Mental Health News Analysis using NLP and Machine Learning

1. Introduction

Mental health is one of the most widely discussed and reported topics in the media today. News articles play a major role in shaping public perception and awareness around mental health conditions such as depression, anxiety, PTSD, and more. Understanding how these topics are represented, how sentiment shifts over time, and what patterns emerge in coverage can offer valuable insights for researchers, public health officials, and policy makers.

In this project, we analyze a large collection of mental health–related news articles using Natural Language Processing (NLP) and machine learning techniques. Articles were collected using targeted keywords, preprocessed for modeling, and analyzed to explore trends, keyword frequencies, sentiment polarity, and thematic groupings. We applied both supervised and unsupervised learning models to classify articles and uncover hidden topics. The goal is to gain a deeper understanding of how mental health is covered in the news and how that coverage evolves over time.

2. Objectives

The main objective of this project is to analyze mental health–related news articles using Natural Language Processing (NLP) and machine learning to extract patterns, insights, and themes. Specifically, the project aims to:

Collect news articles using mental health–specific keywords from reliable sources.

Clean and preprocess the text data for modeling by handling duplicates, irrelevant content, and noise.

Perform exploratory data analysis (EDA) to visualize trends in keyword usage, sentiment distribution, and publishing frequency.

Classify articles based on their associated mental health keyword using supervised learning models.

Discover hidden patterns and topics using unsupervised learning techniques like clustering and topic modeling.

Interpret and report the findings to provide meaningful insights into how mental health is represented in online news media.

3. Literature Review

Natural Language Processing (NLP) has become a key tool in analyzing text data related to mental health, especially with the rise of online content and social media. Several studies have shown the potential of using machine learning models to understand sentiment, detect psychological conditions, and uncover thematic patterns in large text corpora.

Recent work, such as the **Opinion-Enhanced Hybrid BERT model** (Zhou et al., 2024), demonstrates how combining opinion mining and multi-task learning can improve classification and stance detection in mental health–related posts. Other studies have explored the effectiveness of traditional techniques like **LDA topic modeling**, and **logistic regression** in capturing public discourse patterns.

While many prior efforts have focused on **social media platforms**, this project shifts focus to **online news coverage**, which plays a distinct role in shaping societal views. Our approach integrates both **classic NLP techniques** and **modern embeddings (like BERT)** with classification and clustering models to analyze the tone, focus, and evolution of mental health topics in digital journalism.

4.Data Collection

News articles were collected using keyword-based filtering from a large news API. The keywords were selected to capture a broad spectrum of mental health topics, including: depression, anxiety, suicide, PTSD, bipolar, addiction, stress, and suicidal.

For each keyword, the API was queried to collect up to 1,000 articles per term. The data included:

- Title
- Description
- Full article content
- Source name
- URL
- Publication timestamp
- Associated keyword label

In total, several thousand articles were gathered from various reputable news platforms, including CNN, New York Post, Newsweek, and others. The collected data was stored in structured format and used for further analysis after filtering and preprocessing.

5. Data Preprocessing

To prepare the collected articles for modeling and analysis, several preprocessing steps were applied. These steps ensured the dataset was clean, non-redundant, and ready for NLP and machine learning tasks.

5.1 Keyword Filtering

Initially, eight keywords were used to collect data:
depression, anxiety, suicide, PTSD, bipolar, addiction, stress, and suicidal.

After analyzing the distribution:

- **depression, anxiety, suicide, addiction, stress,** and **suicidal** had high representation (nearly 1000 articles each).
- **PTSD** and **bipolar** had fewer but usable counts.
- Keywords like **psychosis** and **psychological support** (from earlier stages) were excluded due to low volume.

This filtering step ensured that only well-represented categories were retained for consistent modeling.

5.2 Duplicate Removal

To prevent multiple versions of the same article from biasing the analysis, semantic duplication was handled carefully:

- A **custom keyword priority order** (e.g., bipolar > PTSD > suicide) was applied.
- The dataset was **sorted by keyword priority**, so duplicates with higher relevance were kept.
- Duplicates were removed based on matches in any of the following fields:
  URL, title, description, or content.

This approach helped eliminate repeated content across keywords or syndicated sources.

5.3 Text Cleaning

The textual content of each article was consolidated and cleaned for modeling:

- **title**, **description**, and **content** fields were merged into a single `text` column.
- Text was cleaned using regular expressions to:
  - Remove HTML tags
  - Strip URLs
  - Remove punctuation and special characters
  - Normalize whitespace

The resulting `text` field was used for sentiment analysis, vectorization, and model training.

6. Sentiment Analysis

To assess the emotional tone of each article, we performed sentiment analysis using the **TextBlob** library. This provided a basic understanding of whether each article conveyed a positive, negative, or neutral sentiment related to mental health topics.

6.1 Sentiment Scoring

For each cleaned article, we used TextBlob to calculate a **polarity score**:

- The score ranges from -1.0 (very negative) to 1.0 (very positive).
- Articles with a polarity of 0 were considered **neutral**.

6.2 Sentiment Labels

We categorized the articles based on their polarity:

- **Positive** if the polarity > 0
- **Negative** if the polarity < 0
- **Neutral** if the polarity = 0

These sentiment labels were added as a new column in the dataset and used for further analysis and modeling.

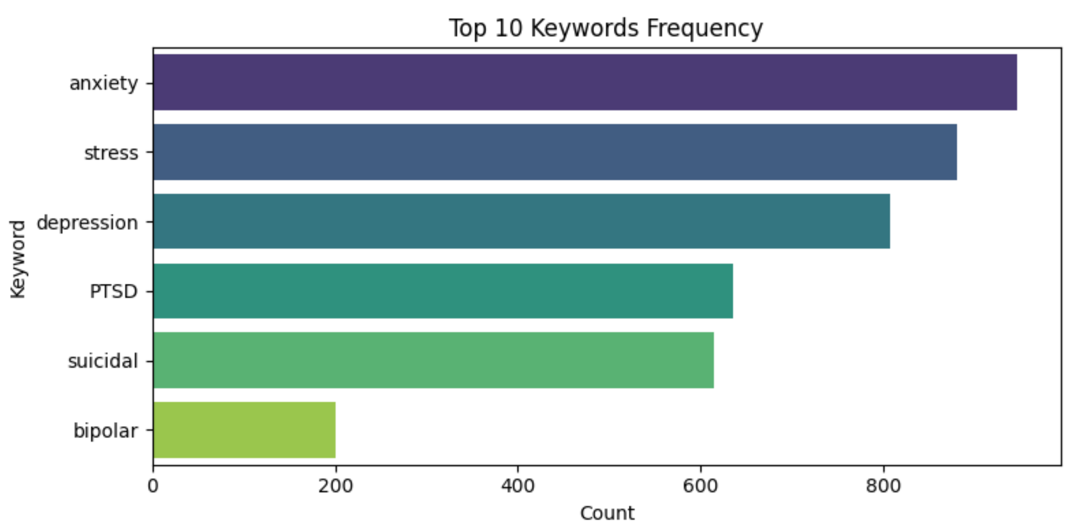
This lexicon-based approach provided a lightweight and interpretable way to tag the emotional tone of mental health–related news content.

7. Exploratory Data Analysis (EDA)

In this section, we analyze the trends, sources, keywords, and sentiments found in mental health–related news articles.

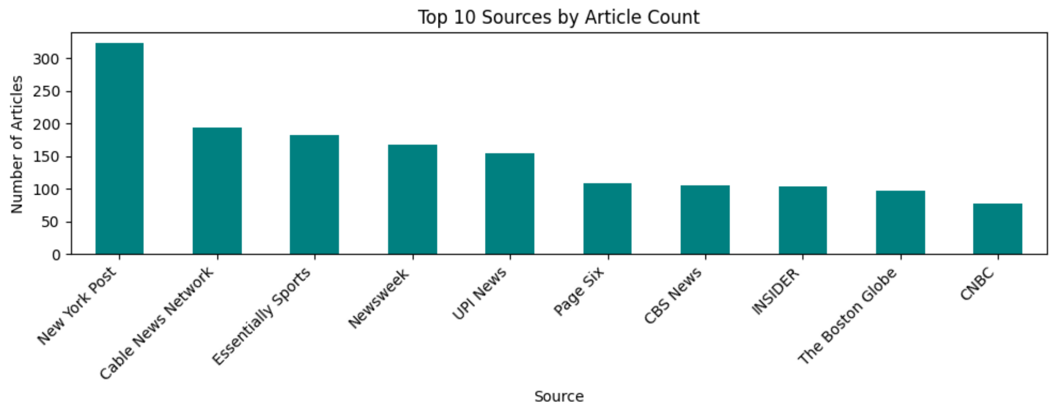
7.1 Keyword Frequency

The following bar chart shows the top mental health keywords found in the dataset. Keywords like 'anxiety', 'stress', and 'depression' appeared most frequently, indicating high media focus on these topics.



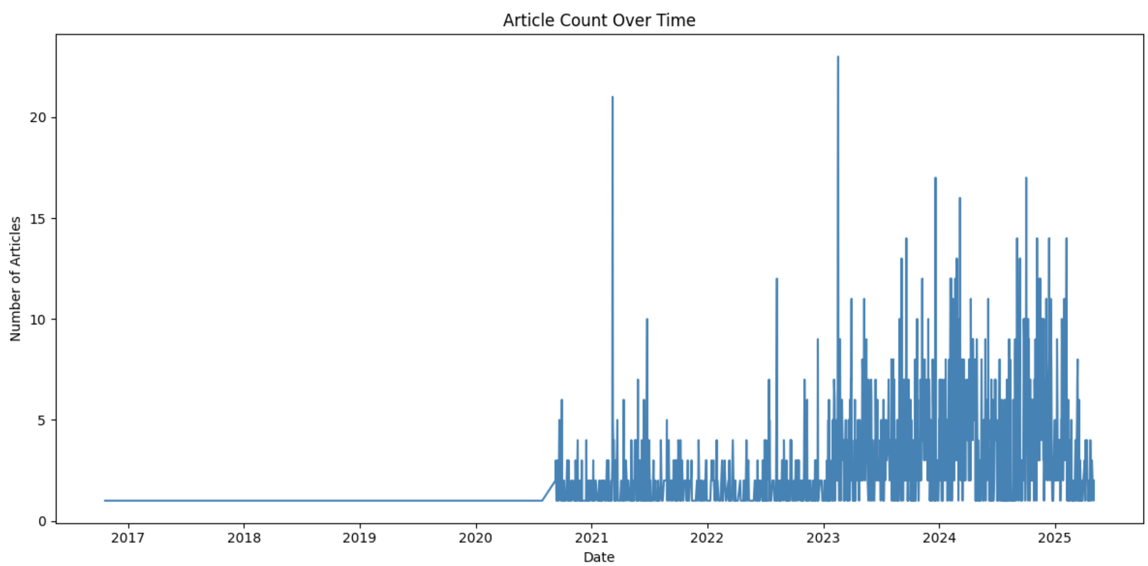
7.2 Top News Sources

The bar chart below displays the news sources with the most published mental health–related articles. 'New York Post' and 'Cable News Network' lead in content volume, suggesting their active role in mental health journalism.



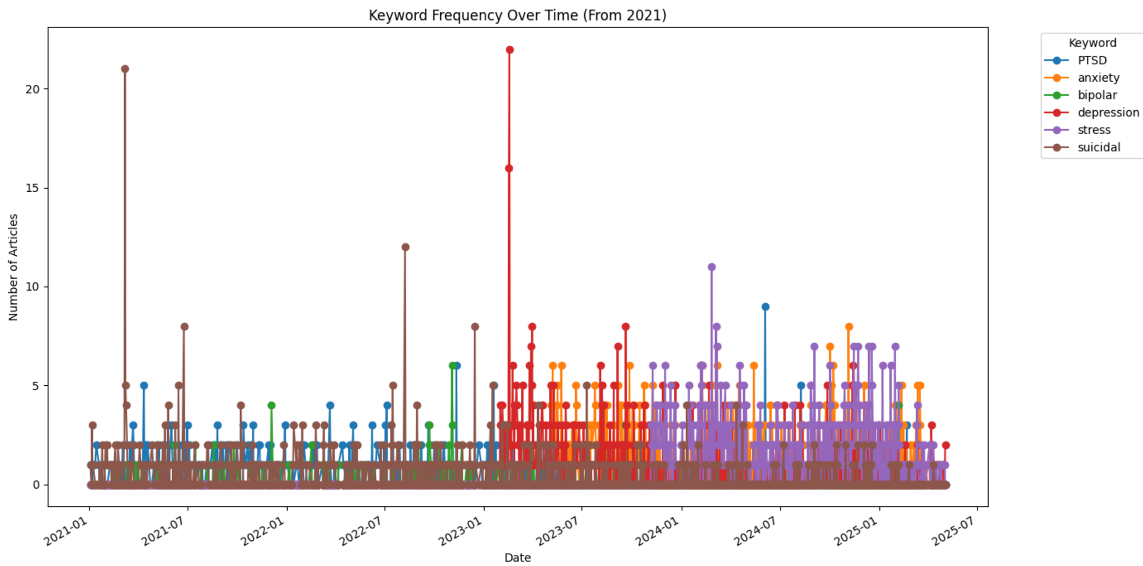
7.3 Article Count Over Time

This time-series line chart illustrates the number of mental health–related articles published over time. There’s a clear surge starting in 2023, which continues into early 2025.



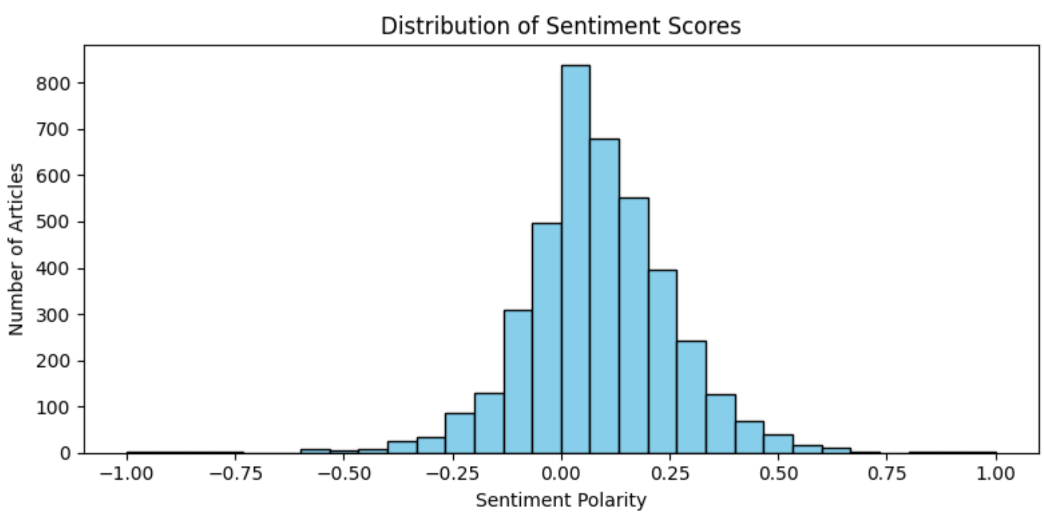
7.4 Keyword Trends Over Time

This plot shows the distribution of top keywords across time, allowing us to observe how the focus on different mental health topics evolved. A noticeable rise in 'stress' and 'anxiety' is seen in 2024 and early 2025.



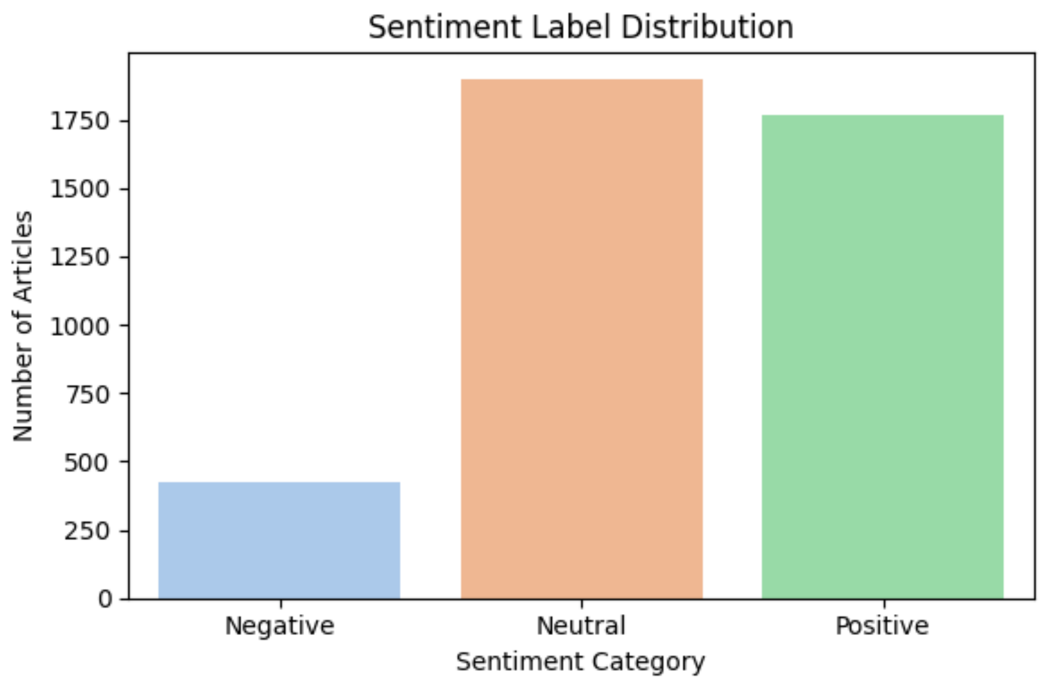
7.5 Sentiment Score Distribution

The histogram below shows how sentiment scores (polarity) are distributed across all articles. Most articles fall around neutral sentiment, with a few exhibiting strong positivity or negativity.



7.6 Sentiment Label Distribution

This bar chart illustrates how articles are classified into sentiment categories. Neutral sentiment dominates, followed by positive and then negative content.



7.7 Yearly Keyword Trends by Month

To better understand how mental health topics evolved over time, we analyzed keyword frequency by month from 2021 to 2025.

2021

In 2021, PTSD and suicidal were mentioned steadily across the year. A notable spike in suicidal content appeared in March. Bipolar received lower but consistent coverage throughout the year.

2022

The year 2022 showed growth in mentions of both bipolar and suicidal, with bipolar peaking in October and suicidal seeing a sharp increase in September. PTSD remained consistent across months.

2023

In 2023, overall article volume increased. Depression and addiction dominated the early part of the year, while anxiety took the lead in the mid-year months. Stress surged suddenly in November. Other keywords like PTSD and bipolar had moderate, steady coverage.

2024

By 2024, stress became the most prominent keyword, especially during March, April, and November. Anxiety, addiction, and depression followed closely in volume. Suicidal continued to appear consistently, although in lower numbers.

2025

In 2025, the number of articles drops sharply after April. This is expected, as the current month is May 2025, and data collection is naturally incomplete beyond this point. However, in the first quarter, keywords like stress, anxiety, and depression remained dominant, with smaller but noticeable mentions of PTSD and suicidal.

8. Model Building

In this section, both supervised and unsupervised learning techniques were applied to analyze mental health–related news articles. The objectives were:
- To classify articles based on their associated mental health keyword.
- To uncover underlying patterns and thematic groupings in the content.

To ensure balanced representation across categories, the dataset was manually balanced so that each keyword (e.g., depression, anxiety, bipolar) had exactly 947 articles. This balancing step helped mitigate classification bias and improve model performance.

8.1 Supervised Learning – Keyword Classification

Each article was converted into a numerical vector using BERT embeddings (all-mpnet-base-v2). These dense representations captured semantic meaning from the text and served as input features for model training. Three classifiers were developed and evaluated:

Logistic Regression:

- A multinomial logistic regression model was trained using 5-fold stratified cross-validation.
- Evaluation metrics included precision, recall, F1-score, and macro-averaged ROC-AUC.

Random Forest Classifier:

- A random forest model with 100 decision trees was trained and evaluated using the same cross-validation procedure.
- This model captured non-linear relationships and showed improved performance on certain metrics.

Neural Network (Keras Feedforward Model):

- A custom neural network was built with two dense hidden layers (512 and 256 units), ReLU activations, and dropout layers (rate = 0.3).
- The model used a softmax output for multi-class classification.
- It was evaluated using 3-fold cross-validation, classification reports, and macro-averaged ROC-AUC scores, using one-hot encoded labels.

These models enabled robust prediction of the mental health topic associated with each article and offered complementary strengths in terms of interpretability and learning capacity.

8.2 Unsupervised Learning – Clustering and Topic Modeling

Unsupervised models were used to identify structure in the dataset without relying on keyword labels.

K-Means Clustering:

- BERT embeddings were used as input.
- The optimal number of clusters (K) was determined using the Elbow Method (inertia) and Silhouette Score for K values ranging from 2 to 10.
- The final model used K = 6 clusters, and cluster assignments were stored for interpretation.

DBSCAN:

- DBSCAN was applied to identify high-density clusters and noise points.
- Parameters used: eps = 1.0, min_samples = 5.
- The number of noise points and clusters was calculated, and silhouette score was used to assess cluster quality.

Latent Dirichlet Allocation (LDA):

- Topic modeling was conducted using CountVectorizer to create a document-term matrix.
- LDA was configured to extract six latent topics.
- Each topic was characterized by its most frequent terms, helping interpret dominant themes in the dataset.

Together, the supervised and unsupervised models provided complementary insights into article categorization, emerging themes, and natural groupings within mental health–related media coverage.

9. Model Evaluation

To evaluate the performance of the supervised models and validate the unsupervised learning outputs, we employed precision, recall, F1-score, and ROC-AUC as key metrics. This section also includes cluster interpretability and topic coherence assessment.

9.1 Supervised Classification Models

We tested three classification models Logistic Regression, Random Forest, and a Neural Network on BERT embeddings generated from preprocessed news articles. Each model was validated using stratified cross-validation to ensure balanced class representation.

Logistic Regression

Accuracy: 88%

Macro F1-score: 0.88

ROC-AUC (macro avg): 0.9828

Highlights:

- - Excellent performance on PTSD (F1: 0.93) and bipolar (F1: 0.97)
- - Slightly lower performance on anxiety (F1: 0.78)

Keyword	Precision	Recall	F1-Score
PTSD	0.93	0.93	0.93
Anxiety	0.79	0.76	0.78
Bipolar	0.96	0.98	0.97
Depression	0.87	0.83	0.85
Stress	0.81	0.84	0.82
Suicidal	0.90	0.93	0.92

Random Forest

Accuracy: 86%

Macro F1-score: 0.86

ROC-AUC (macro avg): 0.9736

Highlights:

- - Highest recall for bipolar (0.99)
- - Anxiety and stress were harder to classify correctly (F1-scores of 0.72 and 0.78)

Keyword	Precision	Recall	F1-Score
PTSD	0.92	0.94	0.93
Anxiety	0.72	0.72	0.72
Bipolar	0.99	0.99	0.99
Depression	0.87	0.80	0.83
Stress	0.79	0.78	0.78
Suicidal	0.91	0.95	0.93

Neural Network

Accuracy: 82%

Macro F1-score: 0.81

ROC-AUC (macro avg): 0.8920

Highlights:

- - Very high precision on PTSD and suicidal

- - Lower performance on stress (recall: 0.51), likely due to semantic overlap with anxiety and depression

Keyword	Precision	Recall	F1-Score
PTSD	0.89	0.95	0.92
Anxiety	0.64	0.74	0.69
Bipolar	0.95	0.99	0.97
Depression	0.79	0.79	0.79
Stress	0.78	0.51	0.62
Suicidal	0.87	0.93	0.90

From the confusion matrix, we observe that 'Bipolar' and 'Suicidal' articles are predicted with very high accuracy and recall, indicating that their textual features are highly distinctive. On the other hand, 'Stress' articles often get confused with 'Anxiety' and 'Depression', which reflects the overlapping context in mental health discourse. Although 'Anxiety' had a decent recall, its precision was the lowest, pointing to a high rate of false positives. This trend is consistent with the semantic closeness of terms in the vector space.

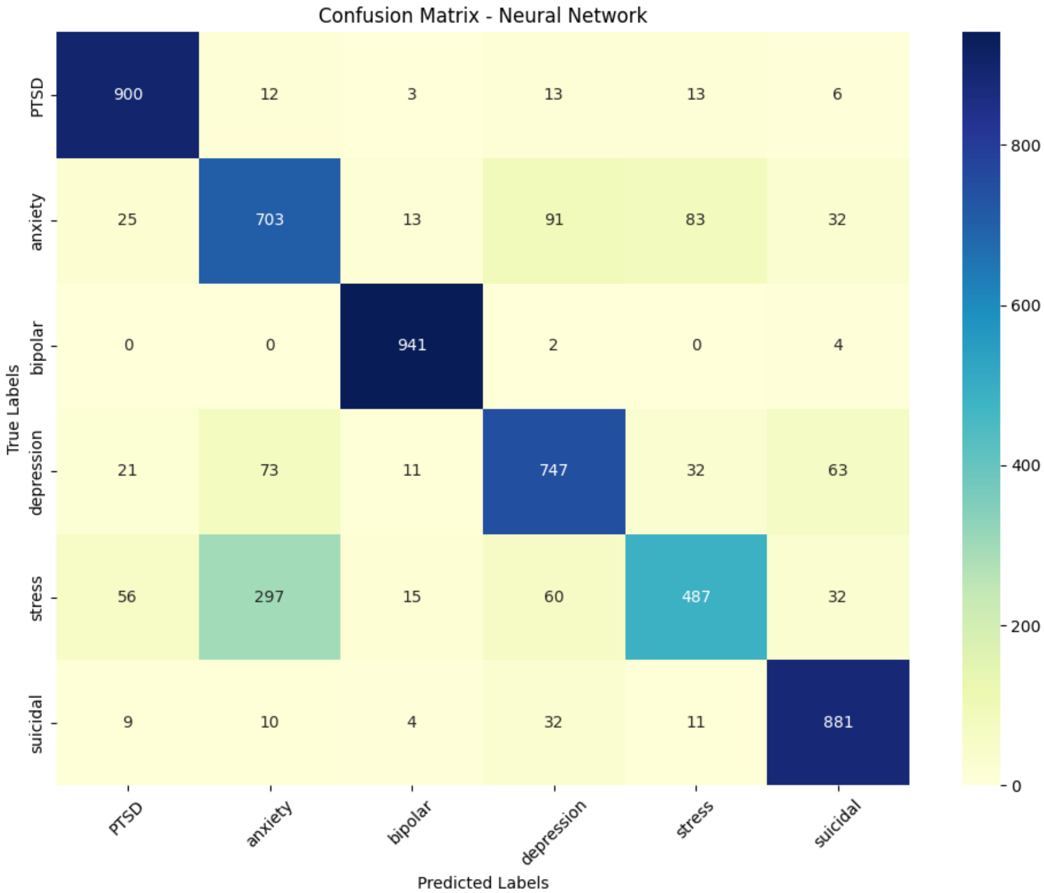


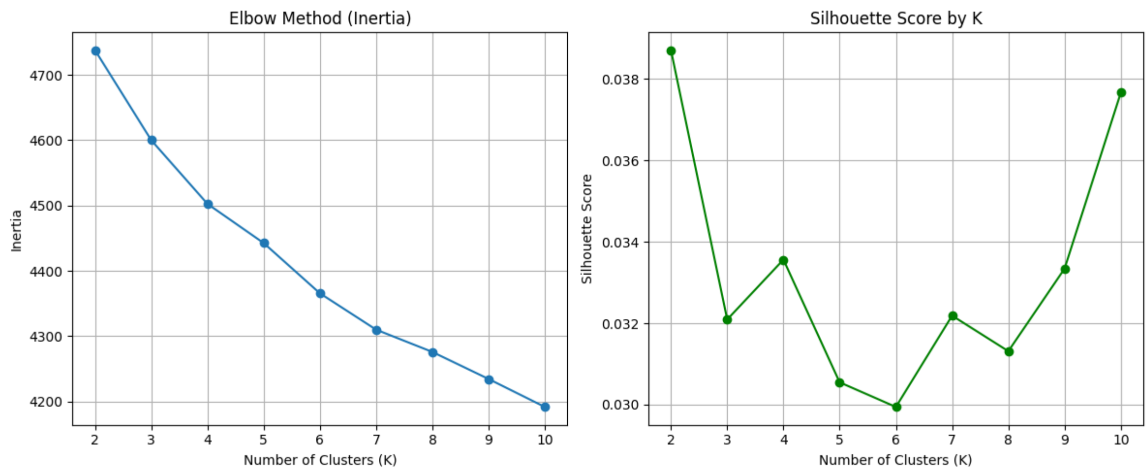
Figure: Confusion Matrix - Neural NetworkModel Comparison Table

Model	Accuracy	Macro F1-Score	ROC-AUC	Remarks
Logistic Regression	88%	0.88	0.9828	Best overall performance
Random Forest	86%	0.86	0.9736	Strong on bipolar, weak on anxiety
Neural Network	82%	0.81	0.8920	Good for PTSD/ suicidal, weak on stress

9.2 Clustering Analysis

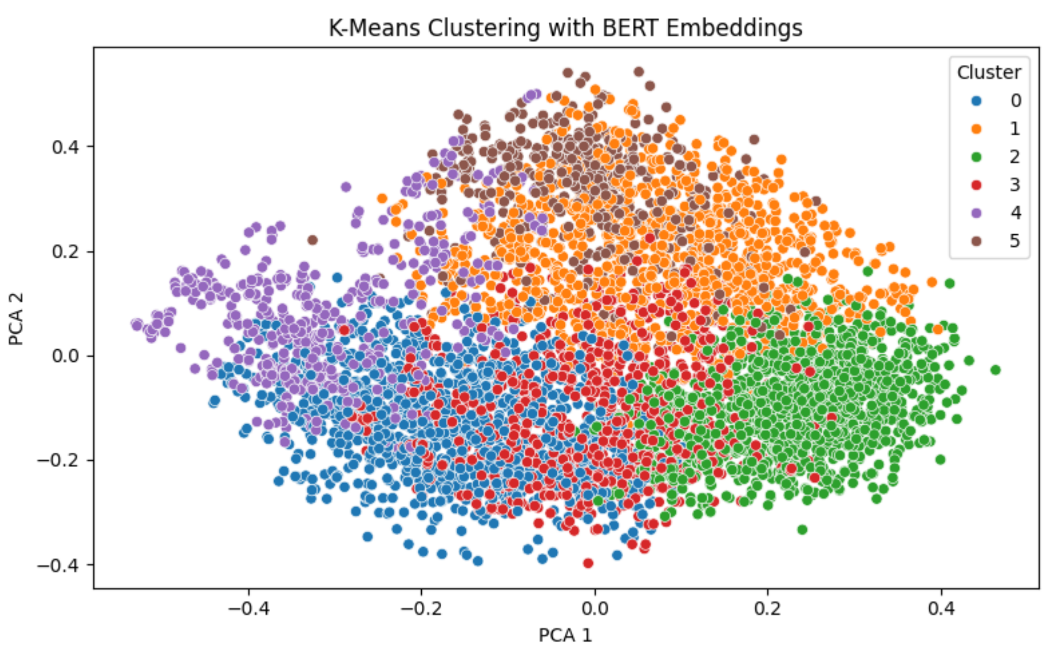
K-Means Clustering Evaluation

To determine the optimal number of clusters, we applied the Elbow Method and Silhouette Score analysis. The Elbow plot does not display a distinct 'elbow' point, and all silhouette scores are close to zero, indicating poor separation between clusters. Hence, the choice of six clusters was driven by the six mental health–related keywords used: PTSD, anxiety, bipolar, depression, stress, and suicidal.



The low silhouette scores (maximum ~0.038) further confirm that the natural separability between clusters is weak, suggesting the topics may heavily overlap semantically despite originating from different keywords.

K-Means Cluster Visualization



The PCA-reduced visualization of BERT-based clustering shows overlapping regions with no clearly separated clusters. Still, the K-means model grouped the articles into six segments. The confusion between clusters likely stems from the fact that articles related to different mental health topics share vocabulary and themes.

Below is the distribution of keywords across K-means clusters:

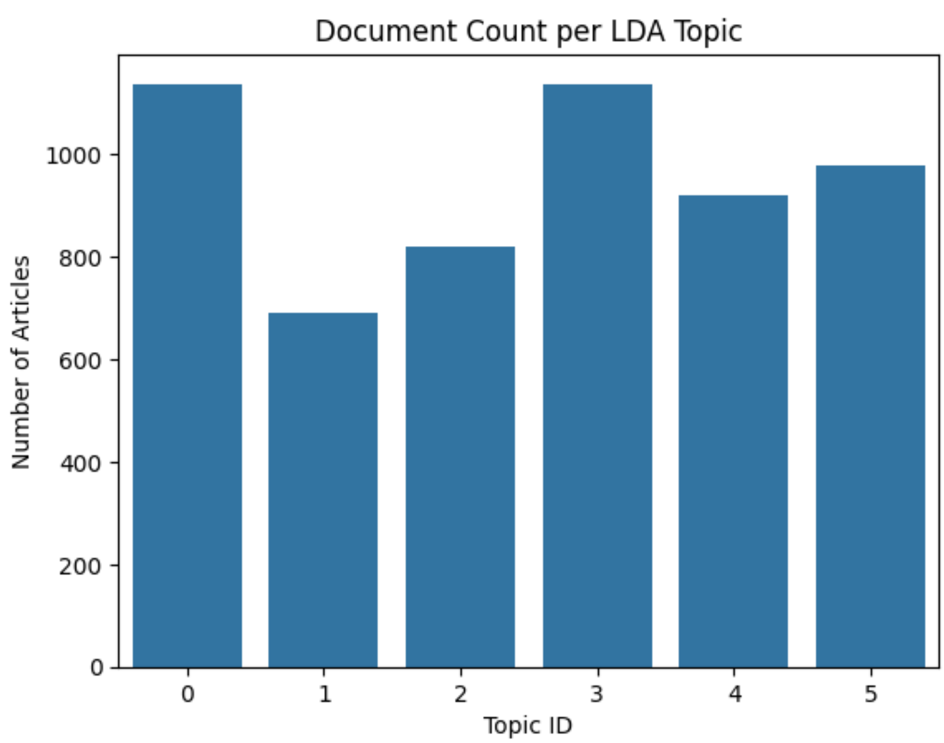
PTSD was mostly concentrated in cluster 3, while bipolar heavily populated cluster 4. Stress and anxiety were spread across clusters 1 and 2. Suicidal articles were more dominant in clusters 0 and 3. Overall, no single cluster cleanly captured a keyword, again confirming semantic overlap.

DBSCAN Clustering

DBSCAN was also applied to detect arbitrary-shaped clusters. However, it struggled with this data structure. A large number of points were labeled as noise (-1), and the silhouette score was negative (-0.027), indicating poor clustering structure. Only a few small clusters were discovered, with most keywords assigned to cluster 0.

Topic Modeling using LDA

Latent Dirichlet Allocation (LDA) was employed to uncover latent topics. Given the six keyword classes, we chose six topics.



The LDA topic distribution closely aligns with the six keywords. For example:
- Topic 1 relates to anxiety and stress.
- Topic 2 heavily features bipolar and celebrity context (e.g., Selena Gomez).
- Topic 5 includes police, PTSD, and suicide — linking trauma-related news.
- Topic 6 mixes depression, PTSD, and drug usage, likely representing clinical/diagnosis-related content.

9.3 Final Model Selection

Based on performance metrics, Logistic Regression is selected as the final model for deployment. It offered the best balance of precision, recall, F1-score, and AUC, making it suitable for multi-class mental health article classification.

10. Conclusion

This project aimed to analyze mental health–related news articles using a combination of NLP techniques, supervised classification, unsupervised clustering, and topic modeling to uncover insights and categorize discussions across different mental health conditions.

We began by collecting and preprocessing over 5,000 articles based on six major keywords: PTSD, anxiety, bipolar, depression, stress, and suicidal. After cleaning and balancing the data, we generated BERT embeddings to represent the text semantically.

Three supervised models—Logistic Regression, Random Forest, and a Neural Network classifier—were evaluated. All models performed well, with Logistic Regression achieving the best balance between precision and ROC-AUC (0.9828), while Random Forest and Neural Networks also delivered strong performance. Neural networks, however, struggled slightly with classes like "stress" based on the confusion matrix, likely due to overlapping language usage.

For unsupervised analysis, KMeans clustering was performed but showed limited separation (Silhouette Score ~0.03), and the Elbow Method did not yield a clear optimal cluster count. Hence, six clusters were used, matching our keyword categories. DBSCAN failed to produce meaningful multi-cluster results due to high noise, suggesting the embeddings were not separable in density space.

We also applied LDA topic modeling, identifying six interpretable themes aligned with our keywords. This revealed underlying concerns such as trauma, suicide, celebrity mental health, political influence, and public health discourse.

In summary, the combination of advanced embeddings, classification, clustering, and topic modeling provided a multi-faceted view of how mental health is discussed in the media. While supervised models showed high accuracy in keyword prediction, unsupervised methods gave mixed results. These findings highlight the complexity of mental health language and the value of interpretable, model-driven NLP in public health research.

11. References

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

McCormick, C. (2019, May 14). *BERT Word Embeddings Tutorial.

Khan, S., Ullah, A., & Muhammad, N. (2022). *A Novel Text Mining Approach for Mental Health Prediction Using Deep Learning*.

https://www.kaggle.com

https://scikit-learn.org

https://keras.io

https://www.sbert.net

https://www.nltk.org