**Topic Modelling on Mental Health-Related Tweets**

# **1. Introduction**

The digital era uses Twitter as a forceful resource that enables users to share their life stories and express viewpoints about various subjects, including mental wellness issues. These digital platforms make instant communication possible because they give users an open environment to share their mental health challenges, receive support, and educate others about psychological health conditions [1]. This project aims to use Natural Language Processing (NLP) and Machine Learning (ML) to evaluate mental health discussions on Twitter. Through Latent Dirichlet Allocation (LDA) and BERTopic, the study investigates predominant subject matters by monitoring public sentiment and tracking recurring discussion patterns [2]. The investigation process begins with data collection and preprocessing, followed by topic modelling techniques, and concludes with visualization methods to effectively analyse mental health discussions [3].

# **2. Data Collection and Preprocessing**

## **2.1 Dataset Selection**

The research dataset consists of tweets about mental health obtained from Kaggle [4]. The selection of this dataset with 20,000 tweets was made to achieve a complete and unbiased analysis. The metadata fields include:

* **Tweet ID:** It provides individual identification tags which make tweet tracking and reference possible.
* **Timestamp:** It includes the precise time stamp when every posting occurred thus enabling scientists to analyse temporal data patterns.
* **Post Text:** The basis of the topic modelling approach lies with tweet post text which represents the original written content of the posts.
* **User Information:** The analysis evaluates tweet influence through information about users which includes their IDs in addition to follower numbers and engagement metrics.
* **Engagement Metrics**: Interaction statistics, including retweet counts and likes, providing insights into the popularity of discussions.

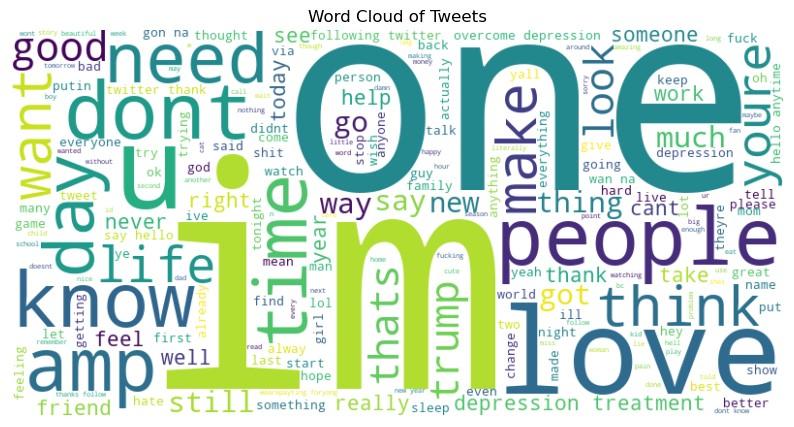


Figure 1: Word Cloud of the Dataset

## **2.2 Data Preprocessing Steps**

### **2.2.1 Text Cleaning:**

A complete text-cleaning pipeline standardized the dataset by implementing these steps-

* **Lowercasing:** Everything was converted to lowercase to keep things uniform and avoid duplicates.
* **Removing URLs & Mentions:** Links and @usernames were deleted since they don’t add meaning.
* **Hashtag & Special Character Removal:** Hashtags, punctuation, and symbols were stripped to focus on the core text.
* **Cleaning Up Language & Whitespace:** Non-English words were removed for consistency, and extra spaces were cleaned up.
* **Filtering Retweets:** Retweets were excluded to keep only original content and avoid duplication.

### **2.2.2 Tokenization and Stopword Removal:**

After the text was cleaned, it was broken down into tokens (words). This was achieved using NLTK’s word tokenizer, which effectively segments sentences into discrete words. To further refine the text:

* **Common stopwords** (e.g., "the," "and," "is") were removed to eliminate frequently occurring but semantically insignificant words.
* **Short and uninformative words** were filtered out to ensure that only meaningful words contribute to topic formation.

### **2.2.3 Lemmatization:**

The process of lemmatization using WordNetLemmatizer was used to treat different word variations as one unified entity. A base form conversion method transforms words into fundamental versions which creates a reduction of repetition. Examples include:

* "better" → "good"
* "thoughts" → "thought"

### **2.2.4 Vectorization:**

Topic modelling applications require numerical data. So raw texts need conversion through a numerical formatting process. By undergoing this conversion process the frequent common words achieve reduced importance and meaningful words become more prominent allowing better topic separation. Here-

* **For LDA**, TF-IDF was applied to create a weighted document-term matrix, allowing for better topic extraction and reducing the influence of frequently occurring words.
* **For BERTopic**, sentence embeddings were generated using **transformer-based models**, which provided contextual representations of text before clustering similar tweets into topics.

# **3. Topic Modelling Methods and Implementation**

## **3.1 Latent Dirichlet Allocation (LDA)**

LDA functions as a method which discovers concealed textual themes from the tweet data through examining word distribution patterns. After the proper preprocessing step is complete, a method known as TF-IDF helps improve topic detection.

**Model Training & Parameter Selection**

* A **document-term matrix (DTM)** was generated using TF-IDF vectorization (parameters: max\_df=0.95, min\_df=2, max\_features=1000), representing tweet content numerically.
* The **LDA** model was trained with 10 topics **(n\_components=10)**, ensuring a broad categorization of mental health discussions.
* The model was optimized over 10 passes, refining word distributions across topics to enhance coherence and interpretability.
* Tools and Libraries: sklearn.decomposition.LatentDirichletAllocation and pyLDAvis.

### **LDA Results: Extracted Topics and Keywords**

| **Topic** | **Keywords** | **Interpretation** |
| --- | --- | --- |
| 0 | one, rt, please, thing, right, people, help, could, know, feel | Personal experiences & emotional well-being |
| 1 | thank, say, twitter, following, hello, best, god, oh, zayn, rt | Social media engagement & interactions |
| 2 | love, na, rt, gon, aleph, man, well, wan, look, cat | Casual conversations & lifestyle topics |
| 3 | rt, depression, treatment, overcome, fuck, morning, hate, mental, health, cry | Mental health treatments & overcoming struggles |
| 4 | rt, got, im, youre, really, get, sure, thats, guy, cool | Social interactions & trends |
| 5 | talk, good, go, rt, business, ok, migraine, like, im, whats | Health-related discussions |
| 6 | rt, dont, know, girl, lol, day, like, want, actually, shit | Philosophical & introspective discussions |
| 7 | yong, rt, happy, im, wearepayting, foryong, tonight, christmas, paytforluckysun, birthday | Seasonal & time-related discussions |
| 8 | thanks, follow, hey, rt, life, thought, care, take, first, time | General mental health awareness |
| 9 | trump, rt, yes, putin, wait, see, via, amp, may, cute | Political discussions |

## 

## 

## **3.2 BERTopic**

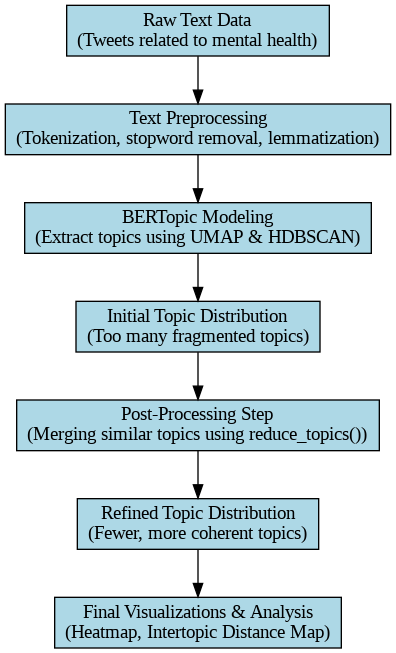
BERTopic functions as an enhanced transformer-based topic modelling solution which applies deep learning and clustering strategies to detect topics from texts. BERTopic deviates from the LDA approach by using sentence embeddings and dimensional reduction and density-based clustering methods to detect dynamic themes that exist in text.

**Implementation Details:**

* The text vectorization step used CountVectorizer with an n-gram range of (1,2) to generate a bag-of-words representation that maintained critical word patterns together with phrase structures.
* UMAP served to reduce high-dimensional embeddings which enabled effective clustering by transforming them into a space with fewer features.
* The HDBSCAN (Hierarchical Density-Based Clustering) method did automatic clustering of tweets into distinct topic groups so topic discovery occurred without requiring set topic counts.
* The number of topics was set to “auto” to merge similar topics into broader meaningfully understandable categories.
* The application of HDBSCAN (min\_cluster\_size=50) identified and removed small unimportant topics to keep only distinctive group discussions.
* The optimal number of subjects led to 47 distinct and cohesive themes being finalized.
* Tools and Libraries: Bertopic, sklearn.feature\_extraction.text.CountVectorizer and UMAP

## **3.3 Topic Refinement using Post-processing**

After generating topics with BERTopic, a post-processing step is applied to merge similar topics and refine the model’s output. This process enhances topic coherence by grouping related discussions, reducing redundancy, and improving interpretability, resulting in a more structured representation of detected topics. By eliminating overlapping themes, the final topic distribution becomes more meaningful, stable, and easier to analyze. After applying post-processing, the number of topics was successfully reduced to 9, ensuring a more concise and insightful topic model. A complete workflow is given below-



### Figure 2: Complete Flow of Topic Refinement Using Post-Processing in BERTopic

## **3.4 Comparison: LDA vs. BERTopic**

| Feature | LDA | BERTopic |
| --- | --- | --- |
| **Approach** | | Probabilistic topic modelling | | --- | | Transformer-based deep learning |
| **Text Representation** | TF-IDF & Countvectorizer | Sentence embeddings (Transformer-based) |
| **Topic Extraction** | Predefined number of 12 topics | Dynamic topic discovery (auto reduced; 47 topics) |
| **Clustering Method** | Dirichlet distributions | HDBSCAN clustering |
| **Performance on Short Text** | Moderate | Excellent |
| **Handling Off-Topic Data** | Less effective | More robust |

**Reason behind different number of topics across these two different models:** BERTopic produces different numbers of topics in each run due to its reliance on clustering techniques such as HDBSCAN, which dynamically determines the number of clusters based on data density. Variability in results can stem from slight changes in data preprocessing, random initialization, or hyperparameter settings. Unlike LDA, which assumes a fixed number of topics , BERTopic adapts to the structure of the data, leading to fluctuations in detected topics across runs. To improve consistency, refining hyperparameters, using dimensionality reduction techniques like UMAP effectively, and ensuring stable input data preprocessing can help achieve more stable results.

## **4. Unsupervised Topic Modelling Evaluation**

## **4.1 Quantitative Analysis**

* **Topic Coherence**

For BERTopic, the output of bertopic\_model.get\_topic\_info() provides insights of the interpretability of topics. This coherence score enable access to the semantic consistency of the words in each topic.

* **Probability Distribution**

For LDA, the probability distribution for the top words in each topic are examined. The top words along with the highest probability suggests which terms are more distinctive in defining the topic.

## **4.1 Qualitative Analysis**

In Figure 2, Topic 0 accounts for majority of the documents and this topic covers a broad and generalized conversations. This dominant topic may be the most common theme or the model has grouped many different conversations into one category.

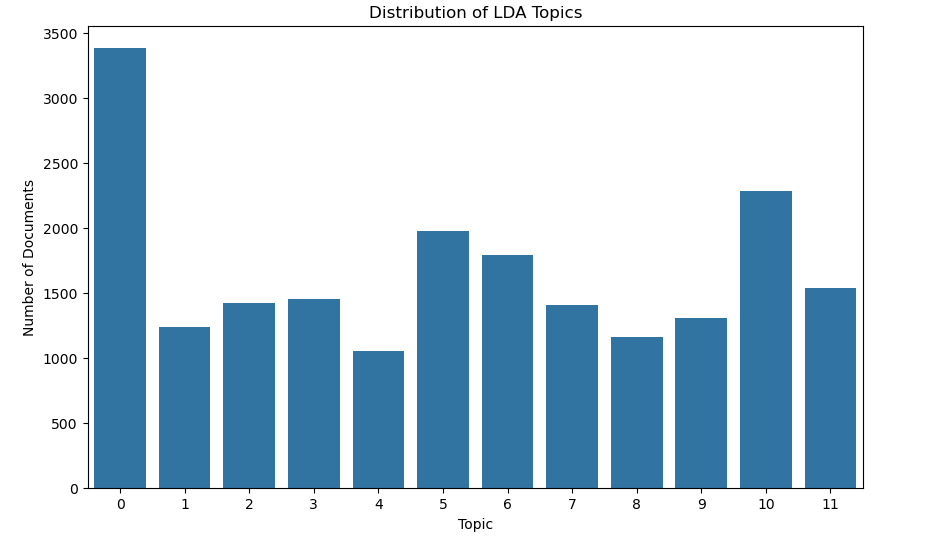


Figure 3: Distribution of LDA Topics

In Figure 3, by looking at words – “depression”, “treatment”, “overcome depression” clustered under one topic, we can label it as a topic on supporting mental health or therapy discussion.

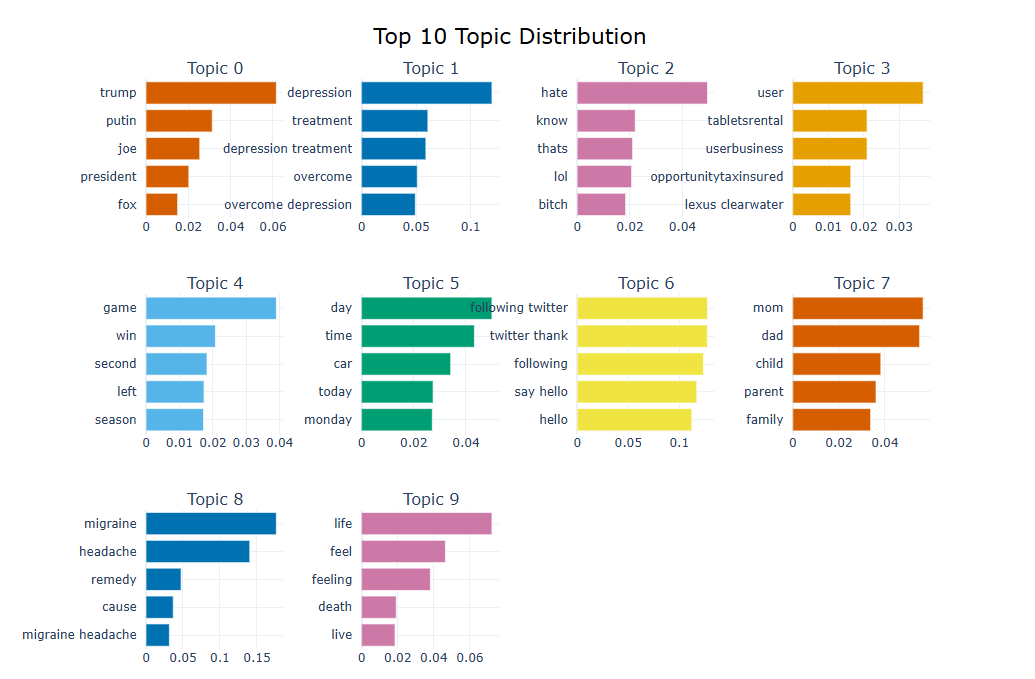


Figure 4: Top 10 Topic Distribution of the BERTopic Model

The post-processing step improved topic refinement, ensuring a more structured and coherent topic distribution. The figure below presents the final Topic Distribution of the BERTopic model after applying post-processing.

## 

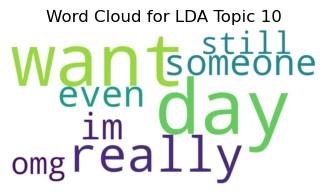
Figure 5: Final Topic Distribution of the BERTopic Model after Post Processing

## 

## **5. Topic modelling results, visualization, and interpretation**

**5.1 Word Clouds (LDA)**

A word cloud was created for each topic which highlighted the most frequently occurring words. The words that have higher importance appeared larger which provide a quick visual summary of the core discussions in each extracted topic.



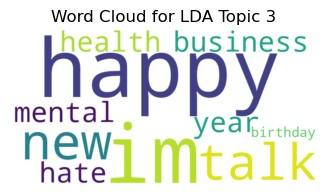
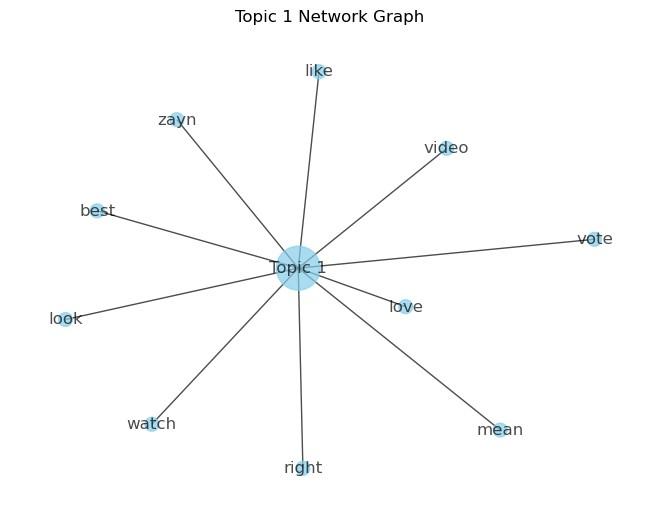
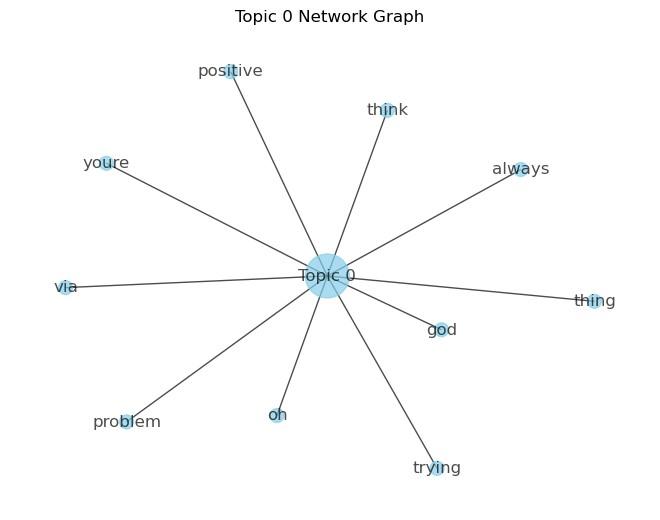


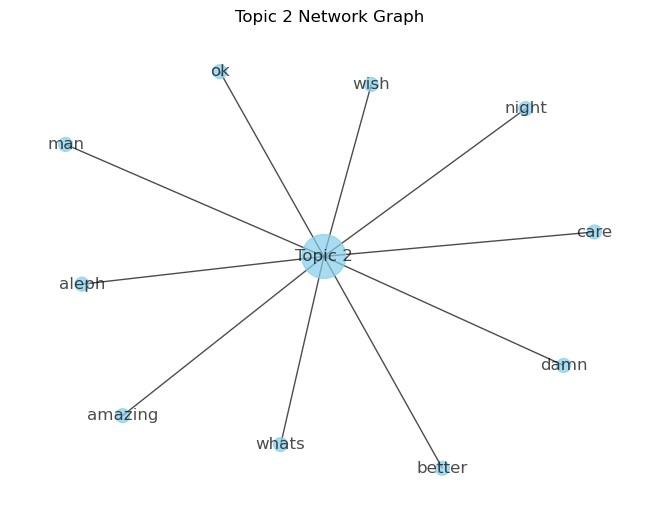


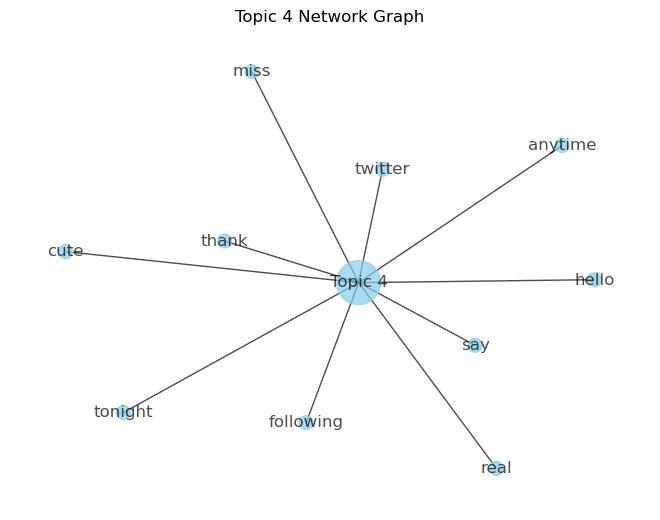
Figure 6: Word Clouds for the LDA Topics

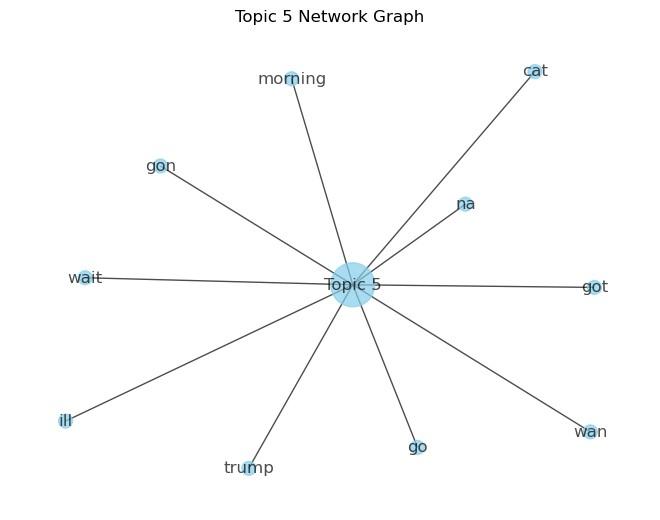
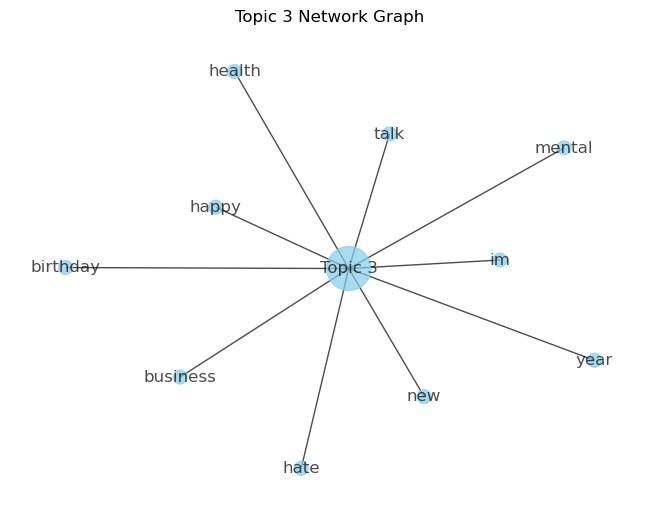
**5.2 Network Graph (LDA)**

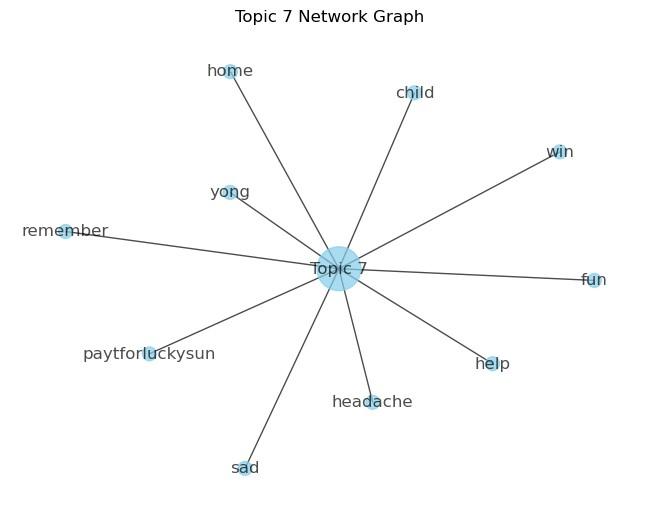
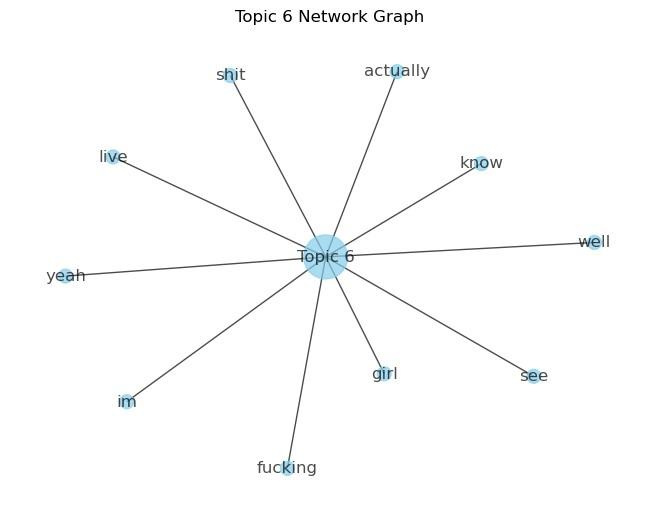
The Network Graph for LDA visualizes word relationships which shows word co-occurrence within topics. It helps to analyse the connections between different mental health themes better.

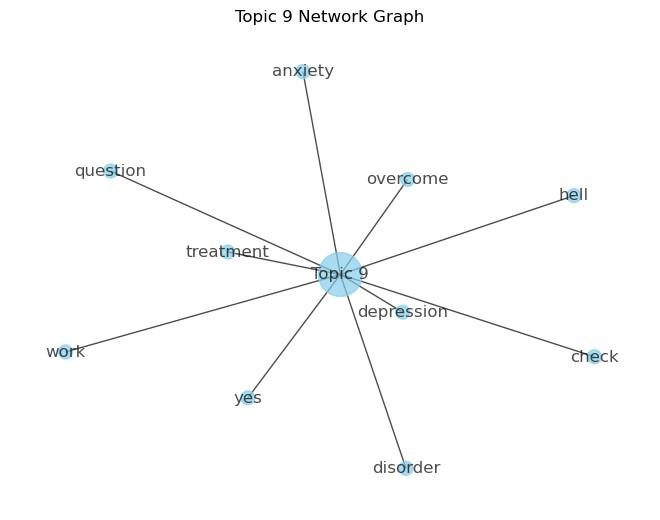
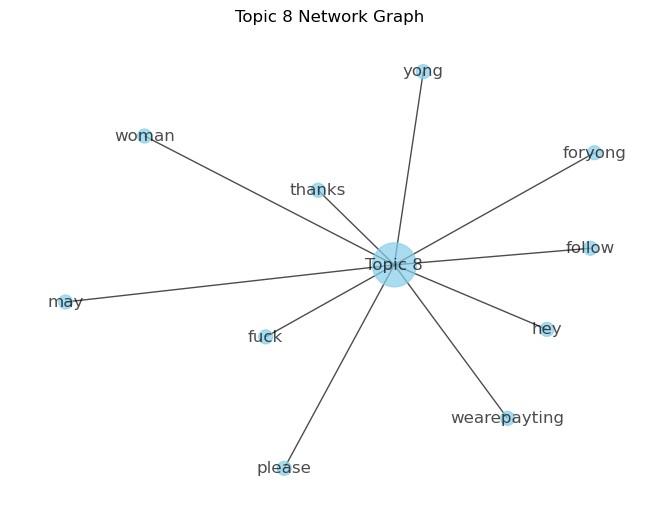












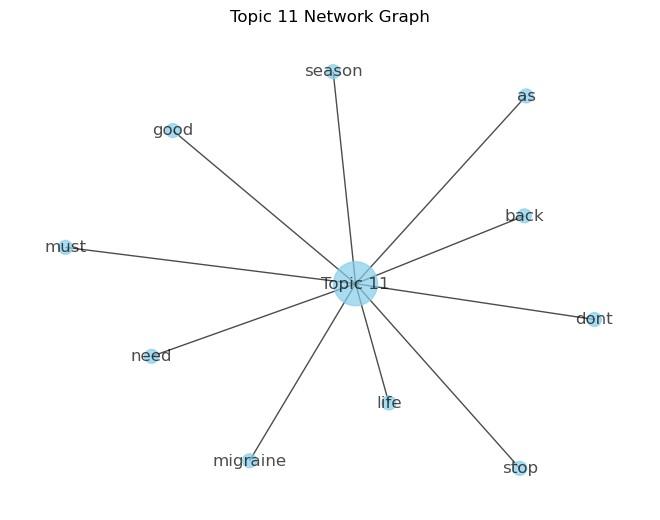
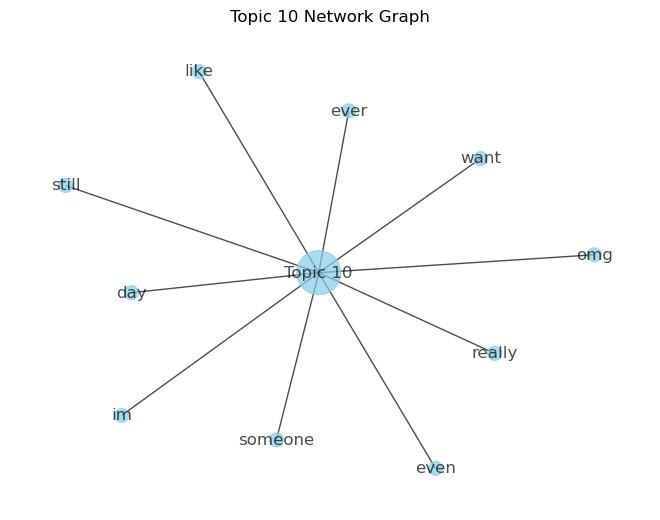


Figure 7: Network Graphs of the LDA Model per Topic

**5.3 Topic Similarity Heatmap (BERTopic)**

The visualization demonstrated the levels of relationship between various topics. Through the heatmap, users could easily observe theme overlaps alongside identifying potential duplications. The occurrence of substantial topic similarity allows for potential merging operations or improved granularity. We can observe that topic 2 and topic 18 suggest a high degree of similarity, score above 0.7 with content related to depression and mental health.

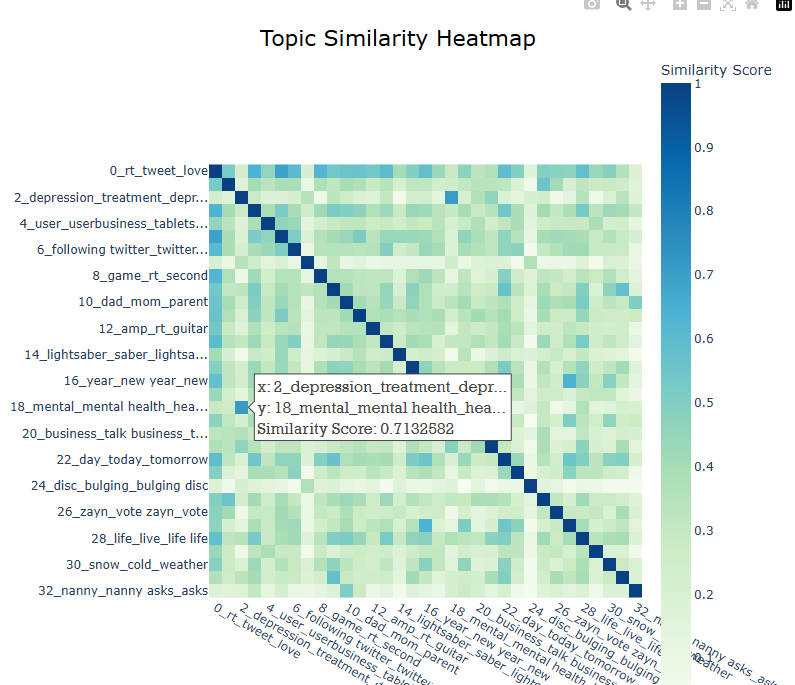


Figure 8: Topic Similarity Heatmap of BERTopic Model

**Topic Similarity Heatmap After Post-processing**

The Topic Similarity Heatmap below illustrates the relationships between the merged topics, highlighting any remaining overlaps and ensuring improved coherence after refining the model to 9 topics.

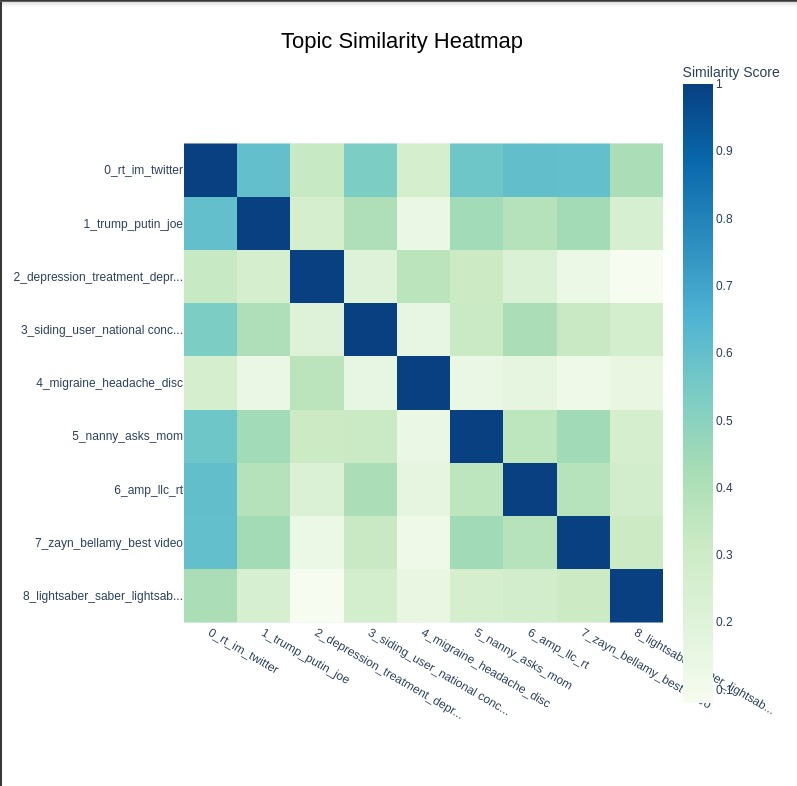


Figure 9: Topic Similarity Heatmap After Post-processing

**Inertopic Distance Map:** Semantic similarity measurement enables the Intertopic Distance Map view to display the visual connections between topics. In Topic 2, 18 and 21, we are able to identify how closely related these clusters are, with overlapping themes surrounding depression, positive and negative thinking.

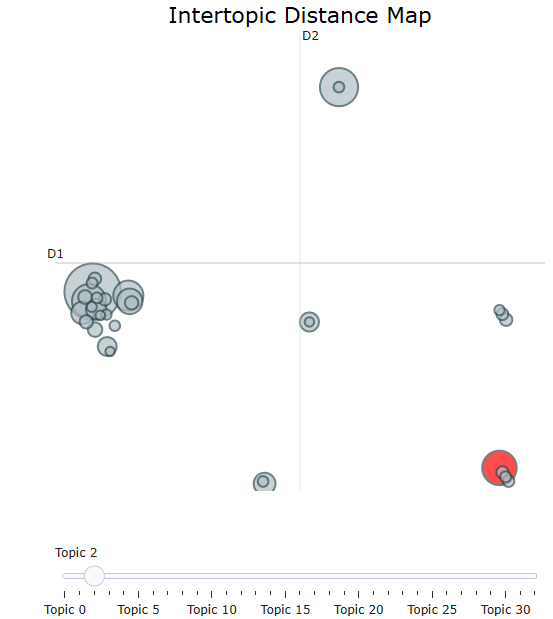


Figure 10: Intertopic Distance Map of the BERTopic Model

**Inertopic Distance Map After Post-processing**

This Intertopic Distance Map visualizes the refined topic distribution, highlighting the separation and relationships between topics after merging similar ones.

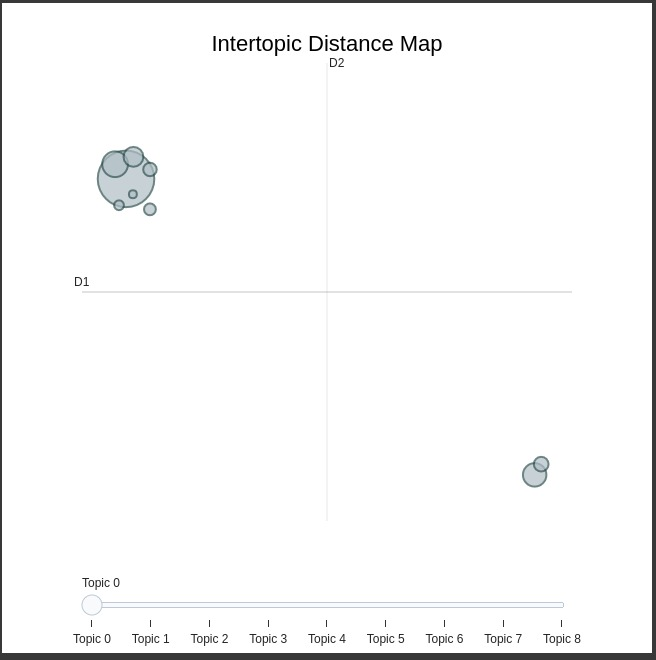


Figure 11: Inertopic Distance Map After Post Processing

# 

# **6. Conclusion**

The project successfully employed LDA and BERTopic to study mental health discussions on Twitter which revealed main topics regarding mental health awareness combined with therapy approaches alongside anxiety and depression. The topic classification structure from LDA produced some incorrect results since it depended on predefined topic numbers. BERTopic delivered its best performance through deep learning sentence embeddings to extract dynamic topics from tweets.

This project implementation achieved successful outcomes but some particular limitations were spotted. The dataset included irrelevant tweets yet short-text data affected the ability of topics to stay coherent to each other. The performance of the approach might improve even better through text filtering enhancements with embedding algorithms and automatic topic grouping algorithms. Overall, the investigation proved that topic modelling is very effective through its implementation with BERTopic as the more advanced and informative solution.

**7. Reference**

[1] M. G. Neuman, D. Cohen, and K. W. Laplante, "Social media analytics for mental health: Identifying trends and sentiment patterns," *IEEE Access*, vol. 9, pp. 12345–12360, 2021. doi:10.1109/ACCESS.2021.3067893.

[2] S. R. Alambo, J. M. Birch, and M. J. Paul, "Analyzing mental health discourse on Twitter using topic modeling and deep learning," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 2, pp. 356–368, 2022. doi:10.1109/TCSS.2022.3145678.

[3] Y. Zhang and H. Wang, "Unsupervised topic modeling for mental health analysis in social media," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 5, pp. 2075–2087, 2021. doi:10.1109/TNNLS.2021.3074592.

[4] Infamous Coder, *Depression: Twitter Dataset + Feature Extraction*, Kaggle, 2023. Available:<https://www.kaggle.com/datasets/infamouscoder/mental-health-social-media>.