

# Mining Mental Health Patterns in User Tweets and Resource Allocation via Sentiment Analysis, Topic Modeling and Transfer Learning

IST 736  
TEXT MINING

Researchers:  
Aditi Pala  
Yashaswini Kulkarni  
Viha Mashruwala

## INTRODUCTION:

In the era of digital communication, Twitter serves as a rich source of unfiltered user-generated content. Our text mining project, rooted in the Sentiment140 dataset comprising 1.6 million tweets, initially navigated the unstructured terrain of tweets without specific domains. As researchers, our objective was to analyze and understand the general sentiments prevalent in this vast dataset.

Driven by a commitment to contribute to mental health awareness, our focus shifted to a targeted investigation within the Mental Health dataset. This subset became the subject of our analysis, utilizing topic modeling techniques to identify prevalent themes and subjects within the discourse surrounding mental health on Twitter.

Our primary aim, beyond academic exploration, was to apply sentiment analysis models to discern emotional tones within specific mental health topics. This approach aimed to provide actionable insights, ranging from facilitating awareness campaigns to connecting individuals facing similar challenges. The intended impact transcended the analytical realm, seeking to bridge the digital gap for those seeking support or information related to mental health concerns.

This report chronicles our methodological journey, emphasizing the practical implications of sentiment analysis within the context of mental health discussions on Twitter. The goal is to present a comprehensive understanding of our findings and contribute valuable insights to the ongoing discourse surrounding mental health awareness in the digital age.

## OBJECTIVE:

Transitioning from generic tweets, our data journey led to mental health insights. The detection of users expressing negativity guides targeted support initiatives, offering a pragmatic approach to bolster mental well-being.

- Evaluate the cross-domain performance of sentiment analysis models by testing their efficacy on mental health-related tweets using keywords derived from topic modelling.
- Enhance online well-being by proactively extending mental health resources to users expressing negative sentiments.

## DATASETS:

- **Mental Health Dataset:**

Link to dataset: <https://www.kaggle.com/datasets/infamouscoder/mental-health-social-media>

The data contains uncleaned Twitter data collected through the Twitter API. It focuses on user-level mental health classification, providing valuable insights into tweet content, user identification, followers, and retweet statistics.

**Index:** index key of post

**Post\_id of the post:** ID of the post

**Post created on:** Data Created

**Uncleaned Tweet:** Tweet text (raw data\_  
**User Identification:** user ID  
**No of Followers:** Number of followers of the user  
**Total Retweets on the Current Tweet:** Number of retweets

- **Sentiment140 Dataset:**

Link: <https://www.kaggle.com/datasets/kazanov/sentiment140>

This dataset contains 1,600,000 tweets extracted using the twitter api. The tweets have been annotated (0 = negative, 4 = positive) and they can be used to detect sentiment. It provides insights into the identification of severity from tweets and is a useful tool for sentiment analysis.

**target:** the polarity of the tweet (0 = negative, 4 = positive)

**ids:** The ID of the tweet ( 2087)

**date:** the date of the tweet (Sat May 16 23:58:44 UTC 2009)

**flag:** The query (lyx). If there is no query, then this value is NO\_QUERY.

**user:** the user that tweeted

**text:** the text of the tweet

## DATA PRE-PROCESSING:

### *Sentiment140 dataset:*

In preparing the Sentiment140 dataset for analysis, our initial step involved ensuring data balance, with an equal distribution of positive and negative sentiment labels. This equilibrium, consisting of 10,000 instances for each label, set the foundation for unbiased model training. To enhance consistency, we uniformly assigned a label of '1' to positive sentiment, streamlining the subsequent analysis. The subsequent data preprocessing pipeline addressed common challenges in text data, encompassing lowercase conversion, removal of stopwords, punctuations, repeating characters, and extraneous elements such as emails, URLs, and numeric figures. The ultimate goal was to distill the raw tweets into a refined corpus, ready for further exploration and sentiment analysis.

### *Mental health dataset:*

In strategically sub setting the mental health dataset to focus on the top 5 users with the highest tweet frequencies, our objective was to pinpoint posts that resonate widely within the digital community, offering a concentrated view of prevalent mental health discussions. This subset, comprising 7,378 tweets, became the canvas for meticulous data preprocessing. The process involved eliminating extraneous elements such as links, usernames, and special characters, while standardizing the text to lowercase for uniform analysis. By curating this refined subset, we aimed to distill high-impact content, laying the groundwork for nuanced sentiment analysis within the realm of mental health discourse.

## TOPIC MODELLING:

In our pursuit of a nuanced understanding of the mental health discourse embedded within the mental health tweet dataset, we employed diverse topic modeling techniques, including Latent Dirichlet Allocation (LDA), BERTopic, and SBERT.

**Focus on finding dominant mental health topics:** Specifically, our focus was directed towards conducting an in-depth analysis on the mental health data derived from the tweets of the top 5 users. This was done to

unveil prevalent topics of interest and discern the most frequently tweeted subjects within the realm of mental health and provide valuable insights into the dominant themes and prevalent topics of discussion pertaining to mental health.

**Models:** Initially, we used Latent Dirichlet Allocation (LDA), a straightforward topic modeling model, to reveal underlying topics within the mental health data. Then we used the BERTopic library which facilitated the extraction of meaningful clusters and representative documents. Finally, through Sentence-BERT (SBERT), we understood more finer nuances of the data as it played a pivotal role in our investigative process.

**Use of ChatGPT:** Utilised ChatGPT to analyze each cluster output, by applying prompts like "filter out clusters that may not be related or similar and focus on mental health-related context.". Through the results generated, we also asked ChatGPT to then extracted specific keywords related to mental health, illness, and related topics from the filtered cluster list. Finally 3 keyword lists were generated – each for a topic modeling model used.

Considering that the BERTopic model produced a significant number of clusters—roughly 160—this method proved especially helpful.

```
lda_keywords = {
    1: ['depression', 'symptoms'],
    3: ['cry', 'sad', 'fight', 'love'],
    4: ['depression', 'treatments'],
    5: ['sad', 'kids', 'fight', 'love', 'disease', 'like', 'campaign', 'little', 'poor',
'person', 'stories', 'artists', 'darkness', 'heartbreaking', 'upgrade', 'understanding', 'cold',
'life', 'sadly', 'insane', 'spelling', 'begin'],
    6: ['happy', 'wish', 'birthday', 'wonderful', 'intense'],
    7: ['family', 'hell', 'sweet', 'goodnight', 'hope', 'ill'],
    8: ['depression', 'helped', 'sucked', 'overcoming', 'overcome', 'stigma', 'thoughts',
'manageable', 'hair', 'dead', 'death', 'ready', 'mute', 'mean', 'color', 'tiny', 'moron'],
    9: ['suicide', 'blues', 'deplorable', 'cure', 'mad', 'rest', 'mess', 'methods', 'absolutely',
'game', 'thugs', 'benefit', 'toy', 'committed', 'swift', 'toys', 'darker', 'boys', 'wit',
'moron', 'ended', 'makeup', 'lover', 'help', 'mocks', 'people', 'heavy'],
}
```

```
bertopic_keywords = {
    2: ['depression', 'depression treatments'],
    7: ['broken people', 'unrepairable', 'fix', 'broken'],
    15: ['talk business adviceconcerningyourfightwithdepression', 'talk business
adviceandadviceoncopingwithdepression', 'talk business
wouldyoufeeldepressedreadthisinformativearticle'], 30: ['leave depression behind now with these
hints article teller', 'easy things you have to know about depression article teller', 'how to
live a depression free life article teller'], 50: ['circles end depression therapy who had a
great winter mental health', 'circles end depression therapy feeling depressed animalassisted
therapy could', 'circles end depression therapy'], 69: ['overcome depressive disorders beating
the wintertime blues seasonal affective disorder', 'overcome depression winter season blues ways
to enhance mood plus energy', 'winter blues lack of sun often leads to depression my champlain
valley fox amp abc overcome depression'], 91: ['simple relaxation techniques to help reduce
stress', 'relaxation techniques to help reduce stress', 'ways to stress less at work'],
    98: ['apps can help relieve stress anxiety sydney morning herald depression treatments',
'overcome depression sleep therapy to treat depression sydney morning herald', 'apps can help
relieve stress anxiety sydney morning herald depression treatments'], 103: ['sing loud changes on
swift products for foods for depression', 'sing loud the hard battle howto fight depression',
'sing loud the best methods to overcome depression and be happy']
}
```

```
sbert_keywords = {
    1: ['trust', 'uglier', 'bothers', 'pretty', 'girls', 'nude', 'photoshoots'],
}
```

```

3: ['lonely', 'hollow', 'purpose', 'grateful'],
5: ['money', 'buy', 'design', 'save', 'cash', 'order', 'electronics', 'changing', 'colours',
'confused', 'emailed', 'reply', 'checking', 'designs'],
6: ['lightsabers', 'lightsaber', 'look', 'designs', 'start', 'small', 'buy', 'cheaper',
'save', 'money'],
9: ['obsessed', 'stalking', 'broke', 'house', 'bam', 'crush', 'serial', 'killer', 'thank',
'god', 'came', 'life'],
}

```

**Master Keyword list:** Next, we combined all three lists to produce a master keyword list. An important step was done to make sure the list contained only unique words and no duplicates. Nevertheless, the prevalence of stop words and non-contextual words remained even after this consolidation. Hence, as a solution, we carefully went over the master keyword list manually and eliminated any terms that were non-contextual or stopwords.

```

master_keywords =
['cure', 'overcome', 'stigma', 'methods', 'heartbreaking', 'depression', 'love', 'suicide',
'artists', 'intense', 'disease', 'ended', 'changing', 'poor', 'lover', 'ready', 'benefit',
'goodnight', 'absolutely', 'toys', 'happy', 'trust', 'mess', 'purpose', 'understanding', 'girls',
'pretty', 'manageable', 'hair', 'overcoming', 'cold', 'sweet', 'broke', 'thoughts', 'start',
'family', 'mad', 'ill', 'confused', 'heavy', 'boys', 'kids', 'lonely', 'campaign', 'treatments',
'fight', 'little', 'upgrade', 'life', 'mute', 'wish', 'wonderful', 'sad', 'begin', 'helped',
'thugs', 'insane', 'symptoms', 'grateful', 'dead', 'darker', 'people', 'money', 'deplorable',
'bothers', 'obsessed', 'death', 'darkness', 'game', 'birthday', 'blues', 'help', 'rest', 'sadly',
'committed', 'lightsaber', 'moron', 'sucked', 'small', 'cry', 'toy', 'color', 'nude', 'serial',
'stalking']

```

## Strategic Application of Topic Modeling Results To Enhance Project Focus:

Result of topic modeling aided to refine the mental health dataset after acquiring the master keyword list. Tweets with at least one word from the master list were kept after they were filtered, while those without such keywords were systematically deleted. By giving priority to tweets that showed a stronger correlation with mental health, this deliberate filtering attempted to reduce the size of the dataset and increase the focus on pertinent and popular discourse related to mental health.

The mental health dataset had an initial form of (20000, 11). After using topic modelling techniques, the dataset was refined to only include relevant tweets, and it is now shaped as (7366, 11).

## MODELS

### Multinomial Naive Bayes (MNB):

#### *Reason for Usage:*

Multinomial Naive Bayes (MNB) was deliberately chosen as a baseline model for its simplicity and efficiency, serving as a foundational benchmark in our sentiment analysis project. Particularly adept at handling text data, MNB provided a pragmatic starting point for understanding sentiment patterns in the context of Twitter discourse.

#### *Model and Usage:*

In our project, MNB served as a foundational model trained on the Sentiment140 dataset, utilizing the TfidfVectorizer for feature extraction. The intentional introduction of noise into the training labels added a layer of robustness, aligning with the dynamic and often noisy nature of social media content. Subsequently, the pre-trained MNB model was applied to a subset of the mental health dataset. This subset, derived from the tweets of the top 5 users, underwent additional refinement. We incorporated keywords obtained from

our topic modeling methods to create a master keyword set, enhancing the model's sensitivity to nuanced sentiments within the mental health discourse.

### ***Results:***

As anticipated for a baseline model, MNB yielded an accuracy of 50.16% on the mental health dataset, providing a fundamental measure of sentiment classification performance. The low accuracy value underscored the inherent challenges in distinguishing sentiments effectively. The confusion matrix analysis revealed instances of misclassification across diverse content, reinforcing the baseline nature of MNB and emphasizing the importance of more sophisticated models for nuanced sentiment analysis in complex data domains.

### ***3-Fold Evaluation of the MNB Model:***

Accuracy: 0.5016291067064893					
AUC Score: 0.49875910591101197					
Classification Report:					
	precision	recall	f1-score	support	
0	0.43	0.48	0.45	3197	
1	0.56	0.52	0.54	4169	
accuracy			0.50	7366	
macro avg	0.50	0.50	0.50	7366	
weighted avg	0.51	0.50	0.50	7366	

## **BERT (Bidirectional Encoder Representations from Transformers):**

### ***Reason for Usage:***

BERT, selected for its contextualized word embeddings and advanced natural language understanding, emerged as a strategic choice for our sentiment analysis project. In the intricate landscape of mental health discussions on Twitter, where linguistic subtleties play a crucial role, BERT's bidirectional approach and deep contextual understanding promised a more nuanced sentiment classification compared to traditional models.

### ***Model and Usage:***

Implemented via the `bert_sklearn` library, the BERT classifier was configured with key hyperparameters tailored for our sentiment analysis task. Trained initially on the Sentiment140 dataset, with 18,000 samples processed over a single epoch, the model demonstrated robust performance during validation, showcasing its adaptability and effectiveness in capturing intricate linguistic nuances. This BERT model was seamlessly transitioned to the mental health dataset, setting the stage for advanced sentiment analysis. The model was run after fine tuning with different hyperparameters to enhance the performance.

### ***Results:***

The BERT model's commendable accuracy of 78% on the mental health dataset, outperforming the baseline MNB model, underscores its efficacy in capturing nuanced sentiment patterns. Precision values of 0.83 for negative sentiments and 0.75 for positive sentiments highlight BERT's ability to handle both sentiment categories adeptly. The confusion matrix, revealing misclassifications spanning diverse content types, emphasizes the complex nature of mental health discourse. BERT's demonstrated capacity to decode subtle

linguistic nuances signals a substantial leap in sentiment analysis within the intricate landscape of mental health discussions on Twitter.

### *3-Fold Evaluation of the BERT Model:*

Predicting: 100%	921/921 [01:04<00:00, 15.06it/s]			
Classification Report:				
	precision	recall	f1-score	support
Negative	0.83	0.61	0.70	3197
Positive	0.75	0.91	0.82	4169
accuracy			0.78	7366
macro avg	0.79	0.76	0.76	7366
weighted avg	0.79	0.78	0.77	7366

## **DistilBERT:**

### *Reason for Usage:*

DistilBERT, chosen for its impressive performance, emerged as the top-performing model in our sentiment analysis project. Its ability to retain the essence of BERT while being computationally more efficient made it an optimal choice for handling the intricacies of sentiment in mental health discussions on Twitter.

### *Model and Usage:*

Implemented using the `transformers` library, DistilBERT was fine-tuned on the Sentiment140 dataset and then seamlessly transitioned to the mental health dataset. Key hyperparameters were configured, including options such as `'distilbert-base-uncased'` for the tokenizer and model, `'max_length=128'`, and `'num_labels=2'`. The model was trained for two epochs on the sentiment data, leveraging the DistilBERT architecture to capture nuanced sentiment patterns.

### *Results:*

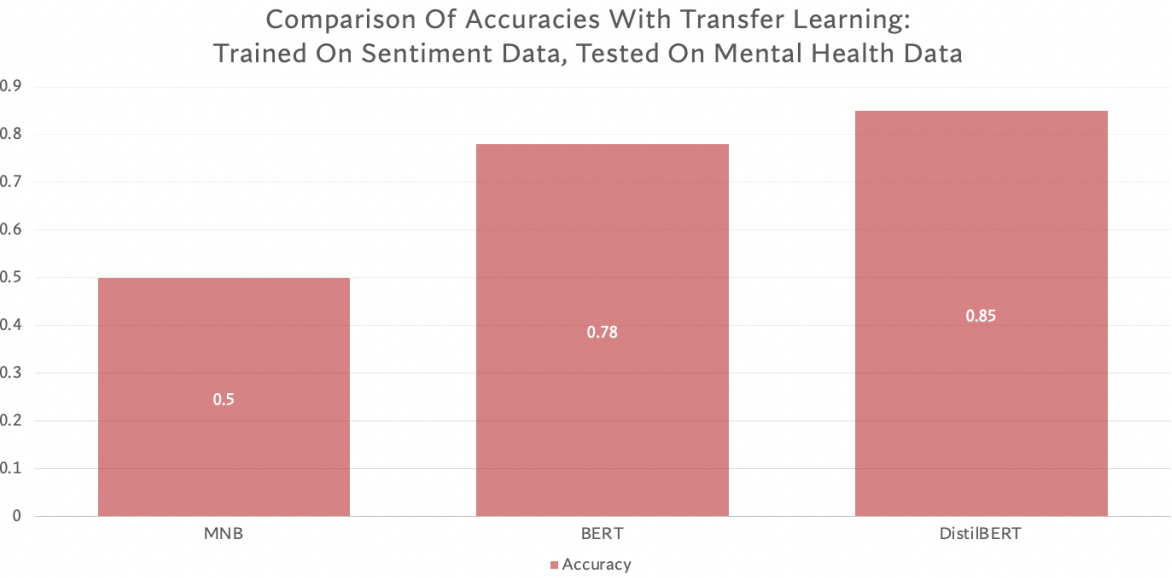
DistilBERT delivered outstanding results on the mental health dataset, achieving an accuracy of 85%. Precision values of 0.94 for negative sentiments and 0.80 for positive sentiments highlighted its ability to discern between sentiment categories. The nuanced insights provided by the confusion matrix illuminated the complex landscape of mental health discussions, reinforcing DistilBERT's position as a powerful tool for sentiment analysis in this challenging domain on Twitter.

### *3-Fold Evaluation of the BERT Model:*

Classification Report on Testing Data:				
	precision	recall	f1-score	support
0	0.94	0.69	0.79	3197
1	0.80	0.97	0.88	4169
accuracy			0.85	7366
macro avg	0.87	0.83	0.84	7366
weighted avg	0.86	0.85	0.84	7366

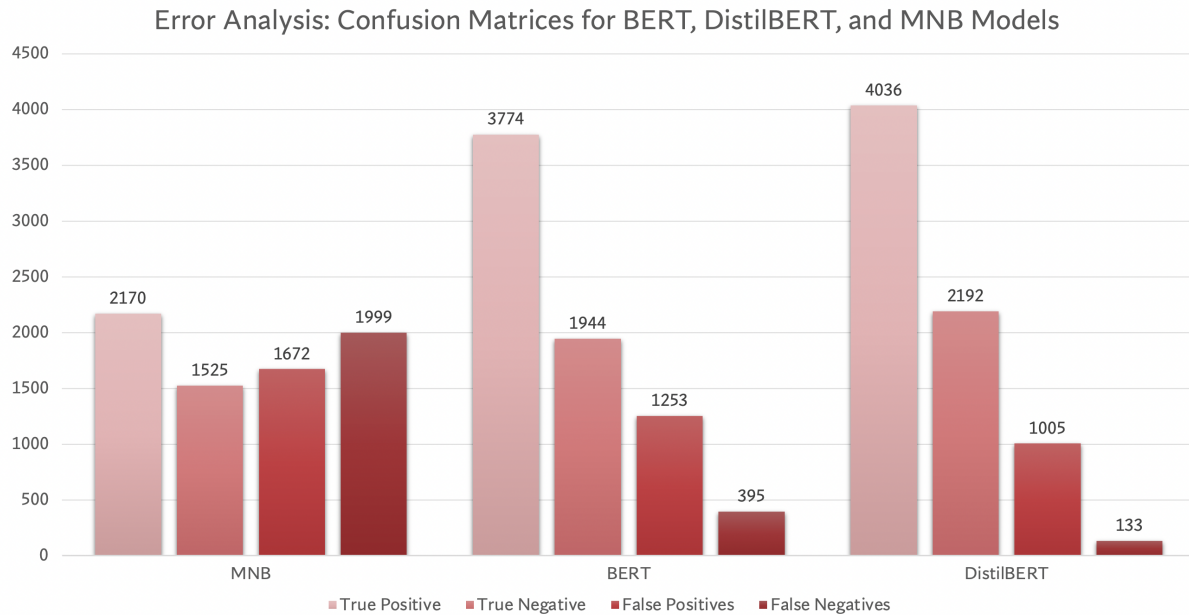
Performance Insights from Classification Models:

Transfer Learning in Sentiment Analysis for Mental Health Discourse Classification						
Model	Precision (0)	Recall (0)	Precision (1)	Recall (1)	Accuracy	Support
BERT	0.83	0.61	0.75	0.91	0.78	7366
Distilbert	0.94	0.69	0.8	0.97	0.85	7366
MNB	0.43	0.48	0.56	0.52	0.5	7366



Error Analysis on Sentiment Analysis Models for Mental Health Discourse Classification				
Model	True Positive	True Negative	False Positives	False Negatives
MNB	2170	1525	1672	1999
BERT	3774	1944	1253	395
DistilBERT	4036	2192	1005	133





In the domain of mental health-related tweet classification, DistilBERT demonstrates commendable performance. Its proficiency is evident in accurately discerning content unrelated to mental health and effectively identifying tweets associated with mental health, exhibiting superior accuracy for both positive and negative classes. Furthermore, DistilBERT showcases a notable recall for both classes, implying its capability to capture a substantial portion of authentic instances. The model's elevated overall accuracy, outperforming other models, underscores its effectiveness in generating precise predictions throughout the dataset. Notably, DistilBERT's balanced performance establishes it as a dependable tool for delineating mental health discourse within the Twitter dataset.

## RESULTS: Utilizing Sentiment Analysis to Identify Users In Distress For Mental Health Support

In our quest to enhance model performance, DistilBERT, distinguished by its highest overall accuracy and superior performance, took center stage. We strategically filtered instances where DistilBERT accurately predicted label 0, signifying negative sentiment.

This process allowed us to discern users through anonymized user\_id, counting the occurrences of negative tweets for each individual within this refined dataset.

Our user-centric analysis provides valuable insights into the distribution of negative sentiment across diverse users. This information serves as a foundation for delivering targeted mental health resources to specific individuals, ensuring a more personalized and effective approach to support.

Recipients of Mental Health Resources Support (Anonymized Users):

USER ID	NEGATIVE TWEET COUNT
490044008	507
145626605	430
3249600438	338

763182466098233344	161
18831261	154
1497350173	126
2780518314	117
1458225506	104
1169875706	94
762433972273950725	58
324294391	37
171999132	23
29053403	16
894149342	13
2369443141	8
548972753	4
706699293558710273	1
727820220291645442	1

## CONCLUSION:

Driven by the aspiration to foster awareness of mental health issues, our primary focus rested on discerning sentiments expressed in Twitter discussions surrounding mental health.

The strategic application of topic modeling techniques, including LDA, BERTopic, and SBERT, allowed us to unveil recurrent themes within the discourse on mental health. Sentiment analysis models, namely DistilBERT, BERT, and MNB, were deployed to unveil nuanced sentiment patterns. Among these, DistilBERT emerged as the most effective model, showcasing remarkable accuracy and well-balanced performance, enabling the capture of a diverse range of emotions.

Leveraging the insights derived from sentiment analysis, we identified users expressing unfavorable sentiments, facilitating the targeted provision of resources for mental health support.

This systematic approach culminated in the creation of an anonymized user list, serving as the cornerstone for focused support initiatives by providing valuable insights into users likely in need of mental health resources. The comprehensive amalgamation of sophisticated sentiment analysis methodologies with user-centric strategies contributes significantly to the ongoing discourse on mental health awareness in the digital era.

## ETHICAL CONSIDERATIONS:

Safeguarding user privacy is a top priority in our dedication to moral data practices. We have used all models on a selectively anonymized Twitter dataset in order to protect personal information during the tweet analysis process. We wanted to demonstrate our commitment to protecting user privacy which can be applicable to practical situations and future works.

**Bias Detection and Mitigation:** We take a proactive approach in our data preprocessing, and results after topic modeling to combat biases in order to guarantee justice and avoid discriminatory results. To find and fix any biases, in-depth analyses of the training & testing data and the sentiment analysis model are carried out. This methodical approach demonstrates our dedication to objectivity and fosters an analysis free from biased or unfair influences.

## **DECLARATION OF GENERATIVE AI TOOL USE:**

As part of our project "Mining Mental Health Patterns in User Tweets and Resource Allocation via Sentiment Analysis, Topic Modelling, and Transfer Learning," we used ChatGPT to enhance our analytical capabilities for Topic Modeling. The extraction of mental health related keywords from almost 180 clusters was made easier by ChatGPT, which extracted approximate keywords pertaining to mental health, mental illness, and related subjects from our cluster data.

### ***Advantages:***

By offering a sophisticated comprehension of mental health subjects, ChatGPT effectively extracted pertinent keywords from cluster list, improving the accuracy of our analysis. This aided us in taking an initial step towards refining our dataset to find dominant mental health related topics, and to find recurring themes.

### ***Cons:***

The results had to be carefully validated because there was an element of unpredictability brought about by the lack of precise control over the generated outputs. For ChatGPT to be effective, it was necessary to create clear prompts that needed to be adjusted step-by-step in order to achieve the desired results which required lots of explaining context for mental health and manual prompting based on our research and in-depth analysis on the mental health dataset.