AMERICAN INTERNATIONAL UNIVERSITYBANGLADESH

FACULTY OF SCIENCE AND TECHNOLOGY

Prediction of suicidal behavior from twitter data using machine learning

A Thesis Presented to the

DEPARTMENT OF COMPUTER SCIENCE

In Partial Fulfillment of the Requirements for the Degree BSc. in CSE

Supervised By

Md. Tohedul Islam

Assistant Professor

Department of Computer Science

Faculty of Science and Technology

Submitted By

Das, Proma – 19-39625-1

Shanto, Majharul Islam – 19-39585-1

Ahmed, Sk Shihab – 19-39575-1

Shahriar, S. M. – 19-39596-1

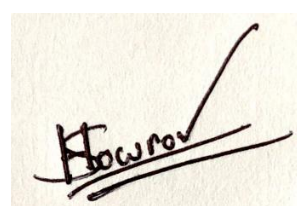
Declaration

We would like to declare that this is our first thesis book and has not been awarded any degree or diploma at a university or other tertiary institution. Information from published and non-published publication is permitted in the text and a reference book is provided.

Das, Proma

Shanto, Majharul Islam

Ahmed, Sk Shihab



Shahriar, S. M.

Approval

The thesis titled "Prediction of suicidal behavior in twitter data using machine learning" has been submitted to the following respected members of the board of examiners of the department of Computer Science in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science on January 2022 by Das, Proma (19-39625-1), Shanto, Majharul Islam (19-39585-1), Ahmed, Sk Shihab (19-39575-1) and Shahriar, S. M. (19-39596-1) has been accepted as satisfactory.

Md. Tohedul Islam

Assistant Professor and Supervisor

Department of Computer Science

American International University-Bangladesh

Prof. DR. Tafazzal Hossain

Pro-Vice-Chancellor, Dean

Faculty of Science & Technology

American International University- Bangladesh

DR. Carmen Z. Lamagna

Vice Chancellor

American International University- Bangladesh

Acknowledgments

We would like to thank our honourable supervisor, Md. Tohedul Islam, sir, for exposing us to this fascinating topic and mentoring us. His extensive expertise in this sector, intense attention, patience, and unwavering support enabled us to complete our task. His directions have made a significant contribution to every part of the thesis. Respected parents, honourable instructors, fellow students, and supportive friends are appreciated for sharing their expertise and thoughts that aided in the preparation of this thesis, as are all who participated in various ways throughout this thesis.

Abstract

In Bangladesh, suicide is among the most prevalent non-natural causes of death. Suicide has an impact on the victim as well as the immediate environment. Analyzing a person's risk of attempting suicide has always been quite perplexing. Due to the large number of social media users in Bangladesh, information gleaned from these platforms and health services can be crucial in determining the likelihood of suicide attempts. Social media data analysis using machine learning offers a viable method for identifying long-term contextual factors that increase a person's risk of having suicidal thoughts and actions. Our goal was to use publicly available Twitter data to forecast future likelihood of suicidal thinking. On Twitter data queried against psychological factors linked to suicide, such as burden, stress, loneliness, hopelessness, sleeplessness, sadness, and anxiety. An automated intelligence system can identify the likelihood of suicide attempts using social media by utilizing natural language processing and machine learning algorithms.

Keywords: contextual, burden, twitter data, psychological factors.

A.Table of Contents

Chapter – 1

1. Introduction

One of the most serious social health problems facing contemporary civilization is suicide. A variety of social and personal circumstances, including trauma or traumatic experiences, physical or mental disease, social isolation, hopelessness, anxiety, and more, can have an impact on suicidal behavior [1]. World Health Organization (WHO) estimates that about 800,000 people die by suicide yearly, perhaps one every 40 seconds [2]. Suicide is the third most common cause of death for people between the ages of 15 - 24, and the second most common cause of death for people between the aged 25 to 34. Despite improvements in the detection and treatment of serious mental diseases, suicide has remained an unsolvable public health issue. The development of suicide screening systems through the access and analysis of social media data is an expanding field [8]. The diversity of those who commit suicide and the lack of accurate, significant predictors of suicide make it difficult to prevent suicide [6]. Several clinical tools have been created recently to identify those who are at a high risk of suicide. The use of social media data for identifying suicide risk has gained attention as social science technology developed [3]. Social networking sites have come under fire for not doing enough to stop suicidal thoughts and actions. The implementation of an intervention that appeared to be hastily designed without adequate consideration of the social and cultural context in which a person experiencing suicidal behavior is embedded has been criticized as part of social media platforms' response to these accusations, which has included strengthening suicide detection mechanisms on the platforms [4]. Over time, it has been found that information from online social media, particularly tweets, contains predictive data for a variety of mental health conditions, such as depression and suicidal thoughts. The data available on Twitter can be used to examine people's suicide ideas [7]. The world has changed because of social media. It now becomes a part of our daily life, Associate in Nursing. Nowadays, a large number of people use popular social media sites like Facebook, Twitter, Instagram, etc. They submit pictures and blog entries on their daily activities, including their meals, clothes, and travels. Some people use social media more frequently than others, and vice versa. It provides copious data at a respectable speed, and services further adjust to user needs. For social communication, social engagement (between people or groups of people), and content creation, these will turn to the internet. social media sites like Facebook and Twitter are more frequently linked to events like harassment, cyberbullying, or even suicide. Therefore, in order to increase online suicide prevention, it is crucial to identify potential victims as soon as possible. For instance, the rappers Capital Steez and American Freddy E posted comments about suicide on their Twitter accounts. Social networks' ability to extract emotional ideas and depressive moods is, in fact, one of their best traits. Because of this, a lot of researchers use social networks to research suicide. For instance, Twitter has grown to be a highly well-liked social network where millions of people exchange their thoughts and emotions through brief texts called tweets that contain semantic expressions like emoticons, hashtags, special characters, etc. As a result, Twitter offers a wealth of data for text mining [5].

1.1. Problem statement

The goal of this study was to determine whether it was possible to reliably determine the level of concern for people's Twitter messages, often known as "tweets," that contained explicit or implicit textual or audio-visual references to suicidality. Human coders attempted to accomplish this using only the content of the tweet itself using a set of guidelines and classifications. After going through this procedure, this study sought to create and put into use an automatic computer classifier that could match the precision of the human coders. It was planned to use recall and precision measurements to assess the viability of this automated prediction.

1.2. Objective

This paper aimed to outline a method for evaluating users' risk of suicide on social media. Our goal was to investigate the behavioral, relational, and multimodal data that was gathered from Twitter and create machine learning models to identify at-risk people.

1.3. Research Questions

The objective of this paper is determined by visualizing several difficulties in the problem statement section, and this study reflects the response to the following questions:

1. How can we predict suicidal behavior in twitter data using machine learning?
2. What attribute should consider to predict suicidal behavior in twitter data using machine learning?

Chapter 2

2. Literature review

The issue of creating computers that learn automatically through use is addressed by machine learning. The convergence of computer science and statistics, as well as the foundation of artificial intelligence and data science, make it one of the technical domains with the fastest growth rates today. Machine learning has advanced recently as a result of the creation of new learning theories and algorithms as well as the continual explosion in the accessibility of online data and low-cost processing. Science, technology, and business have all adopted data-intensive machine-learning techniques, which has increased the use of evidence-based judgment in numerous fields such as marketing, manufacturing, health care, and financial modeling.

Leveraging Twitter to better identify [6] …., More than 305 million people use Twitter on a regular basis, and there are more than 500 million tweets posted per day as of 2015. They retrieved (571,995) dangerous tweets from 396,574 Twitter users between January 1, 2014, and April 15, 2015 using Twitter developer APIs. They also included 500 of each user's most recent publicly available tweets using Twitter REST API. They found the unsafe tweets by using search queries that included phrases or key words linked to the 12 suicide risk factors that were identified, such as "depressive feelings," "drug misuse," "self-harm," "suicide thoughts," "bullying," and "previous suicide attempts."

Suicide Prediction in Twitter Data using [9] ….., A Survey more tightly bound network than non-suicidal systems are suggested by the proposed result, which showed that while normal user availability measurements appear to indicate a standard network, the exchange of either following/follower connections or "mutual" connections between suicidal users is significantly greater (up to 73% rather than 42% in another survey). According to the reported path length retweets, the average shortest path value is larger than the preceding work with respect to suicidal contents, according to the proposed work. According to author Gualtiero B. Colombo et al., results were in the range of 5 as the average, although other authors reported values in the vicinity of 2 and 4.8. Results showed a normal of 5, but other writers' results varied between 2 and 4.8. These findings indicated more successful research in the area of suicidal thoughts.

Bridging big data and qualitative methods in the [10] …., They selected five high-profile suicide deaths that received a lot of press, either because the person was famous before they died or because of the circumstances surrounding their death, to analyze public discourses on social media. They were interested in the variety of responses—from lamentation and tributes to advocacy and actions—that these deaths provoked in open Twitter chats. 20 days after each death, they gathered five datasets of relevant Twitter posts. They used every label with consensus among the coders, yielding a dataset of 7.1k tweets. In a 10-fold cross-validation with around 50 training epochs in each experiment, the models achieved an average accuracy of 71%, and there were only slight increases after that. examining the model's capacity to forecast particular message classes They observe that the accuracy ranged from 60% for predicting "negative acts" to 89% for discriminating "activism," with an average accuracy of over 70% across all classifications. A maximum recall of 79% was attained for the class "activism," with the model being able to capture 69% of instances of each class on average.

Detecting suicidality [11] …., Twitter provides a public API that allows for the automatic collecting of tweets as they are sent out and filtered according to predetermined parameters. This API was employed by a tool created by the CSIRO between 18 February and 23 April 2014 to scan Twitter for any of the English terms or phrases listed below that are typical of the lingo used to describe suicidal ideation. "I'm suicidal; I'm going to kill myself; this is my suicide note; this is my suicide letter; I'm going to end my life; never wake up; it's not worth living; I'm ready to jump; I'd rather be dead; better off without me; better off dead; suicide plan; suicide pact; I'm tired of living; I don't want to be here; die alone; let me sleep forever." ………. ETC. 2000 (14%) of the 14,701 tweets that coincided with the suicide-related search phrases throughout the data collection period were chosen at random for human coding. 9% (n = 178) of the data were discarded or known, and were therefore omitted. In the combined data sets, A and B, 14% were classified as "very worrying," 57% as "potentially concerning," and 29% as "safe to ignore." The researchers classified the tweets and found that Set A had considerably more strongly worrisome' tweets than Set B (2 = 22.67, p .001).

Analysing the connectivity and communication [12] …, Twitter is a publicly available microblogging platform with 255 million active users worldwide who post an estimated 500 million Tweets every day. This leads in an incredibly noisy environment where posts cover a wide range of topics, which makes Twitter a viable source of data for a study into connectedness and dissemination of suicide ideation but also creates a noisy atmosphere. In order to consider a sufficient number of postings that can be identified as containing suicidal ideation, the data retrieved must first be pre-filtered. During four million posts were gathered from Twitter over a six-week period beginning on February 1st, 2014. As a complementary activity, they kept an eye on traditional media during the same time frame to find the names of young people in England who had attempted suicide (focusing on the adolescent range of 11–18 years old), and they then searched for and retrieved information from Twitter that included the name and surname of the deceased. Two leading experts on suicide used the "names" dataset to analyze the characteristics of the

Tweets and create a coding scheme for memorials, campaigns, information and support, and news reporting in addition to suicidal thoughts and thinking (which includes statements of complete despair).

Suicide ideation detection from online [1]...., A total of 188,704 English-language tweets using these search phrases were gathered from 2000 users using the Twitter APIs. Out of these tweets (37740/188704), tweets relating to 400 people were set aside for testing purposes. Let AU=1, 2..., n represent a collection of user U's social media sessions, where n represents the total number of sessions and I represent the session in question. Each social media session AU is made up of a variable indication and a series of posts designated by the letter P. $\theta\alpha$ = [0,1] where = 0 designates a typical session and = 1 signifies suicidal thoughts. The postings are listed in the following order: Ps = p 1, t 1 >, p 2, t 2 >, ... p m, t m >.

Advanced Daily Prediction Model for National Suicide [3] …, The national suicide rate over time Annual national suicide rates increased during the course of our study's 7 years (suicides per 100,000 people: 26 in 2008, 31 in 2009, 31.2 in 2010, 31.7 in 2011, 28.1 in 2012, 28.5 in 2013, and 27.3 in 2014), mirroring a general higher trend over the 20 years prior to 2010. (Figure 2). With a range of 16 to 72, the mean absolute daily suicide rate for this study was 38.83 (standard deviations=8.51).

Prediction of suicide using [5] …., The Twitter Dataset is retrieved. For their project, they leverage the Twitter API dataset. Then Pre-processing was used. After that, they ran an algorithm on a dataset from the Twitter API. Special Letters (), #tags (), and URL are removed (). With the use of the algorithm, the dataset's key information may be found by deleting special characters, URLs, and hashtags. They English terms taken from the dataset. They swap out emoji for words. The employ algorithms. Words like joyful, sad, and angry can be used in place of emojis to indicate the user's emotions. They use the dictionary words' roots to replace synonyms. This step's primary goal is to determine whether or not the post has any suicide-related content. They next used acronyms. GM stands for "Good Morning," etc. The N-gram Model was then used. next classified using classification.

Data Mining Approach to the Detection of Suicide [13] …., They used Reddit's API to accomplish data crawling. They used PRAW, a popular Python package, and a Reddit API wrapper to write Python programs to search and crawl posts and comments from Subreddit "/r/Singapore".385 posts and over 21,000 comments referencing depression and suicide-related phrases were obtained from Reddit. The information covers the years 2010 through 2018. They added new terms to this corpus, largely local phrases like "imh" (Institute of Mental Health in Singapore) and terms relating to Singapore's National Service. They then performed topic modeling using Latent Dirichlet Allocation (LDA), a method that automatically identifies topics. For subject modeling, we utilized the Genism module in Python. They can learn more about where to seek for disturbed people who need help by analyzing group discussions on depression and suicide. A big body of text was quickly divided into 14 distinct themes by topics modelling, each of which is individually identified by a collection of keywords.

Mining Twitter for Suicide [14] …, They found that nine topics—sadness/psychological harm, mental condition, despair, fear, loneliness, description of the third suicide attempt, insults, cyberbullying, and anorexia—are frequently brought up in conversations by suicidal people. They have collected whole utterances from suicides that have been shown to have taken place, such as "I want to die" and "You would be better off without me." Tweets were manually divided into two categories: dangerous tweets and non-risky tweets. We compared the performances of six classifiers (JRIP, IBK, IB1, J48, Naive Bayes, and SMO) in WEKA4 using a Leave One Out validation (LOO) and a 10-fold Cross-Validation (10-CV), with the results being averaged over 10 iterations. Results presentation through a web interface. The goal is to make it possible for psychiatrists to look for tweets that can indicate suicidal dangerous behavior, identify the profile's most recent tweets, and modify statistics.

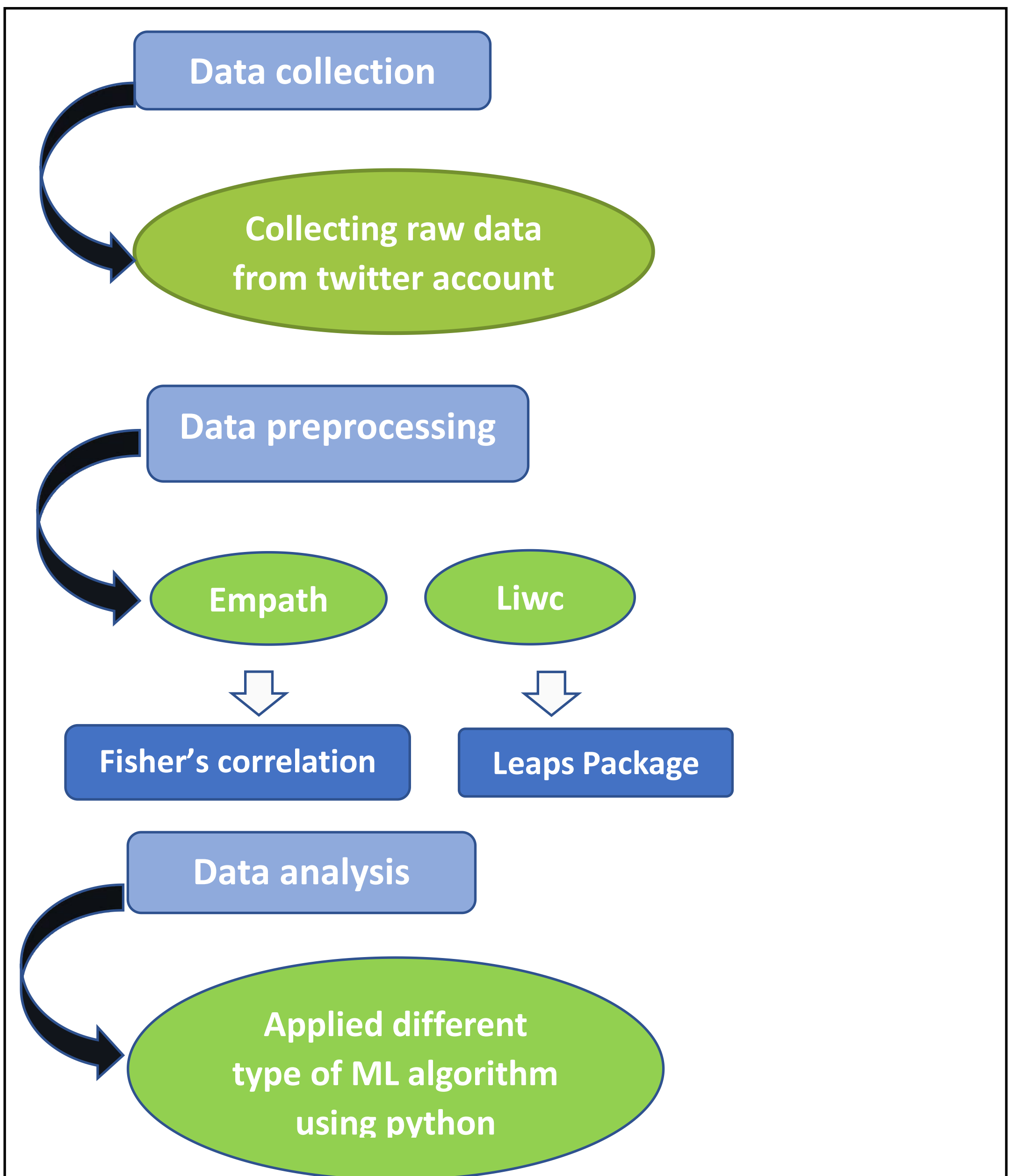
Numerous scholars have made predictions and conducted analyses based on these research publications using a variety of data analysis tools and approaches. To predict suicide, most researchers use analytical methods from different social media data. These articles have their distinctive data from different social media sites. Few have tried to provide some solution to predict suicide.

In these previous research works most of the data were collected from various social media sites. Python is used for data pre-processing. Some of these article data cleanings were missing. Data cleaning is primarily used to improve the quality of data. The main goal of data cleansing is to raise the caliber of the data. The majority of these papers utilized data directly. Additionally, some of these articles analyzed other prior research rather than gathering data. It is unable to make precise forecasts. The data in this research have first been pre-processed. The investigation was then correctly finished by meticulously annexing additional necessary measures.

Chapter 3

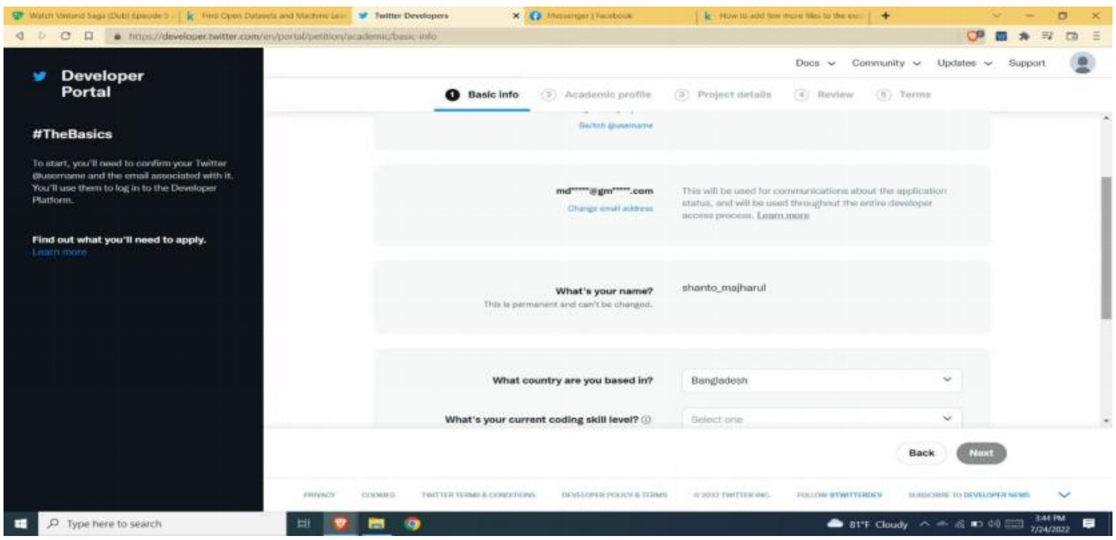
3. Methodology

The data needed to be collected first. Since our research focused on using Twitter data to predict suicide, As a result, we use Twitter's raw suicide data to compile the statistics. Thus, we seek assistance from the Twitter developer account. We make every effort to collect correct data. After a specific amount of time, Twitter approves us and gives us four types of keys: api key, api key secret, access token, and access token secret. With the use of this key, we were able to access the Twitter developer account for our implementation.



3.1 Data Collection:

We have gathered information about specific users who have committed suicide through our Twitter account. We first create a Twitter developer account for it. The next procedures include filling out the basic information, academic profile, project specifics, review, and terms.



They next ask for certain academic data to confirm that we genuinely want it for scholarly purposes.

developer-accounts <developer-accounts@twitter.com>　Thu, Jul 28, 11:48 PM

to me

Hello,

Thanks for your interest in building on Twitter!

To finish our review of your application, please reply to the following questions:

- What's your key use case or business purpose for using the Twitter APIs?
- Do you intend to analyze Tweets, Twitter users, or their content? If yes, how will you run this analysis?
- Will your use case involve Tweeting, Retweeting, or liking content? If yes, how will you interact with the Twitter accounts?
- Will you display Twitter content outside of Twitter? If yes, how and where will you display the content? Will Twitter content be displayed at row level or aggregated?

We appreciate your help!

Thanks,

Twitter

Hello,

Thanks for your response. We still need some more details for our review of your Twitter developer account application.

The information we still need includes:

- The core use case, intent, or business purpose for your use of the Twitter APIs.
  - Please note, "business purpose" in this context includes uses not necessarily connected to a commercial business. We require information about the problem, user story, or the overall goal your use of Twitter content is intended to address.
  - If you are a student, learning to code, or just getting started with the Twitter APIs, please provide details about potential projects, or areas of focus.
- If you intend to analyze Tweets, Twitter users, or their content, please share details about the analyses you plan to conduct, and the methods or techniques.
  - Note that "analyze" in this context includes any form of processing performed on Twitter content. Please provide as detailed and exhaustive an explanation as possible of your intended use case.
- If your use involves Tweeting, Retweeting, or liking content, share how you will interact with Twitter users or their content.
- If you'll display Twitter content off of Twitter, please explain how, and where, Tweets and Twitter content will be displayed to users of your product or service, including whether Tweets and Twitter content will be displayed at row level, or aggregated.

To provide the information, please respond to this email. Where possible, please share links to illustrations, or sample work products.

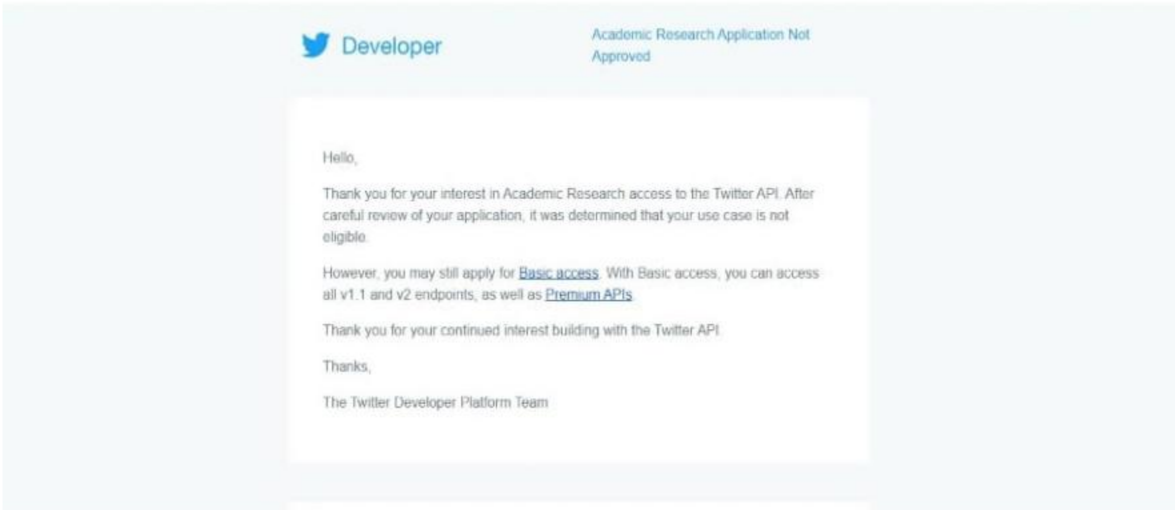
And keep in mind, we can't view attachments.

If we don't receive the information we need, your application will not be accepted.

We appreciate your help!

Thanks,

Finally, they gave us the authority to access their account.



Developer

Academic Research Application Not Approved

Hello,

Thank you for your interest in Academic Research access to the Twitter API. After careful review of your application, it was determined that your use case is not eligible.

However, you may still apply for Basic access. With Basic access, you can access all v1.1 and v2 endpoints, as well as Premium APIs.

Thank you for your continued interest building with the Twitter API.

Thanks,

The Twitter Developer Platform Team

We apply the Twitter API key, API key secret, access token, and access token secret after receiving approval from Twitter to access those user data tweets.

	username	text
0	Esta84M	Anthony Bourdain's Hungry Ghosts WS6QRRS https://t.co/iWZUuKyHIP https://t.co/BWjZUPqyMV
1	coffee6727	Leonard Cohen Anthony Bourdain David Bowie Or Prince if he's bringing pancakes https://t.co/Z1n2utqvYC
2	expertsmap	Morgan Fallon — 10 Years on the Road with Anthony Bourdain, 9 Emmy Nominations, Lessons from Michael Mann, Adventures with Steven Rinella, High Standards, Wisdom from West Virginia, and More (#597) https://t.co/kgk1ubFHNr
3	uncleDopey	Joe That guy Anthony Bourdain (the rapper) My drummer My useless housekeeper My parole officer This old folks home This doctor
4	lastralive	Anthony Bourdain treats you to cuisine from all parts unknown on https://t.co/pPE9Qb5y34 Updated every day Watch on any device from anywhere Travel Channel on https://t.co/bOghjiTreg Commercial Free #OnlineTV https://t.co/R6HNOzgthj
5	james80480694	If not the perfect meal, this was, in many ways, a perfect one. Good food, good company, exotic ambience, and an element of adventure. --Anthony Bourdain
6	Eric_Arroyo_88	@Cobratate Robin Williams,Anthony Bourdain, Chester Bennington...the list is long of rich people that didn't found happiness, and yet i have met people like my grandparents that live in Puerto Rico back in 1930s in houses without floor and where the happiest people i ever met.
7	BendlerJoy	@CPhenom69 @AdamSchiff Ask Anthony Bourdain! Oh right he " suicided" himself 🤔
8	patleenichols	Everybody, if you're a fan of Anthony Bourdain, you'll likely be familiar with his quote about Cambodia; wanting to strangle Kissinger with your bare hands. Now, most of us aren't in a position to, so why not support Tabitha, a foundation for helping the poor, victims, and water.
9	KeremAvclBaba	Exploring the Far West Texas | Anthony Bourdain: Parts Unknown | All Documentary 2023 New #Documentary https://t.co/cP4Ankv7gs
10	BlackFlagPilled	@atzmusshrugged 8 and 11 and I'm eating with Anthony Bourdain and its whatever food and alcohol is brought to the table
11	KWholesaler	@WineJerk (Giving the devilish little Anthony Bourdain grin) Fuck it. Bottoms up!
12	splashfalcon3	anthony bourdain is the sexiest human alive
13	orphancore	@PalmerReynolds anthony bourdain was french not italian 😔
14	playboizachti	@theymerRights this is really funny and it's anthony bourdain
15	LadPsycho	Anthony Bourdain clears Gordon Ramsay https://t.co/JSZ8A1aXjT
16	kamalakkannan	Faves of 2022: Books — 'Down and Out in Paradise: The Life of Anthony Bourdain'. Also, @maggieNYT's 'Confidence Man'. @SebastianEPayne's 'The Fall of Boris Johnson'. (There's clearly a theme).
17	BcdaGR81	RIP Anthony Bourdain…I pray I dont have the same fate 🙌🏾
18	deedeedaddle	It has come to my attention that these are not all photos of Anthony Bourdain https://t.co/4a6SORX7W6
19	reveraissance	anthony bourdain; 11H cancer sun • cancer is associated with the stomach • 11H is the space of the unknown, unfamiliar (public arena) • prolific career as a chef + author of international culture and cuisine even includes the show 'parts unknown'
20	4thBabyPapa	@wethincracker69 Anthony Bourdain tried it once he said it was fetch
21	megantheevalium	Described him as "Anthony bourdain meets Joe rogan"
22	locuradesabado	@yyentaa My old landlord looked like Anthony bourdain but like the Greek version.
23	WorldPonzi	@confucian_the The tasteless displays of wealth was accurate but they left out the abuse. Anthony bourdain in SG showed a glimpse into what the average light skin Chinese Singaporean thinks of their SE Asian neighbors and how they treat them like domestic slaves to be used for their liking.
24	RootsOfCombat	Anthony Bourdain & Diaz Brothers https://t.co/N6G4x0nrnj
25	wormoffabean	im litteraly Anthony Bourdain
26	FreeSpeech0517	@Elex_Michaelson @AdamSchiff @GOPLeader @RepMTG Visiting Epsteins Island, Anthony Bourdain, Lying in the corrupt halls of Congress, 256 emails between you and Jeffrey, putting our National security at risk - all good reasons to remove you Adam Shifty Schiff.
27	thefishhasturnd	AND she wants to watch anthony bourdain with me???
28	thewondering1	Some truth from Anthony Bourdain. If you need anything from me at 4AM I will most likely be fairly useless. https://t.co/gVLLDfQsPF
29	Yugisfriend	Bobby flay julia child anthony bourdain guy fieri thomas keller emeril lagasse david chang paula deen - cope https://t.co/8HbVaNe8Zf
30	formerscenegirl	Anthony Bourdain was my ideal man https://t.co/RP5zij659l

Fig 1: Collected data from twitter developer account

3.2 Data Pre-Processing:

Preprocessing is required following data collection. Data must be cleaned up and standardized because datasets typically contain unnecessary information. Our data focuses on user text, comments, and statuses related to suicide.

```
In [26]: import pandas as pd
         import copy
         import numpy as np
         import matplotlib.pyplot as plt
         import seaborn as sns
         from sklearn.metrics import classification_report

In [41]: dataset = pd.read_csv('./Data1/YES-for-fishers.csv')

In [3]:  dataset.info()

         <class 'pandas.core.frame.DataFrame'>
         RangeIndex: 50 entries, 0 to 49
         Columns: 195 entries, help  to class
         dtypes: float64(192), int64(2), object(1)
         memory usage: 76.3+ KB

In [42]: dataset.head(10)

Out[42]:
```

	help	office	dance	money	wedding	domestic_work	sleep	medical_emergency	cold	hate	...	children	monster	ocean	giving
0	0.000000	0.000000	0.164608	0.000000	0.079193	0.039608	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.0	0.178122
1	0.111742	0.047619	0.000000	0.000000	0.000000	0.219048	0.000000	0.095238	0.037037	0.000000	...	0.250534	0.000000	0.0	0.000000

Fig 2: Pre-processing part 1

```
In [3]:  dataset.info()

         <class 'pandas.core.frame.DataFrame'>
         RangeIndex: 50 entries, 0 to 49
         Columns: 195 entries, help  to class
         dtypes: float64(192), int64(2), object(1)
         memory usage: 76.3+ KB

In [42]: dataset.head(10)

Out[42]:
```

	help	office	dance	money	wedding	domestic_work	sleep	medical_emergency	cold	hate	...	children	monster	ocean	giving
0	0.000000	0.000000	0.164608	0.000000	0.079193	0.039608	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.0	0.178122
1	0.111742	0.047619	0.000000	0.000000	0.000000	0.219048	0.000000	0.095238	0.037037	0.000000	...	0.250534	0.000000	0.0	0.000000
2	0.000000	0.000000	0.450617	0.000000	0.075103	0.000000	0.075103	0.000000	0.075103	0.150206	...	0.225309	0.000000	0.0	0.075103
3	0.127852	0.000000	0.042617	0.085235	0.298321	0.085235	0.085235	0.042617	0.213087	1.235902	...	0.596642	0.085235	0.0	0.383556
4	0.000000	0.045390	0.045390	0.000000	0.090780	0.000000	0.045390	0.000000	0.000000	0.090780	...	0.000000	0.000000	0.0	0.000000
5	0.056785	0.000000	0.227141	0.000000	0.056785	0.000000	0.227141	0.000000	0.113571	0.000000	...	0.056785	0.000000	0.0	0.170356
6	0.000000	0.000000	0.053241	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.0	0.000000
7	0.000000	0.000000	0.000000	0.000000	0.047945	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.047945	0.000000	0.0	0.000000
8	0.171044	0.057015	0.085522	0.085522	0.142537	0.114029	0.000000	0.028507	0.000000	0.114029	...	0.142537	0.000000	0.0	0.313581
9	0.128463	0.128463	0.171285	0.299748	0.042821	0.042821	0.042821	0.042821	0.042821	0.128463	...	0.128463	0.042821	0.0	0.256927

Fig 3: Pre-processing part 2

Therefore, we first analyze our text utilizing the empath tool for preparation. Empath is a tool that can be used to analyze text using several lexical categories (similar to LIWC) and also create new lexical categories that may be used in the analysis.

```
[ ] from empath import Empath
    lexicon = Empath()

lexicon.analyze("@untbaqal @LabaikaGroup Yes, I am ready for the adventure in Gulmarg. Adjust me for the year end."
"Very soon, micro robots will be delivering medication or killing the target cancerous tissue. They will be programmed or guided externally towards tl
"@tariqtramboo If right we must follow"
"@tariqtramboo Because their r 90% of heating instruments r fake in the market which becomes easily reason of fire."
"@tariqtramboo Bt I had heard that Prophet Muhammad (PBUH) didn't know how to write or read..Probably written by some Sahaba"
"Choose wisely ðŸ˜œ https://t.co/IyG1hTBKSx"
"@tariqtramboo Itz  mostly bekoz of heating equipments with high wattage used in sockets with wrong size of wiring even I saw this type of short circu
"@tariqtramboo Why does Kashmir has more respiratory infections and COPD cases even if the air quality is good?"
"The main cause of fire during winters in Kashmir ( what I have observed) is due to carelessness!"
"Maximum fire incidents I have witnessed are because the electricity was shut and the inhabitants forgot to switch off the heating gadgets!"
"(Room heaters and blowers, particularly)! https://t.co/PGeEoFYik2"
"@sara_stone87 Nasty situation. Thanks to #Zelensky"
"These vegetables mostly are being produced at the cost of high fertilizers usage. Point is how healthy is that. https://t.co/zgLbV3jRkE"
"@tariqtramboo Where is potato ðŸ˜ªðŸ˜ª"
"@tariqtramboo Problem is #Kashmiris in general will never acknowledge the reality. The reason behind it is #Pakistan. But #Kashmiris will never ackno
"@tariqtramboo I can only see capsicums ðŸ˜¬ðŸ˜¬ðŸ¤¢ðŸ¤¢"
"Situation is worsening day by day.. https://t.co/SUgyr0PCfD"
"@tariqtramboo Boil tomatoes ðŸ¤¢"
```

Fig 4: Analyzing suicide user text using python

masculine	prison	health	pride	dispute	nervousn	government	weakness	horror	swearing_	leisure	suffering	royalty	wealthy	tourism	furniture	school	magic
0	0	0	0	0	0	0	0	0	0.052632	0.059608	0	0	0.07493	0	0	0	0
0.08465609	0.047619	0.157738	0	0	0.037037	0.051587302	0	0.020833	0	0.037037	0.020833	0	0.022222	0.084656	0.084656	0.081046	0.038462
0.07510288	0	0	0	0	0.075103	0	0	0	0.150206	0	0.150206	0	0	0.150206	0	0.075103	0
0.21308653	0.042617	0	0	0.170469	0.213087	0	0.042617	0.042617	0.298321	0.170469	0.511408	0	0	0.042617	0	0.085235	0.042617
0	0	0	0.09078	0	0	0	0	0	0	0.04539	0.09078	0	0	0	0.04539	0.04539	0
0.05678537	0	0	0	0	0.056785	0	0	0.056785	0	0.056785	0.170356	0	0	0	0.056785	0	0
0	0	0	0	0	0	0	0	0	0	0	0.053241	0.106481	0	0	0	0.212963	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.02850737	0.028507	0.028507	0.057015	0.285074	0.028507	0.142536835	0.028507	0.114029	0	0.199552	0.171044	0.057015	0.085522	0.028507	0.028507	0.285074	0.028507
0.12846348	0.085642	0	0	0.085642	0.042821	0.042821159	0	0.171285	0.085642	0.128463	0.085642	0.042821	0.128463	0	0	0.214106	0
0	0	0.226415	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0.004065	0	0.004065	0	0.004065041	0	0	0	0	0.00813	0.004065	0	0	0	0	0
0	0.098765	0	0	0.049383	0	0.049382716	0	0	0	0	0	0	0	0	0	0.049383	0
0	0	0	0	0	0	0.051587302	0	0	0	0	0	0.051587	0	0	0	0	0
0	0	0.040541	0	0	0	0	0	0	0	0	0.081081	0	0.040541	0	0	0	0
0	0	0.11544	0	0.05772	0.11544	0.057720058	0.05772	0.11544	0.2886	0	0.519481	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0.05	0	0	0	0	0

Fig 5: Getting value by analyzing suicide user text

real_estate	home	divine	sexual	fear	irritability	superhero	business	driving	pet	childish	cooking	exasperat	religion	hipster	internet	surprise	reading
0	0	0.130252	0	0	0	0	0.135714	0	0	0	0	0	0	0	0	0.058824	0.105062
0.12136752	0.194383	0.181818	0.027778	0	0	0	0.027778	0	0	0.194444	0.710879	0	0.090909	0	0	0	0.044643
0	0.075103	0.150206	0.075103	0	0	0	0.075103	0	0	0	0.075103	0	0	0.075103	0.075103	0	0.150206
0	0.213087	0.127852	0.213087	0.042617	0	0	0.042617	0.170469	0.127852	0.255704	0	0	0	0.127852	0.340938	0.042617	0.213087
0	0	0.09078	0	0	0	0	0.04539	0	0	0	0	0	0.13617	0	0	0.04539	0
0	0.056785	0	0	0	0	0	0.113571	0.113571	0.056785	0.056785	0.170356	0	0	0	0	0	0.113571
0	0.106481	0	0	0	0	0	0.053241	0	0	0	0	0	0	0.053241	0.053241	0	0.053241
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.05701473	0.256566	0.114029	0.028507	0.142537	0	0.02850737	0.370596	0.028507	0	0	0.057015	0	0.057015	0.028507	0.114029	0	0.142537
0.25692695	0.042821	0.085642	0	0.128463	0	0.04282116	0.342569	0.042821	0	0.042821	0.042821	0	0.042821	0.042821	0.085642	0.128463	0.171285
0	0	0	0	0	0	0	0	0	0	0.45283	0	0	0	0	0	0	0
0	0	0.004065	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.004065
0	0	0.049383	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.049383
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0.040541	0	0	0.040541	0	0	0	0
0	0	0	0.23088	0.11544	0	0	0.11544	0.05772	0.05772	0.05772	0.05772	0	0	0.05772	0	0.11544	0.05772
0	0	0	0	0	0	0	0	0	0	0	0	0	0.046875	0	0	0	0
0	0.046931	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.046931
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig 6: Getting value by analyzing suicide user text

surprise	reading	worship	leader	independ	movemer	body	noise	eating	medieval	zest	confusion	water	sports	death	healing	legend	heroic	celebratic
0.058824	0.105062	0	0.052632	0	0.119216	0.066667	0	0.066667	0	0	0	0	0.145833	0	0	0	0.0625	0.170102
0	0.044643	0.090909	0	0	0	0.060847	0	0.512902	0	0	0.05787	0	0.022222	0.209957	0.139423	0.02381	0.0625	0
0	0.150206	0	0	0	0.150206	0.075103	0.075103	0.075103	0	0	0	0	0.075103	0.225309	0.075103	0	0.150206	0.225309
0.042617	0.213087	0	0.127852	0	0.085235	0.255704	0.042617	0	0	0	0.085235	0	0.042617	0.63926	0.085235	0.085235	0.127852	0.255704
0.04539	0	0.09078	0.04539	0	0	0	0	0	0	0	0.04539	0	0.04539	0.13617	0	0	0.04539	0.04539
0	0.113571	0	0.056785	0	0	0.113571	0.170356	0.170356	0	0	0.113571	0	0.056785	0.170356	0.056785	0	0	0.056785
0	0.053241	0	0	0	0	0	0	0	0	0	0.053241	0	0	0.159722	0	0	0.159722	0.106481
0	0	0	0	0	0	0	0	0	0	0	0.047945	0	0	0.047945	0	0	0	0
0	0.142537	0.114029	0.256566	0	0.085522	0	0	0.085522	0	0.028507	0.199552	0	0.171044	0.057015	0.142537	0.057015	0.057015	0.228059
0.128463	0.171285	0.042821	0.171285	0	0.256927	0	0	0.042821	0	0.042821	0.214106	0	0.128463	0.171285	0.128463	0.042821	0.128463	0.214106
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0.004065	0.004065	0.004065	0	0	0.004065	0	0	0	0	0	0	0	0.00813	0.004065	0	0.004065	0
0	0.049383	0	0.049383	0	0.098765	0	0	0	0	0	0.049383	0	0	0.049383	0	0	0.049383	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.051587	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.162162	0	0	0.162162	0
0.11544	0.05772	0	0.11544	0	0.05772	0.05772	0	0.17316	0	0.05772	0.05772	0	0.05772	0.2886	0.05772	0	0.05772	0
0	0	0	0	0	0	0	0	0	0	0	0.046875	0	0	0	0	0	0	0
0	0.046931	0	0	0	0	0	0	0	0	0	0	0	0	0.046931	0	0	0.093863	0
0	0	0	0	0	0	0	0	0	0.05	0	0	0	0	0	0	0	0	0

Fig 7: Getting value by analyzing suicide user text

help	office	dance	money	wedding	domestic_	sleep	medical_emergency	cold	hate	cheerfuln	aggressio	occupatio	envy	anticipati	family	vacation	crime
0	0	0.164608	0	0.079193	0.039608	0	0	0	0	0	0	0	0	0	0	0	0.045455
0.111742	0.047619	0	0	0	0.219048	0	0.095238096	0.037037	0	0	0	0.194444	0	0	0.137897	0.037037	0.075397
0	0	0.450617	0	0.075103	0	0.075103	0	0.075103	0.150206	0	0	0.075103	0	0	0.225309	0.075103	0.075103
0.127852	0	0.042617	0.085235	0.298321	0.085235	0.085235	0.042617305	0.213087	1.235902	0.042617	0.170469	0.042617	0.170469	0	0.383556	0.170469	0.042617
0	0.04539	0.04539	0	0.09078	0	0.04539	0	0	0.09078	0	0	0.04539	0.04539	0	0	0	0
0.056785	0	0.227141	0	0.056785	0	0.227141	0	0.113571	0	0.056785	0.056785	0	0	0.056785	0.056785	0	0
0	0	0.053241	0	0	0	0	0	0	0	0	0	0.053241	0	0	0.106481	0	0.053241
0	0	0	0	0.047945	0	0	0	0	0	0	0.047945	0	0	0	0.047945	0	0.047945
0.171044	0.057015	0.085522	0.085522	0.142537	0.114029	0	0.028507367	0	0.114029	0	0.057015	0.114029	0.028507	0	0.142537	0.028507	0.228059
0.128463	0.128463	0.171285	0.299748	0.042821	0.042821	0.042821	0.042821159	0.042821	0.128463	0.042821	0.171285	0.085642	0	0	0.042821	0.085642	0.085642
0	0	0	0	0	0	0	0.226415094	0	0	0	0	0	0	0	0	0	0
0.004065	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00813
0	0	0.049383	0	0	0	0	0	0.049383	0	0	0.098765	0	0	0	0.049383	0.049383	0.098765
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0.040541	0	0	0	0	0	0.040541	0	0	0.040541	0	0
0.05772	0	0.11544	0.05772	0.05772	0.05772	0	0.115440115	0.11544	0.2886	0.05772	0.05772	0.05772	0	0	0.17316	0	0.05772
0	0	0	0	0	0	0	0	0	0.046875	0	0	0	0.046875	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.046931	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig 8: Getting value by analyzing suicide user text

help	office	dance	money	wedding	domestic	sleep	medical_(	cold	hate	cheerfuln	aggressio	occupatio	envy	anticipati	family	vacation	crime	attractive
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1
0	0	2	0	1	0	0	0	0	1	0	0	1	0	0	2	0	0	0
1	2	1	7	1	3	0	1	0	0	1	1	1	0	0	3	1	3	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	10	0	1	1	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	5	2	0	0	0	0	0	0	1	1	0	0	0	2	1	0
7	0	0	13	2	6	0	0	1	2	0	1	4	1	0	8	0	3	1
3	2	0	1	5	0	5	0	0	1	1	3	3	0	0	15	2	3	3
0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	2
0	0	1	0	0	0	2	0	0	1	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	1	2	2	1	1	0	2	0	1	0	0	1	4	1	2	0
0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	3	0	2	2
0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	2	0	1	3	1	2	2	3	3	0	0	0	0	0	7	1	1	4
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0

Fig 9: Getting value by analyzing non- suicide user text

swearing	leisure	suffering	royalty	wealthy	tourism	furniture	school	magic	beach	journalisr	morning	banking	social_me	exercise	night	kill	blue_colla	art
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1
1	1	1	1	3	2	0	2	0	0	0	0	5	3	0	0	1	0	2
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	2	0	1	1	0	2	0	2	1	0	6	0	0	1	1	0	0
0	1	0	2	4	1	0	1	0	0	3	0	13	5	0	0	0	1	1
1	2	2	4	2	1	1	0	1	0	0	5	1	3	0	2	2	0	1
0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	1	0	0	0
0	1	1	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	3	5	0	0	0	1	1	1	0	2	1	1	0	0	1	2	0	0
0	1	0	0	0	1	0	4	0	0	4	0	1	4	0	0	2	0	1
1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	3	1	0	1	0	2	0	0	2	2	0	5	0	0	0	0	0
0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1

Fig 10: Getting value by analyzing non- suicide user text

stealing	real_esta	home	divine	sexual	fear	irritability	superhero	business	driving	pet	childish	cooking	exasperat	religion	hipster	internet	surprise	reading
0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2	2	0	0	0	0	0	7	3	2	2	2	0	1	0	3	0	3
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	2	0	0	0	0	0	1	2	1	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
1	1	0	0	1	0	0	0	3	0	0	0	0	0	1	0	0	0	2
3	7	5	1	1	2	1	0	9	0	0	0	2	1	2	1	5	0	3
2	2	4	2	0	0	0	0	16	0	0	5	0	0	1	1	2	0	2
0	0	0	0	0	0	0	0	0	0	0	3	0	0	2	0	6	0	1
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1
1	0	2	1	2	3	0	0	0	0	2	1	2	0	0	0	1	0	1
2	1	2	0	1	0	1	0	3	1	0	0	0	1	0	0	1	0	8
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	8	1	3	0	0	6	1	0	2	1	0	6	3	5	0	1
0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0

Fig 11: Getting value by analyzing non- suicide user text

healing	legend	heroic	celebratic	restauran	violence	programr	dominant	military	neglect	swimming	exotic	love	hiking	communi	hearing	order	sympathy	hygiene
0	1	1	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1	1	0
0	0	2	2	1	2	2	0	0	0	0	0	0	0	7	3	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	2	0	0	0	0	0	0	1	0	1	0	0	0	0
0	0	0	0	0	2	1	0	0	1	0	0	0	0	0	0	0	0	0
0	1	0	1	0	1	0	1	0	0	0	1	0	0	5	2	0	0	0
2	3	4	2	2	2	3	5	2	0	0	0	2	0	10	4	2	0	1
3	2	6	3	0	2	1	19	1	0	0	0	2	0	1	1	18	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	1
1	0	0	0	0	0	0	0	1	1	0	0	0	0	1	2	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	2	2	2	2	0	0	0	1	0	2	0	2	1	0	2	0
0	3	1	2	0	0	1	2	2	0	0	0	0	0	8	2	2	0	0
0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	1	0	1	3	1	0	1	1	0	0	7	0	2	3	2	2	0
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig 12: Getting value by analyzing non- suicide user text

listen	urban	shopping	disgust	fire	tool	phone	gain	sound	injury	sailing	rage	science	work	appearan	valuable	warmth	youth	sadness
0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	1	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1
5	0	0	0	3	1	3	5	4	1	0	0	0	0	1	5	1	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	11
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
2	0	0	0	0	0	1	2	0	0	1	1	0	0	0	3	0	0	0
3	2	2	1	0	0	3	9	1	0	0	0	0	5	2	8	0	1	2
1	1	1	0	0	0	1	0	1	0	0	0	11	14	3	0	0	3	1
0	0	1	0	0	0	2	0	0	0	0	0	1	0	2	0	0	0	0
2	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	1	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0
0	0	1	0	0	0	1	0	0	2	0	0	1	0	2	1	0	1	4
2	0	1	0	1	0	2	1	0	1	1	0	1	1	1	0	1	2	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	2	1	2	2	0	0	2	5	3	0	0	6	4
0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0

Fig 13: Getting value by analyzing non- suicide user text

Using this empath tool helps us understand the meaning of the text we're analyzing. There are 194 attributes in an empath. As a result of using Python, each user can now see what type of text is being used. Both suicide attempters and non-suicide attempters used the same algorithm. Finally, we use Fisher's correction to determine the coefficiency. The outcome was then revealed.

Classification Function Coefficients		
	class	
	.00	1.00
help	-27.224	6.680
office	-21.813	26.738
dance	17.471	0.751
money	11.579	-9.287
wedding	1.814	4.280
domestic_work	56.450	-6.966
sleep	5.342	-0.787
medical_emergency	-2.505	-2.975
cold	-40.648	5.853
hate	31.227	0.189
cheerfulness	2.962	-6.727
aggression	9.708	-3.121
occupation	-4.068	20.731
envy	-71.812	23.457
anticipation	17.166	-13.298
family	-13.619	-12.422
vacation	7.372	-10.105
crime	33.306	8.247
attractive	30.250	2.563
masculine	-9.502	0.324
prison	-43.770	9.186
health	0.103	1.694
pride	-4.058	4.402
dispute	-24.122	-2.563
government	6.927	2.945
weakness	-19.799	-4.925

Fig 14: Getting value by after analyzing Fisher's correlations

Classification function coefficient		
	class	
	0	1
health	0.103	
masculine		0.324
hate		0.189
sleep		-0.787
dance		0.751

Fig 15: Getting coefficient value by after analyzing Fisher's correlations

By examining the fisher's coefficient value, we can determine that it will be effective if the value is between -1 and 1. Finally, we discovered the truth: issues with masculinity, hatred, sleep, and dancing are some of the factors that contribute to suicide. On the other side, those who do not utilize suicide only have health problems.

LIWC Tool:

After analyzing our text utilizing the empath tool for preparation then we used LIWC tools. The LIWC acronym stands for "Linguistic Inquiry and Word Count". LIWC is a tool that is used as a text analysis tool to study language and its relationship to psychological and emotional states. It quantifies the usage of various linguistic categories in a given text, such as words related to effect, cognition, time, and more, and can be used in a range of applications, such as social media analysis, sentiment analysis, and mental health research.

After using the text, we got from Twitter we apply those text to the LIWC tool and then we find this information for one user in figure-12.

Filename	WC	WPS	Qmarks	Unique	Dic	Sixltr	funct	pronoun	ppron	i	we	you	she
Alexander	1116	34.88	0.27	44.09	40.05	24.19	23.75	5.47	3.14	1.97	0	0.27	

Figure 16: LIWC output for one user.

We collect output for all 25 user who commit suicide and we also collect output for another 25 users who didn’t commit suicide. After collecting all users output, we merged all users in one file who commit suicide which given in figure 13.

class	WC	WPS	Qmarks	Unique	Dic	Sixltr	funct	pronoun	ppron	i	we	you	shehe	they	ipron
1	1116	34.88	0.27	44.09	40.05	24.19	23.75	5.47	3.14	1.97	0	0.27	0.09	0.81	2.3
1	95	23.75	2.11	66.32	65.26	25.26	45.26	13.68	11.58	6.32	0	2.11	2.11	1.05	2.1
1	1466	61.08	0.41	44.41	38.27	18.62	21.62	6.55	4.84	3.68	0.2	0.55	0.41	0	1.7
1	2808	29.56	0.89	35.93	61.25	21.94	40.38	11.57	7.76	1.14	0.25	2.78	2.81	0.78	3.8
1	883	44.15	0.23	34.2	51.19	24.46	23.9	4.64	2.83	1.02	0.23	0.45	1.02	0.11	1.8
1	1278	35.5	0.16	42.72	47.89	27	23.4	5.63	4.07	2.74	0.16	0.55	0.39	0.23	1.5
1	542	45.17	0.55	40.77	50.74	18.45	25.28	7.01	3.51	1.66	0.18	0.37	0.74	0.55	3.5
1	183	30.5	1.64	69.95	45.36	28.42	29.51	7.65	2.19	0	0	0.55	1.09	0.55	5.4
1	3651	25.35	0.77	33.14	56.78	25.36	38.48	10.55	6.27	1.42	1.07	2.47	0.63	0.68	4.2
1	2434	34.77	0.29	46.47	41	23.13	22.47	5.79	3.99	1.68	0.41	1.03	0.62	0.25	1.8
1	179	179	0	22.91	62.01	16.2	41.9	16.76	3.35	3.35	0	0	0	0	13.4
1	655	29.77	2.75	31.3	51.45	27.33	25.95	3.97	3.21	0	0.31	0	2.6	0.31	0.7
1	580	29	0.17	50.52	49.14	21.21	33.1	9.48	6.03	2.24	1.21	0.34	2.07	0.17	3.4
1	311	311	0	27.01	10.93	15.43	5.14	1.29	0.64	0	0	0	0	0.64	0.6
1	756	50.4	0	35.85	32.54	25.26	16.8	4.1	2.91	1.72	0	0	1.19	0	1.1
1	848	42.4	0.35	50.47	51.06	25.24	32.9	6.13	2.71	0.94	0.12	0.47	0.83	0.35	3.4

Figure 17: Combine all users LIWC output who commit suicide.

Leaps Package:

Then we move to our next step which is the Use of the Leaps Package. The LEAPS (Linear Estimation and Prediction in Social Science) package is a collection of functions for conducting statistical analysis in R. The package also provides a range of visualizations, such as partial residual plots, that can help researchers better understand the relationships between variables in their data. The package is widely used in social science research, and its functions have been validated in a range of studies and applications.

We apply our LIWC output data in Leaps Package so that we could find out which word the users who commit suicide use most. Here we change the nvmax value (the maximum number of predictors to incorporate in the model) so that we could find a different result. We set our nvmax value from 4 to 10 to check the result. In Figures 14 and 15, examples are provided where the value of nvmax is 4 and 8 respectively.

```
1 subsets of each size up to 4
Selection Algorithm: exhaustive
         WC  WPS Qmarks Unique Dic Sixltr funct pronoun ppron i   we  you shehe
1  ( 1 ) "*" " " " "    " "    " " " "    " "   " "     " "   " " " " " " " "  
2  ( 1 ) "*" "*" " "    " "    " " " "    " "   " "     " "   " " " " " " " "  
3  ( 1 ) "*" "*" "*"    " "    " " " "    " "   " "     " "   " " " " " " " "  
4  ( 1 ) "*" "*" "*"    "*"    " " " "    " "   " "     " "   " " " " " " " "  
         they ipron article verb auxverb past present future adverb preps conj
1  ( 1 ) " "  " "   " "     " "  " "     " "  " "     " "    " "    " "   " " 
2  ( 1 ) " "  " "   " "     " "  " "     " "  " "     " "    " "    " "   " " 
3  ( 1 ) " "  " "   " "     " "  " "     " "  " "     " "    " "    " "   " " 
4  ( 1 ) " "  " "   " "     " "  " "     " "  " "     " "    " "    " "   " " 
         negate quant number swear social family friend humans affect posemo
1  ( 1 ) " "    " "   " "    " "   " "    " "    " "    " "    " "    " "   
2  ( 1 ) " "    " "   " "    " "   " "    " "    " "    " "    " "    " "   
3  ( 1 ) " "    " "   " "    " "   " "    " "    " "    " "    " "    " "   
4  ( 1 ) " "    " "   " "    " "   " "    " "    " "    " "    " "    " "   
         negemo anx anger sad cogmech insight cause discrep tentat certain
1  ( 1 ) " "    " " " "   " " " "     " "     " "   " "     " "    " "    
2  ( 1 ) " "    " " " "   " " " "     " "     " "   " "     " "    " "    
3  ( 1 ) " "    " " " "   " " " "     " "     " "   " "     " "    " "    
4  ( 1 ) " "    " " " "   " " " "     " "     " "   " "     " "    " "    
```

Figure 18: Using Leaps Package where nvmax=4.

```
1 subsets of each size up to 8
Selection Algorithm: exhaustive
         WC  WPS Qmarks Unique Dic Sixltr funct pronoun ppron i   we  you shehe
1  ( 1 ) "*" " " " "    " "    " " " "    " "   " "     " "   " " " " " " " "  
2  ( 1 ) "*" "*" " "    " "    " " " "    " "   " "     " "   " " " " " " " "  
3  ( 1 ) "*" "*" "*"    " "    " " " "    " "   " "     " "   " " " " " " " "  
4  ( 1 ) "*" "*" "*"    "*"    " " " "    " "   " "     " "   " " " " " " " "  
5  ( 1 ) "*" "*" "*"    "*"    "*" " "    " "   " "     " "   " " " " " " " "  
6  ( 1 ) "*" "*" "*"    "*"    "*" "*"    " "   " "     " "   " " " " " " " "  
7  ( 1 ) "*" "*" "*"    "*"    "*" "*"    "*"   " "     " "   " " " " " " " "  
8  ( 1 ) "*" "*" "*"    "*"    "*" "*"    "*"   "*"     " "   " " " " " " " "  
         they ipron article verb auxverb past present future adverb preps conj
1  ( 1 ) " "  " "   " "     " "  " "     " "  " "     " "    " "    " "   " " 
2  ( 1 ) " "  " "   " "     " "  " "     " "  " "     " "    " "    " "   " " 
3  ( 1 ) " "  " "   " "     " "  " "     " "  " "     " "    " "    " "   " " 
4  ( 1 ) " "  " "   " "     " "  " "     " "  " "     " "    " "    " "   " " 
5  ( 1 ) " "  " "   " "     " "  " "     " "  " "     " "    " "    " "   " " 
6  ( 1 ) " "  " "   " "     " "  " "     " "  " "     " "    " "    " "   " " 
7  ( 1 ) " "  " "   " "     " "  " "     " "  " "     " "    " "    " "   " " 
8  ( 1 ) " "  " "   " "     " "  " "     " "  " "     " "    " "    " "   " " 
         negate quant number swear social family friend humans affect posemo
1  ( 1 ) " "    " "   " "    " "   " "    " "    " "    " "    " "    " "   
2  ( 1 ) " "    " "   " "    " "   " "    " "    " "    " "    " "    " "   
3  ( 1 ) " "    " "   " "    " "   " "    " "    " "    " "    " "    " "   
4  ( 1 ) " "    " "   " "    " "   " "    " "    " "    " "    " "    " "   
5  ( 1 ) " "    " "   " "    " "   " "    " "    " "    " "    " "    " "   
6  ( 1 ) " "    " "   " "    " "   " "    " "    " "    " "    " "    " "   
7  ( 1 ) " "    " "   " "    " "   " "    " "    " "    " "    " "    " "   
8  ( 1 ) " "    " "   " "    " "   " "    " "    " "    " "    " "    " "   
```

Figure 19: Using Leaps Package where nvmax=8.

After finish our full process we found our results. In figure-15 the given words are most use of suicidal users.

NVMAX	Words
4	Word count, Words/sentence, Question Marks, Unique
5	Word count, Words/sentence, Question Marks, Unique, Dictionary words
6	Word count, Words/sentence, Question Marks, Unique, Dictionary words, Six-letter words
7	Word count, Words/sentence, Question Marks, Unique, Dictionary words, Six-letter words, Total function words
8	Word count, Words/sentence, Question Marks, Unique, Dictionary words, Six-letter words, Total function words, Total pronouns
9	Word count, Words/sentence, Question Marks, Unique, Dictionary words, Six-letter words, Total function words, Total pronouns, Personal pronouns
10	Word count, Words/sentence, Question Marks, Unique, Dictionary words, Six-letter words, Total function words, Total pronouns, Personal pronouns, 1st pers singular

Figure 20: most useable words by suicidal users.

3.3 Classification:

We employed a machine learning algorithm after creating training sets. Four classification techniques were used to evaluate and analyze the dataset: Naive Bayes (NB), IBk (KNN), Decision Tree, Logistic Regression, and SVM (Support Vector Machine). These classification models ensured that the correctness of the training dataset was evaluated. Using a dataset of suicide users, these classification algorithms are utilized to forecast suicidal behavior.

Naïve Bayes Classifier:

The Naive Bayes Classifier is one of the most fundamental and effective classification algorithms now in use, and it aids in the development of quick machine learning models that can make accurate predictions. It is a supervised learning strategy that relies on the Bayes theorem to address classification problems. Simple Bayes or independent Bayes are some names for it. In order to calculate the likelihood of each particular classification, the Nave Bayes method integrates prior and conditional probabilities into a single calculation.

IBk (K-NN) Classifier:

A fundamental supervised machine learning method that may be applied to solve classification and regression problems is the k-nearest neighbors (KNN) algorithm. One of the most fundamental yet important categorization methods in machine learning is K-Nearest Neighbors. It is the most popular supervised learning methodology. It is simple to comprehend and apply the K-NN method. The K-NN approach examines the entire dataset for K nearest neighbors in order to categorize a new data point. The ability to use K-NN for both classification and regression problems is a bonus. When creating a K-NN model, the user of the K-NN approach can select the distance. In this case, it is measured using a distance function. These functions include Euclidea, Manhattan, Minkowski, and Hamming distance. Euclidea, Manhattan, and Minkowski are used to represent continuous variables. Applying the most recent Hamming distance to categorical variables.

SVM (Support Vector Machine):

A supervised machine learning approach called Support Vector Machine (SVM) is employed for classification and regression analysis. It operates by identifying the hyperplane that maximizes the margin between the classes and best divides the data into distinct classes. When there are several features, SVM is very helpful since it helps to prevent overfitting and achieve effective generalization. Using a method known as the kernel trick, SVM can handle both linear and non-linear data by translating the data into higher dimensional spaces.

Decision Tree:

In machine learning, a decision tree is a particular kind of algorithm that is used for both regression and classification problems. It is a model of decisions in the form of a tree that depicts potential outcomes, such as utility, resource costs, and chance event outcomes. The decision tree algorithm divides the data into subsets based on the most important traits or features in a recursive manner until a stopping requirement is satisfied. The final structure resembles a tree, with each internal node standing in for a test on an attribute, each branch for a test result, and each leaf node for a prediction or a class label. Decision trees are a popular choice for both data exploration and outcomes presentation since they are easy to understand and depict.

Logistic Regression:

A statistical technique called logistic regression is used to examine datasets with only one dependent binary variable. It is used to simulate the likelihood that a particular class or event, such as yes/no, pass/fail, true/false, would occur. Contrary to linear regression, which represents the relationship between the dependent and independent variables as an equation, logistic regression represents the relationship using probability. A logistic function that may be used to predict the likelihood of the dependent variable given a set of independent factors is the outcome of logistic regression. For situations involving binary classification, the logistic regression approach is frequently utilized, particularly when the relationship between the independent and dependent variables is essentially linear.

Chapter 4

4. Result and Descriptive Analysis

From chapter 3, we learned that suicidal users sent us various types of text, comments, and statuses. We utilize an empath tool to analyze the test results, which focuses primarily on reading suicide user texts to determine what kind of text it is. There are 194 attributes in empath including [help,office,dance,money,wedding,domestic_work,sleep,medical_emergency,cold,hate,cheerfulness,aggression,occupation,envy,anticipation,family,vacation,crime,attractive,masculline,prison,health,pride,dispute,nervousess':government':weakness',horror':swearing_terms',leisure':suffering':royalty':wealthy':tourism':furniture':school':magic':beach':journalism':morning':banking':social_media':exercise':night':kill':blue_collar_job':art':ridicule':play':computer':college':optimism':stealing':real_estate':home':divine':sexual':fear':irritability':superhero':business':driving':pet':childish':cooking':exasperation':religion':hipster'internet':surprise':reading':worship':leader':independence':movement':body':noise':eating':medieval':zest':confusion':water':sports':death':healing':legend':heroic':celebration':restaurant':violence':programming':dominant_heirarchical':military':neglect':swimming':exotic':love':hiking':communication':hearing':order':sympathy':hygiene':weather':anonymity':trust':ancient':deception':'fabric':,air_travel':fight':dominant_personality':music':vehicle':politeness':toy':farming':meeting':war':speaking':listen':urban':shopping':disgust':fire':tool':phone':gain':sound':injury':sailing':rage':science':work':appearance':valuable':warmth':youth':sadness':fun':emotional':joy':affection':traveling':fashion':ugliness':lust':shame':torment':economics':anger':politics':ship':clothing':car':strength':technology':breaking':shape_and_size':power':white_collar_job':animal':party'terrorism':smell':disappointment':poor':plant':pain',beauty':timidity':philosophy':negotiate':negative_emotion':cleaning':messaging':competing':law':friends':payment':achievement':alcohol':liquid':feminine':weapon':children':monster':ocean':giving':contentment':writing':rural':positive_emotion':musical':].According to our analyzing ,From this 194 attribute we got 4 coefficient attribute which are [**masculine, sleep, dance, hate**] for this 4 issue a suicide user get involved in suicide. Firstly, we can therefore conclude that sleep is more crucial to our mental wellness. Many of us are aware that "a good night's sleep" makes us feel better and that lack of sleep makes us irritable or foggy. Furthermore, there is now substantial evidence demonstrating the importance of sleep for both our physical and mental wellbeing. It has been discovered that insufficient or poor sleep decreases good feelings and increases negative emotional reactions to stimuli.

Although further investigation is required to fully comprehend the mechanisms underlying the link between sleep and mental health, it is known that sleep is crucial for many bodily and mental processes involved in processing daily experiences and controlling emotions and actions. Poor sleep can make it very difficult to handle even relatively minor stresses and can even affect our capacity to see the world effectively. Sleep helps preserve cognitive skills, such as attention, learning, and memory. Absolutely. Poor or insufficient sleep might increase the chance of developing mental health problems. It is increasingly understood that sleep issues can also contribute to the beginning and worsening of several mental health problems, including sadness, anxiety, and even suicidal thinking. Insomnia can be a sign of psychiatric diseases, such as anxiety and depression.

Studies on sleep deprivation demonstrate that even normally healthy people might become more anxious and distressed after a night of inadequate sleep. Chronic sleep issues are more common in people with mental health illnesses, and these issues are likely to exacerbate psychiatric symptoms and even raise suicide risk. Secondly This "man up" mentality develops into a denial of other people's feelings of empathy and reframes maternal instincts as "weak." Additionally, if kids are not taught how to effectively handle their emotions, anxiety may start to grow. Even when hurt or experiencing emotional distress, toxic masculinity rejects care for mental and physical illness. Then we applied different type of classifier to find out the result.

Logistic Regression

```
# importing the necessary package to use the classification algorithm
from sklearn.linear_model import LogisticRegression # for Logistic Regressio
model_lr = LogisticRegression(solver='lbfgs', max_iter=1000)
model_lr.fit(X_train, y_train) #train the model with the training dataset
y_prediction_lr = model_lr.predict(X_test) #pass the testing data to the tra
# checking the accuracy of the algorithm.
# by comparing predicted output by the model and the actual output
score_lr = metrics.accuracy_score(y_prediction_lr, y_test).round(4)
print("---------------------------------")
print('The accuracy of the LR is: {}'.format(score_lr))
print("---------------------------------")
# save the accuracy score
score.add(('LR', score_lr))
```

```
---------------------------------
The accuracy of the LR is: 0.6
---------------------------------
```

Figure 21: Logistic regression classifier

Decicion Tree

```
# importing the necessary package to use the classification algorithm
from sklearn.tree import DecisionTreeClassifier #for using Decision Tree Alg
model_dt = DecisionTreeClassifier(random_state=5)
model_dt.fit(X_train, y_train) #train the model with the training dataset
y_prediction_dt = model_dt.predict(X_test) #pass the testing data to the tra
# checking the accuracy of the algorithm.
# by comparing predicted output by the model and the actual output
score_dt = metrics.accuracy_score(y_prediction_dt, y_test).round(4)
print("---------------------------------")
print('The accuracy of the DT is: {}'.format(score_dt))
print("---------------------------------")
# save the accuracy score
score.add(('DT', score_dt))
```

```
---------------------------------
The accuracy of the DT is: 0.8
---------------------------------
```

Figure 22: Decision Tree classifier

Support Vector Machine(SVM)

```
from sklearn import svm
from sklearn import metrics
model_svm = svm.SVC(kernel = 'linear', random_state= 1) #select the algorithm
model_svm.fit(X_train, y_train) #train the model with the training dataset
y_prediction_svm = model_svm.predict(X_test) # pass the testing data to the
# checking the accuracy of the algorithm.
# by comparing predicted output by the model and the actual output
score_svm = metrics.accuracy_score(y_prediction_svm, y_test).round(4)
print("---------------------------------")
print('The accuracy of the SVM is: {}'.format(score_svm))
print("---------------------------------")
# save the accuracy score
score = set()
score.add(('SVM', score_svm))
```

```
---------------------------------
The accuracy of the SVM is: 0.6
---------------------------------
```

Figure 23: Support Vector machine classifier

Naive Bayes

```
# importing the necessary package to use the classification algorithm
from sklearn.naive_bayes import GaussianNB
model_nb = GaussianNB()
model_nb.fit(X_train, y_train) #train the model with the training dataset
y_prediction_nb = model_nb.predict(X_test) #pass the testing data to the tra
# checking the accuracy of the algorithm.
# by comparing predicted output by the model and the actual output
score_nb = metrics.accuracy_score(y_prediction_nb, y_test).round(4)
print("---------------------------------")
print('The accuracy of the NB is: {}'.format(score_nb))
print("---------------------------------")
# save the accuracy score
score.add(('NB', score_nb))
```

```
---------------------------------
The accuracy of the NB is: 0.6
---------------------------------
```

Figure 24: Naïve Bayes classifier

KNN Classifier

In [25]:
```
# importing the necessary package to use the classification algorithm
from sklearn.neighbors import KNeighborsClassifier # for K nearest neighbour
#from sklearn.linear_model import LogisticRegression # for Logistic Regressi
k_score=[]
for k in range(1,25):
 model_knn = KNeighborsClassifier(n_neighbors=k) # 12 neighbours for putt
 model_knn.fit(X_train, y_train) #train the model with the training datas
 y_prediction_knn = model_knn.predict(X_test) #pass the testing data to t
 # checking the accuracy of the algorithm.
 # by comparing predicted output by the model and the actual output
 score_knn = metrics.accuracy_score(y_prediction_knn, y_test).round(4)
 k_score.append(score_knn)

k = k_score.index(max(k_score))+1
#Train the model with the best k value
model_knn = KNeighborsClassifier(n_neighbors=k) # 12 neighbours for putting
model_knn.fit(X_train, y_train) #train the model with the training dataset
y_prediction_knn = model_knn.predict(X_test) #pass the testing data to the t
# checking the accuracy of the algorithm.
# by comparing predicted output by the model and the actual output
score_knn = metrics.accuracy_score(y_prediction_knn, y_test).round(4)
```


```
 model_knn.fit(X_train, y_train) #train the model with the training datas
 y_prediction_knn = model_knn.predict(X_test) #pass the testing data to t
 # checking the accuracy of the algorithm.
 # by comparing predicted output by the model and the actual output
 score_knn = metrics.accuracy_score(y_prediction_knn, y_test).round(4)
 k_score.append(score_knn)

k = k_score.index(max(k_score))+1
#Train the model with the best k value
model_knn = KNeighborsClassifier(n_neighbors=k) # 12 neighbours for putting
model_knn.fit(X_train, y_train) #train the model with the training dataset
y_prediction_knn = model_knn.predict(X_test) #pass the testing data to the t
# checking the accuracy of the algorithm.
# by comparing predicted output by the model and the actual output
score_knn = metrics.accuracy_score(y_prediction_knn, y_test).round(4)
```


```
k_score.append(score_knn)
print("---------------------------------")
print('The accuracy of the KNN is: {}'.format(score_knn))
print("---------------------------------")
# save the accuracy score
score.add(('KNN', score_knn))
```

```
---------------------------------
The accuracy of the KNN is: 0.6
---------------------------------
```

Figure 25: KNN classifier

Chapter 5

5. Future work and conclusion

5.1 Future work

It is very important to identify suicidal behaviors of a person at early stage. In this case, applying data mining techniques to obtain a decent result from data analysis can be beneficial.

This survey was conducted on 50 persons using their twitter data. Where 25 persons committed suicide and other 25 persons did not. We got an accuracy rate of 80% using 50 persons twitter data. But in future we want to work with more data for increasing our accuracy rate. After that we also want to come up with a solution after predicting suicidals behaviors of a person, how we can help that person to overcome this mental health issue.

5.2 Conclusion

The major purpose of this study is predicting suicidal behavior using their twitter data. Many persons share their feelings in their social media platforms. Despite improvements in the detection and treatment of serious mental diseases, suicide has remained an unsolvable public health issue. The development of suicide screening systems through the access and analysis of social media data is an expanding field. The diversity of those who commit suicide and the lack of accurate, significant predictors of suicide make it difficult to prevent suicide. Social networking sites have come under fire for not doing enough to stop suicidal thoughts and actions. The implementation of an intervention that appeared to be hastily designed without adequate consideration of the social and cultural context in which a person experiencing suicidal behavior is embedded has been criticized as part of social media platforms' response to these accusations, which has included strengthening suicide detection mechanisms on the platforms. After analyzing their social media data, we can predict suicidal behavior of a person. For this study we have collected data from twitter accounting using twitter API. Then using empath algorithm, liwc tools for preprocessing. Then we applied different type ML algorithm using python. We got an accuracy rate of 80%. And we can increase the accuracy rate if we use more person twitter data.

References:

[1] M. Chatterjee, P. Kumar, P. Samanta and D. Sarkar, "Suicide ideation detection from online social media: A multi-modal feature based technique," *International Journal of Information Management Data Insights*, 2022.

[2] N. Wang, F. Luo, Y. Shivtare, V. D. Badal, K. P. Subbalakshmi, R. Chandramouli and E. Lee, "Learning models for suicide prediction from social media posts," 2021.

[3] K. S. Lee, H. Lee, W. Myung, G. Y. Song, K. Lee, H. Kim, B. J. Carroll and D. K. Kim, "Advanced Daily Prediction Model for National Suicide Numbers with Social Media Data," pp. 344-354, 2018.

[4] K. L. Celedonia, M. C. Compagnucci, T. Minssen and M. L. Wilson, "Legal, ethical, and wider implications of suicide risk detection systems in social media platforms," *Journal of Law and the Biosciences*, pp. 1-11, 2021.

[5] N. Dasari, R. Mittapalli, A. Dhule and S. Wagh, "Prediction of suicide using social media data," *International Research Journal of Modernization in Engineering Technology and Science,* vol. 4, no. 3, 2022.

[6] S. Fodeh, J. Goulet, C. Brandt and H. Al-Talib, "Leveraging Twitter to better identify suicide risk".

[7] R. Haque, N. Islam, M. Islam and Md. M. Ahsan, "A Comparative Analysis on Suicidal Ideation Detection Using NLP, Machine, and Deep Learning," *Technologies*, 2022.

[8] A. Roy, K. Nikolitch, R. McGinn, S. Jinah, W. Klement and Z. A. Kaminsky, "A machine learning approach predicts future risk to suicidal ideation from social media data," *Digital Medicine*, May 2020, doi: 10.1038/s41746-020-0287-6

[9] E. R. Kumar and A. K. V. S. N. R. Rao, “Suicide Prediction in Twitter Data using Mining Techniques: A Survey,” 2019, doi: 10.1109/iss1.2019.8907987

[10] D. Karamshuk, F. Shaw, J. Brownlie and N. Sastry, “Bridging big data and qualitative methods in the social sciences: A case study of Twitter responses to high profile deaths by suicide,” *Online social networks and media,* vol. 1, pp. 33-43, June 2017, doi: 10.1016/j.osnem.2017.01.002

[11] B. O’Dea, S. Wan, P. J. Batterham, A. L. Calear, C. Paris and H. Christensen, “Detecting suicidality on Twitter,” *Internet Interventions*, vol. 2, no. 2, pp. 183-188, May 2015, doi: 10.1016/j.invent.2015.03.005

[12] G. B. Colombo, P. Burnap, A. Hodorog and J. Scourfield, “Analysing the connectivity and communication of suicidal users on twitter,” *Computer Communications*, vol. 73, pp. 291-300, Jan 2016, doi: 10.1016/j.comcom.2015.07.018

[13] J. H. K. Seah and K. J. Shim, “Data Mining Approach to the Detection of Suicide in Social Media: A Case Study of Singapore,” *2018 IEEE International Conference on Big Data (Big Data)*, pp. 5442-5444, Dec 2018.

[14] A. Abboute, Y. Boudjeriou, G. Entringer, J. Aze, S. Bringay and P. Poncelet, “Mining Twitter for Suicide Prevention,” *International Conference on Applications of Natural Language to Data Bases/Information Systems*, pp. 250-253, 2014.