Analyzing Mental Health Disorders in Social Media Using NLP

BYGB-7977-002 - Group 3
Group Members: Shuxuan Jiang, Shuo Shen, Yuqi Chen, Yizhen Lu
Date: May 3rd, 2023

1. Executive Summary

This report presents the findings of our text analytics group project focused on analyzing mental health disorders in social media using Natural Language Processing (NLP) techniques. The goal of the project is to leverage text analytics to assist groups like policymakers, professionals, and advocacy groups to gain insights into mental health trends and patterns, identify potential at-risk individuals, and contribute to the development of targeted advertisement interventions and support.

The project utilized a dataset sourced from Kaggle to conduct the analysis, incorporating sentiment analysis, text classification, and topic modeling to unravel the results. From the outcome, we are able to identify underlying depressed individuals and categorize them into three different classes: Long Term Depression, Periodic Depression, and Occasional Depression. We have also captured the five most relevant topics based on the data. In conclusion, our project findings can substantially increase the probability of capturing abnormal posts and provide a reliable source of information as a tool to detect mental health disorders.

2. Business Goal Analysis

Mental health disorder is a growing concern worldwide, with increasing numbers of people seeking help and support. Mental health professionals and researchers rely on various methods, including surveys and clinical assessments, to identify and address mental health challenges. However, these methods can be time-consuming, expensive, and subjective. To address this challenge, we propose using text analytics to analyze social media posts related to mental health disorders. By leveraging natural language processing (NLP) techniques, such as sentiment analysis and text classification, we can gain insights into mental health trends and patterns at both individual and group levels.

In addition, due to the stigma and misinformation surrounding mental health, individuals may not seek help or access appropriate treatment. Social media platforms have become a space where people can express their thoughts and feelings freely, including their mental health challenges. By analyzing social media posts, we can obtain a better understanding of how mental health disorders manifest in online communication and identify potential at-risk individuals. However, analyzing large volumes of text data manually is time-consuming and prone to errors. This text analytics project using NLP techniques can automate the process and provide accurate and reliable insights into mental health trends and patterns.

This project has the potential to achieve several goals in addressing mental health challenges. First, it can help mental health professionals identify potential at-risk individuals by analyzing individual sentiment and time series data. This information can inform targeted interventions and support for people who may be struggling with mental health challenges. Second, it can help researchers and advocacy groups gain a better understanding of mental health trends and patterns by analyzing monthly emotion trends and classification results. This information can inform the development of new mental health policies and interventions to improve mental health outcomes at the population level. Finally, social media platforms can use text analytics to create targeted advertising campaigns to increase awareness of proper mental illness treatments and interventions. By partnering with mental health professionals and organizations, social media platforms can take a proactive role in addressing mental health challenges and reduce the stigma associated with mental health disorders.

3. Dataset Description

The dataset employed in this project is sourced from Kaggle (https://www.kaggle.com/datasets/infamouscoder/mental-health-social-media), compiled and shared by collaborator InFamousCoder on August 6th, 2022. It consists of 20,000 English-labeled tweets from a diverse array of users, spanning a time period from August 2009 to January 2017. The dataset is organized into 11 columns, each providing specific information about the tweets and their respective users:

Column Name	**Description**	**Example**
Index	A sequential numerical identifier for each entry in the dataset	2
post_id	A unique identifier for each individual tweet	637749345908051968
post_created	The date on which the tweet was posted	Sat Aug 29 22:11:07 +0000 2015
post_text	The unprocessed content of the tweet	Awake but tired. I need to sleep but my brain has other ideas...
user_id	A unique identification number assigned to each user	1013187241
followers	The total number of followers for the user who posted the tweet	84
friends	The total number of friends (or users being followed) for the user who posted the tweet	211
favorites	The total number of tweets favorited by the user who posted the tweet	251
statuses	The cumulative count of tweets and retweets posted by the user who posted the tweet	837
retweets	The total number of retweets garnered by the specific tweet	0
label	A binary classification label for each user, with 0 representing non-depressed users and 1 indicating depressed users	1

Table 1: Dataset Description

In this dataset, each user is assigned a unique user_id, ensuring that every individual is distinctly identifiable. Similarly, every tweet is documented with a non-repeating post_id, allowing for the accurate tracking of each post, even if a single user has contributed multiple tweets. The binary labels assigned to each user (1 for depressed and 0 for non-depressed) are utilized to categorize the users based on their mental health status and serve as the basis for the mental health-related language analysis conducted in this project.

Since our primary focus in this text analytics project is on the content of the tweets, we did not utilize the other columns (followers, friends, favorites, statuses, and retweets) within the scope of this project. However, these columns can be valuable for further machine learning analyses, such as logistic regression, classification trees, and others, to enhance the understanding of mental health disorders in social media and develop more effective interventions.

4. System Design

In order to achieve the most ideal results, we divided the project into 4 phases: Data Collection/ Preprocessing, Text Classification, Sentiment Analysis, and Topic Modeling. After obtaining the dataset from Kaggle, we performed a series of cleaning and formatting tasks to set a clean foundation for subsequent analyses, including but not limited to filtering the number of posts per user, tokenization, stop words removal, etc.

With the preprocessed data, we move on to the text classification phase. Feature extraction and machine learning are incorporated in this step. Relevant text features are extracted using BoW (Bag of Words) and TF-IDF (Term Frequency-Inverse Document Frequency); three classifiers – Naive Bayes, Decision Tree, and Random Forest are used to distinguish the polarity of each post.

The third phase is sentiment analysis. In this phase, we focused on analyzing both individual and group level tweet performances, first using methods like TextBlob and Vader to calculate scores and then conducting a time series analysis. To further capture and compare the trends among each individual, a monthly emotion graph will be created. Based on the visualizations, individuals will be categorized into three classes: Long Term Depression, Periodic Depression, and Occasional Depression.

Finally, the last phase is topic modeling. This step distinguishes the 5 most relevant topics using techniques like Latent Dirichlet Allocation (LDA) and Latent Semantic Analysis (LSA) aiming to assist professionals discover the underlying trends and patterns.

5. System Implementation

5.1. Data Preprocessing

To ensure the statistical rigor and robustness of our analysis, we implemented a threshold criterion to exclude posts from users who had less than 100 tweets within the designated time frame of our dataset. This strategic approach enabled us to concentrate on users who were comparatively more active, consequently providing more valuable insights into mental health disorders on social media.

After performing data preprocessing, we obtained a data frame consisting of 19,062 rows of distinct texts and 42 unique users.(Figure 1)

	level_0	index	post_id	post_created	post_text	user_id	followers	friends	favourites	statuses	retweets	label
0	0	0	637894677824413696	Sun Aug 30 07:48:37 +0000 2015	It's just over 2 years since I was diagnosed w...	1013187241	84	211	251	837	0	1
1	1	1	637890384576778240	Sun Aug 30 07:31:33 +0000 2015	It's Sunday, I need a break, so I'm planning t...	1013187241	84	211	251	837	1	1
2	2	2	637749345908051968	Sat Aug 29 22:11:07 +0000 2015	Awake but tired. I need to sleep but my brain ...	1013187241	84	211	251	837	0	1
3	3	3	637696421077123073	Sat Aug 29 18:40:49 +0000 2015	RT @SewHQ: #Retro bears make perfect gifts and...	1013187241	84	211	251	837	2	1
4	4	4	637696327485366272	Sat Aug 29 18:40:26 +0000 2015	It's hard to say whether packing lists are mak...	1013187241	84	211	251	837	1	1
...	...	...	...	...	...	...	...	...	...	...	...	...
19057	19995	19995	819336825231773698	Thu Jan 12 00:14:56 +0000 2017	A day without sunshine is like night.	1169875706	442	230	7	1063601	0	0
19058	19996	19996	819334654260080640	Thu Jan 12 00:06:18 +0000 2017	Boren's Laws: (1) When in charge, ponder. (2) ...	1169875706	442	230	7	1063601	0	0
19059	19997	19997	819334503042871297	Thu Jan 12 00:05:42 +0000 2017	The flow chart is a most thoroughly oversold p...	1169875706	442	230	7	1063601	0	0
19060	19998	19998	819334419374899200	Thu Jan 12 00:05:22 +0000 2017	Ships are safe in harbor, but they were never ...	1169875706	442	230	7	1063601	0	0
19061	19999	19999	819334270825197568	Thu Jan 12 00:04:47 +0000 2017	Black holes are where God is dividing by zero.	1169875706	442	230	7	1063601	0	0

19062 rows × 12 columns

Figure 1 Data frame after data preprocessing

Moreover, we conducted data cleaning operations to refine the text corpus, which involved removing non-alphabetic characters, tokenizing the text into individual words, applying stemming with the PorterStemmer algorithm to reduce words to their root form, and subsequently applying lemmatization with the WordNetLemmatizer algorithm to further standardize words into their base form.

5.2. Text classification

5.2.1. Model Design

In this study, we performed feature engineering using BoW (Bag of Words) and TF-IDF (Term Frequency-Inverse Document Frequency) to extract relevant features from the text data. We then applied machine learning on two different approaches to classify the mental health disorder of each post. To accomplish this, we trained three different classifiers on the extracted features: Naive Bayes, Decision Tree, and Random Forest. We evaluated their performances on the test data using metrics such as accuracy, precision, recall, and F1-score. Moreover, we took into account the cost scenario of misclassifying a post from depression to healthy, which has a cost of 50, and misclassifying a post from healthy to depression, which has a cost of 10.

In addition, on an individual level, we calculated the percentage of depressed posts out of all posts posted by each person and defined the percentage greater than 30 percent as having underlying mental problems. These people are labeled as 1, meaning they are the people doctors should pay more attention to.

5.2.2. Evaluation

Our findings suggest that the Naïve Bayes classifier outperformed the other two classifiers, achieving the highest accuracy (86.25%). Besides, considering the cost scenario, using TF-IDF to select features and Naive Bayes as classifier was the best option, as it had the lowest cost of misclassification (cost 19840). Overall, our study highlights the importance of considering the cost scenario in addition to model performance when selecting a classifier for mental health disorder classification.

5.3. Sentiment Analysis

5.3.1. Model Design

The sentiment analysis was carried out using TextBlob and VADER libraries to derive polarity and compound scores respectively. The overall sentiment score distribution was visualized using a histogram to provide an overview of the dataset's overall sentiment. The tweets were grouped by user_id, and the average sentiment score for each user was calculated. To better understand the variation of sentiment for each individual, the sentiment score distribution was plotted for each user using a histogram. The TextBlob and VADER histograms were compared to determine the library that generated a better distribution of sentiment scores (Figure2 & Figure3).

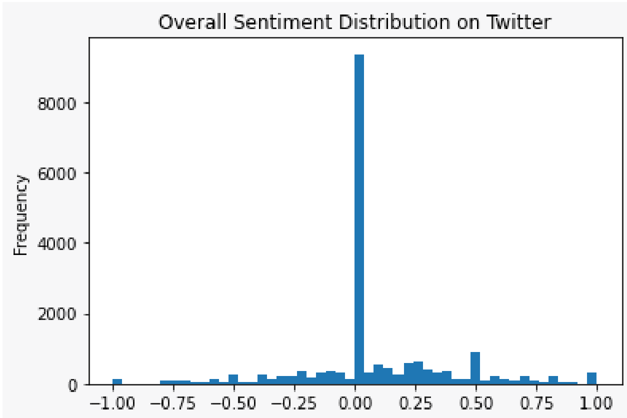

Figure 2 TextBlob Sentiment Distribution

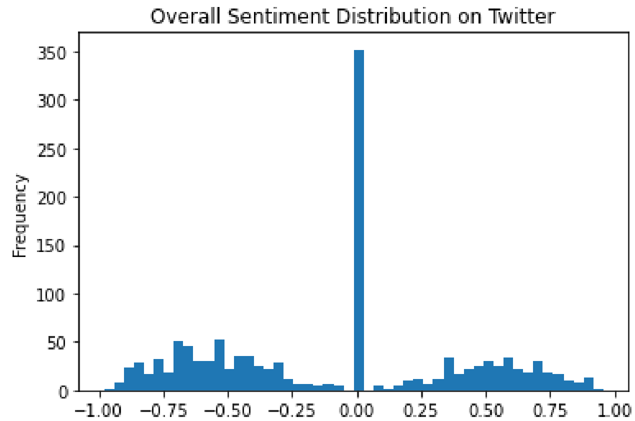

Figure 3 Vader Sentiment Distribution

Based on the two sentiment score distribution plots, it is evident that TextBlob classified nearly 50% of the texts as neutral with a sentiment score of zero, indicating a significant limitation in detecting sentiment. On the other hand, VADER's sentiment score distribution exhibited more discernible trends that align better with our original datasets. Consequently, we have opted to utilize VADER for subsequent analyses.

5.3.2. Results & Visualization

For the people who post more than 30 percent of negative posts and are labeled 1, We did an individual-level time-series sentimental analysis of people with underlying mental health disorder, trying to put the severity of their depression into different classes so we're able to tailor a treatment plan to each person. We present the outcomes of our analysis of individual-level time-series sentiment analysis utilizing bar charts. Each chart represents the sentiment trend for a specific individual labeled as 1 over time, where the X-axis denotes the time in days, and the Y-axis exhibits the average sentimental score for each day. Based on these charts, we were able to classify individuals having underlying mental health disorders into different mental state classes.

For instance, the first bar chart portrays a person with a consistently low mood, indicating long-term depression.(Figure 4) The second bar chart showcases a person with distinct stages of depression and happiness, occurring with equal frequency, which we refer to as periodic depression.(Figure 5) The final bar chart illustrates a person with a few days of relatively severe depression, indicating occasional depression.(Figure 6) Based on these classes we are able to provide the best treatment for different people according to their depression level.

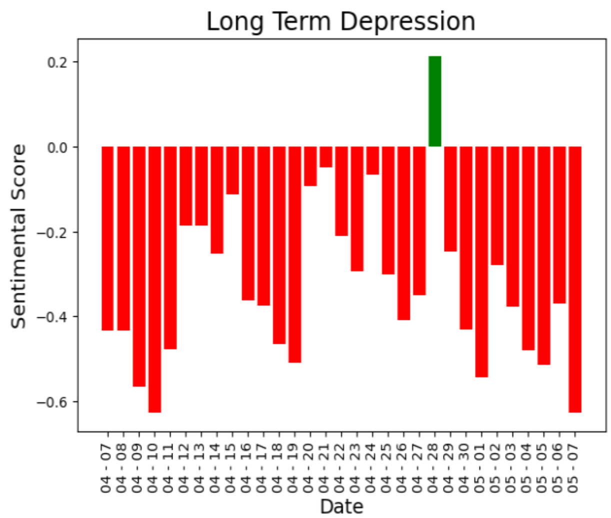

Figure 4 User_id = 1616997456

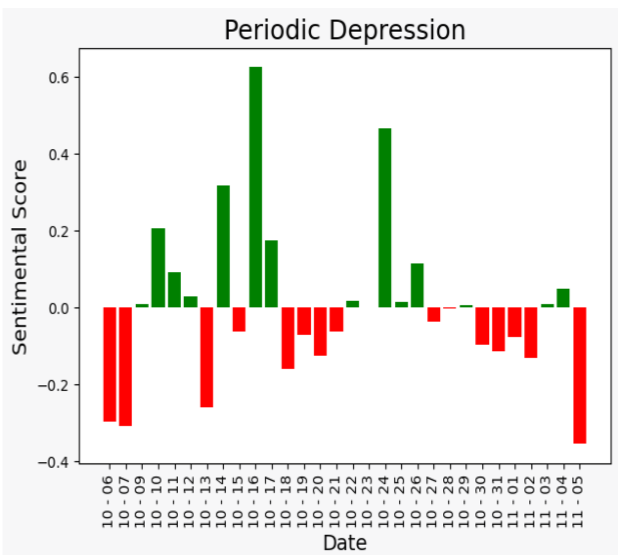

Figure 5 User_id = 1052121847

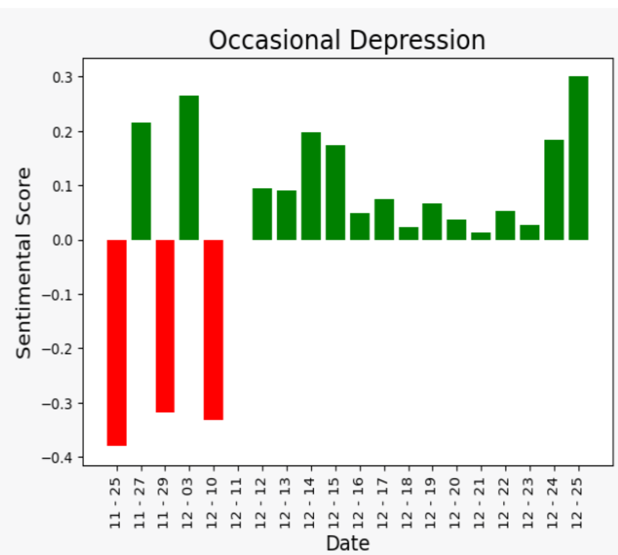

Figure 6 User_id = 20118423

Furthermore, we computed the mean sentiment score for all Twitter posts in each month. (Figure 7) We observed positive overall sentiment scores in most months except April. We speculate several rational reasons for why people's level of depression peaked in April.

- Seasonal affective disorder (SAD): April is often associated with the transition from winter to spring, and some people may experience symptoms of seasonal affective disorder (SAD) during this time. SAD is a type of depression that is related to changes in the seasons, and it is more common in the fall and winter months. However, some people may also experience SAD during the spring or summer months. If many people in your dataset are experiencing SAD during April, it could be contributing to the lower mean sentiment score.

- End of the school year: For many students, April marks the end of the school year and can be a stressful time due to upcoming final exams and projects. This stress can contribute to negative emotions and potentially lower sentiment scores in social media posts related to mental health.
- Tax season: In many countries, April is tax season, which can be a stressful and anxiety-inducing time for some people. The financial stress associated with tax season can also contribute to negative emotions and potentially lower sentiment scores in social media posts related to mental health.

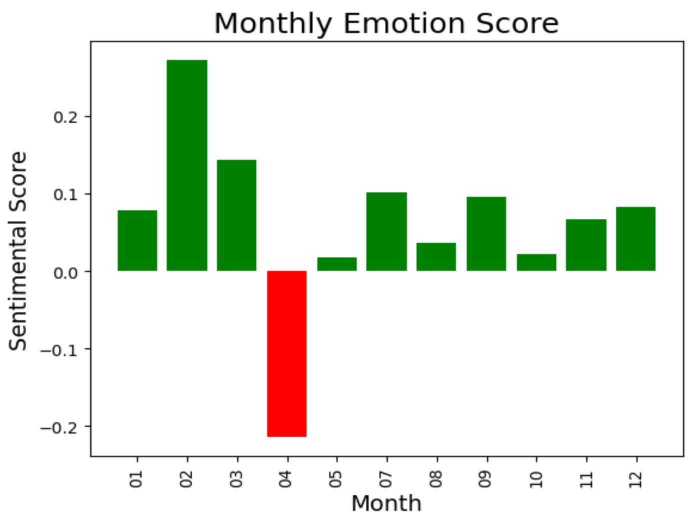

Figure 7 Monthly Sentiment Score

Additionally, we identified the top 30 words that appeared in tweets labeled as 1, such as 'depression', 'help', 'treatments', and 'life,' providing us with an overview of atypical posts.(Figure 8) By analyzing these words and their context, researchers or analysts can gain a better understanding of the language and content used in tweets related to depression, which can inform the development of interventions and strategies to address mental health issues on social media.

Figure 8 Word Cloud

5.4. Topic Modeling

5.4.1. Model Design

In this project, we conducted topic modeling on the posts of users labeled as depressed (label "1") to uncover the underlying themes and discover latent patterns in the content. We employed Latent Dirichlet Allocation (LDA), Correlated LDA Topic Model, and Latent Semantic Analysis (LSA/LSI) to identify the most 5 relevant topics within the text data. In order to eliminate the noises from social media posts, we further expanded our customized stop words list with some unmeaningful words such as 'tauri3l', 'liz_smith333', 'ly', 'co', 'allthxngsozzie', 'bit', and etc.

5.4.2. Results and Analysis

After comparing the results, we found that the LDA model performed the best in identifying coherent and distinct topics. The top 10 words for each topic identified by the LDA model and their corresponding themes are as follows:

<u>Topic #1</u>
Terms: "depression", "overcome", "overcome_depression", "health", "mental", "yes", "anyone", "anxiety", "family", "year"
Theme: *Coping with depression and related mental health issues*. This topic revolves around users discussing their experiences with depression and anxiety, as well as their efforts to overcome these challenges. It also highlights the importance of mental health and support from family members.

<u>Topic #2</u>
Terms: "like", "get", "one", "day", "make", "time", "know", "good", "need", "really"
Theme: *Daily experiences and personal struggles*. This topic captures users' daily lives, emotions, and personal battles. It reflects their struggles with everyday situations and the need for understanding and support from others.

<u>Topic #3</u>
Terms: "shit", "fuck", "stop", "yeah", "night", "positive", "tweet", "thinking", "gt", "also"
Theme: *Emotional outbursts and seeking positivity*. This topic is characterized by users expressing strong emotions and the use of profanity, while also seeking positive thoughts and perspectives to cope with their feelings.

<u>Topic #4</u>
Terms: "love", "go", "life", "much", "still", "never", "say", "even", "talk", "someone"
Theme: *Relationships, communication, and life challenges*. This topic encompasses users discussing their relationships, the importance of communication, and the difficulties they face in various aspects of their lives.

Topic #5:
Terms: "treatment", "depression", "depression_treatment", "help", "back", "headache", "oh", "amp", "going", "pain"
Theme: *Seeking treatment and managing physical symptoms*. This topic focuses on users seeking professional help and treatment for depression, as well as managing physical symptoms like headaches and pain.

5.4.3. Evaluation

The results of our topic modeling analysis, specifically the LDA model, provide valuable insights that can support various business functions and goals related to addressing mental health challenges. In this section, we will evaluate the effectiveness of the identified topics in addressing the aforementioned goals and present an example of how the analysis results can be helpful.

Identifying potential at-risk individuals.

The LDA model's ability to uncover themes such as coping with depression, daily experiences and struggles, and seeking treatment can enable mental health professionals to identify potential at-risk individuals. For instance, by monitoring social media posts and analyzing the frequency of users discussing topics related to depression and anxiety, professionals can prioritize support and interventions for those who seem to struggle the most with these issues. This targeted approach can help ensure that individuals receive timely and appropriate care, ultimately improving mental health outcomes.

Example: A mental health professional might observe that a user frequently posts about their struggles with coping mechanisms (Topic #1) and daily life experiences (Topic #2). This information could prompt the professional to reach out to the user and offer resources, support, or referrals to appropriate mental health services.

Understanding mental health trends and patterns.

The identified topics can help researchers and advocacy groups understand mental health trends and patterns on a broader scale. By examining how the prevalence and discussion of these topics change over time or differ between demographics, researchers can identify areas where mental health policies or interventions may need improvement.

Example: Researchers might find that discussions related to seeking treatment (Topic #5) have increased significantly in a specific demographic group or region. This information could suggest a growing need for mental health services in that area, prompting policymakers to allocate more resources or develop targeted interventions.

Creating targeted advertising campaigns for mental health awareness.
Social media platforms can leverage the insights gained from our topic modeling analysis to create targeted advertising campaigns that promote awareness and understanding of mental health disorders. By focusing on the themes and language patterns identified in the analysis, platforms can develop content that resonates with users and encourages them to seek help or support when needed.

Example: A social media platform could use the language patterns identified in Topic #3 (emotional outbursts and seeking positivity) to create an advertising campaign that targets users who frequently express strong emotions in their posts. The campaign could promote resources, such as mental health apps or helplines, that offer support and coping strategies for managing emotions.

6. Conclusion

The Naive Bayes classification can classify posts as depressed or non-depressed. In order to find people who have an underlying mental health disorder and need medical care, we collect all posts of each person and calculate the percentage of depressed posts. If the percentage is higher than 30%, we give mental health professionals and researchers a list of whom they should provide medical care with.

For those people who have an underlying mental health disorder, our individual-level time-series sentimental analysis will give a visual picture of their mental state. Doctors can offer more specific and personal treating advice for them based on their level of depression. Besides, the bar chart of sentiment score for all Twitter posts in each month can help researchers and advocacy groups better understand trends and patterns in mental health by analyzing monthly mood trends and disaggregated results. This information can inform the development of new mental health policies and interventions to improve mental health outcomes at the population level.

The topic modeling results obtained through the LDA analysis have proven effective in supporting various business functions related to addressing mental health challenges. By identifying themes and topics in social media posts, our analysis can help mental health professionals, researchers, and social media platforms make informed decisions and develop targeted interventions to improve mental health outcomes.

7. Future Direction

The present project has concentrated on examining social media data for the identification and analysis of mental health disorders and conditions. Despite the valuable insights gained, there remain numerous avenues for further exploration in future research endeavors.

One potential direction for continued investigation involves adapting the developed models and methodologies to scrutinize mental health-related language on alternative social media platforms or within other online contexts. These may include online forums, chat rooms, and virtual therapy sessions. By broadening the scope of analysis, we could acquire a more comprehensive understanding of mental health-related language and behavior across diverse online settings.

Another promising avenue for future research is the implementation of multi-label classification to recognize and categorize social media posts associated with various mental health disorders or conditions. Given that the current project focused primarily on binary classification of tweets as either indicative of depression or not, it is essential to acknowledge that mental health is a multifaceted and intricate subject involving multiple disorders and conditions. Employing multi-label classification techniques can offer a more detailed comprehension of the mental health-related language utilized in social media.

Moreover, while the VADER library employed for sentiment analysis in this project provides valuable insights into the overall mood of mental health-related social media posts, its capacity to capture the subtleties and complexities of mental health-related language may be limited. As a result, future research could benefit from more refined sentiment analysis approaches using customized libraries that account for specific emotions or affective states pertinent to mental health. This would facilitate a deeper understanding of the emotions and attitudes expressed in mental health-related language.

Lastly, the current project conducted individual-level time-series analysis to examine the fluctuations in sentiment over time for each user. An intriguing direction for subsequent research would be to investigate how sentiment changes over time at a population level. This might involve identifying trends and patterns in sentiment across various regions or demographic groups, ultimately providing a more comprehensive understanding of mental health-related language and behavior among diverse populations.

8. References

Al-Mosaiwi, M., & Johnstone, T. (2018). In Search of Subjective Well-Being: Identifying Precursors to Subjective Well-Being in Mental Health Disorders Using Machine Learning. Frontiers in Psychology, 9, 1711.

Cohan, A., & Goharian, N. (2018). Detecting Depression Symptoms in Twitter Users Using a Hybrid Approach. Journal of Biomedical Informatics, 83, 103-111.

De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting Depression via Social Media. Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media, 128-137.

Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2019). Detecting Depression and Mental Illness on Social Media: An Integrative Review. Current Opinion in Behavioral Sciences, 31, 51-59.

Huang, X., Li, Y., Zhang, Y., Liu, Y., & Zhang, X. (2018). Detecting Depression in Tweets Using

Multi-view Convolutional Neural Networks. Journal of Biomedical Informatics, 88, 59-67.

Reece, A. G., Danforth, C. M., & Langer, E. J. (2017). Psychological Language on Twitter Predicts County-Level Heart Disease Mortality. Psychological Science, 28(2), 138-150.

Tsugawa, S., Kikuchi, K., Kishino, F., Nakajima, K., & Kuroiwa, S. (2015). Monitoring Depressive Symptoms Using Speech and Language Information: A Preliminary Study. Journal of Medical Internet Research, 17(1), e10.