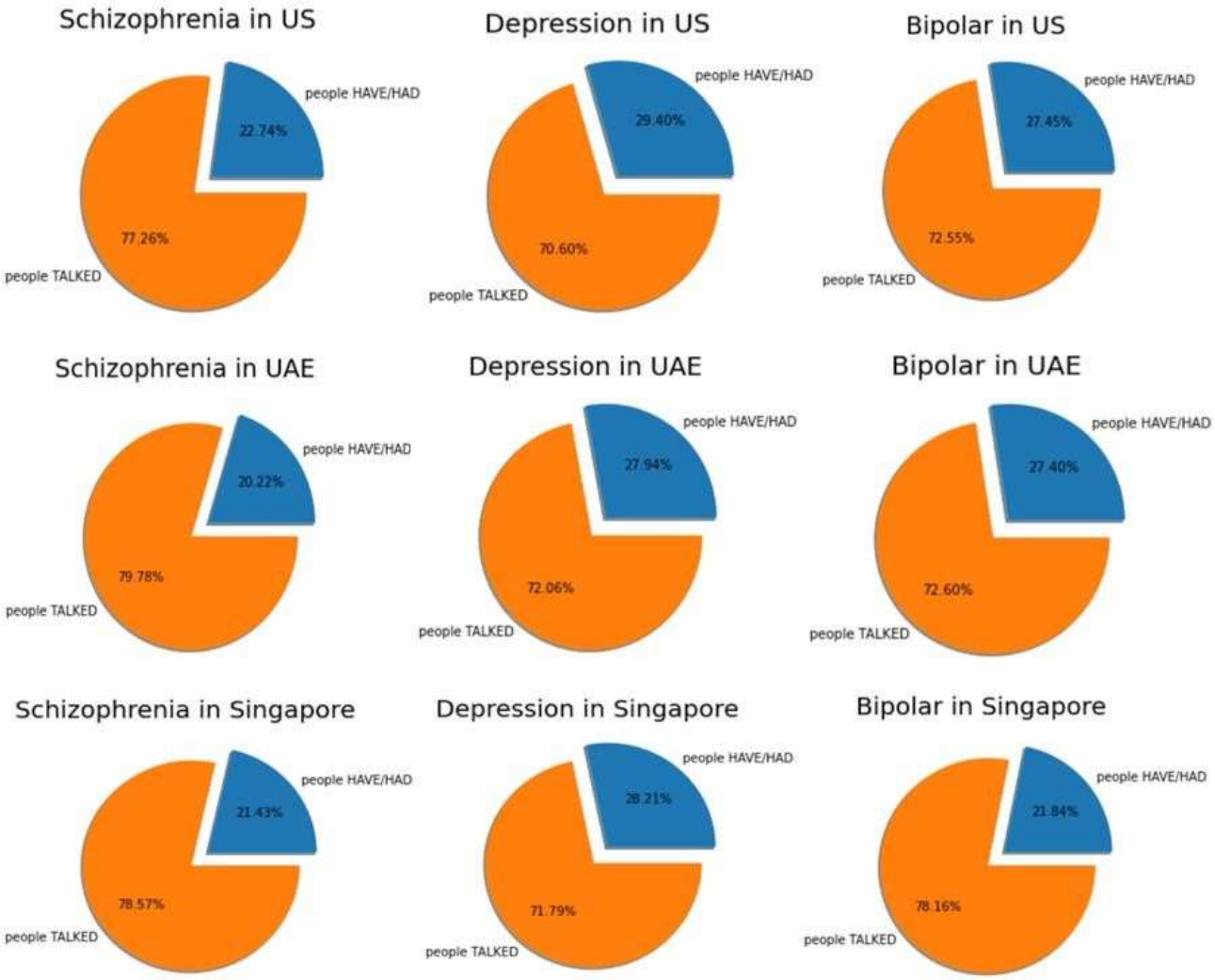
There is a huge population gap between US compared to UAE and Singapore. This population gap is one of the reasons for large difference in total tweets of US, UAE and Singapore. According to this graph, if US population is higher, then eventually number of twitter users, mental health awareness and online communication will be higher too.

Number of Filtered tweets from each year



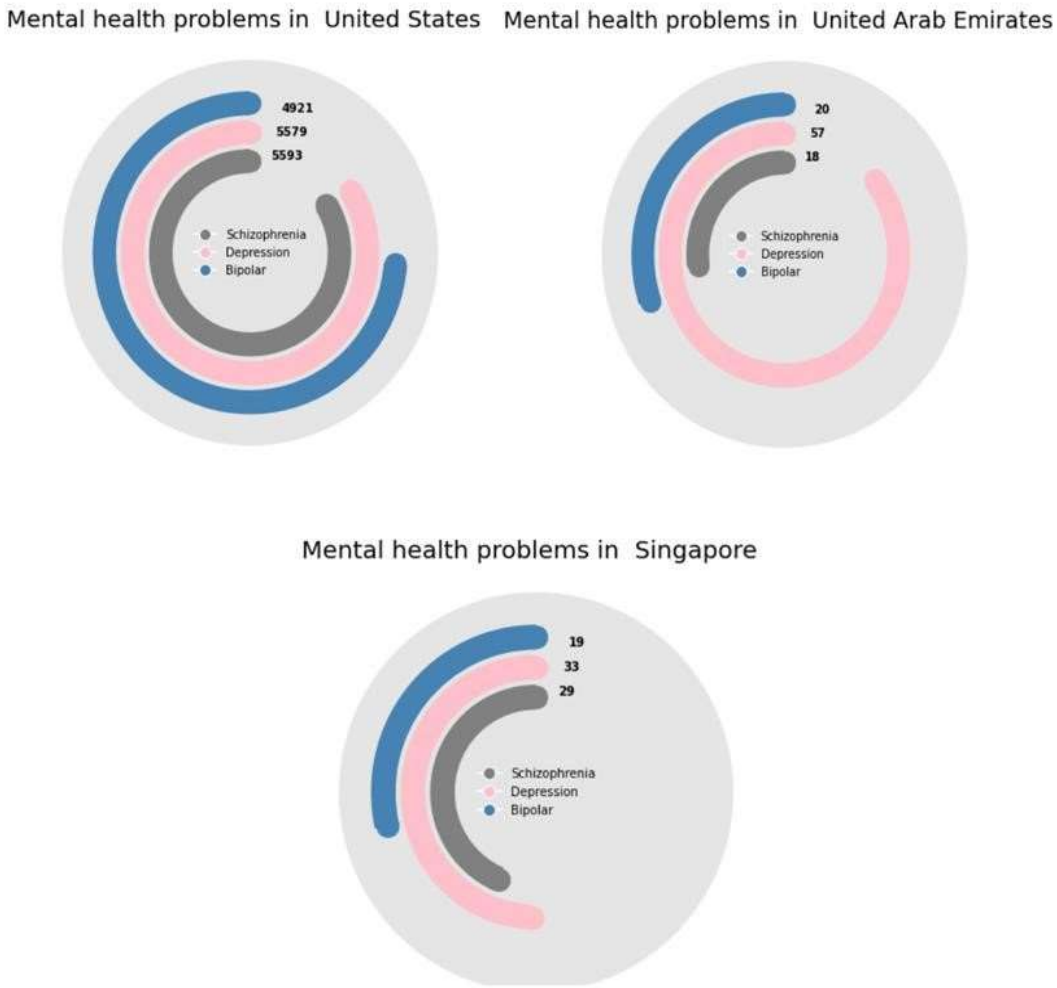
The Scatter plot shows that highest number of tweets is in year 2021 and lowest in 2022. Maximum number of tweets in US is above 1500(Depression and Schizophrenia), maximum number of tweets in UAE is 18(Depression) and maximum number of tweets in Singapore is 11(Depression).

People who tweeted about mental health problems vs people who actually said they have mental health problem



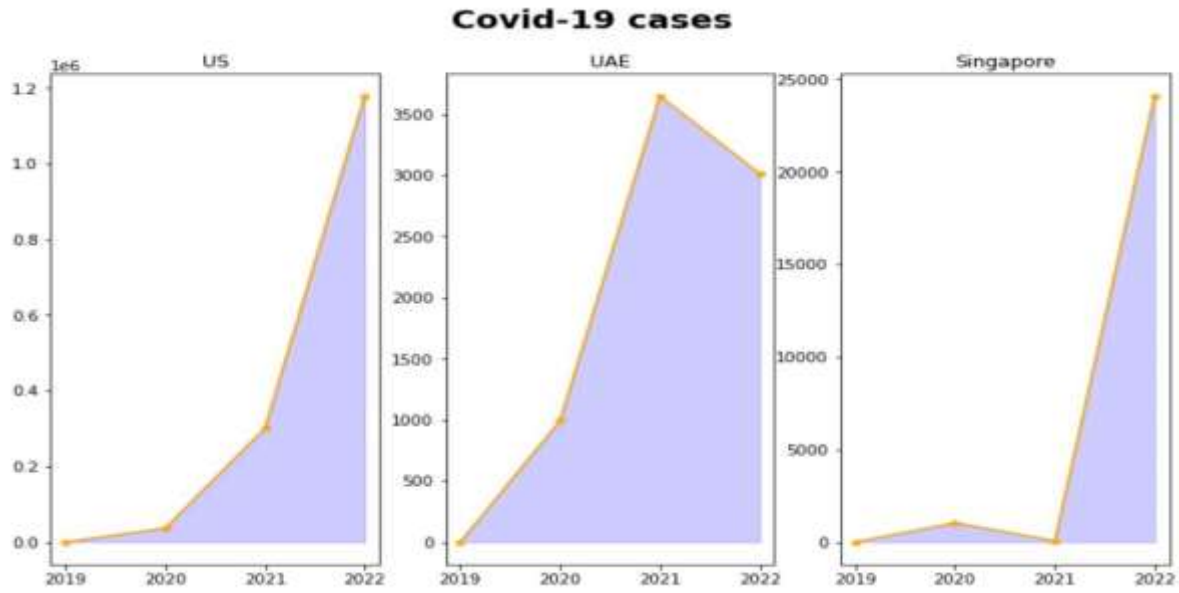
The pie chart shows the number of total tweets extracted that contained the word depression, schizophrenia or bipolar disorder vs the number of filtered tweets of people actually declared having/ had depression, schizophrenia or bipolar disorder. This displays that awareness of these mental health problems are common in countries but very few people suffers from them. We cannot compare US, UAE and Singapore in this chart because numbers of total tweets extracted from every country are different. We can compare each mental health problem in a country at a time. This graph also shows that rate of people having depression in each country are higher than Schizophrenia and Bipolar disorder.

Total filtered tweets of Depression, Schizophrenia and Bipolar disorder according to countries



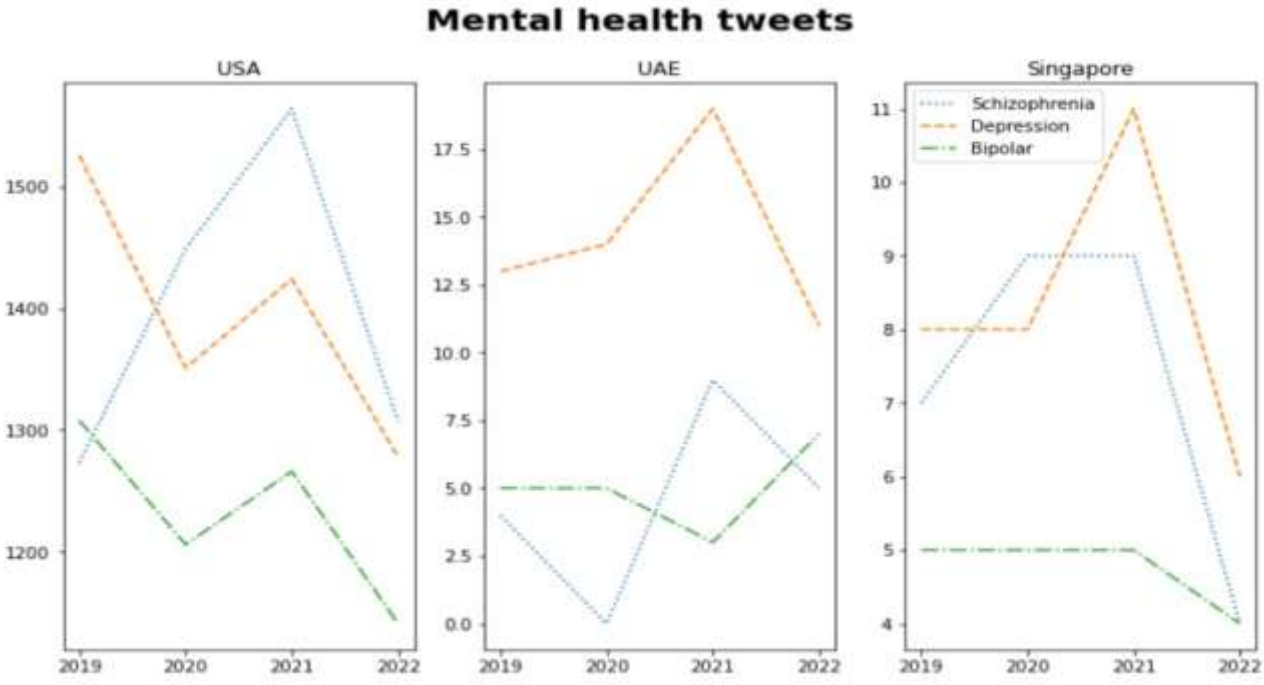
In these circular bar charts, depression seems one of the leading mental health problems in UAE and Singapore where as Schizophrenia is slightly higher than depression in US. Bipolar disorder is relatively less than other two mental health problems in every country.

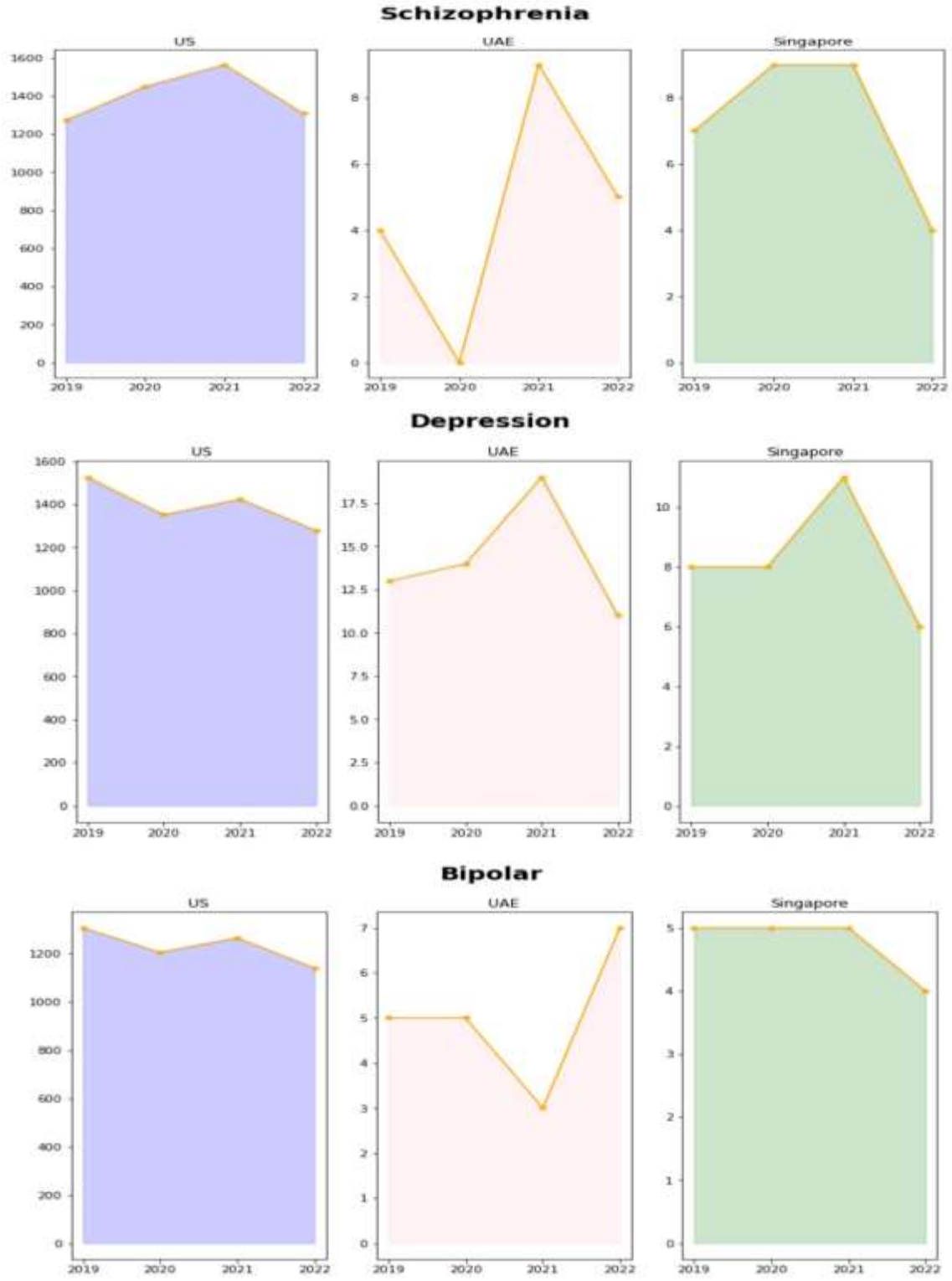
Peaks of Covid-19 in four years



Data of Covid_19 cases has been taking from 4 highest peaks of Covid_19 in 4 years. Each year shows the highest number of cases in 1 days of every peak. Y-axis shows the number of cases. These statistics are taken from Google statistics which is published by New York Times. Spike of Covid-19 cases in year 2021 is only in UAE where spike of Covid-19 cases in US and Singapore was in year 2022. There is a huge jump of Covid-19 cases in Singapore in year 2021-2022. These stats are shows to gain insight of Covid-19 effects on people.

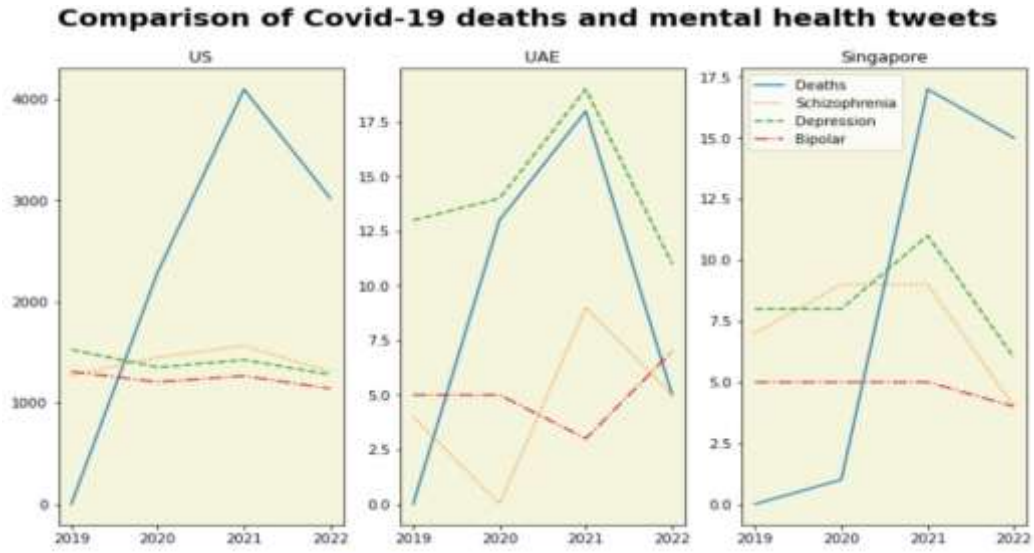
Mental health problems according to year





The line graph shows the number of filtered tweets of US, UAE and Singapore. This graph shows the increase in depression of each country in year 2021 and a sudden decrease in year 2022. US also show increase of Schizophrenia and Bipolar disorder in year 2021 where as UAE and Singapore has few ups and downs in the graph. Proper Line graph visualization of UAE and Singapore was unachievable because of lesser amount of data from these two countries.

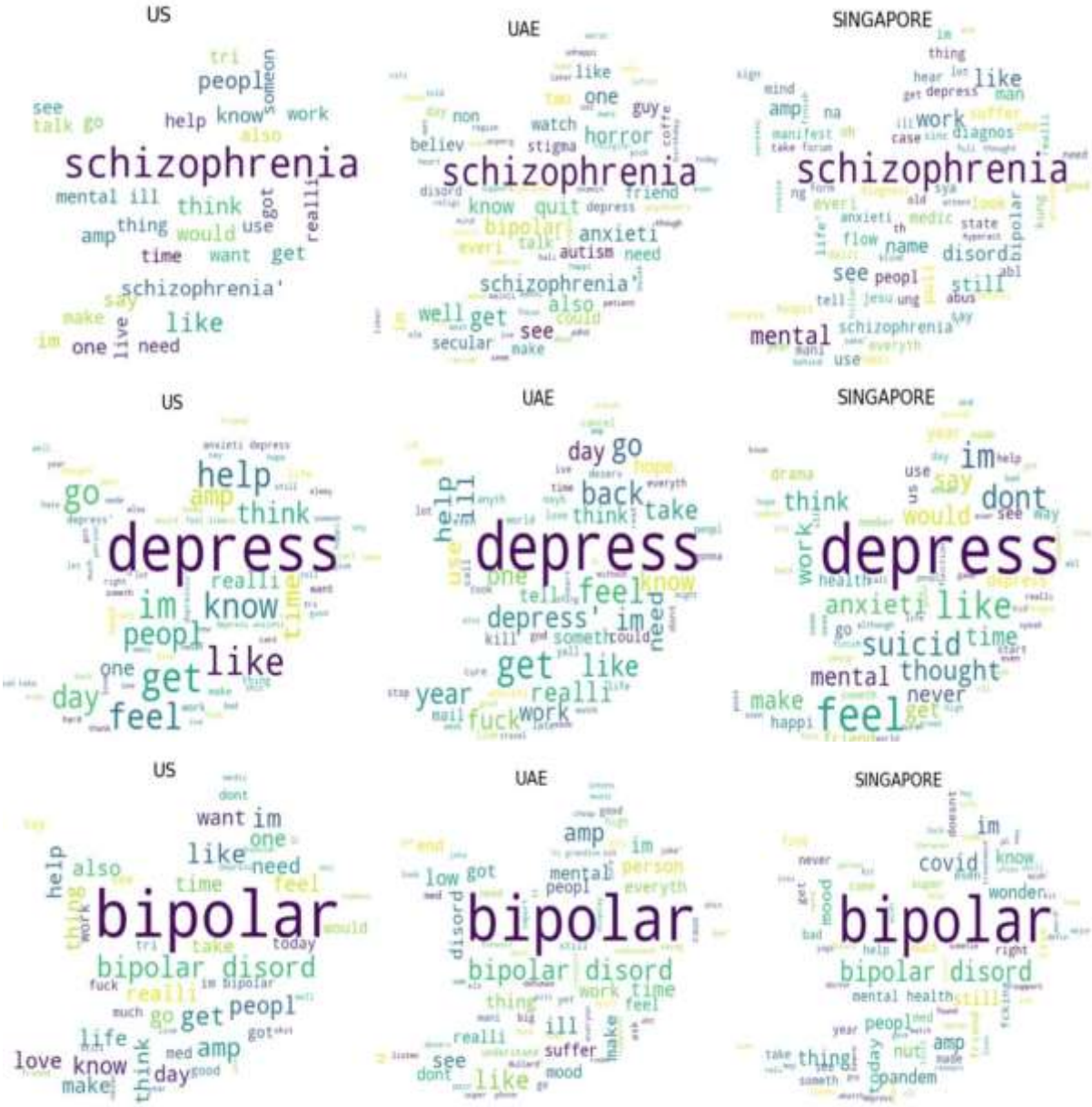
Relation of Covid-19 with mental health of people



The line graph displays all the mental health problems along with Covid-19 deaths in each country. As there is an increase in deaths due to Covid_19, there is an increase in tweets related to people declaring that they are suffering/ed from mental health problem. Covid-19 deaths were taken from same time frame from which tweets were extracted. The deaths peaks in the graph represent maximum number of deaths in a day within 3 month time period. These three months are same from which twitter data is extracted. The Covid-19 death data is taken from Google statistics which is published by New York Times. According to this graph, Bipolar seems low in every country where as depression increases in all three countries. This graph also gives a rough idea that Covid-19 deaths had affected people's mental health negatively.

WordCloud of most used words in each dataset

Word Cloud emphasizes on most common words used in text. It gives an idea how people express their feelings. Each dataset is tokenized and displayed as a word cloud.

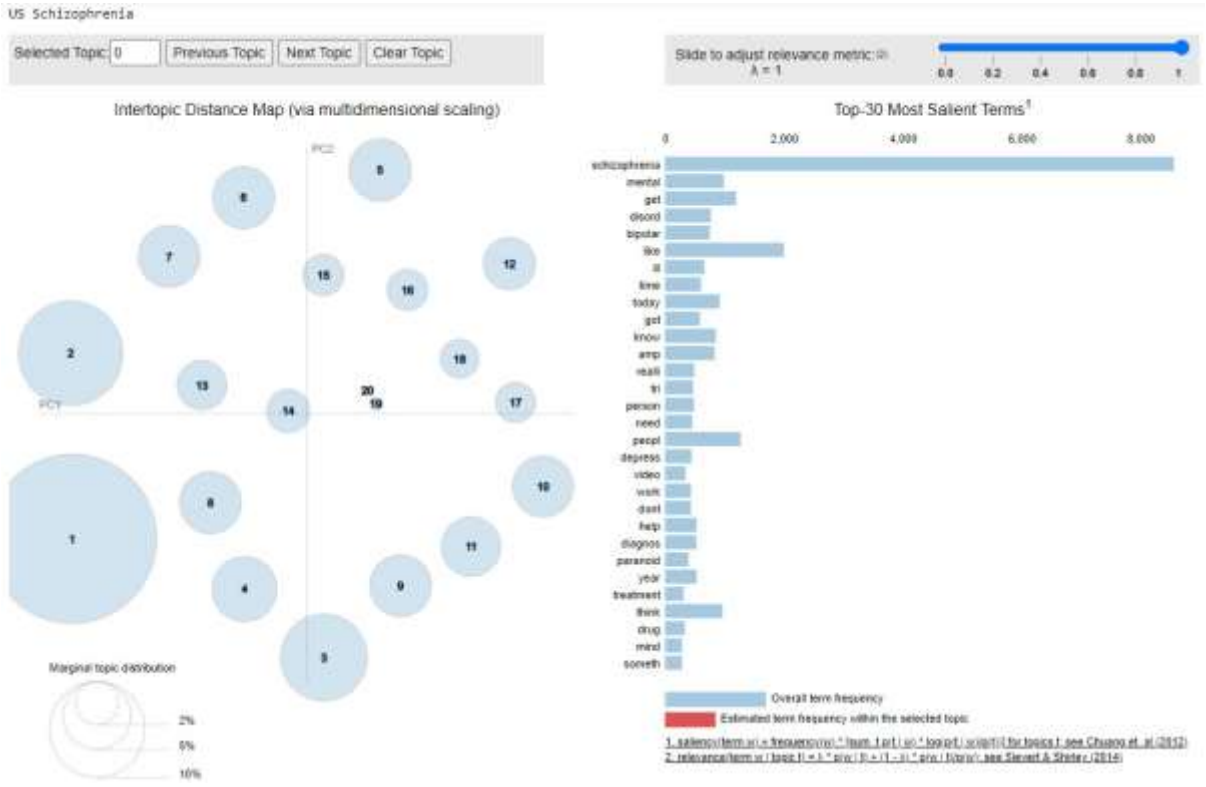


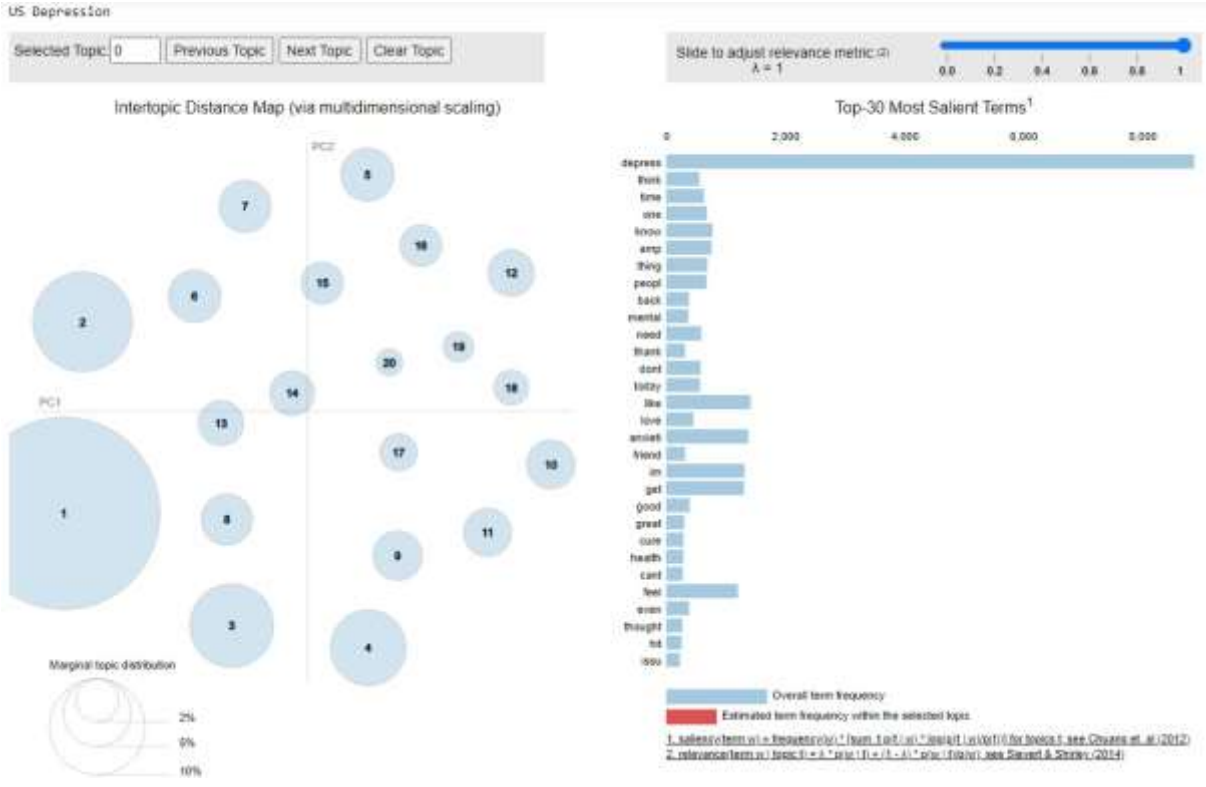
Most common word in each dataset is the mental health problem name. Apart from that, if we look closely, we will see words like horror, stigma, reality, autism, anxiety, bipolar, depression, believe, quit, unhappy, worst, think, suffer, thought, mind, sign, stress, ill and hear in Schizophrenia datasets. Words like anxiety, depress, never, feel, sad, always, help, kill, hope, need, suicide, health, drama, thought and bad are displayed in Depression. And in Bipolar, we can find words like depress, want, joke, day, help, feel, year, suffer, mood, sympathy, time, pain, covid, today, pandemic, bad, treatment, and still.

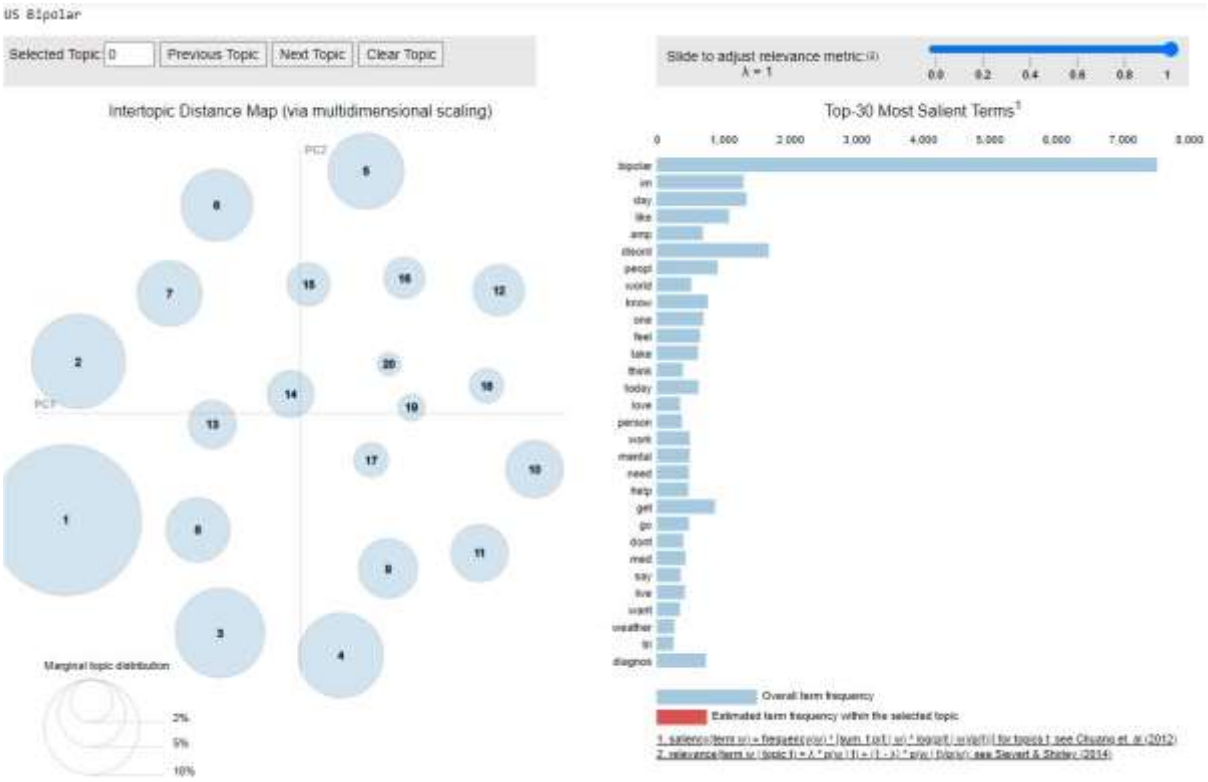
Few words are trimmed or reduced in stemming process. This word cloud does not gives a full meaning to a sentence but gives a quick view on most emphasized words in the tweets. A reality is that, we don't know people who tweeted were actually suffering from a mental health problem or it was just sarcasm or joke. But this wordcloud shows few words that can describe a mental health problem and its symptoms.

Word count of filtered tweets

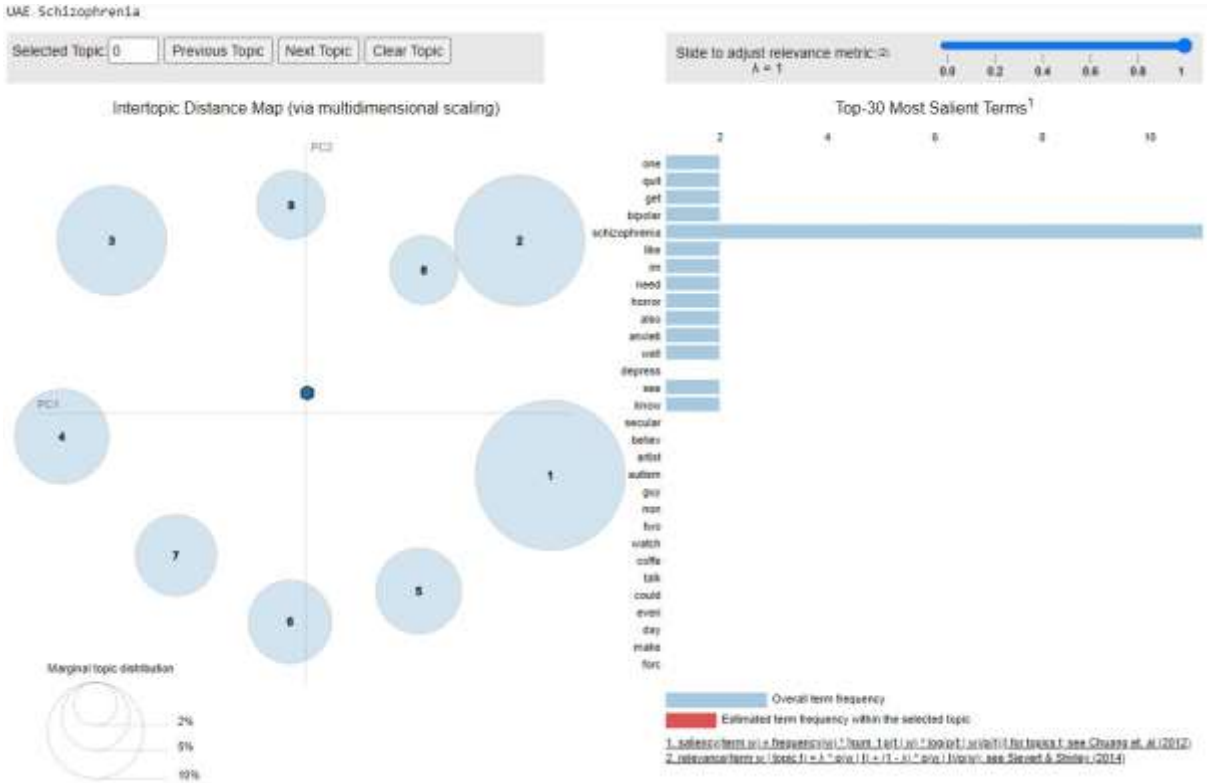
US datasets

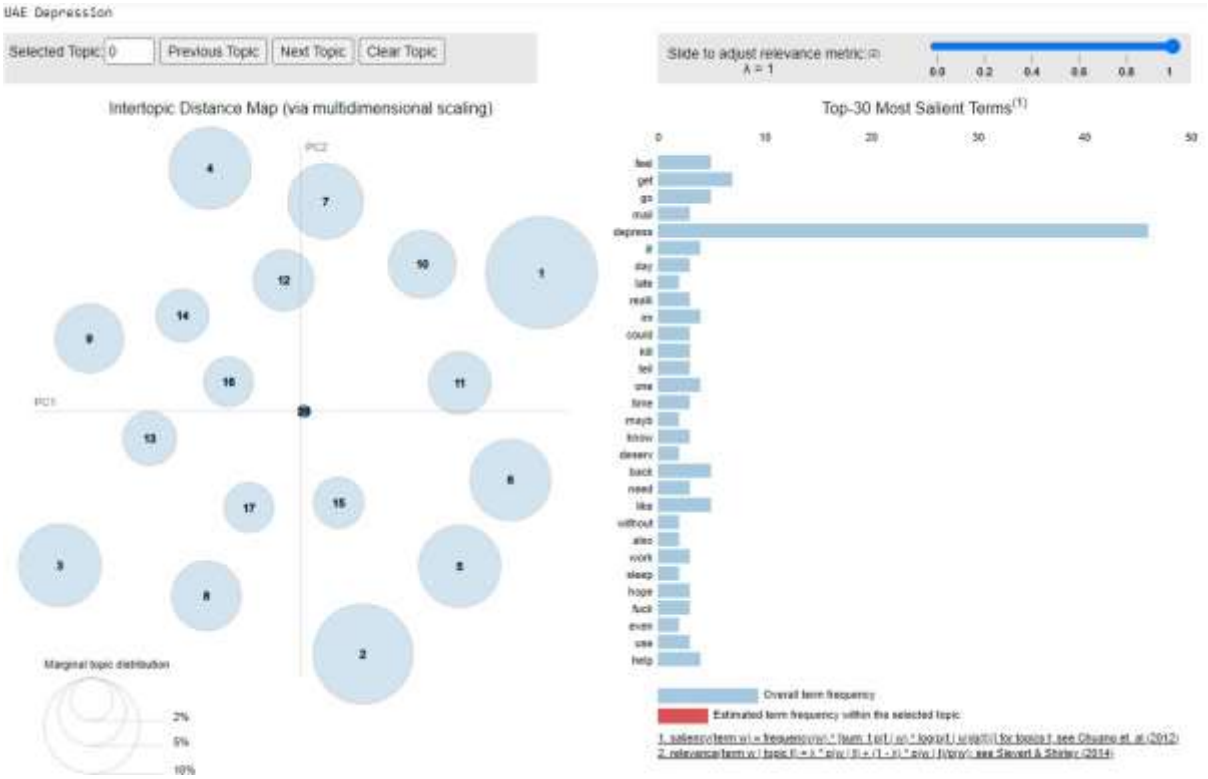

US Schizophrenia
Selected Topic: 0
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:
λ = 1
Intertopic Distance Map (via multidimensional scaling)
PC1
PC2
Marginal topic distribution
Top-30 Most Salient Terms
schizophrenia
mental
get
disord
bipolar
like
ill
time
today
got
know
amp
realli
hi
person
need
peopl
depress
video
work
dont
help
diagnos
paranoid
year
treatment
think
drug
mind
someth
Overall term frequency
Estimated term frequency within the selected topic

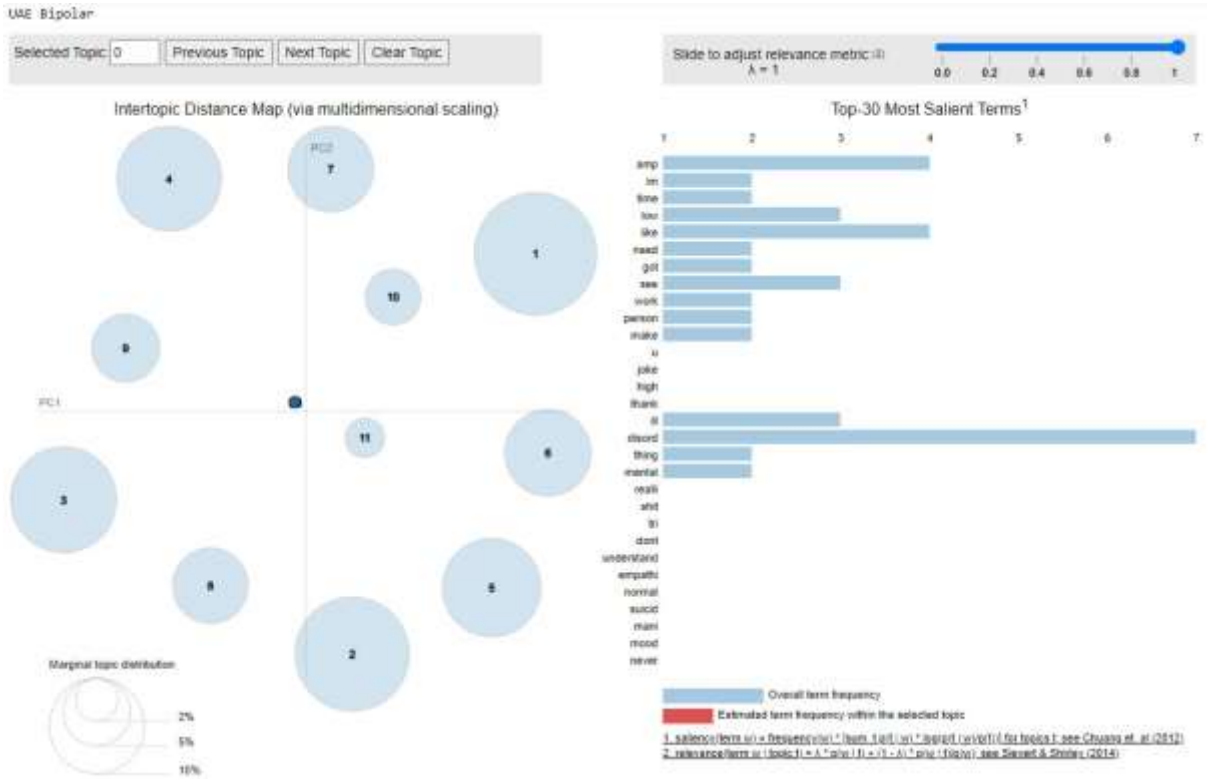

US Depression
Selected Topic: 0
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:
λ = 1
Intertopic Distance Map (via multidimensional scaling)
PC1
PC2
Marginal topic distribution
Top-30 Most Salient Terms
depress
think
time
one
know
amp
thing
peopl
back
mental
need
thank
dont
today
like
love
anxiet
friend
im
get
good
great
cure
health
cant
feel
even
thought
lot
now
Overall term frequency
Estimated term frequency within the selected topic



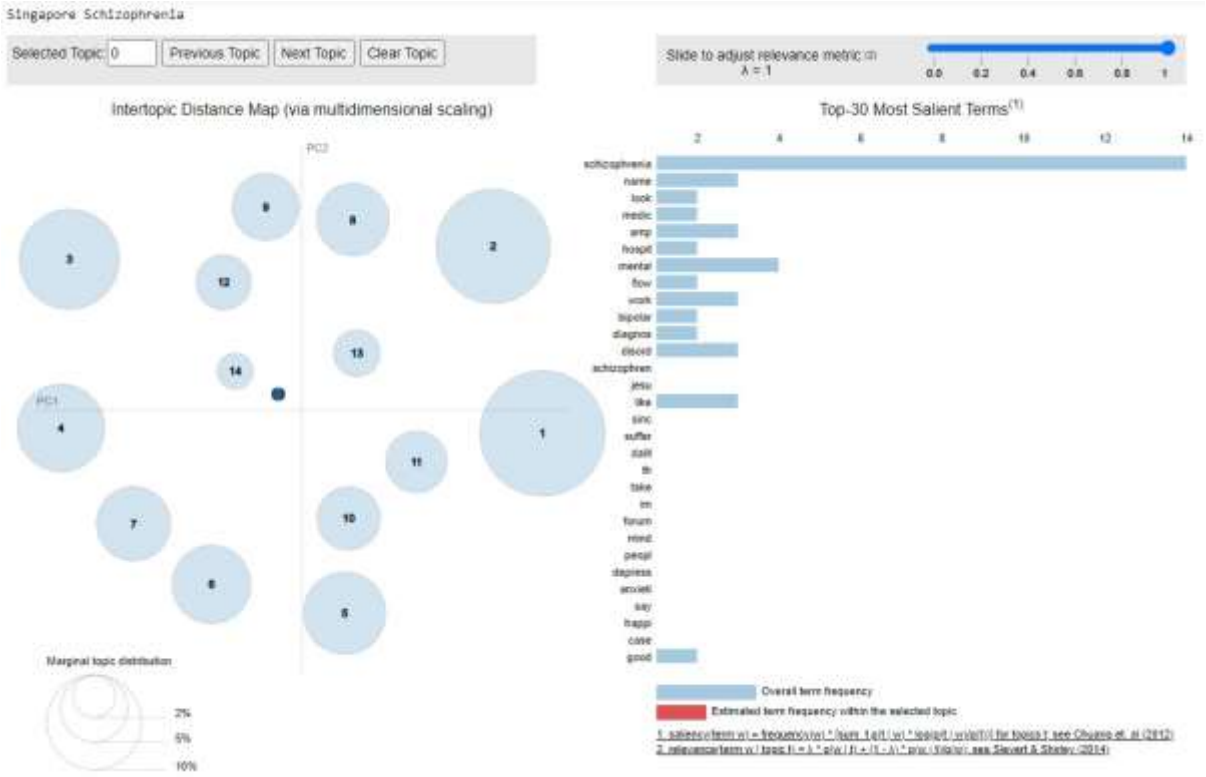
UAE datasets

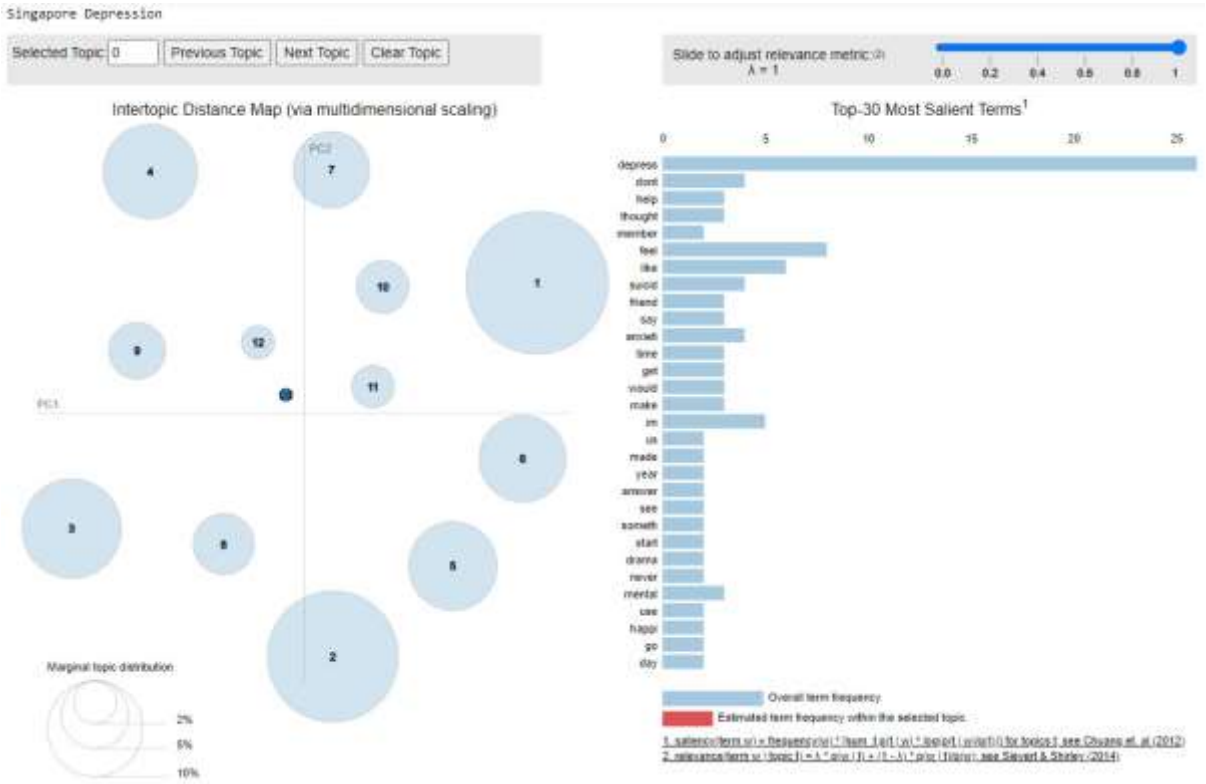


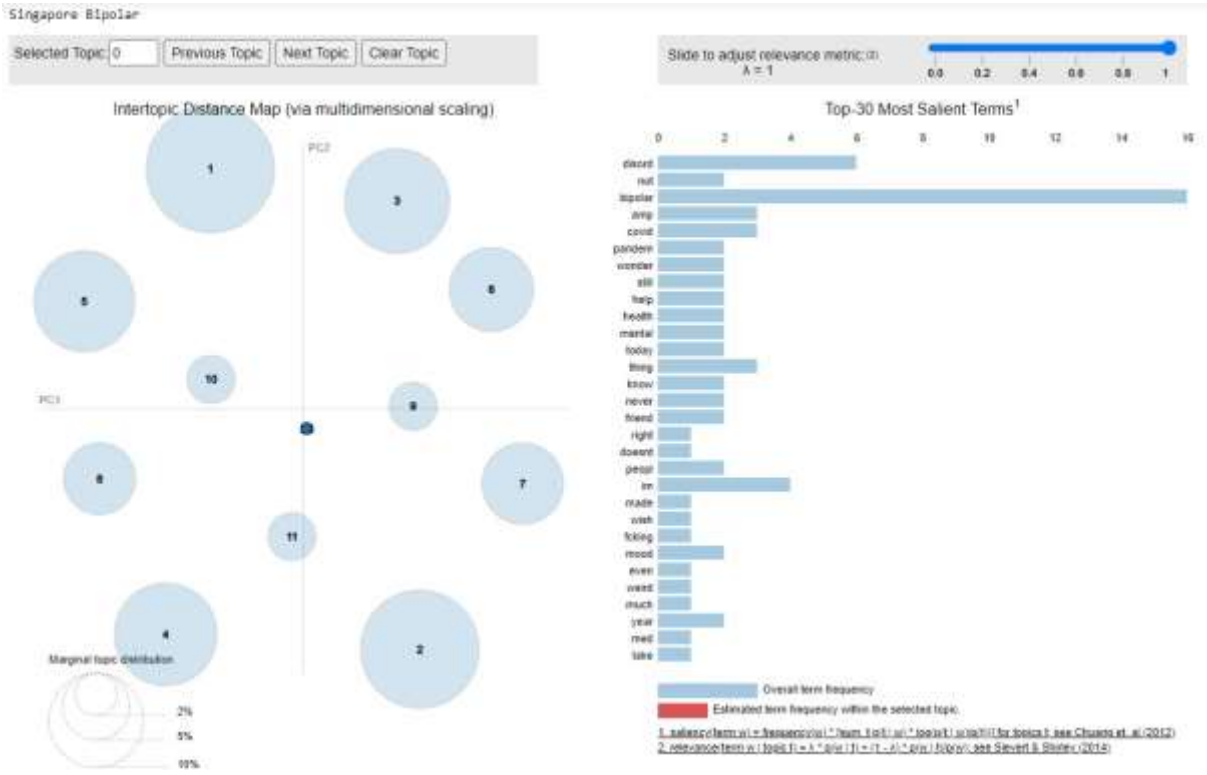




Singapore datasets


Singapore Schizophrenia
Selected Topic: 0
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:
λ = 1
Intertopic Distance Map (via multidimensional scaling)
Top-30 Most Salient Terms
Marginal topic distribution
Overall term frequency
Estimated term frequency within the selected topic

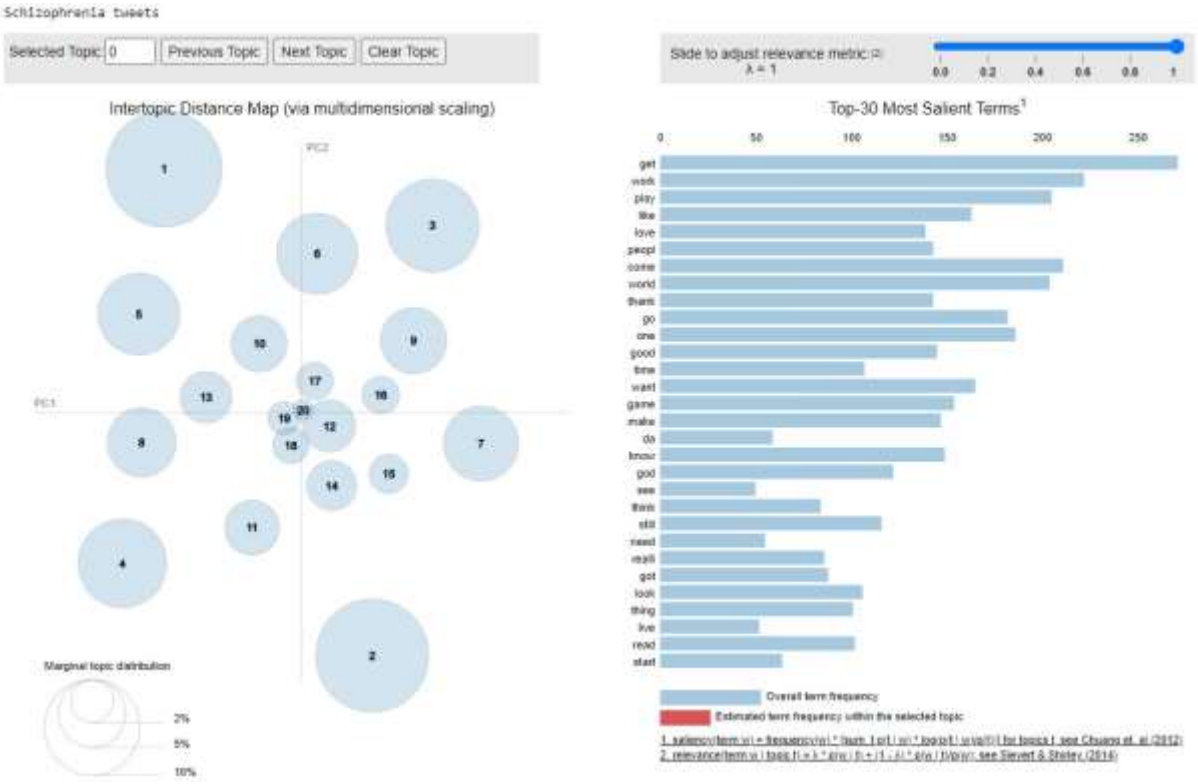

Singapore Depression
Selected Topic: 0
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:
λ = 1
Intertopic Distance Map (via multidimensional scaling)
Top-30 Most Salient Terms
Marginal topic distribution
Overall term frequency
Estimated term frequency within the selected topic



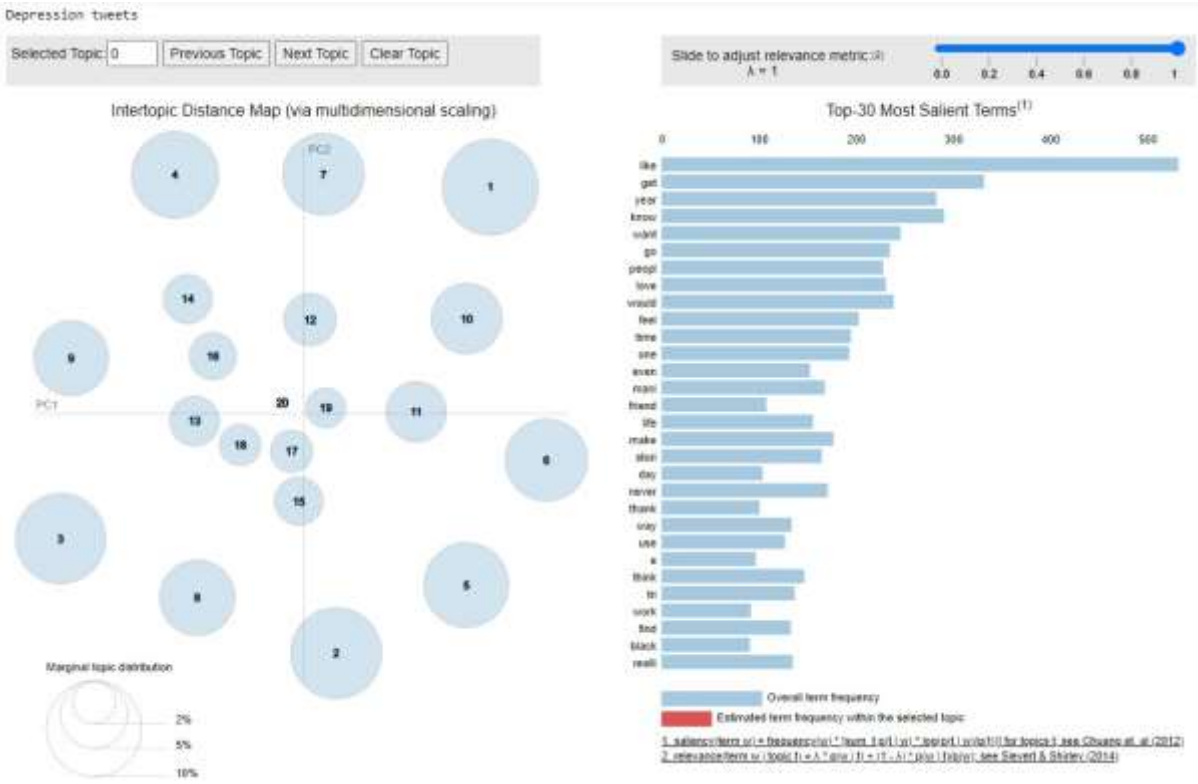
Another representation of words is word count frequency. This graph shows the repetition of words and gives a count of each word presented in the dataset. The intertopic distance map categorizes tweets as topics. Word count of any topic can be viewed by hovering over any circle.

Word count frequency of selected users

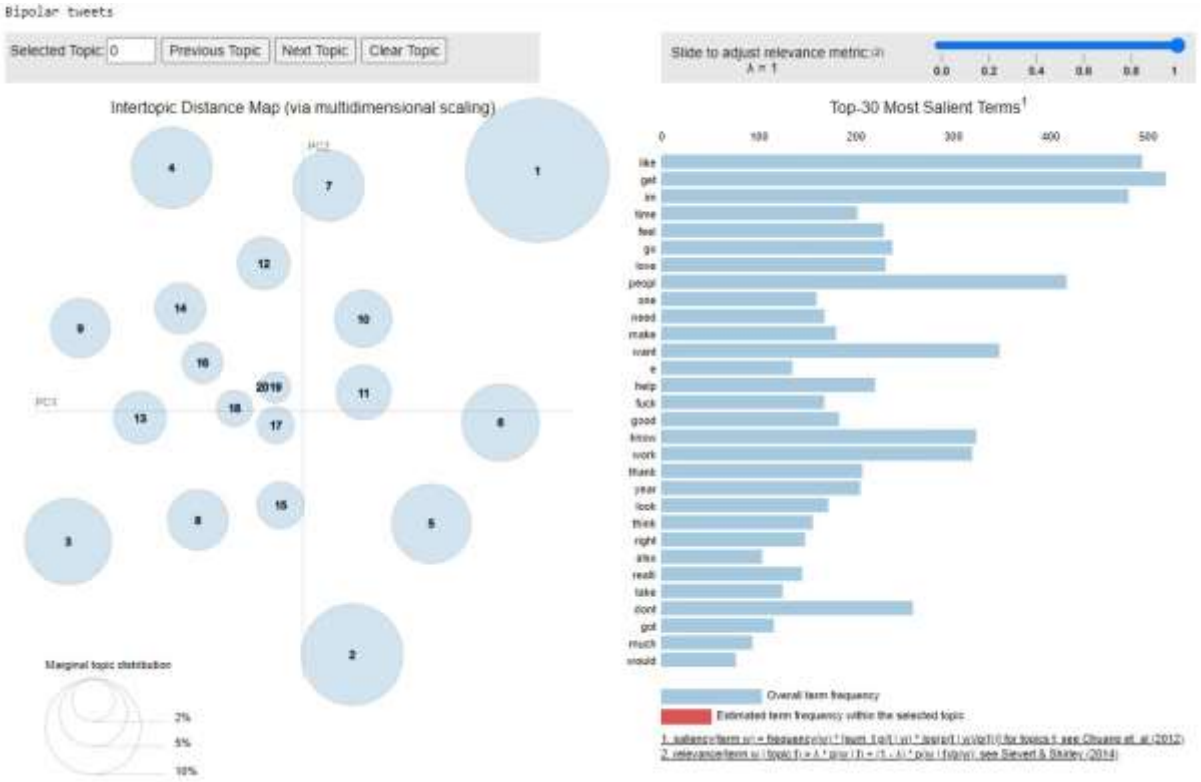
Schizophrenia



Depression



Bipolar



These graphs present tweets of selected users from filtered dataset. All URLs and username were removed in cleaning tweets. Another extraction was done for US, UAE and Singapore to extract tweets along with user names. The process of cleaning and filtering were done on only three mental health problems of US. One reason for not displaying UAE and Singapore user's tweets was they had limited number of tweets. Other reason was limited time for project. 6 users were manually selected from each dataset of US to extract 500 tweets of each user. Separate files were created for Schizophrenia, Depression and Bipolar users.

Conclusion

This research clarifies the concept of mental health and its problems. The research is supported by some references used to prove the hypothesis that mental health is an immerging and dangerous problem which should be of concern to the upcoming youth. Machine learning on the other hand, assists in identifying such problems without any face to face interaction in fact it classifies the disease with only texts and posts of people. Furthermore, the research investigates three mental health problems named depression, bipolar disorder and schizophrenia. It describes the problem, causes, symptoms, effects and treatments of each mental health problem. Models created using such data can be useful in predicting suicidal rates or pre diagnose a disease in a country. Information related to the

project, its methods and steps of achieving the goal are explained in the report. The snapshots of all the results are displayed and explained. The significance of the model, if future work is done, is explained by providing points on how that model will be beneficial to psychologists or health administration of countries.

Moreover, it's important to follow a project plan in order to accomplish all the aims and objectives. Project plan of the overall project is explained and displayed by Gantt chart. Also it's better to perform a risk analysis on the project; this project contains a risk register which defines all the risks according to their category, priority, effects and ways to mitigate.

Appendix

Abbreviations

NLTK	Natural Language Tool Kit
EDA	Exploratory Data Analysis
API	Application Program Interface
PCA	Principal Component Analysis
SVD	Singular Value Decomposition
TF –IDF	Term Frequency and Inverse Document Frequency
SVM	Support Vector Machine
LIWC	Linguistic Inquiry and Word Count
CLI	Command Line Interface
MH	Mental Health
NLP	Natural Language Processing

Bibliography

Akram(psychologist), S. (2022, 2 10). mental health problems . (S. Ghaffar, Interviewer)

Association, T. A. (2022, 1 1). */www.psychiatry.org*. Retrieved 3 15, 2022, from American Psychiatric Assosiation: https://www.psychiatry.org/patients-families/bipolar-disorders/what-are-bipolardisorders

Avenaim, J. (2021, 5 1). *Mental Health Statistics 2021.* Retrieved 4 8, 2022, from Mental Health Foundation: https://mentalhealthfoundation.org/mental-health-statistics-2021/

Beck, M. (2020, 12 3). How to Scrape Tweets With snscrape. *Better Programming* , 1.

Berry, N., Lobban, F., Belousov, M., Emsley, R., Nenadic, G., & al., e. (2017). #WhyWeTweetMH: Understanding Why People Use Twitter to Discuss Mental Health Problems. *Journal of Medical Internet Research , 19*, ISS 4.

C. C., S. Y., J. W., B. G., M. F.-S., & F. V. (2021). Indications of Depressive Symptoms During the COVID-19 Pandemic in Germany: Comparison of National Survey and Twitter Data. *JMIR Publications , 8*, no. 6.

Coppersmith, G., Dredze, M., & Harman, C. (2014). Quantifying Mental Health Signals in Twitter. *Human Language Technology Center of Excellence* , 51-60.

Dean, B. (2022, jan 5). *Backlinko*. Retrieved 2 13, 2022, from Backlinco.com: https://backlinko.com/twitter-users#twitter-users

JAIN, K. (2021, 6 5). *Twitter Sentiment Analysis LSTM.* Retrieved 3 25, 2022, from kaggle: https://www.kaggle.com/code/kritanjalijain/twitter-sentiment-analysis-lstm/comments

KAIROS. (2018, 6 26). *Machine Learning for Mental Health.* Retrieved 3 25, 2022, from Kaggle: https://www.kaggle.com/code/kairosart/machine-learning-for-mental-health-1/notebook

KALE, S. S. (2015). *TRACKING MENTAL DISORDERS ACROSS TWITTER USERS .* Athens : University of Georgia.

Kumar, A. (2020, 3 25). *Twitter Sentiment Analysis.* Retrieved 3 20, 2022, from Analytics Vidhya: https://medium.com/analytics-vidhya/twitter-sentiment-analysis-b9a12dbb2043

Mental Health Conditions. (2022, 1 1). Retrieved 3 2, 2022, from NAMI: https://www.nami.org/About-Mental-Illness/Mental-Health-Conditions

MR_KNOWNOTHING. (2020, 4 14). *Twitter sentiment Extaction-Analysis,EDA and Model.* Retrieved 3 20, 2022, from Kaggle: https://www.kaggle.com/code/tanulsingh077/twitter-sentiment-extactionanalysis-eda-and-model/notebook

OpenStax. (2022, 1 1). *Introduction to Psychology*. Retrieved 3 20, 2022, from Lumen Learning: https://courses.lumenlearning.com/wmopen-psychology/chapter/introduction-to-mental-health/

Organization, W. H. (2022, jan 1). *World Health Organization*. Retrieved 2 15, 2022, from

www.who.int: https://www.who.int/health-topics/depression#tab=tab_3 Shah, P.

(2020). Basic Tweet Preprocessing in Python. *Towards Data Science* , 1.

shanthababu. (2021, 8 1). Effective Data Visualization Techniques in Data Science Using Python. *Data Science Blogathon* , 1.

Smith, K. (2014). 515180a. *www-nature-com.ezproxy1.hw.ac.uk/articles/515180a* , 10.

Staff, M. C. (2021). COVID-19 and your mental health. *Mayo Clinic* , 1.

Stanford University, S. C. (2015). Mining Twitter Data to Improve Detection of Schizophrenia. *AMIA Jt Summits Transl Sci Proc.* , 122-126.

Steven Gangloff, B. H. (2021). How #Epilepsy is Viewed on Social Media (4204). *Home , 96*, no 15.

Tiller, J. W. (2013). Depression and anxiety. *THE MEDICAL JOURNAL OF AUSTRALIA* , V199.

USA.gov. (2022, 04 06). *Centers for Disease Control and Prevention .* Retrieved 04 20, 2022, from cdc.gov: https://www.cdc.gov/suicide/facts/index.html

Weerasinghe, J., K. M., & R. G. (2019). “Because... I was told... so much”: Linguistic Indicators of Mental Health Status on Twitter. *Proceedings on Privacy Enhancing Technologies* , 152 - 171.

WHO. (2022, 1 1). *World Health Organization*. Retrieved 3 1, 2022, from www.who.int: https://www.who.int/news-room/fact-sheets/detail/schizophrenia

The End