SENTIMENT-BASED ANALYSIS & PREDICTION OF MENTAL HEALTH

ABSTRACT

Mental health disorders such as depression, anxiety, and PTSD affect millions of people worldwide. Traditional diagnosis methods rely on self-reporting and clinical assessments, which may lead to delayed intervention. Understanding these emotions can shed light on psychological health issues, the effects of various guidelines, and the reaction to the public. Feelings related to financial texts can be systematically categorized, whether headlines or detailed articles. A well-known digital media company, Guardian serves as an essential platform for monitoring public mental health and mood through mood analysis in financial news.

This study provides a detailed methodology for investigating and predicting mental health conditions, including depression, suicidal delicacies, fear, bipolar disorder, stress, personality disorders, etc., by using acid analysis. This study uses techniques for advanced natural language (NLP) processing such as NLTK, TextBlob, Vader, and Spacy to assess the mood of various text antiwords, including research data, social media posts, and underlying emotional states. The extracted mood reviews are integrated into additional related features for training robust machine learning models, such as logistics regression, accidental forests, support vector machines (SVM), and deep learning (BERT), which are intended to implement mental health outcomes. To develop more effective intervention strategies for mental health professionals. Although data visualization tools such as Matplotlib and Seeborn are used to explain the results, tightening is used to ensure the accessibility and ease of use of the developed system. The purpose of this study is to contribute to a data-controlled framework for mental health assessment, allowing individuals and experts to make appropriately determined decisions regarding mental health.

INTRODUCTION

Mental health disorders are a growing global concern, affecting individuals across all age groups, cultures, and socioeconomic backgrounds. Conditions such as depression, anxiety, bipolar disorder, PTSD, and personality disorders significantly impact personal relationships, work performance, and overall well-being. Despite advancements in healthcare, mental health remains underdiagnosed and untreated due to stigma, lack of awareness, and limited access to professionals. Traditional diagnosis relies on self-assessment and clinical interviews, which, while useful, are time-consuming, subjective, and prone to bias.

With the rise of Artificial Intelligence (AI) and Machine Learning (ML), researchers are now exploring computational methods to support mental health diagnosis. ML models can analyze large-scale data from sources like social media, surveys, and electronic health records to detect early signs of mental distress. Natural Language Processing (NLP), a branch of AI focused on understanding human language, plays a key role by extracting emotional cues from text, enabling sentiment analysis and mood evaluation.

ML techniques, including supervised learning (e.g., SVM, Random Forest, ANN) and unsupervised learning (e.g., K-Means clustering), have shown promise in identifying and predicting mental health conditions. While supervised models offer accurate classification, unsupervised approaches provide insights into hidden patterns, though their clinical use remains limited due to interpretability challenges.

This study aims to develop an AI-powered system that leverages NLP and ML to detect and predict mental health issues like depression, anxiety, stress, and personality disorders. By analyzing user-generated texts for emotional states, the system can enable early intervention and scalable mental health monitoring. Additionally, the research addresses concerns around model accuracy, data privacy, and ethical use while proposing future improvements in AI-driven mental healthcare.

LITERATURE REVIEW

This paper presents a systematic literature review, critical analysis, and summary of machine learning techniques used to predict, diagnose, and identify mental health issues. It highlights existing challenges, limitations, and explores opportunities for future research in this growing field. By identifying research gaps and potential advancements, the study contributes to the ongoing development of AI-based mental health solutions and offers a roadmap for future exploration.

While several existing reviews discuss the use of machine learning in mental health, most provide general overviews without addressing recent gaps or offering targeted research directions. This study addresses that gap by critically examining recent advancements, evaluating various machine learning and deep learning models—including SVM, Random Forest, ANN, BERT, RoBERTa, and GPT—and comparing their effectiveness based on accuracy, sensitivity, specificity, and AUC metrics.

Transformer-based models have shown particular promise in analyzing nuanced language associated with mental distress. Moreover, emerging multimodal approaches that combine text, audio, and facial expressions are being explored to further enhance diagnostic accuracy. However, challenges such as data scarcity, computational requirements, and ethical concerns remain.

Ultimately, this paper serves as a valuable resource for researchers, clinicians, and developers, guiding them toward more effective, interpretable, and scalable AI-driven solutions for mental health prediction and monitoring.

RESEARCH GAP

The rise in mental health issues—especially post-pandemic—has highlighted the need for intelligent systems to analyze and classify mental health conditions from text. While ML and NLP have advanced, key research gaps still exist.

- Hybrid Preprocessing Techniques Underused:
  Most studies rely solely on either traditional NLP tools (like NLTK) or advanced models (like SpaCy or transformers). Our approach combines both, extracting surface-level and semantic features for better classification.
- Lack of Multi-Input Modal Systems:
  Existing research often stops at model evaluation without practical tools. Our solution supports varied input formats (CSV, text, handwritten notes, etc.) for real-world applications beyond just social media.
- Class Imbalance Ignored:
  Datasets often lack representation for rare disorders like schizophrenia. We use SMOTE to handle imbalanced data, ensuring fair learning across categories.
- Poor Interpretability and Visualization:
  Many models are black boxes. We include visual tools like box plots, density plots, and word clouds to make insights clearer and more explainable.
- Narrow Evaluation Metrics:
  Most studies rely on accuracy or F1-score. Our model uses confusion matrices, classification reports, and confidence visualizations for a more complete performance view.

Additionally, many models ignore multimodal data (like voice, facial cues, or physiological signals), suffer from data bias, and lack interpretability and ethical safeguards. Bridging these gaps is essential for building trustworthy, clinically relevant, and inclusive mental health tools.

PROBLEM FORMULATION

- The primary objective of this research is to develop a robust machine learning framework capable of analyzing textual data to identify and categorize various mental health conditions. Given a textual input, the system is expected to perform the following tasks:

- Text Classification: Automatically classify the input text into predefined mental health categories such as Depression, Anxiety, Post-Traumatic Stress Disorder (PTSD), Obsessive Compulsive Disorder (OCD), or Normal.

- Keyword Extraction: Identify and extract significant keywords or phrases that are closely associated with specific mental health conditions. These keywords serve as interpretable indicators and features for classification.

- Visualization: Generate intuitive visual representations, such as word clouds or heatmaps, to reveal frequently occurring terms and underlying emotional patterns within each mental health category. These visualizations offer deeper insights into the linguistic features contributing to each label.

- Class Imbalance Management: Address the issue of imbalanced data distribution among mental health categories using synthetic data augmentation techniques like SMOTE (Synthetic Minority Over-sampling Technique), to ensure the model performs well across all classes.

Mathematical Formulation:

Let the dataset be represented as:

- $D = \{ (x_1, y_1), (x_2, y_2), \ldots\ldots (x_n, y_n) \}$

- where $x_i \in X$ denotes the i^{th} text sample, and $y_i \in Y$ is its corresponding label from the set $Y = \{depression, anxiety, PTSD, OCD, normal\}$

The task is to learn a mapping function $f : X \rightarrow Y$, such that the predicted output $\hat{y} = f(x)$ closely matches the true label y.

Optimization goal:

The model's performance is enhanced by:

- Feature extraction: employing term frequency-inverse document frequency (TF-IDF) to convert the raw text into weighted feature vectors that highlight the importance of words within the corpus.

- The classification algorithm involved training a support vector machine (SVM) classifier on the transformed data, aiming to achieve high classification accuracy and ensure that the model can generalize well to new, unseen data

DATASET TO BE USED

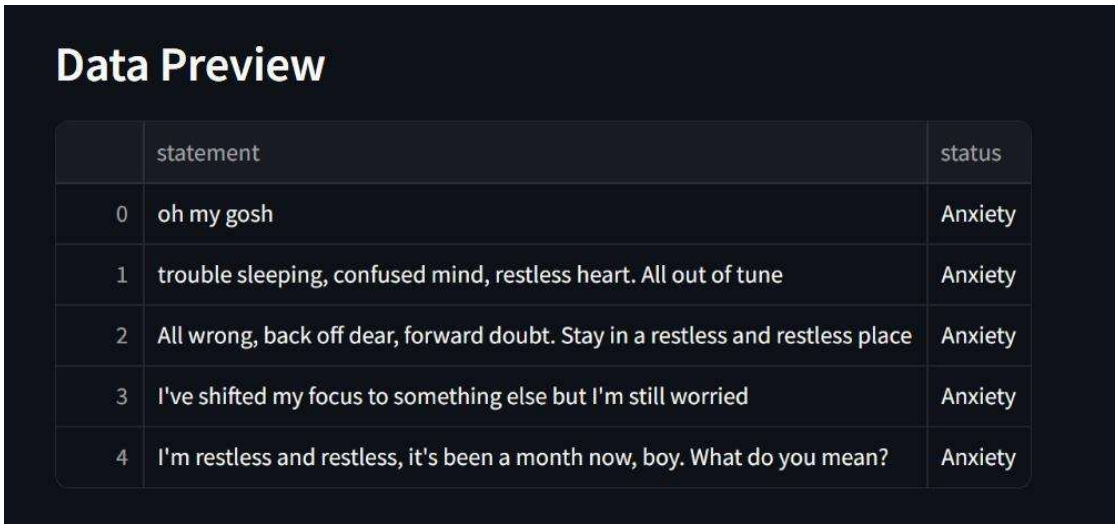
This dataset consists of textual statements labeled with mental health statuses, making it suitable for NLP-based sentiment and condition classification tasks in mental health analysis.

Features:

- statement: Textual entries expressing personal thoughts, feelings, or mental states.
- status: Mental health condition associated with the statement. The dataset includes seven

categories:

- Normal
- Depression
- Suicidal
- Anxiety
- Stress
- Bi-Polar
- Personality Disorder



Data Preview

	statement	status
0	oh my gosh	Anxiety
1	trouble sleeping, confused mind, restless heart. All out of tune	Anxiety
2	All wrong, back off dear, forward doubt. Stay in a restless and restless place	Anxiety
3	I've shifted my focus to something else but I'm still worried	Anxiety
4	I'm restless and restless, it's been a month now, boy. What do you mean?	Anxiety

URL of Dataset: https://www.kaggle.com/datasets/suchintikasarkar/sentiment-analysis-for-mental-health/data

METHODOLOGY

The methodology consists of several key stages:

1. Data Preprocessing

   Techniques Used:

   - Text Normalization: Converting text to lowercase and removing punctuation.
   - Tokenization: Breaking down sentences into individual words.

- Stopword Removal: Removing common words (e.g., "the", "and") to reduce noise.
- Lemmatization: Converting words to their root forms.
- Enhanced Preprocessing using spaCy: Handling Named Entity Recognition (NER) and linguistic features.

Mathematical Representation:

Let S be the input sentence.

Tokenization can be represented as:

$S = \{ w_1, w_2, \ldots\ldots\ldots, w_n \}$
where each w_i is a word token.

Stopword removal applies a filter function $f : S' = \{ w_i \mid w_i \notin \text{Stopwords} \}$
where S' is the filtered text.

Lemmatization applies a function $f : S'' = \{ g(wi) \mid wi \in S' \}$
where S'' is the final preprocessed text.

2. Feature Extraction Using TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency) transforms text into numerical features.

Mathematical Formula:

$TF(w) = n_w / N$
where:
- n_w = number of times word w appears in the document
- N = total number of words in the document

$IDF(w) = \log D / d_w$
where:
- D = total number of documents
- d_w = number of documents containing word w

$TF - IDF(w) = TF(w) \times IDF(w)$

3. Handling Class Imbalance using SMOTE

Synthetic Minority Over-sampling Technique (SMOTE) generates synthetic samples for the minority class.

Mathematical Formulation:

Given a minority class sample x, SMOTE selects a nearest neighbor x_{nn} and generates a synthetic sample as:

$x_{new} = x + \lambda * (x_{nn} - x)$
where λ is a random number in [0,1].

4. Classification Using SVM

Support Vector Machine (SVM) separates classes using a hyperplane.

Mathematical Representation:

Given a dataset $D = \{(x_i, y_i)\}$
where:

- x_i is a feature vector
- y_i is the class label (e.g., mental health condition)

SVM finds the optimal hyperplane:

$w . x + b = 0$
where

- w = weight vector
- b = bias term

The objective function to minimize:

$1/2 \; ||w||^2$

subject to : $y_i (w . x_i + b) >= 1, \; \forall I$

This ensures correct classification with maximum margin.

5. Model Evaluation

Metrices Used:

- Precision = TP / TP + FP
- Recall = TP / TP + FN
- F1-Score = 2 × (Precision + Recal) / (Precision × Recall)

where:

TP = True Positives

FP = False Positives

FN = False Negatives

ALGORITHM

Input: Raw mental health-related text data (collected from datasets or extracted from images using OCR)
Output: Predicted mental health condition labels for input text (like Anxiety, Depression, Stress, etc.)

Step 1: Data Acquisition

- The dataset consisting of mental health-related textual data and corresponding labels is collected from reliable sources.

- This dataset may include user comments, patient feedback, social media posts, or clinical text.

- The data is loaded using Python's pandas library.

Step 2: Text Preprocessing

- The raw textual data is unstructured and may contain noise like punctuation, stop words, symbols, etc.

- The following preprocessing operations are performed sequentially:

  a) Lowercasing: Convert the entire text to lowercase to ensure uniformity.

  b) Removal of Punctuation: Remove all special characters and punctuation symbols using regular expressions.

  c) Tokenization: Split text into individual words (tokens) for better manipulation.

  d) Stop Words Removal: Eliminate commonly used words (like "is", "the") that do not contribute meaningful information.

  e) Lemmatization: Reduce words to their base/root form to standardize different variations of words (e.g., 'running' → 'run').

  f) Named Entity Recognition (NER): Utilize the spaCy library to detect and retain important named entities (like names, places, diseases, etc.) from the text.

Step 3: Feature Extraction using TF-IDF

- Convert the cleaned text data into numerical features using the Term Frequency-Inverse Document Frequency (TF-IDF) technique.

- TF-IDF measures the importance of a word relative to a document and the entire corpus.

Step 4: Handling Class Imbalance using SMOTE

- Real-world datasets often suffer from class imbalance, where some classes have significantly fewer samples.

- To address this, the Synthetic Minority Oversampling Technique (SMOTE) is applied to generate synthetic samples for under-represented classes.

Step 5: Dataset Splitting

- The processed data is split into training (80%) and testing (20%) subsets using stratified sampling to ensure proportional representation of classes.

Step 6: Model Selection and Training

- The Support Vector Machine (SVM) classifier is chosen for its robustness in handling text classification problems.

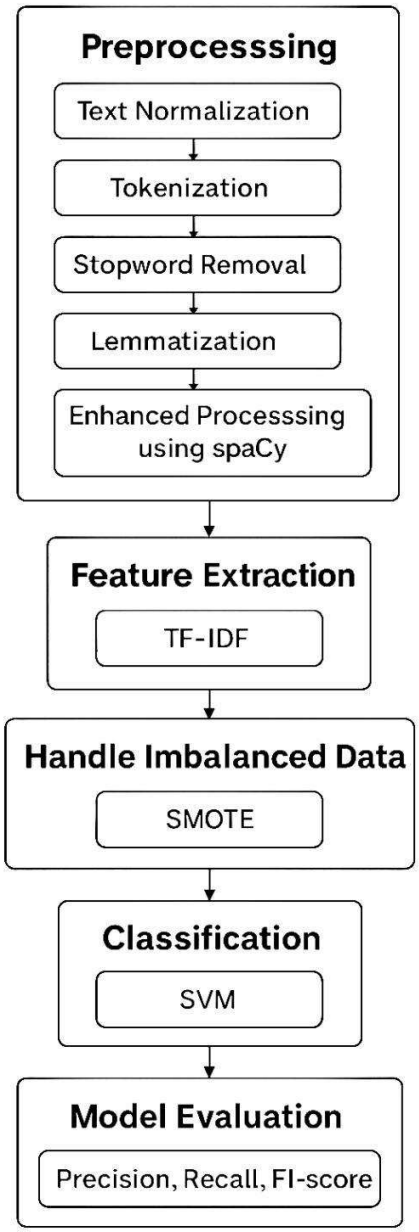
- The SVM model is trained on the training subset using the feature vectors generated in Step 3.

Step 7: Prediction

- The trained SVM model predicts the mental health labels for the unseen testing data.

Step 8: Model Evaluation

- The model's performance is evaluated using: Accuracy, Precision, Recall, F1-Score, Confusion Matrix: - These metrics give a comprehensive understanding of the model's correctness, sensitivity, and overall performance.

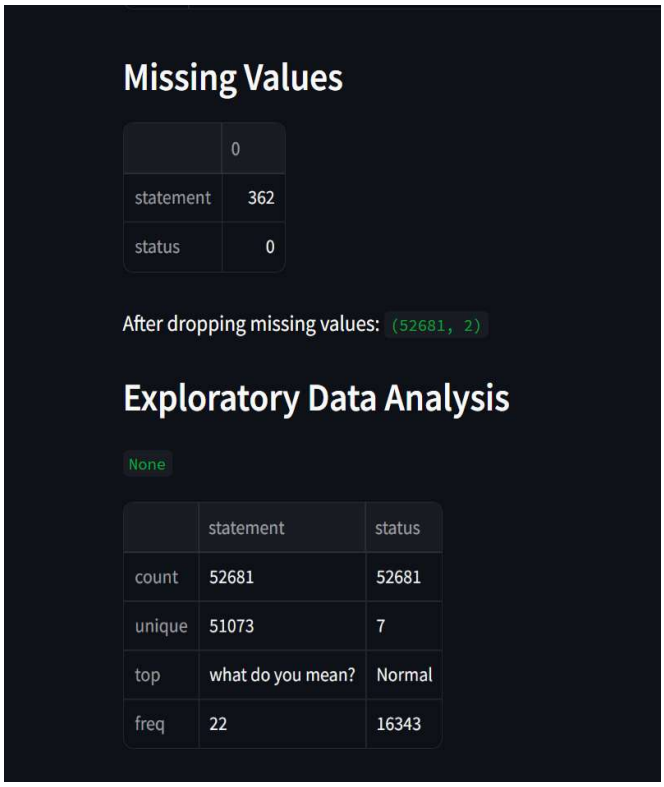


Flow Diagram Following the Methodology & Algorithm

RESULT

Missing Values and Exploratory Data Analysis

The dataset initially had 362 missing statement values, which were dropped, resulting in 52,681 valid entries for analysis and most entries belong to the "Normal" and "Depression" categories, with "Suicidal" also significantly represented, indicating a diverse mental health distribution.



Missing Values

	0
statement	362
status	0

After dropping missing values: `(52681, 2)`

Exploratory Data Analysis

None

	statement	status
count	52681	52681
unique	51073	7
top	what do you mean?	Normal
freq	22	16343

1. Sentiment Counts

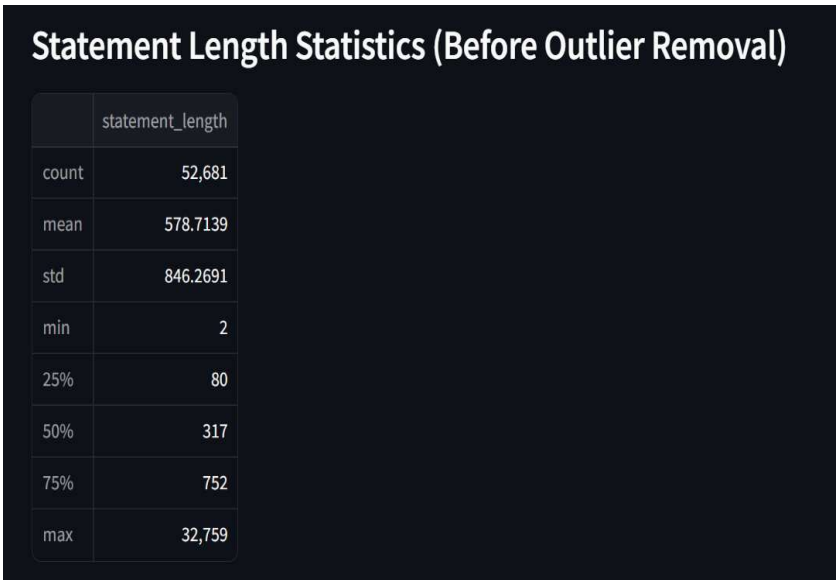
Most entries belong to the "Normal" and "Depression" categories, with "Suicidal" also significantly represented, indicating a diverse mental health distribution.

Sentiment Counts:

status	count
Normal	16,343
Depression	15,404
Suicidal	10,652
Anxiety	3,841
Bipolar	2,777
Stress	2,587
Personality disorder	1,077

2. Statement Length Statistics (Before Outlier Removal)

The statement lengths vary widely, with a mean of ~579 words and a long tail of outliers extending up to over 32,000 characters.

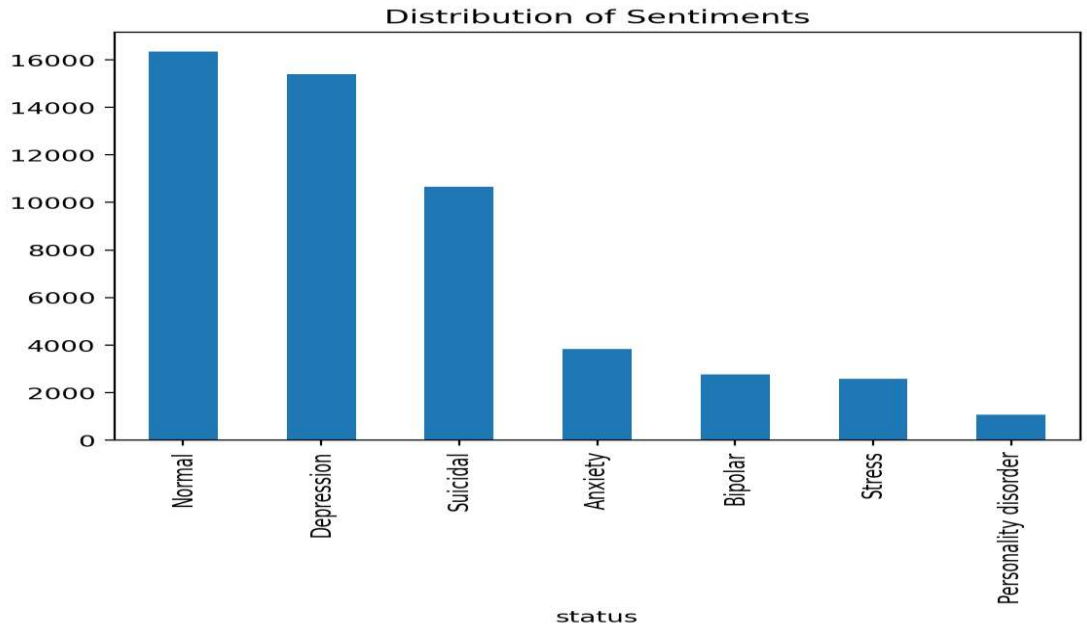


3. Outlier Removal

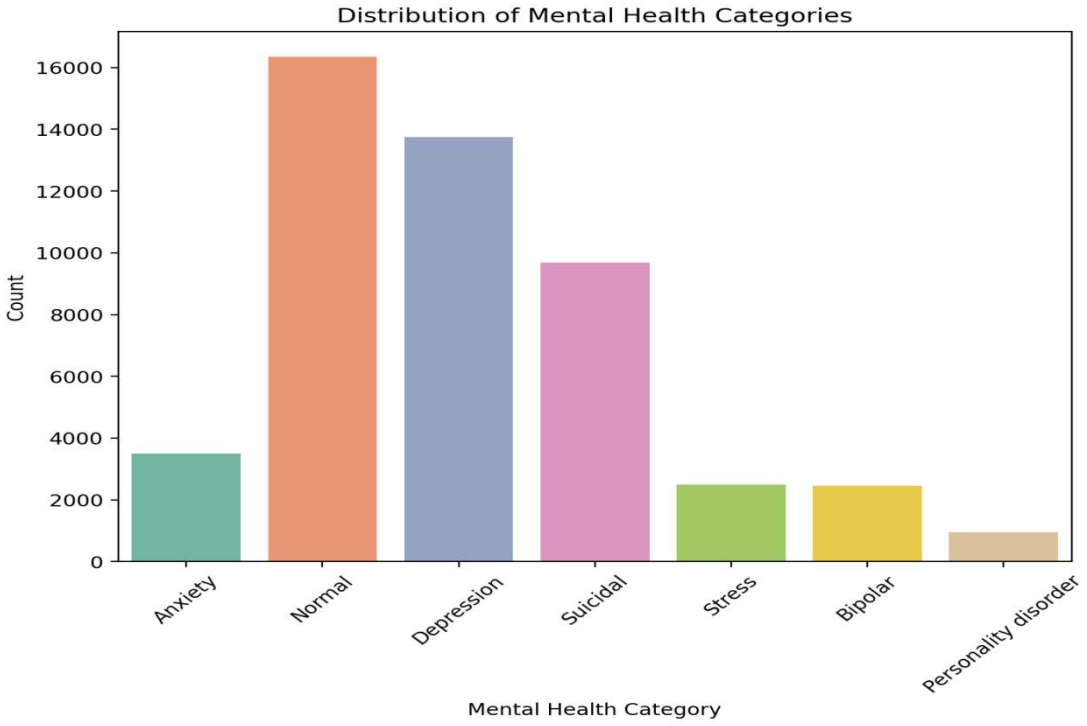
3,512 outliers were removed based on extreme statement lengths, reducing the dataset to 49,169 rows for more consistent analysis.

4. Distribution of Sentiments (Bar Plot) (Before Outliers)

"Normal", "Depression", and "Suicidal" categories dominate the dataset, highlighting the need for focused mental health classification.

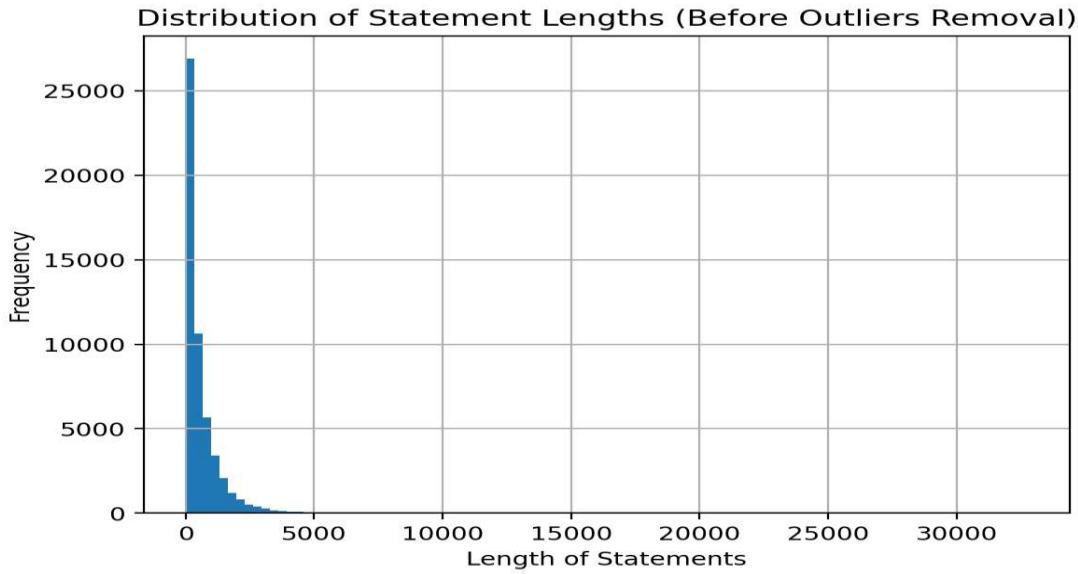


5. Distribution of Sentiments (Bar Plot) (After Outliers)

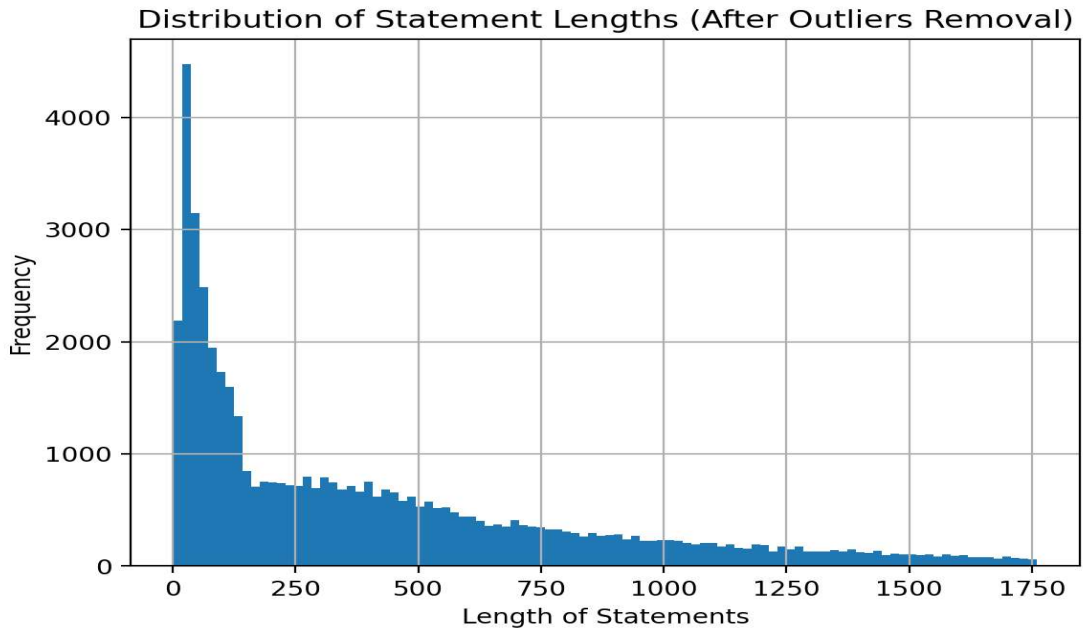


6. Distribution of Statement Lengths (Histogram) (Before Outliers)

Most statements are short, with frequency sharply dropping as length increases, indicating a highly skewed distribution.

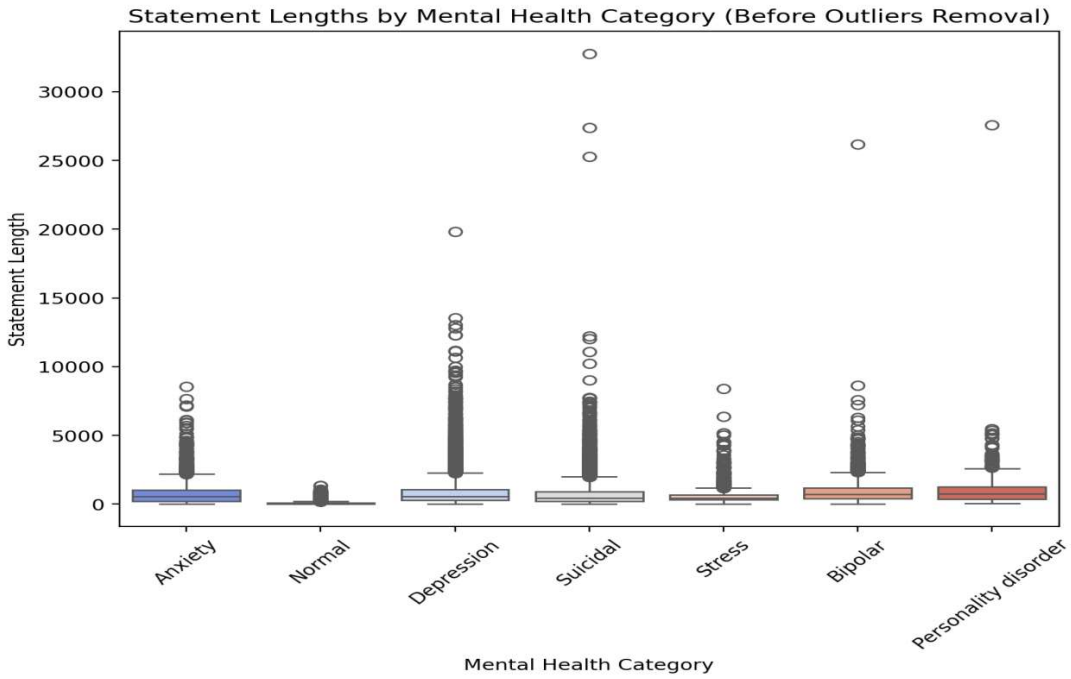


7. Distribution of Statement Lengths (Histogram) (After Outliers)

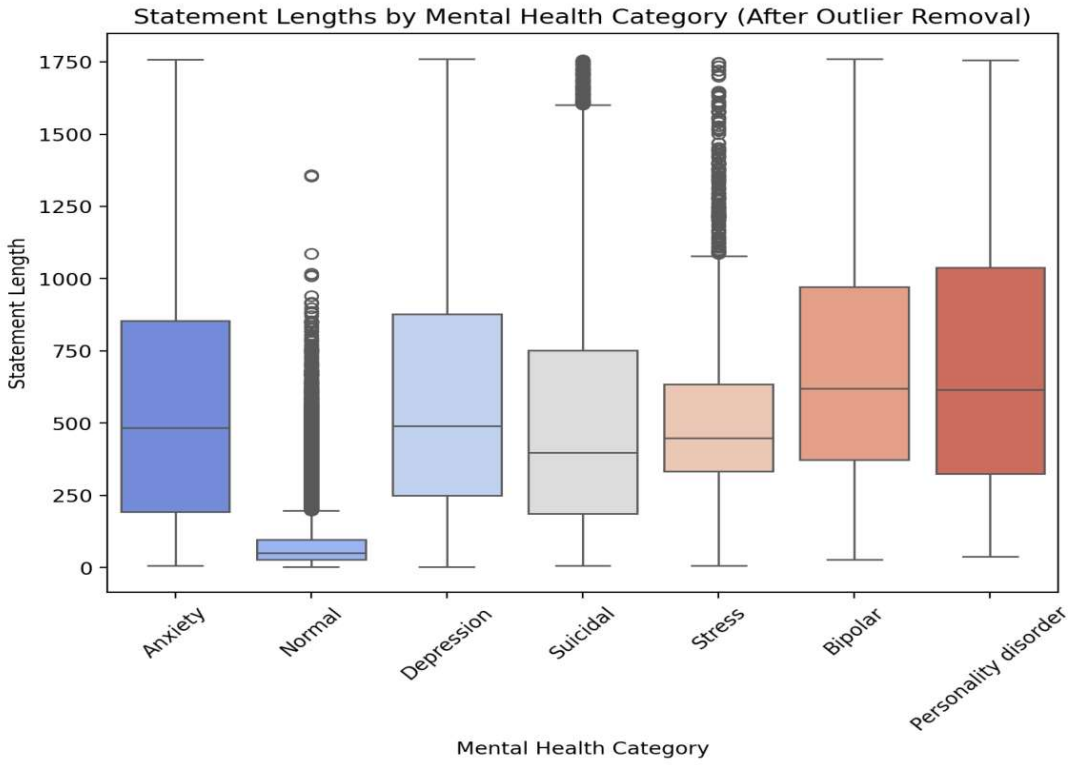


8. Box Plot: Statement Lengths by Mental Health Category (Before Outliers)

Most categories have outliers with long statements, particularly in "Suicidal" and "Depression", showing variability in emotional expression.

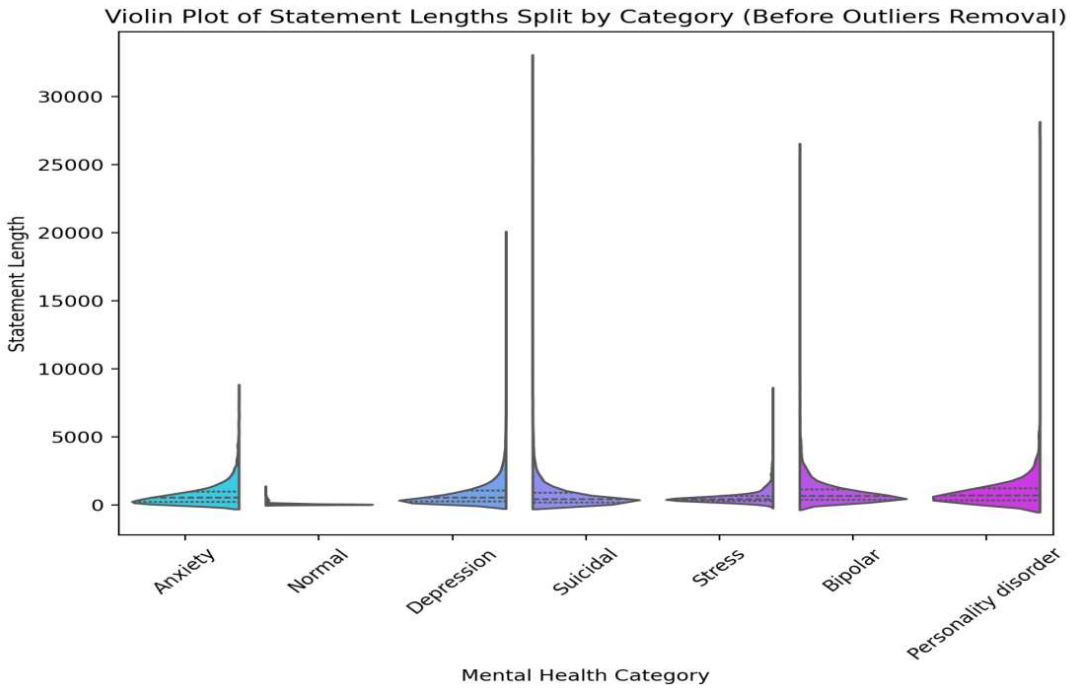


9. Box Plot: Statement Lengths by Mental Health Category (After Outliers)

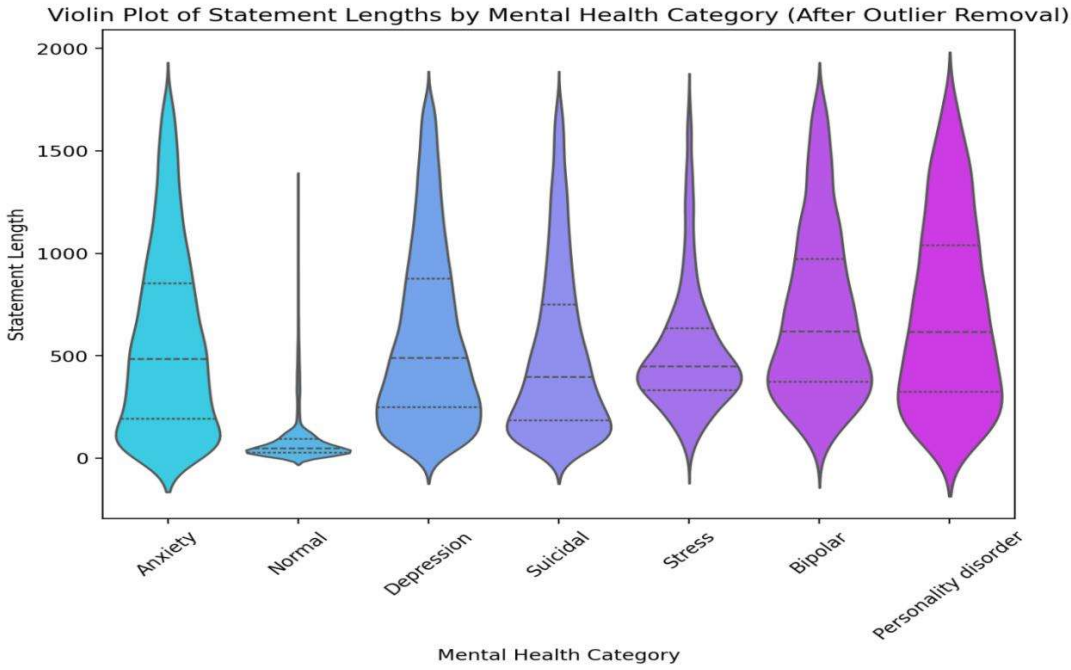


10. Violin Plot: Statement Lengths by Mental Health Category (Before Outliers)

Emotional categories exhibit wider and heavier tails in distribution, especially "Suicidal" and "Personality disorder", suggesting longer and more varied input lengths.

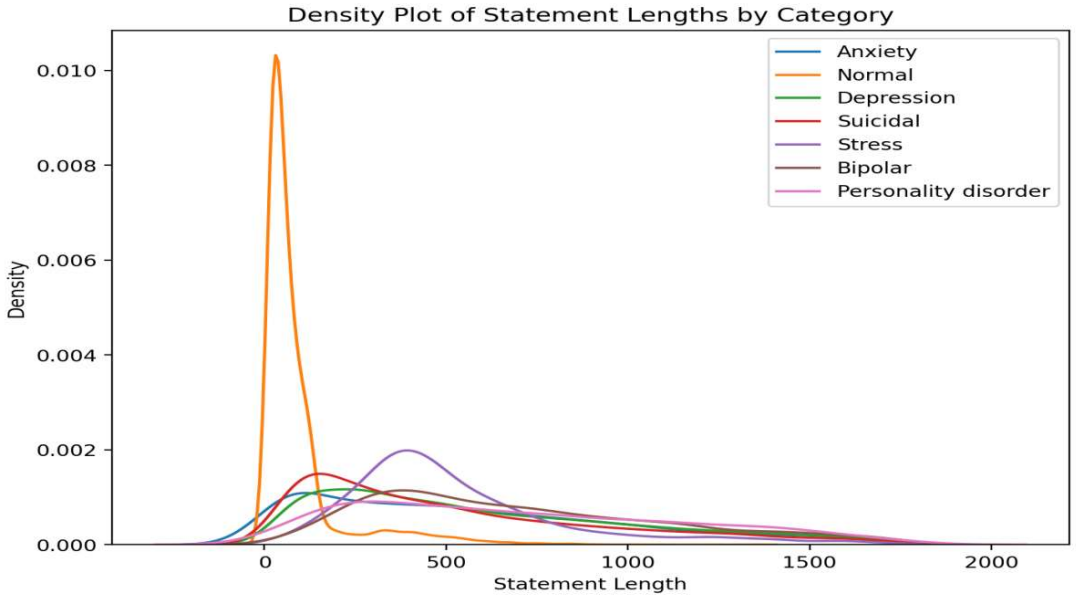


11. Violin Plot: Statement Lengths by Mental Health Category (After Outliers)



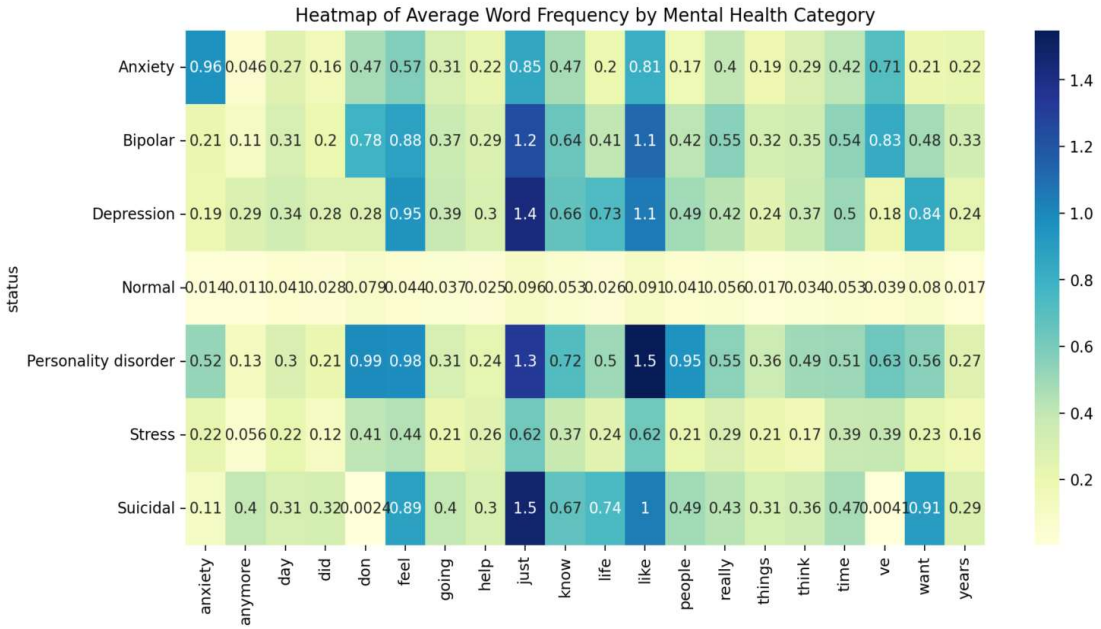
12. Density Plot of Statement Lengths by Category

"Normal" statements are significantly shorter than others, while "Depression", "Suicidal", and "Bipolar" show broader distributions of longer lengths.



13. Heatmap of Average Word Frequency by Mental Health Category

Emotionally expressive words like "just", "help", "feel", and "life" appear more in mental health categories than in the Normal class, suggesting strong lexical signals for classification.

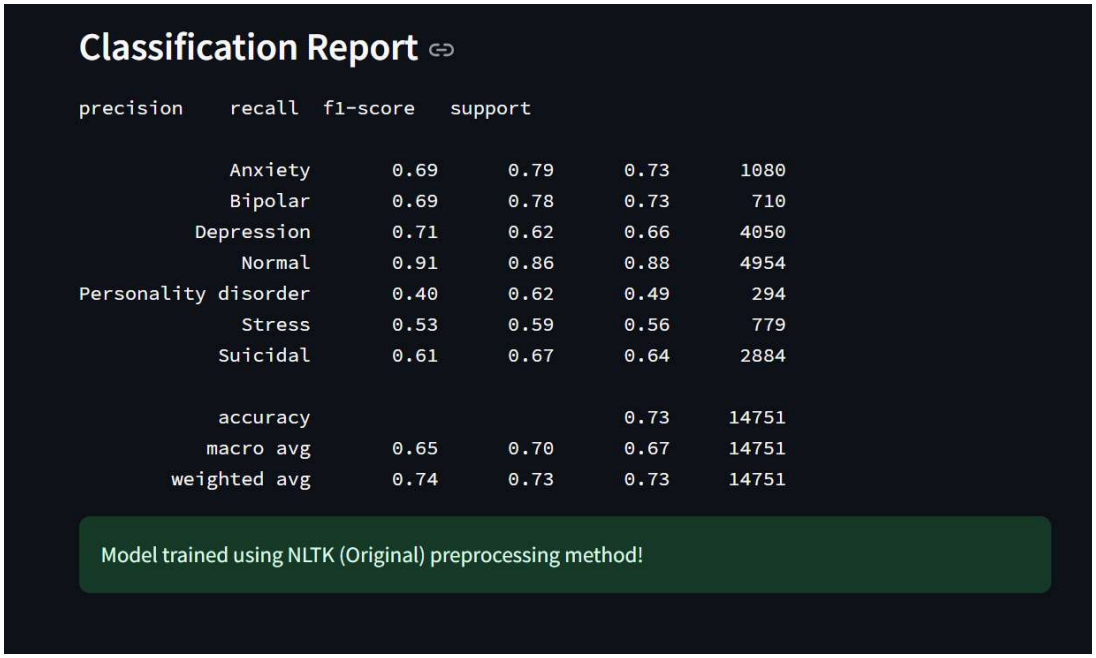


14. Classification Report (for 25k rows only)

- The model achieved an overall accuracy of 73%, with particularly strong performance in the Normal category (F1-score: 0.88), suggesting high reliability in distinguishing non-mental-health-related statements.
- Anxiety and Bipolar categories also performed well, each with an F1-score of 0.73, indicating the model's capability to detect these conditions with reasonable consistency. Depression and Suicidal showed moderate performance, with F1-scores of 0.66 and 0.64, respectively.

- Performance for Stress (F1-score: 0.56) and Personality disorder (F1-score: 0.49) was comparatively lower, likely due to fewer samples and overlapping language cues that made distinction challenging.

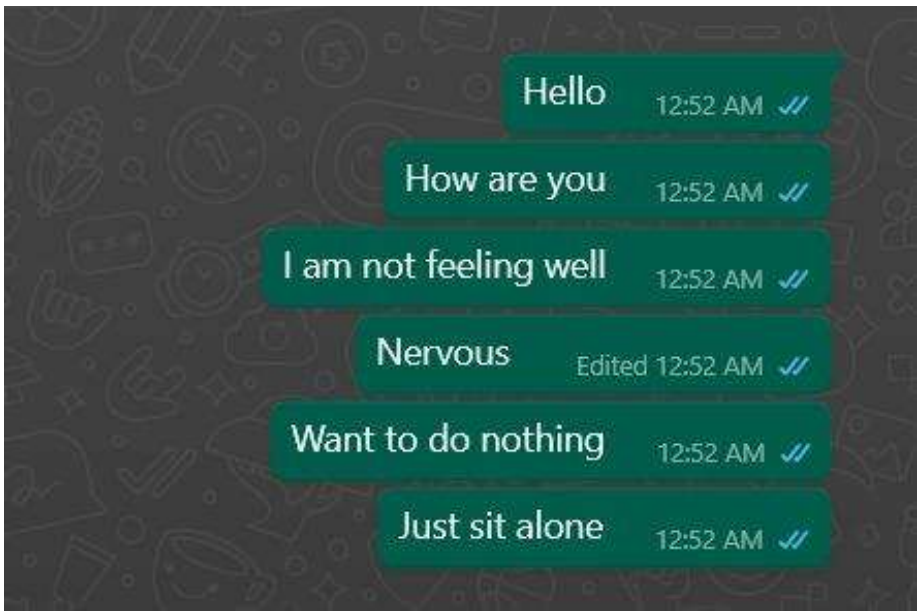
- The macro average F1-score of 0.67 reflects the model's relatively balanced performance across all categories, despite noticeable class imbalance.



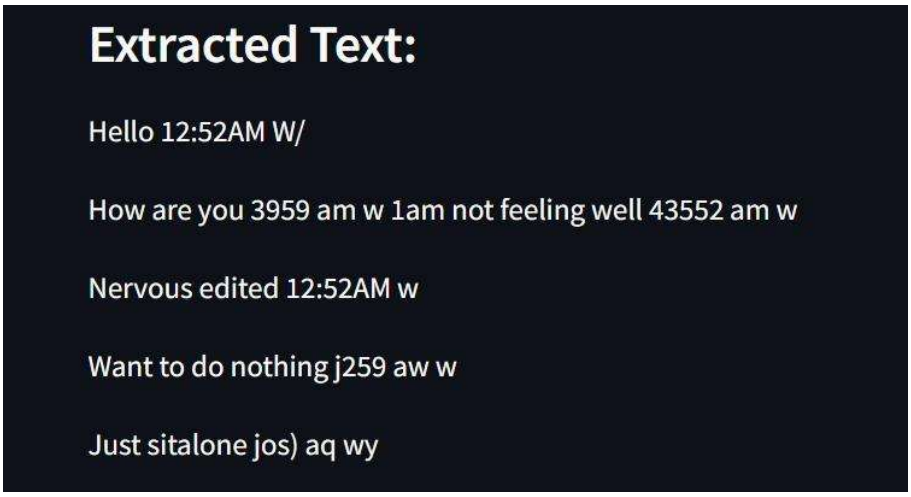
15. Working

A) Through OCR

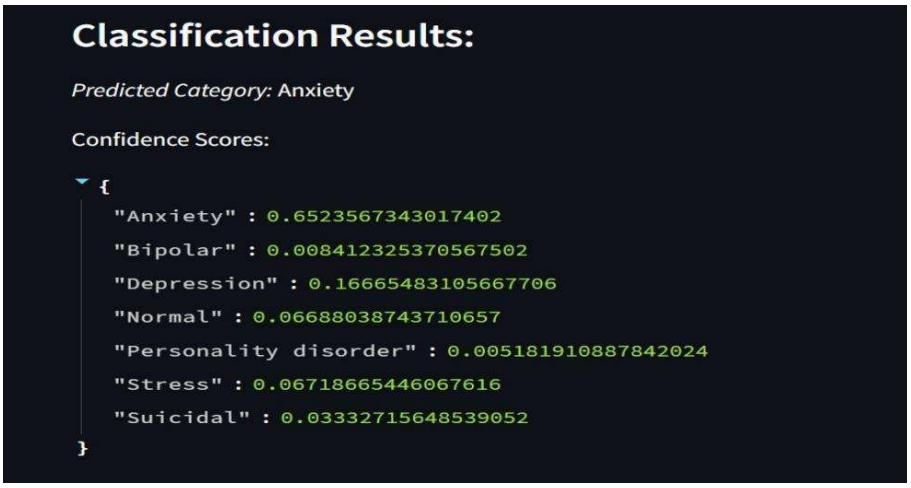
1) Input Data (WhatsApp Conversation Screenshot)



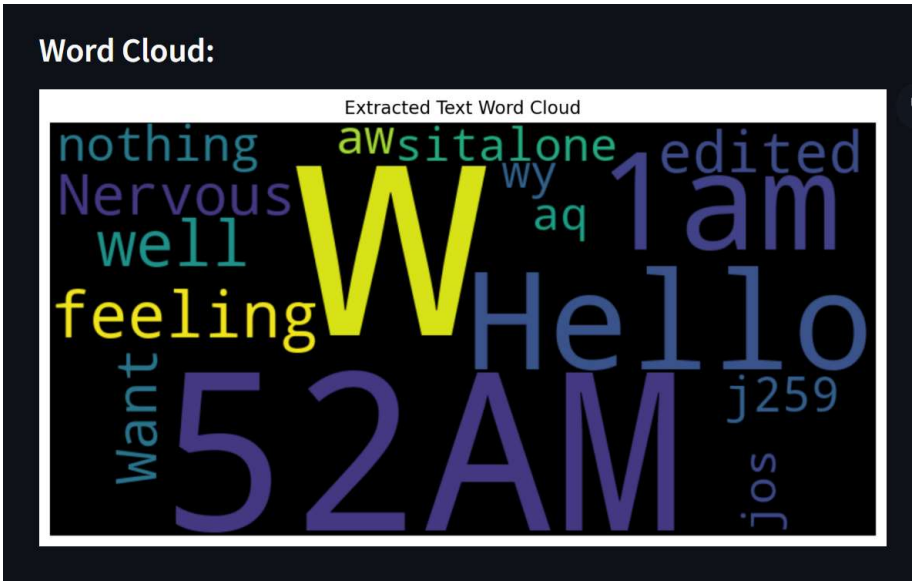
2) Text Extraction Results



3) Classification Results

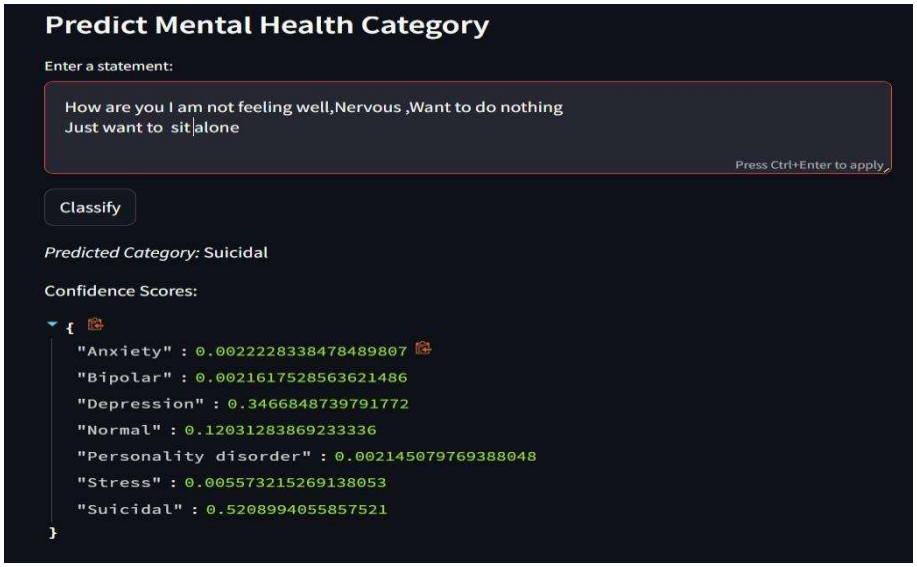


4) Word Cloud

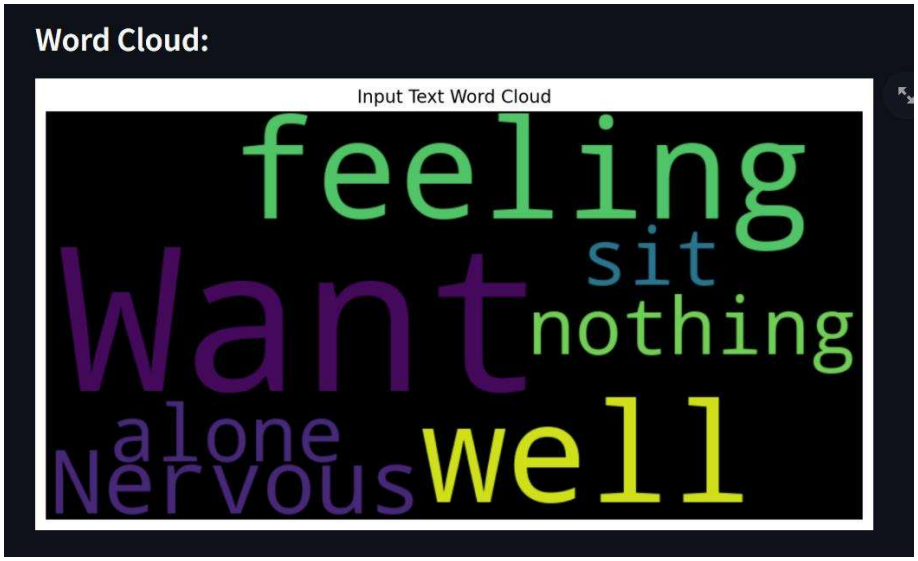


B) Through Text

1) Input Data and Classification results



2) Word Cloud



CONCLUSION

This Streamlit application provides a comprehensive solution for mental health text classification with several key features:

1) Multi-Input Support: The app accepts three types of inputs:
   a. CSV files containing mental health statements for batch processing
   b. Images containing text through OCR extraction
   c. Direct text input for quick classification

2) Advanced Text Processing: The application implements multiple preprocessing approaches:
    a. Traditional NLTK-based processing
    b. Enhanced spaCy linguistic analysis
    c. A combined approach leveraging both NLTK and spaCy

3) Robust Machine Learning Pipeline: The system includes:
    a. Comprehensive EDA with visualizations
    b. Outlier detection and removal
    c. Text vectorization with TF-IDF
    d. Class balancing using SMOTE
    e. SVM classification with probability outputs

4) Visual Analytics: The app provides multiple visualization tools:
    a. Distribution plots of text lengths
    b. Word clouds for different mental health categories
    c. Heatmaps of word frequencies
    d. Violin and box plots for comparative analysis

Entity Recognition: The system identifies mental health-related terms in the text using spaCy's linguistic capabilities.

Performance Optimization: The implementation includes parallel processing for efficient text preprocessing on larger datasets.

Future Work

While this research presents a promising approach to mental health text classification using NLP and SVM, future enhancements can make the system more robust, inclusive, and real-world applicable.

1. Leveraging Transformer-based Models

Incorporate advanced models like BERT, RoBERTa, or DistilBERT to capture deeper contextual and semantic meaning from mental health-related texts. These models significantly outperform traditional techniques in NLP tasks due to their attention mechanisms.

2. Cross-Language & Cultural Mental Health Analysis

Extend the system to support multilingual datasets to handle mental health discourse across different cultures and languages, improving inclusivity and global relevance.

3. Real-time Social Media Monitoring

Develop an AI-based real-time system to analyze user-generated content from platforms like Twitter, Reddit, and Instagram for early detection of mental health risks such as depression, anxiety, or suicidal ideation.

4. Emotion, Sentiment & Sarcasm Detection

Enhance the model to detect emotional states (e.g., sadness, anger, fear, joy) along with sarcasm, irony, or vague expressions, which are often present in mental health discourse. This provides a deeper understanding of user well-being and improves classification accuracy.

5. Personalized User Profiling & Recommendations

Build user profiles based on historical data and classification outcomes to offer tailored suggestions, such as mental health resources, support groups, or crisis helpline information.

6. Intelligent Conversational Agent

Deploy the model within a mental health chatbot or virtual assistant that offers empathetic conversation, mindfulness resources, and referrals to professional help when necessary.

7. Dataset Expansion & Generalization

Increase dataset size through web scraping or crowdsourcing, ensuring diversity and real-world noise. This will improve the model's generalizability across scenarios.

8. Optimization with GridSearch

Apply GridSearchCV for hyperparameter tuning of classifiers (like SVM or transformer-based heads), improving performance metrics like F1-score and recall, especially for underrepresented classes.

9. Continuous Learning via Feedback Loop

Implement a feedback mechanism where user responses and corrections help the model improve over time. This makes the system adaptive to evolving language patterns and user needs.

References

- Dataset: https://www.kaggle.com/datasets/suchintikasarkar/sentiment-analysis-for-mental-health/data
- Jurafsky, D., & Martin, J. H. (2023). Speech and Language Processing (3rd ed.). Pearson.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016)
- Coppersmith, G., Dredze, M., & Harman, C. (2014).
- Spacy NLP Documentation: https://spacy.io/usage