Chapter 1: Introduction, Motivation, Problem Statement.

It is no secret that college students face mental health challenges. Whether it be separation from family, increased workloads, or anxiety about living in a new place with unfamiliar faces, many studies have observed increased mental health issues amongst students. The Covid-19 pandemic exacerbated these problems on college campuses around the world, the University of Notre Dame being no exception. In the midst of the pandemic the University brought students back to campus and imposed an extremely rigorous protocol. Midway through the semester, the administration administered a survey of the student body in order to assess how they were handling campus life during the pandemic. According to an Observer article by Dane Sherman, “A Notre Dame mental health survey found 18% of students to be under ‘severe’ mental distress” (Sherman, 2020, para. 2), an astounding number as this equates to 1,572 students.

Restrictions imposed in the Fall of 2020 upended campus life. These restrictions were put in place to protect students’ physical health from the SARS-CoV-2 virus, however they were detrimental to the mental health of many students. Such restrictions included forbidding students from visiting residence halls other than their own, forbidding students from going into a dorm room other than their own, removing all outdoor volleyball nets and basketball rims, requiring social distancing and mask-wearing in all locations other than students’ own rooms, and limited food selections at the dining halls with to-go dining only. Restrictions that harmed students’ mental health the most were those that limited socialization. Students reported that they “struggled . . . to hang out with friends and meet new people”, and were “feeling confined and feeling isolated and not being able to make new friends” (Guffey, 2021, paras. 20,23). This adjustment for students was not easy, however the restrictions were only a part of what led to the deterioration of the overall mental health of the student body.

It can be argued that the response from the University regarding mental health concerns has been inadequate. According to the director of McWell, the student center for wellbeing, “Feeling lonely and isolated is normal, especially during this pandemic… Students who feel isolated or homesick should reach out to old friends, new acquaintances, family, and hall staff. Connecting with others and talking about our feelings helps us to cope” (Sherman, 2020, para. 11). Telling students to simply “reach out”, will not solve the problem – resources, activities, and real action are required. In the opinion of Elaine Carter, a Junior at Notre Dame, who is president of a student-led club that tackles mental health issues at the University, “The hardest thing in getting through all this is the immense lack of authenticity on the side of the administration in terms of talking to us as a student body,... People saw things on the sidewalk that said ‘you’re doing what many people thought was impossible.’ Seeing something like that for me just disregards the struggle of it. It glosses over the difficulty a little bit” (Sherman, 2020, para. 16). It is clear that the University could benefit from more specific feedback like this when making decisions that involve student well-being.

Our project aims to analyze the data collection methods used by the University of Notre Dame during the time in question and to propose alternative methods that could have given the University a better picture of the student body’s overall mental health and well-being. Specifically, we want to determine whether or not publicly available social media data, collected from platforms such as Reddit or Twitter, could inform administrative decisions regarding student well-being on campus. Our analysis aims to provide interested University administrators insight into the practicality, usefulness and ethical implications of analyzing publicly available student social media posts. We regard the University
administration (or other institutional leadership bodies) as the stakeholders and the data as publicly available social media posts related to mental health issues or policies on campus.

Chapter 2: Ethics and Design Statement

Our data collection efforts focused on scraping online social media posts from platforms such as Twitter and Reddit. While this information is publicly available and we are not violating any privacy laws or standards, questions still need to be raised regarding the ethics of this type of data collection. These questions revolve around the effects this method of surveillance might have on the students, even though they willingly share this information through social media channels. Our team has identified three major ethical considerations.

Firstly, a major concern is that students may not be comfortable, or even aware of, the university checking public social media platforms to look at student posts. Although many types of social media are public and students may understand the ramifications of posting publicly (i.e. not posting material that may lead to negative consequences from their employer, friends, family, etc.), they may not expect the University to be conducting this kind of surveillance, and it may cause anxiety. If students are made aware of this surveillance research they may refrain from sharing certain types of, or perhaps any, information on public social media channels related to their college experience.

Secondly, when collecting public social media data the possibility of bias being introduced must be considered. Available social media posts may only represent certain demographics or student groups that are more likely to post on these social media channels. Location-based scraping may miss students who are attending the university virtually, students from different home cities or countries may favor different social media platforms, and students who are anxious about their public presence may conceal information related to their feelings. To ensure the broadest possible representation such biases must be kept in mind. In light of these concerns, social media data should be considered as supplemental to information that has been collected under more rigorous controls.

Thirdly, we need to keep in mind that we are trying to provide improved insights to decision-makers regarding student well-being by analyzing publicly available posts and data. This goal falls into an ethics of care approach, using empathy as a driving force for change in response to documented struggles by students due to anxiety, stress, and major events such as the covid-19 pandemic. The methods outlined in this paper are to be used only for this goal and not for malicious purposes such as unwarranted monitoring of students for financial or social gain. Transparency is essential. So-called “black box” models are becoming more widely used in areas such as law enforcement and loan recommendations and can cause confusion and frustration to those negatively affected by them (Kenton, 2022). The methods employed by the university should be well documented and disseminated to students, as they are the ones being primarily affected by the policies and decisions derived from the data.

These three main ethical concerns must be considered by stakeholders who may be either university or institution officials of higher education. Students stand to benefit from policies instituted as a result of our methods. Incorporating the types of data previously mentioned may pose some concerns from the students as noted above, but the difficulties faced by students may provide grounds for this implementation.

Chapter 3: Related Work
As more and more people have taken their thoughts to social media researchers have attempted to leverage this information to tackle difficult real-world problems. Social media data has been used to gauge everything from the level of disease activity and spread (Signorini, 2011) to measuring the success of a college admissions campaign (Bolat, 2017). Valuable information can be extracted from public tweets, as they tell us what a user is up to in real time. Twitter traffic from Signorini’s study was used to track user interest and concern for the H1N1 influenza. Since geographical data is embedded into tweets, the study also demonstrated that Twitter traffic can estimate disease activity in real time, approximately one to two weeks faster than the practice from a decade ago allowed. In fact hygiene and mask interest increased after messages from the CDC about the H1N1 outbreak.

Additionally, college admissions committees have begun using big data from social media to predict the academic success of individuals. Ithaca College has found that using this data in their admission decisions has increased the retention rate of students from their first to the second year in college. At Wichita State University and Sarah Lawrence College, big data is being used to help save money by determining interest in the colleges and increasing yield rates. There are some concerns, however, with the use of social media data in admissions decisions including how it might affect applicants who aren’t active on social media or who do not have access to social media due to lack of Internet access (Felton, 2015).

The study our team was most interested in, however, was the usage of social media scraping to attempt to measure depression, stress, or feelings about a certain situation. A prominent study conducted by researchers at the University of California, Los Angeles attempted to use sentiment analysis to “investigate whether students’ perceived levels of stress were associated with the sentiment and emotions of their tweets”(Liu, 2021, p. 80). Sentiment analysis is a fast-growing field of NLP defined by Oxford Languages as “the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral” (“Sentiment Analysis”, n.d.)” In other words, these researchers from UCLA were attempting to look at Tweets from university students and determine if their overall sentiment was at all correlated with their levels of stress in the real world. They recruited 181 UCLA freshmen to complete a questionnaire covering their current level of stress and emotional state during the past 7 days. Tweets participants had posted during this time period were also collected. These 1879 tweets were manually labeled either positive, negative, or neutral using a basic sentiment analysis approach. They found that higher weekly stress was in fact correlated with “a greater percentage of negative sentiment tweets”(Liu, 2021, p. 80). The team finished the study by concluding “Sentiment and emotions expressed in the tweets have the potential to provide real-time monitoring of stress level and emotional well-being in college students”(Liu, 2021, p. 80). Our methods are similar to theirs. One important item to note here is that even though the study observed that there could be a correlation between Tweets and levels of stress, these participants knowingly provided their Tweets for use in the study. As mentioned in the previous section, it is important to keep in mind the ethical implications of scraping social media sites for public student posts.

De Choudhury et. al. used postings from Twitter as a measure of depression. Similar to the goals of our current research, this study was focused on how social media can complement traditional survey techniques in understanding the mental health of populations. In their study, a crowdsourcing methodology was used to obtain Twitter posts from clinically depressed individuals. Using this data, they developed an SVM classifier which predicted with 70% accuracy whether a Twitter post was “depression-indicative”. Several features of each Twitter post are used in the classifier model including...
features of emotion, time, linguistic style, user engagement, and ego-network. These models were then used to create a social media depression index (SMDI) to analyze “depressive-indicative” posts across different populations, including geography, gender, time, and seasonal patterns. The data collected across populations using the SMDI were comparable to data from the CDC, indicating the success of the ability to identify “depression-indicative” posts using this model (De Choudhury, 2013, p. 48).

The mental health of college students, in particular, has been a topic of research since the start of the covid-19 pandemic. Along with using social media to monitor mental health, many researchers have been focusing on what measures universities can take to better support the mental health of their students. Zhai and Du, researchers from Pennsylvania State University cite events such as students losing jobs due to the pandemic, loss of stable room and board, health-related anxiety, amongst others, as issues that could potentially harm students' mental health. They suggest courses of action, such as offering telehealth counseling services and public health messaging, that could be used in conjunction with social media monitoring in order to attend to student mental health as quickly and efficiently as possible (Zhai, 2020).

Chapter 4: Data (describe the sources, collection mechanism, etc)

Over the past few years, the amount of publicly available data has skyrocketed. Along with this drastic increase in data has come the amount of free, reliable, open-source methods for collecting this data quickly and at a large scale. According to Statista, in August 2020 about 480 thousand people interacted in some way with Reddit content every single minute (“Reddit Usage”, 2022). Twitter has been reporting equally if not more impressive stats with over 192 million daily active users worldwide (“Twitter”, 2022). Not only do these platforms have a massive number of users, but many of these users are in our target demographic. For Reddit, of their total 52 million daily active users, 21 percent fall in the 18 to 24 age range (Lin, 2021). Twitter’s user base comprises 37 percent of people in the 18 to 29 age range (Digital Communications Team, n.d.). These platforms provide a wealth of information specific to our target demographic that can be easily scraped in real-time.

We initially planned to scrape both Reddit and Twitter information but decided to focus solely on Twitter because it is free and has an extensive Developer API, as well as the fact that Reddit boards are inherently public, with no guarantee a post from the Notre Dame board is posted by a member of the community. Our target data includes posts by college students and more specifically, posts by students at the University of Notre Dame during the time period of January 2019 through March 2022. With our platform chosen, we explored two of the several widely used methods for extracting subgroups of Tweets from a collection.

The first method we attempted to implement was to filter Tweets by hashtag. The Twitter API allows programmers to scrape millions of Tweets based on certain attributes such as hashtags or keywords. This method has been used in various projects such as Kevin Gimpel’s project “Beating the NFL Football Point Spread” where he compiled lists of hashtags that were associated with certain NFL teams in order to pinpoint fans of the team and what they were saying about each team (Gimpel, 2006). We found that attempting to implement this method was tricky for our usage since we ended up getting a lot of Tweets from fans of the University of Notre Dame sports teams using the hashtags “#Notre Dame”, “#UND”, “#FightingIrish”, etc. Separating fan and student Tweets was very difficult to do manually, so we explored an alternative route of scraping Tweets by location.

According to the Twitter Developer API, there are two main approaches to scraping Tweets by location. The first method is referred to as scraping “Tweets with a specific latitude/longitude “Point” coordinate”
and the second is “Tweets with a Twitter ‘Place’”, which is the name or title of a city, town, landmark, etc. (“Filtering Tweets”, n.d., “Tweet Locations”). The former method proved to be more useful than the latter due to the lack of Tweets that users manually tagged “The University of Notre Dame” or “South Bend, IN.” To more accurately gather Tweets that were coming from University Students, we landed on the Latitude/Longitude coordinate-based location search. After some experimentation, we decided to use the coordinates of Main Building (41.7, -86.2) and a range of 2km to cover the entirety of Notre Dame’s campus with a little extra room for off-campus students living relatively close to campus as shown below in figure 1. (“Draw a Circle”, n.d.)

![Figure 1. Tweet Location Filter Range](image)

Before implementing our scraping program we needed to once again consider the ethical implications of our methods. We understand that by using this location method we may miss students living far from campus or those with physical or mental health issues that elected to complete the Fall 2020 semester virtually and were not on campus. We also may have missed students who do not use or have access to Twitter and those who only post Tweets privately. Additionally, the United States is known to have the highest number of Twitter users of any country in the world, so we need to ensure we are taking into account the possibility of missing the opinions of international students who may be utilizing social media forms other than Twitter (Fish, 2021). All of these groups must be included when the university makes decisions and policies, so we again note that this data is to be considered supplemental to other sources.

With our approach decided on, we began collecting Tweets corresponding to our target location and time period. Over the course of several days, we were able to collect 25,000 tweets. Once the Tweets were collected, we set about creating a process for cleaning the data so that we could use them in our
analysis. Our process was based on the framework set out by Stanford AI Researcher Younes Mourri in his course Natural Language Processing with Classification and Vector Spaces. To prepare each Tweet to be ready for sentiment analysis, we first converted all words in the Tweet to lowercase to remove any ambiguity regarding case. Then we removed all stop words such as “a”, “is”, “are”, etc which do not provide useful information in determining the sentiment of a Tweet. Next, we removed hyperlinks that may be included in Tweets which also do not provide any valuable information in determining sentiment. Lastly, we removed the individual hashtag character from any hashtags to single out the English word specifically to be used for our analysis. The processed Tweets were stored in a large CSV file to be used in our analysis.

Chapter 5: Method (describe the data science, system/platform, etc)

With our Tweets collected and ready for analysis, we researched several methods of Natural Language Processing in order to determine the best course of action to turn Tweets into useful information about student feelings on campus. The most appropriate approach to our problem statement we decided would be applying sentiment analysis to the Tweets we scraped in order to determine a score for the overall well-being of students on campus. As mentioned in our chapter covering relevant work, this process can be applied to almost any field of research with currently widespread usage in forecasting human opinion on political topics such as elections, sports, and public policy just to name a few.

A major benefit of this method is the ability to examine exactly how an individual Tweet is scored and how each word in the sentence contributes to the overall score. This means that explaining the reasoning behind a particular score is extremely straightforward allowing for transparency when it comes to announcing results. This benefit aligns perfectly with one of our ethical goals outlined earlier, to avoid having a “black-box” model where it is not easy to determine the details of how a particular score was calculated. Transparency with the stakeholders is paramount for our method and sentiment analysis aligns perfectly with this goal.

Another benefit of sentiment analysis is the ability to translate Tweets from text into a single numerical value. Vectorization of words and sentences is what allows for this and is extremely helpful for machine learning with large datasets. Neptune.ai describes vectorization as “a step in feature extraction. The idea is to get some distinct features out of the text for the model to train on, by converting text to numerical vectors” (Jha, 2021, para. 7). It would be impossible for a human to read through the TODO # TWEETS manually and record the overall sentiment of each Tweet, then aggregate all of the sentiments to one overall score to represent the feelings of students on campus. Even with modern-day machine learning models, it is impossible for computers to understand language without first converting words and sentences into a numerical representation using vectorization.

Fortunately for our project, there are many Python libraries that tackle Natural Language Processing tasks like this that are made available for free public use. One of the most common libraries right now is the Natural Language Toolkit developed by the NLTK Project. Their website proudly claims to provide a “suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum” (“NLTK”, 2022, para. 1). This library is the gold standard for NLP tasks as it aggregates over 50 corpora and resources into one single package that can be easily installed on any computer running Python. As discussed in the data collection section, we needed to follow a detailed cleaning procedure to ensure that our data was of high quality and the Natural Language Toolkit was a major help in implementing that
process. Another major benefit of this toolkit is the active community and vast sea of examples to base our own project on.

The Valence Aware Dictionary for Sentiment Reasoning known as VADER is one of the available tools in the Natural Language Processing Toolkit for Python. This tool allows for determining the positivity/negativity of a particular Tweet as well as the strength of that emotion in either direction. The tool outputs a score that tells how positive or negative the sentiment of a tweet is. Using this toolkit, we wrote a Python notebook to load all of the Tweets we had scraped and cleaned previously and then performed VADER sentiment analysis on each one, recording the scores along the way. We then aggregated the scores and calculated the average sentiment score for each day, week, and month. A graph of the 30-day rolling average is shown below in Figure 2:

![Figure 2. 30 Day Rolling Average Compound Sentiment](image)

As you can see, there are some notable fluctuations in the average 30-day rolling sentiment that could be mapped to certain changes in University policy as well as important time periods in the academic calendar.

Chapter 6: Analysis and Societal Impact (describe how your project will inform society)

Using our sentiment data gathered from the tweets, we were able to show some promising results from this method of using social media to monitor mental health. Figures 3 and 4 show two histograms of the compound score for all tweets gathered from before the pandemic and after the pandemic, respectively. We used March 15th, 2020, as the “start date” of the pandemic. While Covid-19 was circulating before then, March 15th, 2020 is the day Notre Dame students received the message they would be sent home, and the day lockdowns started throughout most of the United States. We also removed all tweets with a compound score of 0 here, as completely neutral tweets do not give us insight into one’s mental health. This led us to 8,009 “pre-pandemic” tweets, and 7,250 “post-pandemic” tweets.
These two histograms take a big picture look at sentiment before and after the pandemic started. The date ranges include times when students were not on campus, which can account for some noise in the data. Despite this, we see this is a representative sample, as neither period accounts for student absences on campus. The average sentiment for the “pre-covid” tweets was 0.325612, with a standard deviation of 0.500719. The average sentiment for the “post-covid” tweets was 0.249255, with a standard deviation of 0.507488. To determine if there is a statistically significant difference between these two sets of tweets, we performed an unpaired T-Test between the two sets. This T-Test provided us with a P-Value less than 0.0001, and a 95% confidence interval of the difference between pre-pandemic sentiment and post-pandemic sentiment of -0.09240753 to -0.06030647. This is a promising result, and aligns with our initial hypothesis, as well as the survey data published by the University of Notre Dame. We can see the difference between the two sets when we plot both histograms on the same chart after normalizing, which is shown in Figure 5.
We can see a noticeable difference between the post-covid and pre-covid histograms when they are aligned with each other. The pre-covid histogram features more “positive” sentiment tweets, and fewer “negative” tweets. Analysis like this allows us to compare two different periods in time, and see how the sentiment for a given location, in our case The University of Notre Dame, changes over time.

We also decided to compare the sentiment distributions for both the 2020 Fall Semester and the 2021 Fall Semester, as these are the only two full semesters we have in which we can compare the sentiment distributions. These two distributions are shown in Figures 6 and Figures 7 respectively, and the combined and normalized histogram is shown in Figure 8.
For both of these distributions, we also dropped all tweets with a compound sentiment score of 0, to eliminate neutral tweets. We used 1,984 tweets from the Fall Semester of 2020, and 1,428 tweets from the Fall Semester of 2021. The average compound score for the Fall Semester of 2020 was 0.217148, and the average compound sentiment score for the Fall Semester of 2021 was 0.316217. The standard deviations were 0.506501 and 0.484418 respectively. Doing the same unpaired T-test for both of these distributions, we again found that the P-Value for the difference in distributions was less than 0.0001, and the 95% confidence interval of this difference was from -0.13298130 to -0.06515670. Even though both of these semesters took place during the Covid-19 pandemic, there is a stark difference between the sentiment distributions throughout these semesters. The 2020 Fall semester notably had more virtual classes and more limitations on social gatherings. We can speculate about why mental health was worse across these two semesters, but proving causation is not a point of this project. We aim to be able to analyze mental health throughout time on college campuses, and this project shows that it is possible.

We hope this project provides a baseline for an ethical approach for analyzing social media posts to gauge information about the mental health of a particular locality. Performing data analysis on social
media is a powerful tool for learning about people of a particular community, but it cannot be used without adhering to ethical standards. Throughout this project, ethics was at the forefront of every decision made. Although this data scraping was used strictly for research purposes, we acknowledged that if institutions implement scraping procedures for assessing mental health on campus, it is critical for the institution to be clear in their exact uses of the data and ensure students’ anonymity so they continue to feel comfortable sharing their thoughts on social media.

This project was also informed of the almost unavoidable bias that comes with this type of data collection. Although the target social media site Twitter is used highly amongst college-aged individuals, it may not be representative of the entire student population, especially students who are not active on twitter, who choose to post privately, or who may post in a language other than English. As previously mentioned, an institution implementing social media scraping must be aware of this bias and only use the scraped data as an added source of mental health information, not the sole source. Institutions are also encouraged to analyze multiple social media platforms to provide a better representation of the student body.

During this research, we took an ethics of care approach which we encourage future users of social media data to use as well. Using an ethics of care approach requires one to always think back to the why behind what they’re doing and who will benefit from the data analysis. So for any institutions thinking of implementing social media scraping into their mental health analysis, we encourage them to remember to focus on improving the mental health of their campus to promote wellness and success of their students.

Although these ethical principles may sound simple, they are essential when dealing with data that may reveal personal feelings about a particular community. It is not uncommon for these principles to be overlooked when performing studies on public social media data. It is easy to get caught up in the research and the digging that people can begin veer away from their initial purpose. We encourage future researchers and institutions to keep these ethical principles and to never forget the humanity behind the social media posts when performing research similar to that outlined in this project.

While we want institutions to be cautious with the ethics behind this method, we are also excited by the potential for institutions to use social media to better understand their students’ mental health and how institutional policies may affect the campus. Many universities around the US use some form of an annual or bi-annual survey to help understand the mental health of the campus. While these surveys are beneficial for understanding specific questions, they are static in the sense that they only look at student responses in one moment of time. This leaves room for a lot of potential bias with how the student was feeling at one specific time and doesn’t provide any insight to the changing emotions of the students over the course of the school year. We hope that with the data from this pioneering study, that institutions begin to explore the use of social media scraping to supplement their surveys of mental health data. While keeping the ethical principles in mind, this method would help institutions get continuous feedback on the mental health of their campus. This would allow institutions to better analyze how policies impact their students as well as provide more feedback if changes made to better mental health are actually working. With this information, institutions will be able to actively engage with the mental health of their students and provide better resources for those who may be struggling.

Chapter 7: Conclusion.
Collectively, we believe that this research can serve to help institutions better assess and understand the sentiment of their students, especially in trying situations as we have been experiencing during COVID-19. We are confident that with the ethical considerations we have put in place, we have generated a data collection method that can be easily replicated and utilized by other institutions.

As we have stated above, this project is not intended to replace surveys or traditional methods of data collection, but rather to be a potential supplement to them. We believe that students share things on social media that they may not want to share in a university setting when asked for feedback. Additionally, students’ inboxes are constantly filled with emails asking to participate in a quick questionnaire and it is easy to overlook them. By scraping data from social media sites, universities can eliminate these potential barriers to data collection. However, it is important to re-emphasize that the data must be collected in an ethical, respectful way. Here, we tackled this by anonymizing the scrapped tweets before conducting our analysis.

Our main motivation for this project was to help to improve the way universities can assess the needs of their students. We were all college sophomores when this pandemic hit and were all affected by it in different ways. We felt that when we came back in the fall of 2020 for the school year, there was a lot of miscommunication and misunderstanding between the university and the students. The students felt unheard by the administration and as if they blindly brought us back without a plan to ensure that students could still have a positive experience at the university despite the circumstances. We were asked about our mental health through an email survey and from that, the University concluded that “75% of college students reported feeling increased stress and anxiety due to coronavirus...[and] 18% of students are under “severe” mental distress” (The Observer, 2020). If universities were able to use our proposed tool of social media scraping, they would be able to detect negative sentiment amongst students before it’s too late and it gets to the point where nearly a fifth of the student body is severely suffering. Supplementing traditional surveys with social media data scraping can help universities better understand the current state of their student population and they can more accurately assess how to help.

It is important to emphasize when collecting data on anyone, especially in this setting, universities are transparent with their students about the data they collect. Even if they plan to anonymize the scrapped social media posts, all data collection should be and must be done with the utmost respect to ethical practices. As we have outlined in the previous chapters, we believe strongly in this project because we approached it using the ethics of care foundation. We believe that this method has the potential to improve the everyday lives of university students, but will not work as so unless the data is carefully taken with transparency between both parties. Without maintaining student trust, the university will be unsuccessful in improving the sentiment across campus.

Going forward, we hope to see more research in order to help universities understand and adapt to student feelings and concerns. While we propose here a method of data collection, it is up to the universities to educate themselves on what to do with their findings. We have seen over the past few decades the power that technology has and the driving force that it can be to help advance all aspects of our society. In times such as the COVID-19 pandemic, and even in our regular day lives, it is imperative that universities have the tools and effective training to communicate and understand the mental health of their students.
References


Digital Communication Team. (n.d.) *Social Media Platforms and Demographics*. The London School of Economics and Political Science.


Fish, T. (2021, September 22). *These countries have the most people on Twitter*. Newsweek.


[https://home.ttic.edu/~kgimpel/papers/machine-learning-project-2006.pdf](https://home.ttic.edu/~kgimpel/papers/machine-learning-project-2006.pdf)

https://ndsmcobserver.com/2021/03/letter-to-administration-on-students-mental-health-garners-over-1500-signatures/


https://neptune.ai/blog/vectorization-techniques-in-nlp-guide


https://doi.org/10.2196/mental.5626


