Analyzing the Effects of COVID-19 on Social Media Activity in Mental Health Subreddits using AI: Large Language Models and Time Series Analysis

Hanna Song

*School of Engineering and Applied Science*

*The George Washington University*

Washington, DC, USA

[hanna.song@gwmail.gwu.edu](mailto:hanna.song@gwmail.gwu.edu)

Anh Nguyen

*School of Engineering and Applied Science*

*The George Washington University*

Washington, DC, USA

[apnguyen@gwmail.gwu.edu](mailto:apnguyen@gwmail.gwu.edu)

***Abstract*—During the COVID-19 pandemic, there was a significant increase in social media usage, as it allows individuals to freely express their thoughts and emotions. This study analyzes post trends and engagement with mental health-related posts on social media platforms, specifically Reddit, in order to capture the changes brought about by COVID-19. Rather than focusing on individual users, we examine individual posts from 9 different mental health-related subreddits to understand how trends and engagement have changed before and after the COVID-19 pandemic. Our research utilizes the Large Language Model BERT, DistillBERT, BERTopic by binary classification to analyze post texts, and Prophet Time Series to compare overall post trends and engagement before and after COVID-19.**

***Keywords—Big Data, Health Data Analytics, Mental Health, Natural Language Processing, Large Language Model, Binary Classification, Topic Modeling, Time Series Analysis, Social Media, Text Mining***

# Introduction

The recent global pandemic has brought about various stressors and uncertainties such as health concerns, economic instability, social isolation, and disruptions to daily routines, which lead to increased levels of anxiety, fear and emotional distress among individuals. Furthermore, COVID got us to realize how social isolations can lead to an increased usage of social media platforms, which might act as an early sign of mental illness. The global prevalence of anxiety and depression increased by a massive 25 percent which increased unprecedented stress caused by the social isolation resulting from the pandemic. The survey results from NIH indicates that COVID-19 can impact mental health in the sense that you get COVID-19, you may experience a number of symptoms that affects your brain and mental capacity including brain fog, anxiety, depression, psychosis, and seizures, as well as potentially showcase suicidal tendency. These behaviors are shown from individuals’ social media activities especially during the pandemic, indicating that they have been more likely to engage with social media. As Reddit communities serve as safe spaces for individuals to express their thoughts, emotions and concerns related to mental health, the study will observe how post trends and post engagements on the platform change over the pandemic.

* 1. *Research Question*

Does COVID impact post trends and engagements within mental illness-related subreddits?

* 1. *Hypothesis*

Mental health-related subreddits will exhibit a shift in post trends as well as the number of engagements –such as post text, number of comments, score and upvote ratios– between the pre-COVID and post-COVID periods.

# related works

Prior research in mental health analysis has been quantifying user’s mental health status by their content sharing on social media platforms such as Twitter or Reddit, which are high-volume data based on an individual's online interactions. Previous studies mostly approach the user's mental health prediction by sentiment analysis or topic modeling of posts by users.

Specifically for prior work on Social Media and Natural Language Processing, Reagan et al. 2022 illustrates analysis on Twitter datasets with entailing sentiment analysis, and topic modeling to address the problem of discovering knowledge on mental health issues from social media data on Twitter [6]. Moreover, they also utilized sLDA that is a supervised topic model with conclusions that define a set of documents of each author on a weekly basis rather than aggregating by author [6]. Since topics and emotional statistics vary over time and language samples grouped on a weekly basis, they are likely to have more internal coherence than samples aggregated over long periods [6].

Furthermore, Choudhury et al. 2014 utilized mental health-related subreddit datasets extracted by reddit’s official API [7]. The study observed unigrams in various semantic categories provided by the psycholinguistic lexicon LIWC [7]. Additionally, using response variables of Karma and Comments to measures of social support among the posts. Karma is net votes that a post receives, and comments are also dependent variables of social support [7]. They also illustrate differences between anonymous redditors and non-anonymous to recognize proportions of comments with emotional, informational, instrumental, and prescriptive.

As we could not use a self-diagnosed dataset, we observed previous works on self-diagnosed dataset. Sekulic et al. 2019 illustrated analyzing the SMHD dataset[9] for two words extracted with the highest attention weight as being the most relevant for the classification [4]. Analysis on the dataset is about the users who were identified by constructing patterns for discovering self-reported diagnoses of nine different mental disorders [4]. Analyzing F1 scores of different numbers of posts per user, using Logistic regression and linear SVM achieved higher scores where there is a smaller number of diagnosed users [4]. Additionally, they observed similar patterns in features shown relevant by the HAN and previous research on signals of depression in language [4]. Their conclusion on research is that attention weights on word level suggested similarities to previous studies of depressed authors.

In this study, we will focus on mental health-related subreddits with binary classification of pre- and post-COVID data to exhibit shifts of post trends and engagements through illustrating natural language processing by BERT and DistilBERT, topic modeling by BERTopic, and time series analysis by Prophet.

* 1. *Ethical Considerations*

As utilizing mental health-related data might raise ethical concerns, it is important to clarify that our research does not aim to measure the mental health status of users or use any personal information they might display on the forum. As social media platforms are often publicly broadcast media, much has been written about the perception that users do not necessarily treat social media as a purely public space. [1] As doing so, we will instead analyze how the posts themselves have changed over time. Furthermore, publicly available data may include sensitive data that requires protection [1]. By removing all user information from our dataset, we ensure that we have less indicating individual’s information from our analysis. Moreover, we will also make sure that our use of Reddit data is in compliance with platform policies and guidelines.

# Data

We will proceed the study by data collection and data preprocessing for text classification and time series analysis.

* 1. *Data and Preprocessing*

Our dataset is extracted directly from the official reddit API using the python library PRAW. Upon comparing pre and post COVID, dataset is collected between 2018 to 2022 based on WHO confirmed the start date of COVID, we decided February 2020 would be the divider of the 2 periods: Pre-COVID consists data from January 2018 to January 2020, and Post-COVID consists data from February 2020 to December 2022.

Our study will focus on Reddit posts from 9 specific mental health-related subreddits: depression, ADHD, anxiety, bipolar, PTSD, schizophrenia, eating disorders, brain fog, and suicide watch. The datasets from these subreddits will include title text, post texts, score, upvote ratio, number of total comments, and the post’s creation time. We will analyze these features to observe changes in both post trends and engagement before and after the COVID-19 pandemic. Through text classification and time series analysis, we will transform these features into valuable variables for our analysis.

Our descriptive analysis contains correlation among variables, and log transformation of variables.

Figure 1 and 2 show pre and post COVID correlation matrix using log transformation.

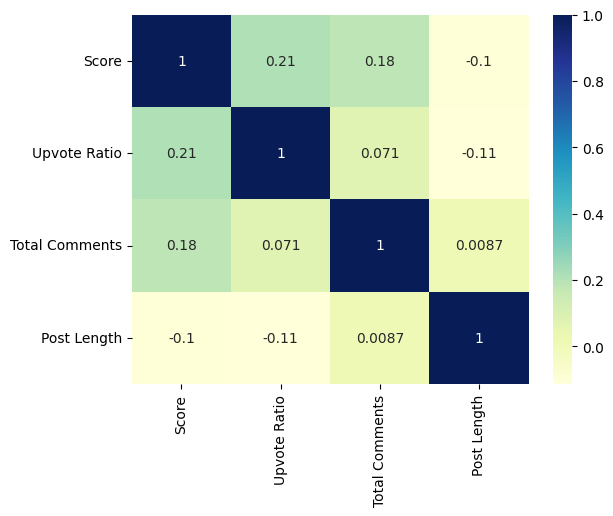


Fig. 1 Pre-COVID correlation matrix

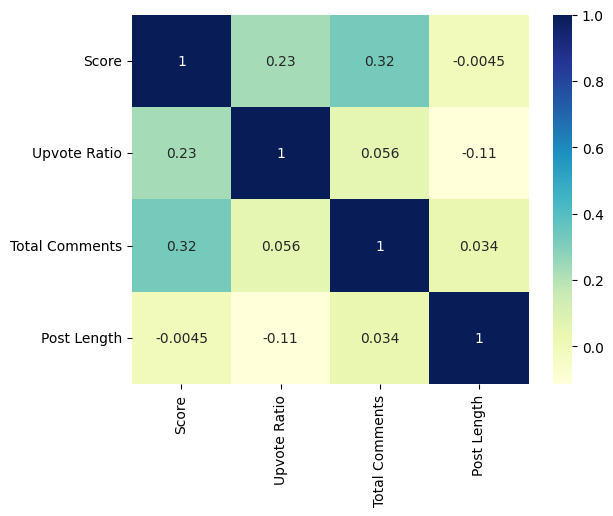


Fig. 2 Post-COVID correlation matrix

Based on the correlation matrices, we can see that there is a stronger correlation between score with upvote ratio and total comment Post-COVID than during pre-COVID. However, both periods show a proper positive correlation so it’s safe to assume that they can be helpful in determining our dataset.

# methodologies and models

## Text Classification using Large Language Model BERT

In this research, we utilized the BERT pretrained model from PyTorch Huggingface for BERT and DistillBERT to proceed text classification on pre and post COVID. BERT (Bidirectional Encoder Representations from Transformers) has become one of the most popular models to solve a wide range of natural language processing tasks with high accuracy. It is developed by Google AI Language which applies the bidirectional training of Transformer, in contrast to previous models that looked at text sequence either from left to right or combined left-to-right and right-to-left training.

In our BERT model, we employed binary classification with balanced samples to compare pre and post-COVID post texts. For instance, imbalance datasets can severely affect the accuracy of class predictions, and thus they need to be handled by appropriate data processing before analyzing the data [11]. Specifically, our classification utilizes a balanced sampling approach, drawing 50 percent of the data from pre-COVID and 50 percent from post-COVID, resulting in 2047 data points for each period. This approach helps mitigate bias that could arise from larger datasets dominating the results, allowing us to more accurately identify differences between the two datasets.

Furthermore, it is important to note that our binary classification model is not solely focused on achieving higher accuracy than binary classification is used in comparison to competing methods [10]. While higher accuracy is desirable, our primary objective is to compare pre and post-COVID post texts. To this end, we also employ topic modeling to identify which topics are prevalent within each dataset.

Steps of BERT model:

1. Concat nine subreddit datasets into two pre and post-COVID dataset
2. Text cleaning by removing nulls, punctuations, emojis, replace \n, and proceed to lowercase
3. Label pre as 0 and post as 1 to proceed binary classification
4. Before concat into one dataset, downsampling post-COVID dataset to balanced two labels - 4094 data points
5. Save it to pickle
6. Define X as post text and y as label from our saved preprocessed dataset
7. Split train and test with test size of 0.2, and random state of 42
8. Import Pre-trained model from PyTorch
9. With max length of 128, define X train tokens, X test tokens, y train tensor, and y est tensor
10. Import Bert for Sequence Classification
11. Training loop with Adam optimizer (2e-5), epochs, and batch size
12. Model evaluation
13. Repeat for better accuracy

## Text Classification using DistillBERT

Further, we continued on the performance of a DistilBERT model with a higher batch size, and used BERTopic to identify the actual topics present in the pre and post COVID periods.

Steps of DistilBERT model:

1. Import model saved into pickle
2. Train and test split with X as post text y as label
3. Import tokenizer for distilbert
4. Class function for X and y train and test
5. Using data loader to split train and test loader
6. Import config and pre-trained with dropout of 0.3 and attention dropout of 0.3
7. Classification neural network function:
   1. dropout of 0.3
   2. linear with (768, 64)
   3. Relu
   4. Linear for output
8. Get number of trainable parameters for model
9. Build a model with epochs of 5 and batch size of 90

## Text Classification using BERTopic

Our last BERT model is topic modeling using BERTopic. Using this technique, we employed how popular topics change over the pandemic. We again used PyTorch embedding to analyze the topics from pre and post-COVID. Our dataset for topic modeling is not preproceed with downsampled datasets. We used the pre and post-COVID nine subreddits

Steps to BERTopic:

1. Preprocess the concat pre and post-COVID datasets
2. Install all necessary libraries
3. Import BERTopic embedding model
4. Visualization of top 8 topics of pre and post-COVID
5. Visualization of Intertopic Distance Map

By topic modeling ,we will illustrate the popular topics that are discussed in two time frames and recognize how they are different or similar.

## Time Series Analysis using Prophet

We utilized the built-in Prophet library in Python, which is a time series forecasting library developed by Facebook (now Meta)'s Core Data Science team. It was designed to forecast time series data, particularly those with multiple seasonality patterns and non-linear trends. Since we used multiple variables to determine the change of each feature–i.e., Score, Upvote Ratio and Total Comments for Post Engagement, and Post Length and Title Length for Post Trend, we decided to utilize the multivariate Prophet time series in order to incorporate different variables into one time series. The model returns a prediction of the forecast variables for the next 2 years, but we will only be focusing on the changes pre and post-COVID.

Steps to Prophet Time Series Analysis:

1. Import the necessary libraries
2. Using our combined, log transformed data, create an instance of the Prophet class and call its fit method
3. Split our data to variables that affect Post Text and ones that affect Post Engagement to train the model on our current data
4. Set the prediction time frame to 2 years (~730 days) to generate prediction for future dates
5. Visualize the results using time series combined with scatterplot

By this time series analysis, we will conduct the analysis on quantitative features of post trends and engagement.

# results

## Results of Text Classifications

**BERT model**

Table 1

Accuracy and model Evaluation

| **Number of Epochs/**  **Batch size** | **Accuracy** | **F1-score** | **Precision** | **Recall** |
| --- | --- | --- | --- | --- |
| **3/16** | 0.58 | 0.52 | 0.67 | 0.43 |
| **6/16** | 0.56 | 0.55 | 0.57 | 0.52 |
| **3/15** | 0.60 | 0.59 | 0.65 | 0.54 |
| **3/8** | 0.61 | 0.61 | 0.66 | 0.55 |

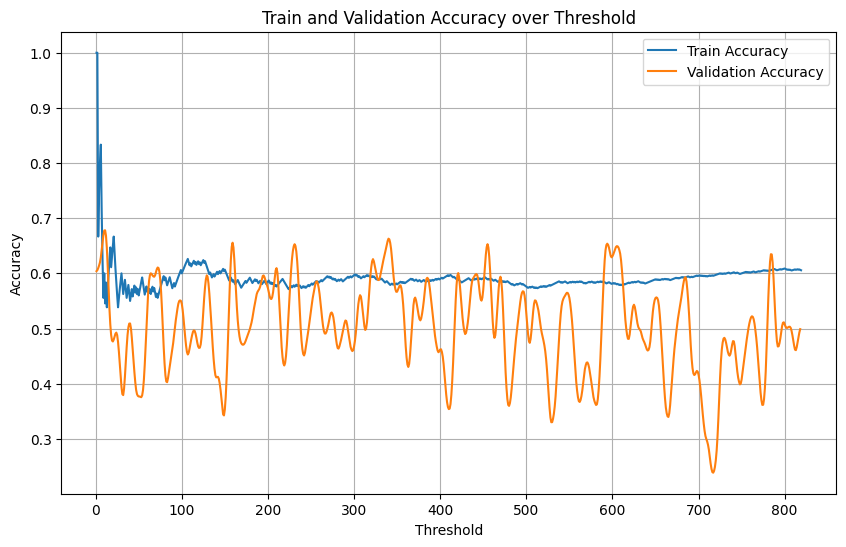


Fig. 3 Train and Validation Accuracy by threshold

Considering that we undersampled our imbalanced dataset to mitigate bias lead to higher accuracy from post-COVID data, we acknowledge that achieving high accuracy is not crucially significant for our results. Instead, our focus lies in accessing how much the context changes and whether the BERT model can detect these changes.

With accuracy of 61 percent, the model predicts pre-COVID with a precision of 0.66 and an F1-score of 0.61, indicating a moderate understanding of the difference between pre and post-COVID contexts. Although there aren’t significant differences in the contexts between the two datasets, we can still observe slight variations from pre to post-COVID.

Given that we only considered mental health-related subreddits, even though we acknowledge potential changes in post-COVID contexts, it is challenging to detect significant differences between pre and post-COVID contexts. However, considering our balanced accuracy assumption (half from pre and half from post COVID), our result of 61 percent accuracy of the BERT model is acceptable for our conclusion.

In case, here is an explanation of 50 percent accuracy is good for binary classification.

**Balanced Accuracy methodology:**

Ideal Accuracy from our BERT model is 50 percent which explains from below example:

Among 1000 data points, we have 500 pre-COVID and 500 post-COVID and we want to clarify if their contexts are different at all.

Table 2

Confusion Matrix

|  | | Actual values | |
| --- | --- | --- | --- |
| Positive 1 | Negative 0 |
| Predicted Values | Positive 1 | TP | FP  (Type 1 error) |
| Negative 0 | FN  (Type 2 error) | TN |

Using this confusion matrix, we will use the null hypothesis that 500 are pre-COVID and 500 are post-COVID.

* True positive = 500 labeled as pre and true prediction
* False positive = 500 is post but predicted as pre
* True negatives = 500 is post and predict as post
* False negatives = 500 is pre but predict as post

**Accuracy** = True Positive / True Positive + False Positive

**Precision** = True Positive + True Negative / Total

**Recall** = True Positive / True Positive + False Negative

For example, If 600 is pre and 400 is post, then Accuracy = 0 + 600 / (1000) = 0.6

Using this theoretical method, we claim that our 61 percent of accuracy is considerable:

Because our sample is balanced of 2047 data points from pre and 2047 data points from post

→ Balanced acc = 0.5 \* (0/2047 + 2047/2047) = 0.5

As a result, 50 percent of accuracy of the model will be ideal for our balanced sampling dataset.

**DistilBERT**

For the next step, we investigated DistilBERT which is 60 percent faster but has less layers than the original BERT model. For this model, we used 90 batch sizes calculated from total parameters.

Our result of DistilBERT found that it is overfitting which is not sufficient to use as our final result. Therefore, we will use our original BERT model for our final result.

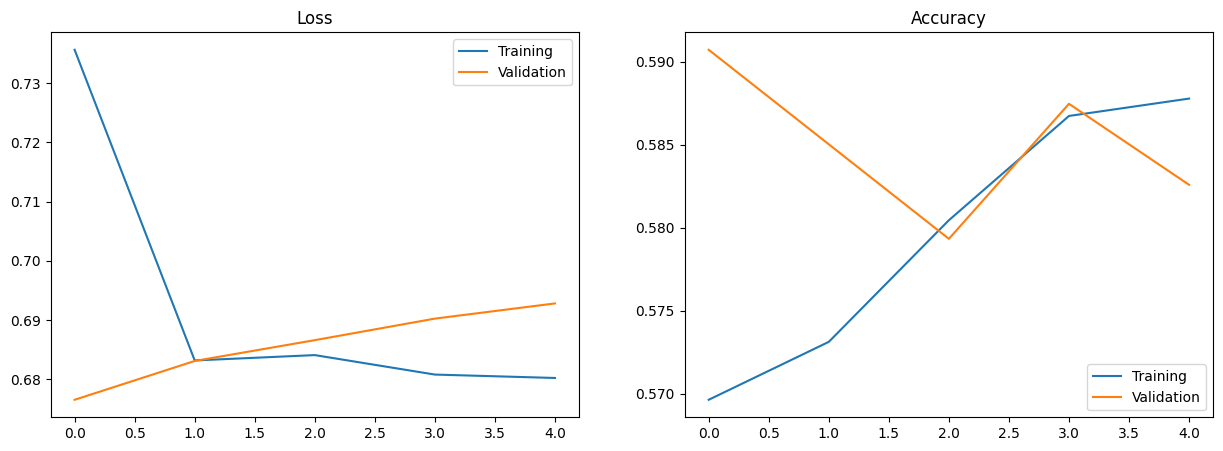


Fig. 4 Train and Validation Accuracy/Loss Graph by epochs

**BERTopic**

Our final model is BERTopic, a topic modeling technique that combines transformers and c-TF-IDF to create dense clusters. As we compare post trends between pre and post COVID, topic modeling is necessary for understanding the contexts of posts which is crucial for tracking how topics evolve during the pandemic. Throughout this model, we can identify whether post topics have changed, the extent of these changes, and the common themes within these topics. Continuing with BERT models, we opted for BERTopic over LDA to stay within the transformer model framework instead of transitioning to other models. Additionally, we chose to analyze title text instead of post text, as it yields fewer topics. We believe that post text contains too much ambiguity, while title text succinctly captures the essence of the posts themselves.

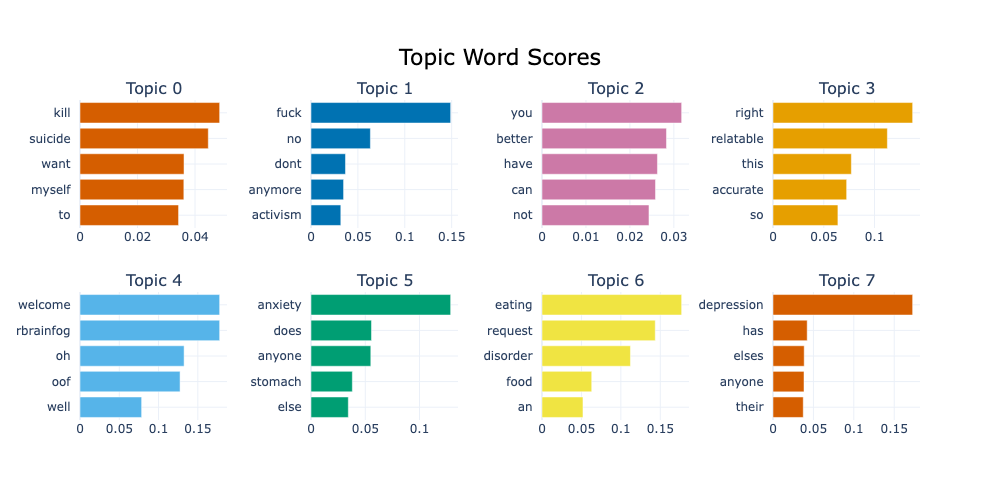


Fig. 5 Pre-COVID topic word scores



Fig. 6 Post-COVID topic word scores

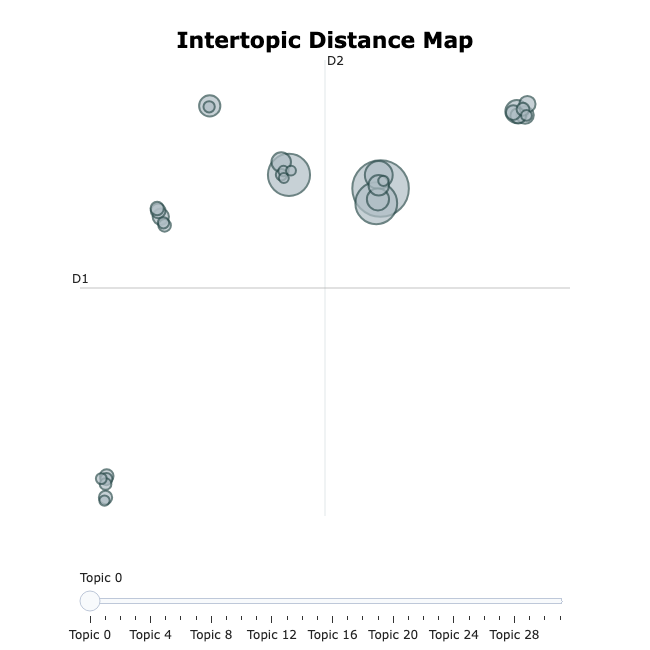


Fig. 7 Pre-COVID intertopic distance map

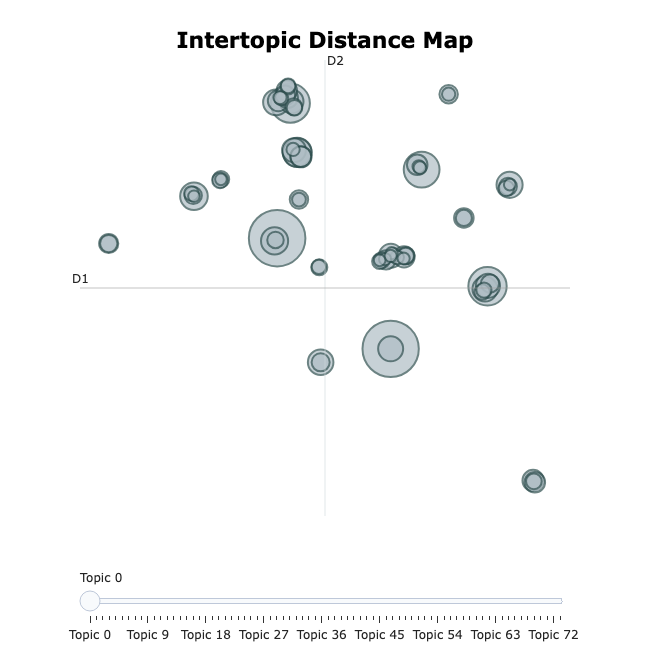


Fig. 8 Post-COVID intertopic distance map

The visualizations of topics indicated that post-COVID encompasses a broader range of topics compared to pre-COVD period. These post-COVID topics exhibit less overlap bubbles and are more widely scattered across the intertopic distance maps than pre-COVID. Given that the data is derived from mental health-related subreddits, we initially anticipated that the insights might be challenging to discern. However, the BERTopic visualizations clearly demonstrate that post-COVID period is characterized by a greater diversity of topics, including those with academic relations that are reflected by the pandemic. Despite some overlapping bubbles, these insights strongly suggest a noticeable difference in the topics discussed before and after the pandemic.

## Results of Time Series Analysis

Since we have multiple variables that impact one feature (Post Text or Post Engagement), we utilized the Prophet time series in order to incorporate different variables into one time series. The Prophet time series also helped us visualize and identify changes and patterns in our combined dataset. First, we conducted the time series analysis on Post Engagement, which is determined by Score, Total Comments and Upvote Ratio. Here we used Score as the forecast variable and Total Comments and Upvote Ratio as regressors.

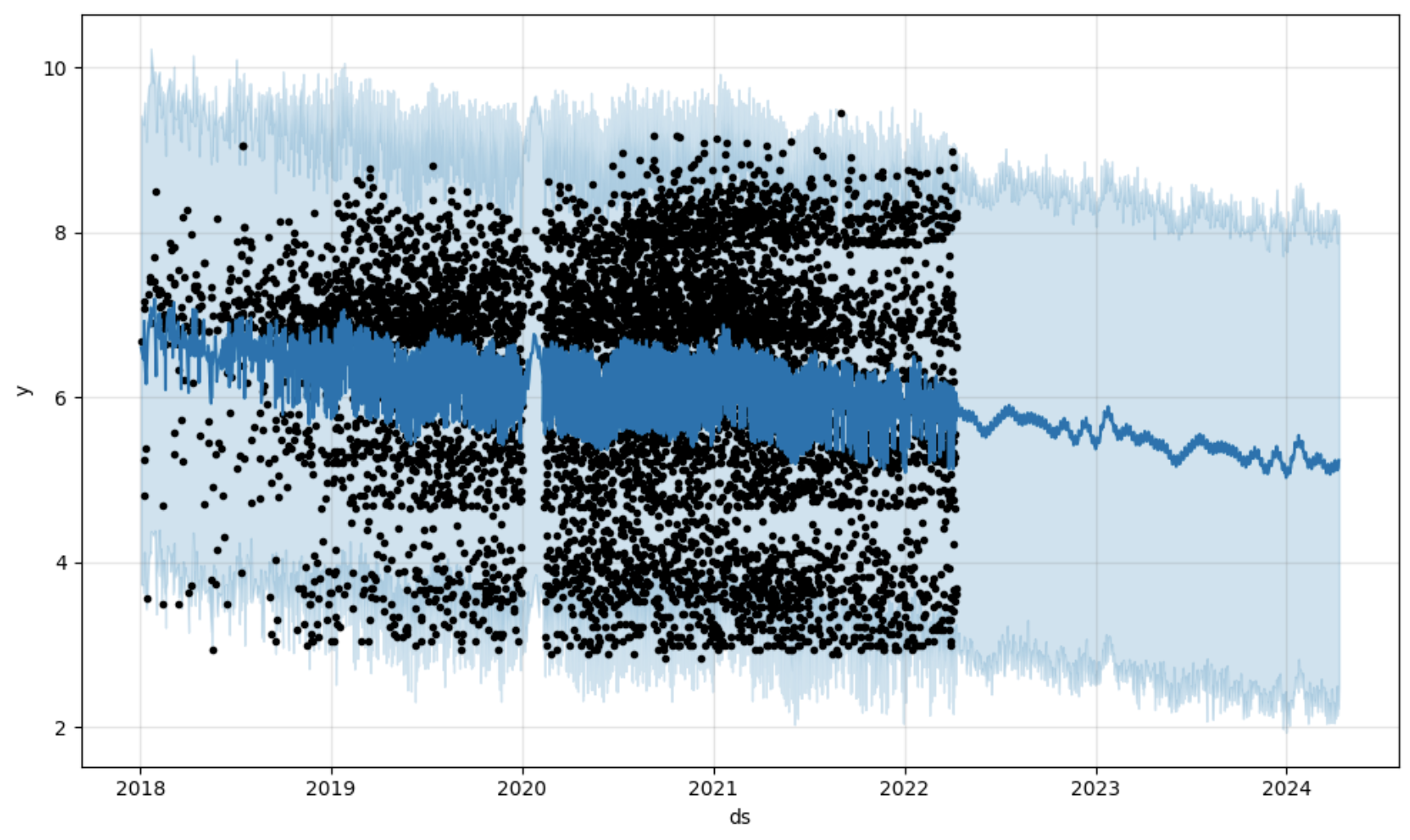


Fig. 9 Post Engagement Time Series

For Post Trend, we used Post Length as the forecast variable and Title Length as the regressor.

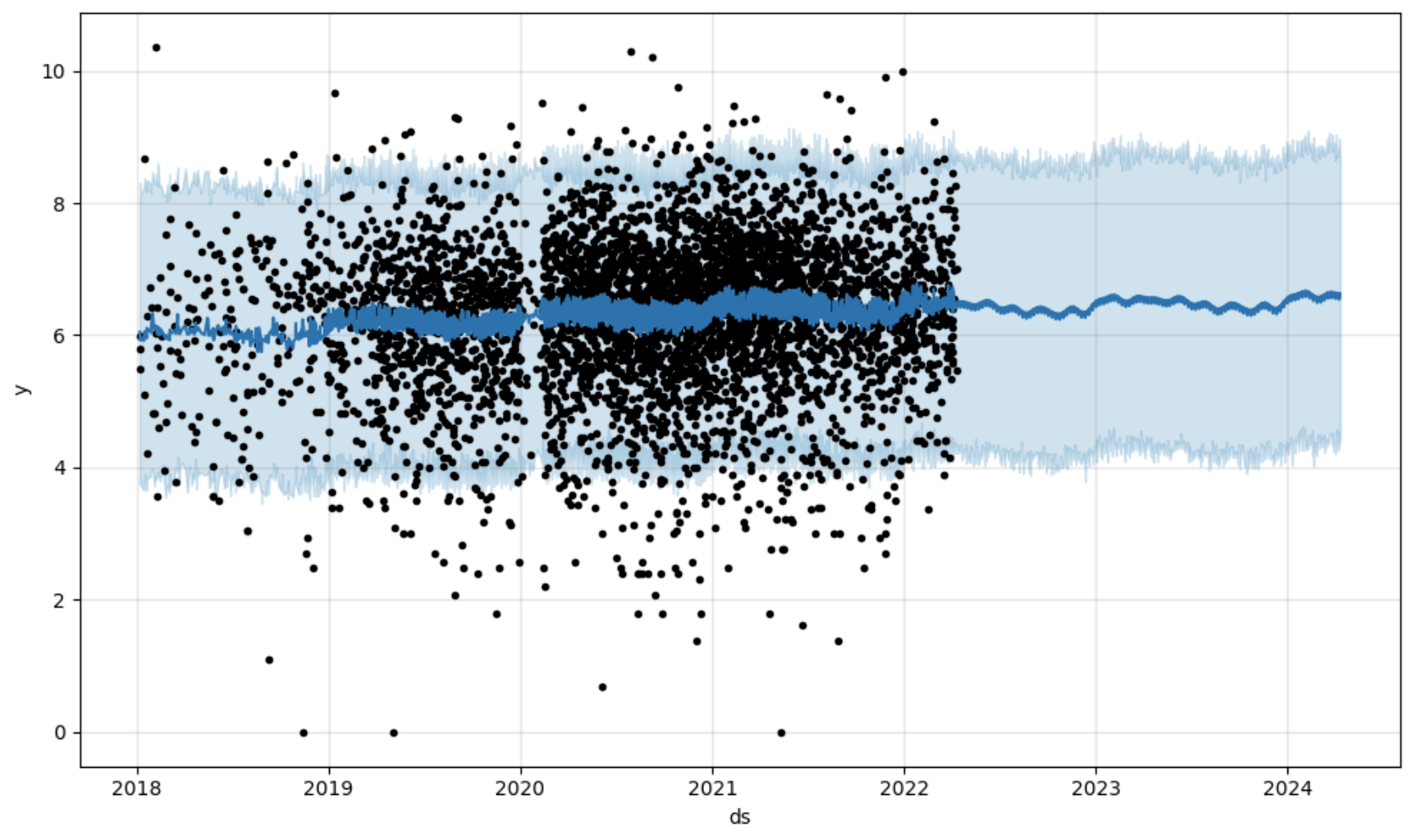


Fig. 10 Post Engagement Time Series

According to our time series analysis, there are some changes between pre and post-COVID both in post engagement and post trend. For post engagement, The pandemic makes it much more robust as more posts show up, but over time (within the 2 years period we predicted) it will slowly decrease to its previous level, if not lower. Post trend also has a dramatic change, which is the increased length of both the actual posts and the titles.

**Result**

Post Engagement: The MAE is 1.0; the MSE is 2.63.

Post Trend: The MAE is 0.85; the MSE is 1.3.

The MAE is relatively high for our model, but since MAE is robust to outliers since it only considers the absolute differences between the predicted and actual values, this could be due to the fact that our data is too spread-out, creating large differences between data. The same thing goes for MSE.

# conclusions and future work

In this work, we have investigated the shifts of post trends and post engagements between pre and post COVID. Our analysis on datasets indicates that the post-COVID period has a higher number of post engagements and an increase in post trend throughout the 9 mental health-related subreddits.

Through exploratory data analysis, we examined the temporal patterns, seasonality, and trends present in our Reddit dataset. This initial analysis allowed us to identify important features and characteristics that informed subsequent modeling decisions, which led to our use of the Prophet time series analysis model. This model shows that during post-COVID period, users tend to post more and with longer posts as well, which agrees with our findings from our text analysis. It also predicts that this trend will slow down and stabilize after a year or two.

Employing text classification with BERT models, we conclude that post texts from pre-COVID and post-COVID exhibit a shift of topics. The accuracy of ~61 percent allows us to conclude that there are differences between post trends between two time frames. Moreover, our BERTopic topic modeling indicates that people tend to discuss a diverse range of topics after the COVID while less overlapping similar topics.

There are some limitations on our analysis that need a larger dataset, and balanced samples of two datasets. We plan to further explore this topic with a bigger set of data within more subreddits included to stabilize and increase the accuracy and reliability of our models.

##### Acknowledgment

We would like to express our sincere gratitude to Professor Maksim and Professor Brownataski for their invaluable guidance and support throughout the preparation of this research. Their insightful advice and constructive feedback have been immensely valuable to us, and we are truly grateful for their assistance.

##### References

1. A. Benton, G. Coppersmith, and M. Dredze, “Ethical research protocols for Social Media Health Research,” *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, 2017. doi:10.18653/v1/w17-1612
2. B. V. Barde and A. M. Bainwad, “An overview of topic modeling methods and Tools,” *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, Jun. 2017. doi:10.1109/iccons.2017.8250563
3. D. M. Low1 et al., “Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during COVID-19: Observational study,” Journal of Medical Internet Research, https://www.jmir.org/2020/10/e22635/ (accessed May 1, 2024).
4. I. Sekulic and M. Strube, “Adapting deep learning methods for mental health prediction on social media,” *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*, 2019. doi:10.18653/v1/d19-5542
5. K. Zeberga et al., “[retracted] a novel text mining approach for mental health prediction using Bi-LSTM and Bert Model,” Computational Intelligence and Neuroscience, https://www.hindawi.com/journals/cin/2022/7893775/ (accessed May 1, 2024).
6. K. Reagan, A. S. Varde, and L. Xie, “Evolving perceptions of mental health on social media and their medical impacts,” *2022 IEEE International Conference on Big Data (Big Data)*, Dec. 2022. doi:10.1109/bigdata55660.2022.10021013
7. M. De Choudhury and S. De, “Mental health discourse on reddit: Self-disclosure, social support, and anonymity,” *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, no. 1, pp. 71–80, May 2014. doi:10.1609/icwsm.v8i1.14526
8. P. Resnik *et al.*, “Beyond lda: Exploring supervised topic modeling for depression-related language in Twitter,” *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 2015. doi:10.3115/v1/w15-1212
9. “Self-reported mental health diagnoses (SMHD) dataset,” ir@Georgetown - Resources - SMHD, https://ir.cs.georgetown.edu/resources/smhd.html (accessed May 1, 2024).
10. T. Saito and M. Rehmsmeier, “The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets,” PLOS ONE, https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0118432 (accessed May 1, 2024).
11. Wonjae Lee a, a, b, and A. many industrial applications, “Downsampling for binary classification with a highly imbalanced dataset using active learning,” Big Data Research, https://www.sciencedirect.com/science/article/pii/S2214579622000089 (accessed May 1, 2024).

**\**