Dizo, Joanne Micaela R.
Montemayor, Annika Nicole F.
Shu Too, Kaitlyn L.
Yumol, Dianne C.
MATH 103.1-B

MindInsight Classifier: Unveiling Mental Health Patterns in Pandemic Discourse through Data-Driven Analysis

I. Introduction

The global pandemic has not only reshaped our physical landscape but has also significantly impacted the mental health of individuals worldwide. Amidst these challenging times, online platforms have become an outlet for many to share their experiences and seek support. Reddit, being a prominent online community, has served as a rich source of real-time insights into the mental health struggles faced by individuals during the pandemic.

The remarkable increase in mental health-related conversations on Reddit during the pandemic presents an invaluable opportunity to utilize data-driven approaches to understand, classify, and potentially assist in identifying various mental health disorders. Our project seeks to utilize natural language processing (NLP) and machine learning techniques to examine an extensive collection of Reddit posts from this period.

By analyzing the linguistic patterns, sentiments, and contextual cues present in these posts, we aim to create a classifier that not only identifies the presence of mental health issues but also distinguishes between different disorders. The significance of this project lies in its capacity to provide valuable insights into the occurrence and manifestation of different mental health disorders amid the pandemic.

II. Discussion on Data and Methods

The chosen dataset was extracted from Kaggle and is titled “Mental Disorders Identification (Reddit)”. It contains 700,000 rows of text data obtained from six (6) subreddits, namely BPD, Anxiety, Depression, Bipolar, Schizophrenia, and Mental Illness. The data was collected using the Reddit API and PushShift API by Kamarul Adha last 2022. The dataset can be accessed online here. Listed below are the columns of the dataset:

Table 1. Dataset Information

Column	Data Type	Description
`title`	String	This is the title of the post on Reddit. 46 posts have no `title`.
`selftext`	String	This contains the contents of the post.

		12% of the dataset has been marked as `[removed]`, while 5% of the dataset contains missing values.
`created_utc`	Integer	This is the time the post was published in UTC format. The posts in this dataset are dated from November 19, 2010 to November 24, 2022.
`over_18`	Boolean	This returns `True` if the post is categorized as "Over 18", which implies that its contents are inappropriate for minors, and `False` otherwise. 3% of the posts are categorized as suitable for audiences above the age of 18.
`subreddit`	String	This is the target variable of the dataset. It contains the subreddit under which the post was published. There are six (6) categories: `BPD`, `bipolar`, `depression`, `Anxiety`, `schizophrenia`, `mentalillness`.

For the purposes of this project, the dataset will be cleaned to eliminate rows with removed or missing values under `selftext`. Moreover, as we wish to spotlight on posts published during the duration of the COVID-19 pandemic, we will be limiting our data to only include posts from March 2020 onwards. A random sample of 1,000 posts will be taken from the dataset for efficiency.

As for our methods, we will utilize the standard natural processing language (NLP) pipeline as seen in the image below:

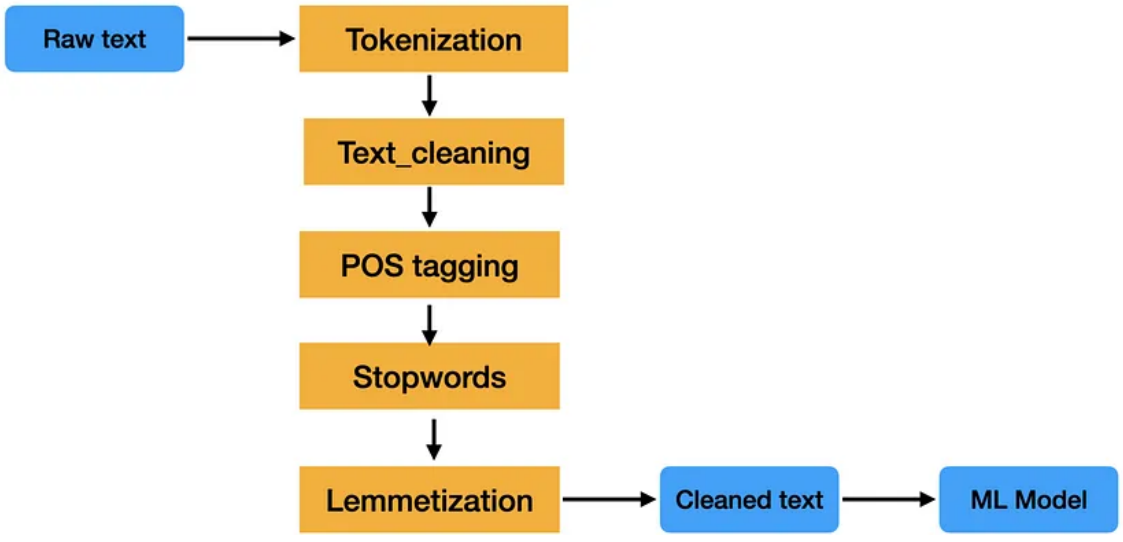

Figure 1. Natural Language Processing Pipeline

After performing data cleaning and conducting a simple exploratory data analysis, we will tokenize our dataset. Tokenization refers to the process of converting a sequence of text into smaller parts, known as tokens. These tokens can be as small as characters or as long as words. The primary reason this process matters is that it helps machines understand human language by breaking it down into bite-sized pieces, which are easier to analyze.

For our project, we will apply word tokenization. This method breaks text down into individual words. It is the most common approach to tokenization and is particularly effective for languages with clear word boundaries like English.

Our next step involves text cleaning, which includes but is not limited to the following procedures:

- Lowercasing the text data
- Removing punctuations
- Removing numbers
- Removing extra space
- Replacing the repetitions of punctuations
- Removing emojis
- Removing emoticons
- Removing contractions

If time permits, we can also utilize part-of-speech tagging or POS tagging. POS tagging is a process where each word in a text is labeled with its corresponding part of speech. This can include nouns, verbs, adjectives, and other grammatical categories.

Removing stopwords is also a crucial part of the NLP pipeline. Stopwords are the most common words in a language like “the”, “a”, “me”, “is”, and etc. These words do not carry important meaning and are usually removed from texts.

Lastly, we will perform lemmatization. Lemmatization consists in doing things properly with the use of vocabulary and morphological analysis of words, to return the base or dictionary form of a word, which is known as the lemma.

After performing the above steps, we will now proceed to data modeling. Both Convolutional Neural Network (CNN) and Long-Short Term Memory (LSTM) networks will be used in this project as our modeling method. First, CNN is a deep-learning neural network that is able to analyze and classify various data types such as image, text, and audio data. While CNN is mainly used for analyzing images, this technique has also been proven effective in Natural Language Processing (NLP) tasks, as it is able to detect and extract complex features in text data.

On the other hand, LSTM network also belongs under deep learning and is a type of Recurrent Neural Network (RNN). LSTM network is able to handle sequential data and capture long-term dependencies. As such, the LSTM network is highly capable and more

accurate when analyzing long strings of text. LSTM network also has various applications in NLP tasks such as text classification, speech recognition and language modeling and translation. In this project, the two deep learning methods will be used in order to analyze and classify the Reddit posts based on the mental disorder being mentioned in the post. Finally, in order to evaluate the model, we will calculate the accuracy, precision, recall, and F1 score of the model.

III. Results

Simple Exploratory Data Analysis

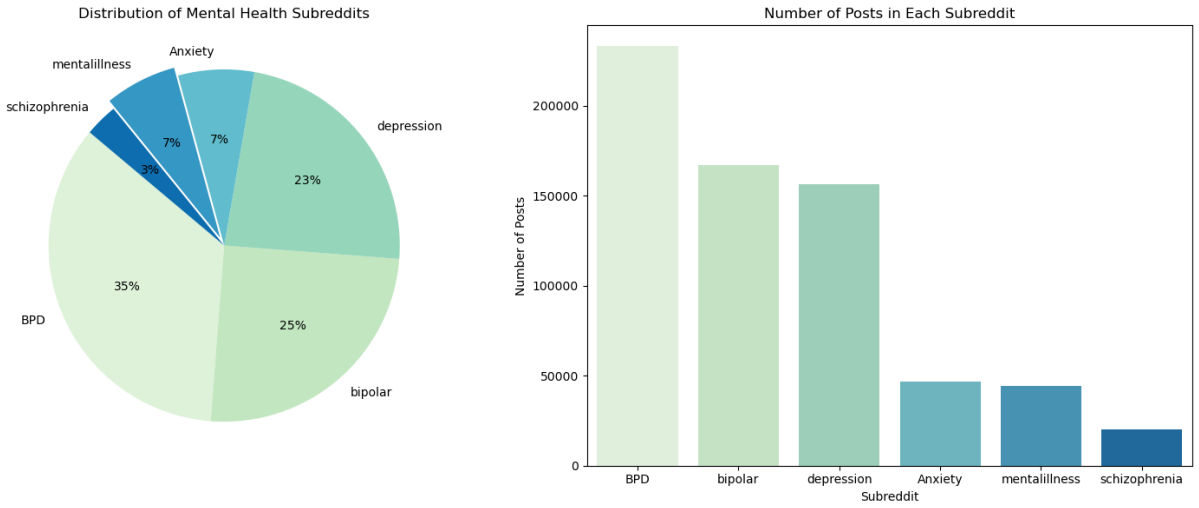

Figure 2. Distribution of Reddit Posts Categories

The figure above illustrates the distribution of over 700,000 Reddit posts across six distinct subreddits: BPD, bipolar, depression, anxiety, schizophrenia, and mental illness. The data reveals that the most prevalent category among these subreddits is BPD, constituting 35% of the posts, followed closely by bipolar and depression, accounting for 25% and 23%, respectively. In stark contrast, the subreddit related to schizophrenia represents the least common category, comprising only 3% of the total posts analyzed. These findings offer a clear insight into the varying levels of engagement and discussion within these mental health-related subreddits

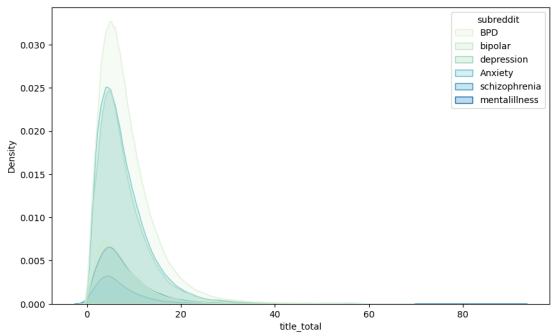

Figure 3. Word Count Distribution in Post Titles by Reddit Posts

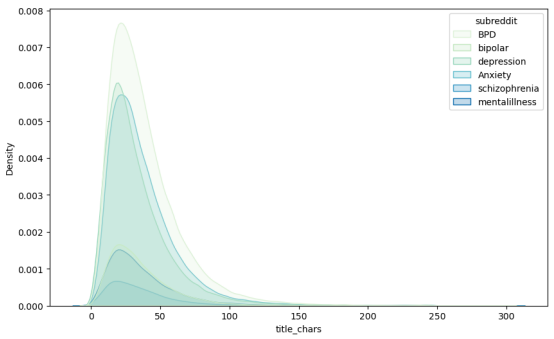

Figure 4. Character Count Distribution in Post Titles by Reddit Posts

Titles, typically limited in length, present a snapshot of the post's content, aiming to capture attention and provide a brief overview of the topic. The distribution analysis of word and character counts in titles reveals a focus on brevity, with an emphasis on attention-grabbing phrases, keywords, or questions to entice engagement. Figures 3 and 4

above illustrate how the distribution of both word count and character count of different subreddits are skewed to the left which means that they tend to be concise.

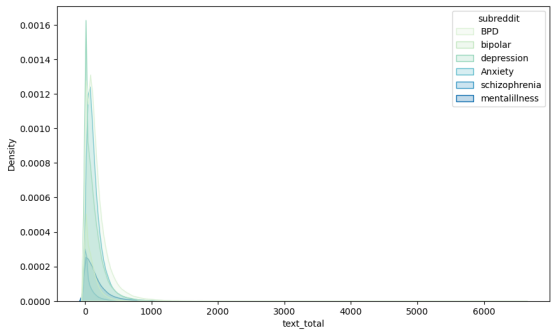

Figure 5. Word Count Distribution in Post Texts by Reddit Posts

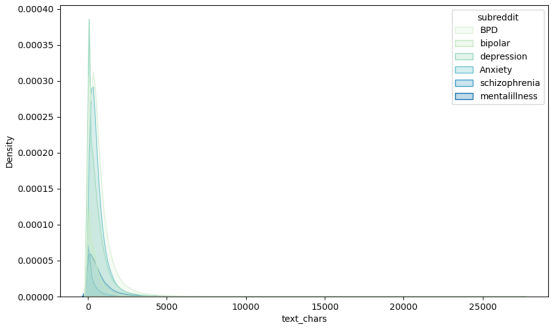

Figure 6. Character Count Distribution in Post Texts by Reddit Posts

Word count and character count analysis in Reddit posts related to mental health disorders provide valuable quantitative insights into these discussions' depth, focus, engagement, and language nuances. Figures 5 and 6 above show the different subreddits' word count and character count distributions. It can be observed that although the subreddits exhibit similar mean counts, their variances differ significantly. This indicates consistent discussion lengths across disorders but varying levels of content variability.

Figure 7. Word Cloud of Post Titles

The word cloud above, derived from sampled Reddit post titles on mental disorders, prominently features terms like “BPD,” “anxiety,” “help,” “feel,” and “depression.” This highlights prevalent themes, emphasizing discussions around specific disorders, seeking support, expressing emotions, and addressing common issues. This could further provide insights into what categories are most common and the similarities among symptoms of different disorders.

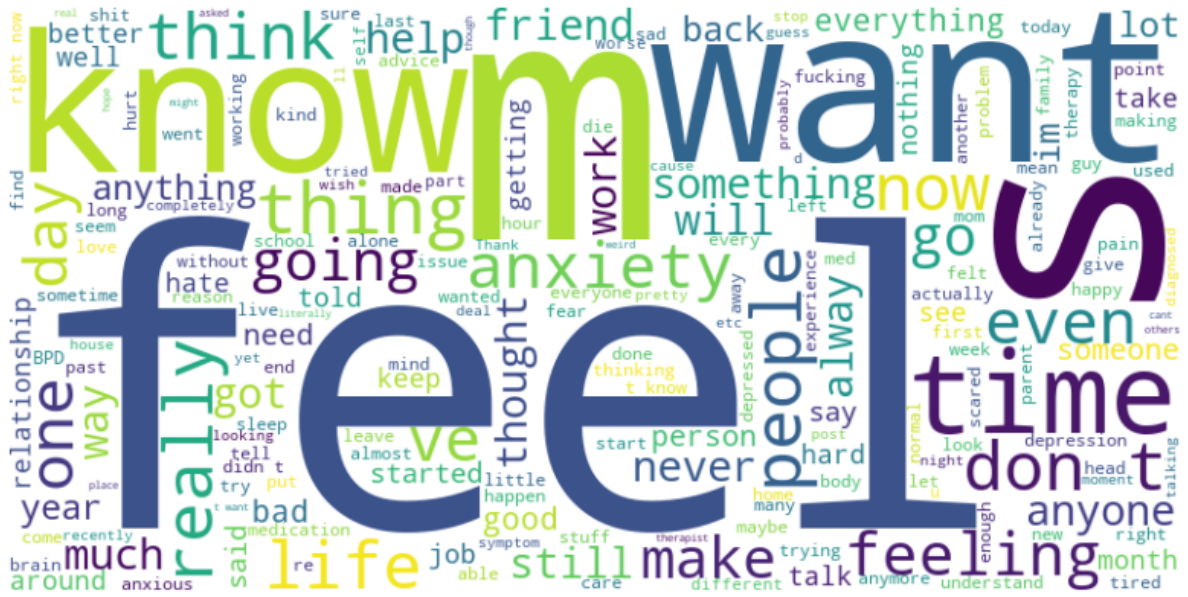

Figure 8. Word Cloud of Post Texts

The word cloud generated from the main texts of the sampled Reddit entries reveals prominent terms such as “feel,” “know,” “want,” “anxiety,” and “work.” These significant words suggest a focus on emotions, personal experiences, desires, and the impact of anxiety within the discussions. Compared to the word cloud from the post titles, this provides a deeper understanding of the context and subjective experiences captured in the posts.

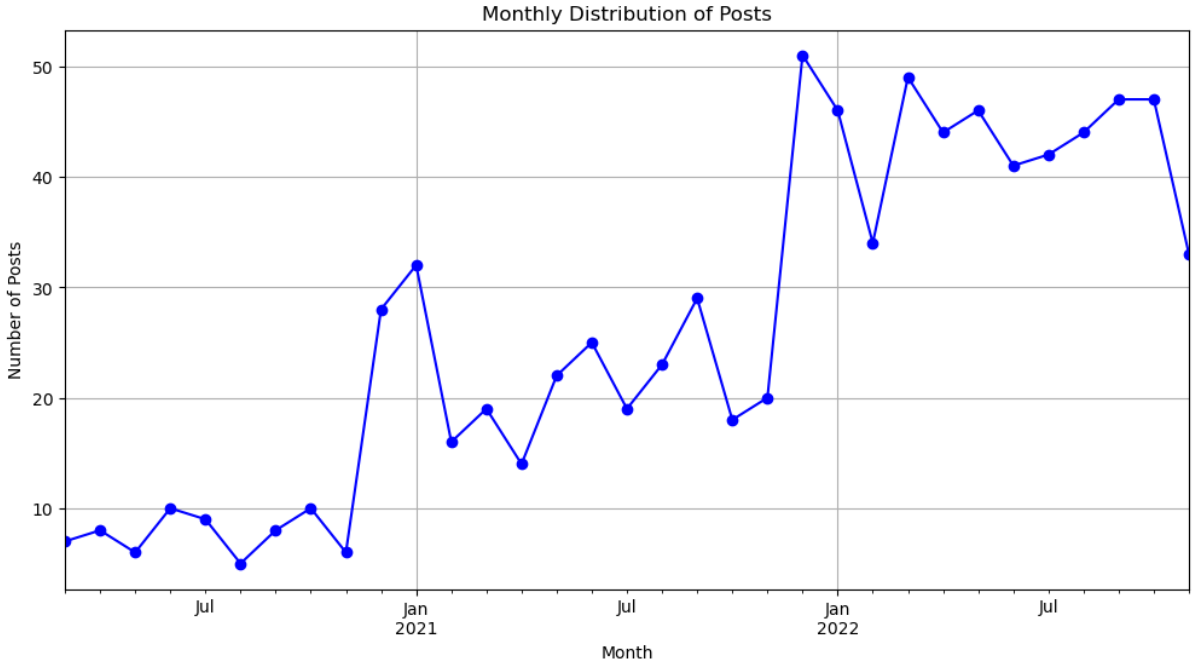

Figure 9. Temporal Analysis of Reddit Posts Starting from March 2020

The graph above shows the monthly distribution of the sampled Reddit posts from March 2020 to November 2022. A temporal analysis of Reddit posts about mental health can provide a nuanced understanding of how discussions evolve, when users are most active, which topics gain traction, and how external events influence online discourse. For instance, it can be observed that posts continue to increase with peaks during the end of the year which can be correlated to the holidays.

Labeling Strategy

Given the multi-class nature of the dataset, we initially attempted to classify six different mental health conditions: BPD, Bipolar, Depression, Anxiety, Schizophrenia, and a category for other illnesses. However, the achieved accuracies across all classes varied significantly. Consequently, a decision was made to refine the scope of the project to focus on classifying posts related to anxiety and schizophrenia, combining the other four classes into a unified "others" category.

The following accuracies were obtained for various class combinations:

Table 2. Accuracies for Various Class Combinations

Class Combinations	Accuracy Score
BPD and Bipolar, the two most common illnesses (and others)	0.73
BPD only (and others)	1.00
BPD and anxiety (and others)	0.57
BPD and depression (and others)	0.54
Depression only (and others)	1.00
Depression and anxiety (and others)	0.53
Schizophrenia only (and others)	1.00
Schizophrenia and anxiety, the two least common illnesses (and others)	0.93

Based on the evaluation, achieving 100% accuracy for posts related to a single illness, such as BPD only, depression only, or schizophrenia only, suggests that the models can effectively distinguish between these specific conditions when they are the sole focus.

Moreover, the combination of BPD and bipolar, being the most common illnesses, yielded a relatively high accuracy of around 70%. This indicates that the models can

differentiate well between these more prevalent conditions. However, Reddit entries inherently offer limited insight into an individual's experiences with mental health conditions like borderline personality disorder (BPD) or bipolar disorder. Users might describe symptoms such as mood swings, impulsivity, or relationship challenges, which are characteristic of both BPD and bipolar disorder.

In spite of that, these descriptions often lack the nuanced context necessary for a precise diagnosis. This poses a challenge for NLP models relying solely on Reddit entries to differentiate between these conditions. These disorders often coexist or share comorbidities; a study showed that 20% of patients with BPD were also diagnosed with bipolar disorder. Conversely, it was also found that 30% of patients diagnosed with bipolar disorder were also diagnosed with BPD (Zimmerman & Morgan, 2022).

Moreover, this pair of mental health conditions also face a significant case of misdiagnosis – it was found that borderline personality disorder is misdiagnosed as bipolar disorder 40% of the time, especially when they exhibit symptoms of instability, anger, impulsivity, and recurrent suicidal behavior (Ruggero et. al., 2010). This further complicates the task of accurate identification through text analysis alone. Therefore, a nuanced understanding of the context is crucial.

However, combining two illnesses, especially if they are less common, resulted in lower accuracies, hovering around 50%. This suggests that distinguishing between two specific mental health conditions is a more challenging task, possibly due to overlapping symptoms or language patterns. Combinations such as BPD and anxiety, and BPD and depression fall under this category.

Studies suggest that BPD and anxiety are difficult to distinguish as the two are often associated with one another and have overlapping symptoms. Furthermore, studies have shown that people that are diagnosed with BPD are often also diagnosed with anxiety disorder (Abraham, 2020). The reason for this is that those with BPD often experience rapid mood shifts that usually entail negative feelings and emotions, such as anxiety (“Borderline Personality Disorder”, 2022).

Similar to anxiety, depression has overlapping symptoms with that of BPD, making it difficult to distinguish between the two mental disorders. As those diagnosed with BPD often have difficulty regulating their intense emotions and mood swings, they often experience very negative thoughts and feelings. Particularly, studies have shown that those with BPD have recurring experiences of self-harming and suicidal thoughts as well as feelings of emptiness and loneliness, which are also symptoms of depression (“Why is Borderline Personality Disorder Hard”, 2018).

Another instance of lower accuracies due to overlapping symptoms is classifying between anxiety and depression. Distinguishing between anxiety and depression presents challenges due to their intertwined nature and shared symptoms. Both conditions exhibit

overlapping features such as difficulty concentrating, irritability, and changes in appetite. Moreover, individuals experiencing depression often report feelings of anxiety and worry. Notably, these disorders can trigger or exacerbate each other, with anxiety frequently preceding depression. Moreover, approximately 60% of people with anxiety experience symptoms of depression, and vice versa.

Surprisingly, classifying the two least common illnesses, schizophrenia and anxiety, yielded a high accuracy of 93%. This could be attributed to the distinct language patterns associated with these less frequent conditions, making them easier to differentiate.

Research has shown that distinguishing between anxiety disorders and schizophrenia is often more straightforward due to the differences in their symptoms and the general nature of each disorder. Anxiety disorders primarily involve restlessness, irritability, and trouble concentrating, often accompanied by physical symptoms like increased heart rate, sweating, and muscle tension. In contrast, schizophrenia is characterized by a profound disruption in thought processes, perceptions, and emotions. Its symptoms may include hallucinations, delusions, and disorganized speech or behavior. This allows NLP models to easily contrast schizophrenia and anxiety in Reddit entries.

In light of these observations, we decided to focus on classifying between anxiety, schizophrenia, and others. Anxiety and schizophrenia represent two critical and distinct categories within the mental health domain. By concentrating on these conditions, our project, MindInsight Classifier, can contribute valuable insights into the language used by individuals experiencing anxiety or schizophrenia, potentially aiding in development of tools for early detection/intervention and support.

Discussion and Analysis

As mentioned above, the objective of this project is to classify Reddit posts into three categories: anxiety, schizophrenia, and others. Two deep learning models, CNN and LSTM, were employed and evaluated based on their training and testing accuracies, as well as precision, recall, and F1 scores. It is crucial to delve into the evaluation metrics that provide a comprehensive understanding of our models' performances.

According to Afeworki (2021), precision is defined as the measure of the actual correctly identified positive cases from all the predicted positive cases. On the other hand, recall is the measure of many of the actual values of a class our model was able to capture. Finally, the F1 score is the harmonic mean of recall and precision, giving a better representation and measure of the incorrectly classified cases (false positives and false negatives).

Upon inspection of the accuracies of both models, evident in Table 2 and Table 3, both CNN and LSTM models exhibit stable training and testing accuracies across different

epochs. However, CNN consistently outperforms LSTM in terms of accuracy, achieving higher testing accuracy throughout.

Table 3. CNN Evaluation Results

Number of Epochs	Training Accuracy	Testing Accuracy
10 Epochs	0.9333	0.9474
50 Epochs	0.9333	0.9474
100 Epochs	0.9333	0.9474

Table 4. LSTM Evaluation Results

Number of Epochs	Training Accuracy	Testing Accuracy
10 Epochs	0.9511	0.8772
50 Epochs	0.9511	0.8772
100 Epochs	0.9511	0.8772

Additionally, CNN demonstrates superior performance across all evaluation metrics as seen in Table 5. It achieves a higher precision score (0.8975) compared to LSTM (0.76947), indicating a better balance between true positives and false positives. Both models show similar recall scores, suggesting their ability to capture relevant instances. However, CNN's overall higher recall contributes to its superior F1 score (0.9218) compared to LSTM (0.81981).

Table 5. Summary of Evaluation Metrics and Results

Evaluation Metric	CNN	LSTM
Accuracy Score	0.9474	0.87719
Precision Score	0.8975	0.76947
Recall Score	0.9474	0.87719
F1 Score	0.9218	0.81981

Insights and Recommendations

Following directly from our results, we have seen how the constructed NLP model was able to distinguish between mental health conditions and is able to categorize texts according to the mental health condition of the author of the Reddit post. As such, this opens up the potential application of artificial intelligence (AI) in detecting moderate to extreme cases of mental health conditions in online spaces, such as social media and search platforms.

Since AI has the ability to make distinctions among common mental health conditions, it can then be used as a tool to introduce interventions upon detecting extreme conditions in posts about one's mental health and well-being by redirecting them to mental health hotlines and other forms of mental health support systems.

However, while the NLP model for this project was shown to be capable of categorizing mental health conditions based on Reddit posts, it exhibited difficulty in distinguishing conditions with comorbidities, that is, commonly associated mental illnesses that have overlapping symptoms. It is possible that the shared language patterns brought about by similar experiences may have affected the model's ability to accurately distinguish these mental health conditions. In this manner, future researchers can look into improving the existing NLP model to meet this need or choose to adopt other machine learning algorithms to further improve the accuracy in categorizing mental health conditions.

Another point of consideration is the choice of data source. The Reddit posts in the dataset are primarily based on the subjective experiences of the authors who were diagnosed with mental health conditions. In this regard, there may be a lack of medical terminologies and other technical information about the authors' condition, which can consequently hinder the model's capabilities of distinguishing the nuances of the mental health conditions being discussed in the Reddit posts.

Hence, we recommend for future researchers to improve the existing model by trying various groupings of comorbidities or mental health conditions, as the group only tested the model on different pairings of mental health conditions. We also recommend using other machine learning methods or algorithms in order to improve the performance of the models and accuracy of predictions. Lastly, we also recommend adopting different datasets from different online platforms, such as other social media platforms. More specifically, datasets with text that contain medical terminology may be used in constructing machine learning models that aim to distinguish among common mental health conditions.

Conclusion

Overall, CNN and LSTM were two algorithms used to predict and distinguish the mental health disorders associated with the Reddit posts in the chosen dataset. It is important to note that the project focused on classifying posts related to anxiety and schizophrenia, combining the other four classes into a unified "others" category, due to the interrelated nature of the other mental illnesses. Across all evaluation scores, CNN performed better in comparison to LSTM. With that being said, these machine learning models open opportunities to distinguish mental illnesses based on texts in online platforms and can potentially become avenues to offer interventions for posts related to moderate and extreme cases of mental health disorders.

IV. References

Abraham, M. (2020, October 10). *Anxiety and borderline personality disorder*. Calm Clinic. https://www.calmclinic.com/anxiety/causes/borderline-personality-disorder

Afeworki, M. (2021, July 19). Precision, recall, accuracy, or F1 score? *Medium*. https://milena-pa.medium.com/precision-recall-accuracy-or-f1-score-c53ca78c03a

Borderline personality disorder: Causes, symptoms & treatment. (2022, May 20). Cleveland Clinic. https://my.clevelandclinic.org/health/diseases/9762-borderline-personality-disorder-bpd

BPD vs. Bipolar disorder: Similarities and differences. (2021, May 11). Psych Central. https://psychcentral.com/disorders/bpd-vs-bipolar-disorder

Convolutional neural network tutorial [update]. (n.d.). Simplilearn.Com. Retrieved December 12, 2023, from https://www.simplilearn.com/tutorials/deep-learning-tutorial/convolutional-neural-network

Casarella, J. (2023, January 18). *Depression vs. Anxiety: Which one do I have?* WebMD. https://www.webmd.com/depression/depression-or-anxiety

Chen, J. Q., Qi, K., Zhang, A., Shalaginov, M., & Zeng, T. H. (2022). Covid-19 impact on mental health analysis based on reddit comments. *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2253–2258. https://doi.org/10.1109/BIBM55620.2022.9995512

elvis. (2020, April 18). Deep learning for nlp: An overview of recent trends. *DAIR.AI*. https://medium.com/dair-ai/deep-learning-for-nlp-an-overview-of-recent-trends-d0d8f40a776d

Gillette, H. (2023, January 25). *Schizophrenia and anxiety: Differences and overlap*. Healthline. https://www.healthline.com/health/schizophrenia/schizophrenia-and-anxiety

Mishra, M. (2020, September 2). Convolutional neural networks, explained. *Medium*. https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939

Monica, D. (2020, September 7). Basic steps in natural language processing pipeline. *MLearning.Ai.* https://medium.com/mlearning-ai/basic-steps-in-natural-language-processing-pipeline-763cd299dd99

Palm, A., & Rodriguez, C. (2023). Borderline personality disorder misdiagnosed as bipolar disorder. In H. Tohid, L. G. Baratta, & H. Maibach (Eds.), The Misdiagnosis Casebook in Clinical Medicine (pp. 459–464). Springer International Publishing. https://doi.org/10.1007/978-3-031-28296-6_62

Reddit could tell us how the coronavirus is affecting mental health. (n.d.). CNET. Retrieved December 12, 2023, from https://www.cnet.com/culture/reddit-could-tell-us-how-the-coronavirus-is-affecting-mental-health/

Ruggero, C. J., Zimmerman, M., Chelminski, I., & Young, D. (2010). Borderline personality disorder and the misdiagnosis of bipolar disorder. *Journal of Psychiatric Research*, *44*(6), 405–408. https://doi.org/10.1016/j.jpsychires.2009.09.011

Shekhar, S. (2021, June 14). What is LSTM for text classification? *Analytics Vidhya*. https://www.analyticsvidhya.com/blog/2021/06/lstm-for-text-classification/

Sherrell, Z. (2022, March 29). *Anxiety vs. depression: Symptoms, treatment, and more.* Medical News Today. https://www.medicalnewstoday.com/articles/anxiety-vs-depression

Thakur, A. (2020, May 28). What is LSTM? Introduction to long short term memory. *Intellipaat*. https://intellipaat.com/blog/what-is-lstm/

Tjornehoj, T. (n.d.). *The relationship between anxiety and depression.* Hartgrove Behavioral Health System. https://hartgrovehospital.com/relationship-anxiety-depression/

What are convolutional neural networks? | definition from techtarget. (n.d.). Enterprise AI. Retrieved December 12, 2023, from https://www.techtarget.com/searchenterpriseai/definition/convolutional-neural-network

Why is borderline personality disorder hard to diagnose? (2018, April 27). Clearview Treatment Programs. https://www.clearviewtreatment.com/resources/blog/why-is-borderline-personality-disorder-hard-to-diagnose/

Zimmerman, M., & Morgan, T. A. (2013). The relationship between borderline personality disorder and bipolar disorder. *Dialogues in Clinical Neuroscience*, *15*(2), 155–169. https://doi.org/10.31887/DCNS.2013.15.2/mzimmerman