

# Analyzing Social Media Posts for Mental Health Disorder Detection

Submitted by

SOUMYADEEP NANDY (13000121033)

PRITHWISH SARKAR (13000121037)

SAGNIK MUKHOPADHYAY (13000121040)

ARKAPRATIM GHOSH (13000121058)

⟨ Group 29 ⟩

Final Year 8<sup>th</sup> Semester

⟨ June, 2025 ⟩

Submitted for the partial fulfillment for the degree of  
Bachelor of Technology in  
Computer Science and Engineering



Techno Main Salt Lake,  
EM 4/1, Salt lake, Sector - V, Kolkata - 700091

Department of Computer Science and Engineering  
Techno Main Salt Lake  
Kolkata - 700 091  
West Bengal, India

## **APPROVAL**

This is to certify that the project entitled "**Analyzing Social Media Posts for Mental Health Disorder Detection**" prepared by **SOUMYADEEP NANDY (13000121033)**, **PRITHWISH SARKAR (13000121037)**, **SAGNIK MUKHOPADHYAY (13000121040)** and **ARKAPRATIM GHOSH (13000121058)** be accepted in partial fulfillment for the degree of Bachelor of Technology in Computer Science and Engineering.

It is to be understood that by this approval, the undersigned does not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn thereof, but approves the report only for the purpose for which it has been submitted.

.....  
(Signature of the Internal Guide)

.....  
(Signature of the HOD)

.....  
(Signature of the External Examiner)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

## **ACKNOWLEDGEMENT**

We would like to express our sincere gratitude to our project guide in the department of Computer Science and Engineering. We are extremely thankful for the keen interest our guide took in advising us, for the books, reference materials and support extended to us.

Last but not the least we convey our gratitude to all the teachers for providing us the technical skill that will always remain as our asset and to all non-teaching staffs for the gracious hospitality they offered us.

Place: Techno Main Salt Lake

Date:

**SOUMYADEEP NANDY (13000121033)**

**PRITHWISH SARKAR (13000121037)**

**SAGNIK MUKHOPADHYAY (13000121040)**

**ARKAPRATIM GHOSH (13000121058)**

## Table of Content

<b>Abstract . . . . .</b>	<b>1</b>
<b>1 Introduction . . . . .</b>	<b>1</b>
1.1 Project Overview . . . . .	1
1.2 Project Purpose . . . . .	1
1.3 Technical Domain Specifications . . . . .	2
1.4 Business Domain Specifications . . . . .	3
1.5 Glossary / Keywords . . . . .	4
<b>2 Related Studies . . . . .</b>	<b>6</b>
<b>3 Problem Definition and Preliminaries . . . . .</b>	<b>9</b>
3.1 Context and Background . . . . .	9
3.2 Objective . . . . .	9
3.3 Challenges . . . . .	9
3.4 Scope . . . . .	10
3.5 Exclusions . . . . .	11
3.6 Assumptions . . . . .	11
<b>4 Proposed Solution . . . . .</b>	<b>12</b>
4.1 Special Contributions . . . . .	12
4.2 Reusable Components . . . . .	14
<b>5 Project Planning . . . . .</b>	<b>14</b>
5.1 Software Life Cycle Model . . . . .	14
5.2 Dependencies and Milestones . . . . .	16
5.3 Scheduling . . . . .	16
<b>6 Requirement Analysis . . . . .</b>	<b>19</b>
6.1 Requirement Matrix . . . . .	19
6.2 Requirement Elaboration . . . . .	19
6.2.1 Functional Requirements . . . . .	19
6.2.2 Non Functional Requirements . . . . .	20
<b>7 Design . . . . .</b>	<b>22</b>
7.1 Technical Environment . . . . .	22
7.2 Hierarchy of Modules . . . . .	24
7.3 Detailed Design . . . . .	26
7.3.1 Data Loading and Preprocessing . . . . .	26
7.3.2 Feature Extraction . . . . .	26
7.3.3 Model Training and Validation . . . . .	26
7.3.4 Prediction . . . . .	27
7.3.5 Testing and Deployment . . . . .	27
7.4 Emotion Detection Functionality . . . . .	29
7.5 Extract Text From Image . . . . .	30
7.6 Image Description . . . . .	31
7.7 Translation to English . . . . .	32
7.8 Prediction to Wellbeing Mapping . . . . .	34
7.9 Audio Mood Analysis . . . . .	35

<b>8 Implementation . . . . .</b>	<b>38</b>
8.1 Features From RM . . . . .	38
8.2 Code Details and Output . . . . .	39
8.2.1 Data Collection . . . . .	39
8.2.2 Data Preprocessing . . . . .	41
8.2.3 Logistic Regression Model for Classification . . . . .	43
8.2.4 Naive Bayes for Classification . . . . .	45
8.2.5 Support Vector Machine for Classification . . . . .	47
8.2.6 Random Forest for Classification . . . . .	49
8.2.7 XGBoost for Classification . . . . .	51
8.2.8 K Nearest Neighbours for Classification . . . . .	53
8.2.9 Long Short Term Memory based Classification . . . . .	55
8.2.10 Hyperparamter Tuning Using RandomizedSearchCV . . . . .	59
8.2.11 Ensemble Model . . . . .	68
<b>9 Results and Analysis . . . . .</b>	<b>82</b>
9.1 Classification Metrics and Confusion Matrix . . . . .	82
9.2 Results of Logistic Regression . . . . .	83
9.3 Results of Naive Bayes . . . . .	85
9.4 Results of Support Vector Machine . . . . .	87
9.5 Results of Random Forest . . . . .	89
9.6 Results of XGBoost . . . . .	91
9.7 Results of KNN . . . . .	93
9.8 Results of LSTM . . . . .	95
9.9 Results of Hyperparameter Tuning . . . . .	97
9.10 Results from Ensemble Model Training and Testing . . . . .	107
9.10.1 Logistic Regression and XGBoost . . . . .	107
9.10.2 Logistic Regression, SVM, Naive Bayes, XGBoost(meta-learner), LSTM . . . . .	110
9.10.3 Logistic Regression(meta-learner), SVM, Naive Bayes, XGBoost, LSTM . . . . .	115
<b>10 Conclusion . . . . .</b>	<b>120</b>
10.1 Project Benefits . . . . .	120
10.2 Future Scope for Improvements . . . . .	121
<b>11 References . . . . .</b>	<b>122</b>
<b>APPENDIX A - Prototype . . . . .</b>	<b>123</b>

# List of Figures

1	Iterative Waterfall Model . . . . .	16
2	Project Plan . . . . .	17
3	Gantt Chart . . . . .	18
4	Requirement Matrix . . . . .	19
5	Project Modules . . . . .	24
6	System Overview . . . . .	25
7	Model Workflow . . . . .	25
8	DFD Level 0 of the System . . . . .	27
9	DFD Level 1 of the System . . . . .	28
10	DFD Level 2 of Emotion Detection Functionality . . . . .	29
11	DFD Level 2 of Text Extraction from Image . . . . .	30
12	DFD Level 2 of Image Description . . . . .	31
13	DFD Level 2 of Translation to English . . . . .	33
14	DFD Level 2 of Prediction to Wellbeing Mapping . . . . .	34
15	DFD Level 2 of Audio Mood Analysis . . . . .	36
16	Features from Requirement Matrix . . . . .	38
17	Obtained Dataset . . . . .	39
18	Data Collection and Preprocessing . . . . .	42
19	Output for LSTM Epochs . . . . .	58
20	Accuracy v/s Epoch . . . . .	58
21	Loss v/s Epoch . . . . .	59
22	Confusion Matrix (Logistic Regression) . . . . .	84
23	ROC AUC (Logistic Regression) . . . . .	84
24	Confusion Matrix (Naive Bayes) . . . . .	86
25	ROC AUC (Naive Bayes) . . . . .	86
26	Confusion Matrix (SVM) . . . . .	88
27	ROC AUC (SVM) . . . . .	88
28	Confusion Matrix (Random Forest) . . . . .	90
29	ROC AUC (Random Forest) . . . . .	90
30	Confusion Matrix (XGBoost) . . . . .	92
31	ROC AUC (XGBoost) . . . . .	92
32	Confusion Matrix (KNN) . . . . .	94
33	ROC AUC (KNN) . . . . .	94
34	Confusion Matrix (LSTM) . . . . .	96
35	ROC AUC (LSTM) . . . . .	96
36	Classification Matrix after Hyperparameter Tuning (Logistic Regression) . . . . .	98

37	ROC AUC after Hyperparamter Tuning (Logistic Regression) . . . . .	98
38	Classification Matrix after Hyperparameter Tuning (KNN) . . . . .	100
39	ROC AUC after Hyperparameter Tuning (KNN) . . . . .	100
40	Classification Matrix after Hyperparameter Tuning (SVM) . . . . .	102
41	ROC AUC after Hyperparameter Tuning (SVM) . . . . .	102
42	Classification Matrix after Hyperparameter Tuning (Naive Bayes) . . . . .	104
43	ROC AUC after Hyperparameter Tuning (Naive Bayes) . . . . .	104
44	Result Comparison of the Algorithms . . . . .	105
45	Result Comparison after Hyperparameter Tuning . . . . .	106
46	Confusion Matrix for Ensemble Model(LR + XGB) . . . . .	108
47	ROC AUC Curve for Ensemble Model(LR + XGB) . . . . .	109
48	Comparison of results from LR, XGB and Ensemble Model . . . . .	110
49	Confusion Matrix for ensemble model (XGB - meta learner) . . . . .	111
50	ROC AUC Curve for multiple models in ensemble model (XGB - meta learner)	112
51	Comparison of results with Ensemble Model (XGB - meta learner) . . . . .	113
52	Confusion Matrix for Ensemble Model (LR - meta learner) . . . . .	117
53	ROC AUC for Ensemble Model (LR - meta learner) . . . . .	118
54	Comparison of results with Ensemble Model (LR - meta learner) . . . . .	119
55	Sequence Diagram of the Application . . . . .	143
56	Website with all options . . . . .	143
57	Entering Text for classification . . . . .	144
58	Text Classification Result . . . . .	144
59	Upload Image . . . . .	145
60	Image Classification Result . . . . .	145
61	Upload Video . . . . .	146
62	Video Classification Result . . . . .	146
63	Reddit User Analysis . . . . .	147
64	Result from Reddit Posts Analysis . . . . .	147
65	Twitter User Analysis . . . . .	148
66	Result from Twitter Posts Analysis . . . . .	148
67	Text Classification and Retrain Model . . . . .	155
68	Result from Text Classification and Model Retraining . . . . .	155
69	Reddit Analysis with Model Retraining . . . . .	156
70	Result from Reddit Analysis and Model Retraining . . . . .	156
71	Result from the emotion analysis of facial expression . . . . .	160
72	Generate Image Caption . . . . .	162

## Abstract

This project aims to analyze social media posts for early detection of mental health disorders. Specifically, it focuses on using machine learning and deep learning algorithms to classify social media text data based on potential mental health issues. The problem lies in efficiently detecting patterns that indicate mental health conditions within large, unstructured datasets. Using methods like Logistic Regression , XGboost the project seeks to enhance the accuracy of detecting mental health concern with a specific probability. Expected results include a robust classifier capable of distinguishing between different mental health concerns with high accuracy. This project could provide a valuable tool for mental health monitoring on social platforms.

## 1 Introduction

### 1.1 Project Overview

Mental health has become a critical global issue, with millions of people affected by various mental disorders, including depression, anxiety, and others. With the increasing use of social media, these platforms have emerged as spaces where people often express their emotions and struggles, sometimes unknowingly revealing signs of mental health challenges. This project focuses on analyzing social media posts to detect mental health disorders using advanced machine learning techniques. By examining language patterns and contextual usage in text data, this project aims to classify posts that potentially indicate mental health issues. Such detection can facilitate early intervention and help direct individuals to appropriate mental health services.

### 1.2 Project Purpose

The main goal of this project is to leverage machine learning and deep learning to identify the best model and create a web application capable of identifying signs of mental health disorders from text, images and social media posts by giving username as input. This aligns with a broader goal of using technology to address public health concerns by enabling early detection through data analysis. Specifically, we use classification models such as Logistic Regression and XGboost to predict mental health issues based on text patterns. The project also addresses the technical challenges of processing large datasets and optimizing algorithms for accurate classification.

## 1.3 Technical Domain Specifications

This project falls within the intersection of natural language processing (NLP) and machine learning (ML), leveraging techniques such as text vectorization, and classification algorithms. Here are the key technical domain specifications:

- **Hardware :** The project does not require specialized hardware beyond a standard machine with adequate processing power. However, for larger datasets or complex model training, a machine equipped with a GPU (Graphics Processing Unit) could significantly reduce processing time. The project can be run on any system with at least 8GB of RAM and a multi-core processor.
- **Operating System :** The project is cross-platform and can be developed and executed on any modern operating system, including:
  - Windows 10/11
  - macOS
  - Linux distributions (Ubuntu, Linux Mint, etc.) A Linux-based system is often preferred in machine learning projects due to its stability and support for tools like TensorFlow, PyTorch, and other libraries used for model training.
- **Software :**
  - **Programming Languages :** Python 3.x will be the primary programming language, given its extensive libraries for machine learning, data analysis, and NLP.
  - **Libraries / Frameworks :**
    - \* **Scikit-learn :** Used for machine learning algorithms (k-NN, SVM) and model evaluation.
    - \* **Pandas :** For data manipulation and preprocessing.
    - \* **NumPy :** To handle large arrays and matrices, which are crucial for efficient numerical computations.
    - \* **NLTK and spaCy :** For text preprocessing and natural language understanding.
    - \* **Matplotlib and Seaborn :** For data visualization.
  - **Development Environment :**
    - \* **Jupyter Notebook :** For interactive development, experimentation, and visualization.
    - \* **Anaconda :** A distribution that simplifies package management and deployment.

- \* **Google Colab** : For cloud-based execution when working with larger datasets or GPU-based model training.

## 1.4 Business Domain Specifications

From a business perspective, this project holds significant value across various sectors, particularly those that intersect with mental health monitoring, public health awareness, and social media governance. With the increasing prevalence of mental health issues globally, organizations within these industries are searching for innovative solutions to mitigate the growing mental health crisis. Leveraging machine learning for early detection of mental health disorders from social media data can revolutionize how mental health is addressed at both individual and societal levels. Below is a detailed exploration of how this project can impact different business domains:

- **Mental Health Services** : Mental health service providers—such as hospitals, therapy centers, and private practices—can greatly benefit from machine learning models capable of identifying early signs of mental health issues from social media data. In the traditional mental health setting, early detection of disorders like depression or anxiety often relies on self-reporting or clinical assessments, which may come too late in the progression of the disorder. By analyzing patterns in social media posts, these services can adopt a more proactive approach, reaching out to potential patients earlier in their mental health journey.
- **Social Media Platforms** : Social media platforms like Twitter, Facebook, Instagram, and others play an integral role in the public's expression of thoughts and feelings, including mental health struggles. These platforms face increasing pressure to safeguard the well-being of their users. This project's machine learning models can enable these companies to offer valuable services to users while adhering to ethical standards.
- **Public Health Organizations** : Public health organizations are tasked with monitoring and improving the mental well-being of the population on a large scale. For these organizations, access to real-time data from social media can provide a comprehensive view of the mental health landscape, identifying emerging trends and enabling data-driven interventions. Understanding how mental health is being discussed online can help public health organizations create more effective mental health awareness campaigns. Tailored messaging based on the language patterns identified by the model can lead to better engagement with individuals suffering from mental health issues.

## 1.5 Glossary / Keywords

Term	Definition
Machine Learning (ML)	A subset of artificial intelligence (AI) that enables computers to learn from data and make predictions or decisions without explicit programming.
Natural Language Processing (NLP)	A branch of artificial intelligence focused on the interaction between computers and humans through natural language, including tasks like text analysis.
Support Vector Machines (SVM)	A supervised learning algorithm used for classification or regression tasks, focusing on finding a hyperplane that best separates different classes.
Vectorization	The process of converting textual data into numerical form (such as a vector) so that it can be used as input for machine learning models.
Classifier	A machine learning model or algorithm that categorizes or labels data points into predefined classes.
Mental Health Disorder	A wide range of conditions that affect mood, thinking, and behavior, including depression, anxiety, schizophrenia, etc.
Data Preprocessing	The process of preparing raw data for analysis by cleaning, normalizing, and transforming it into a usable format for machine learning models.
Cross-validation	A model validation technique used to assess how well a model performs by dividing data into training and testing sets multiple times for better accuracy.
Precision	In the context of classification, precision refers to the accuracy of positive predictions, calculated as the ratio of true positives to the sum of true and false positives.
Recall	In classification, recall measures the ability of a model to identify all relevant instances within a dataset, calculated as the ratio of true positives to the sum of true positives and false negatives.
PRAW	PRAW (Python Reddit API Wrapper) is a Python library that provides a simple interface to interact with Reddit's API, allowing developers to easily access, retrieve, and analyze Reddit data, such as posts, comments, and user information.
TesseractOCR	TesseractOCR is an open-source Optical Character Recognition (OCR) engine that extracts text from images with high accuracy; it is widely used for various applications like scanning documents and digitalizing printed text.

## ASMPFMHDD

<b>Term</b>	<b>Definition</b>
Depression	There is a difference between depression and mood swings or short-lived emotional reactions to daily experiments; A mental state causing painful symptoms adversely disrupts normal activities (e.g., sleeping).
Anxiety	Several behavioral disturbances are associated with anxiety disorders, including excessive fear and worry. Severe symptoms cause significant impairment in functioning cause considerable distress. Anxiety disorders come in many forms, such as social anxiety, generalized anxiety, panic, etc.
Bipolar Disorder	An alternating pattern of depression and manic symptoms is associated with bipolar disorder. An individual experiencing a depressive episode may feel sad, irritable, empty, or lose interest in daily activities. Emotions of euphoria or irritability, excessive energy, and increased talkativeness can all be signs of manic depression. Increased self-esteem, decreased sleep need, disorientation, and reckless behavior may also be signs of manic depression.
Post-Traumatic Stress Disorder (PTSD)	In PTSD, persistent mental and emotional stress can occur after an injury or severe psychological shock, characterized by sleep disturbances, constant vivid memories, and dulled response to others and the outside world. People who re-experience symptoms may have difficulties with their everyday routines and experience significant impairment in their performance.
DeepFace	DeepFace is a Python library for deep learning-based facial recognition and attribute analysis. It supports several pre-trained models and simplifies face recognition tasks, making it suitable for various applications in image analysis.
Transformers Module	The Transformers module in Python, developed by Hugging Face, is a library for natural language processing (NLP) tasks like text classification, translation, and summarization, using state-of-the-art models like BERT and GPT.
Gemini 1.5 Flash	Gemini 1.5 Flash is a cutting-edge AI model developed by Google, capable of performing advanced generative and analytical tasks across text, image, and other modalities.
FFmpeg	FFmpeg is a multimedia framework used for encoding, decoding, transcoding, streaming, and manipulating audio and video files, supporting a wide range of formats and codecs.
Hyperparameter Tuning	Hyperparameter tuning involves selecting the best parameters for a machine learning model to optimize its performance on a given task, using methods like grid search or random search.
Logistic Regression	Logistic Regression is a statistical method used for binary and multiclass classification tasks by modeling the relationship between input variables and probabilities using a sigmoid function.

<b>Term</b>	<b>Definition</b>
K-Nearest Neighbors (KNN)	KNN is a simple, instance-based machine learning algorithm used for classification and regression tasks, predicting outcomes based on the majority class of its k nearest neighbors.
Naive Bayes	Naive Bayes is a probabilistic classification algorithm based on Bayes' Theorem, assuming independence between features, making it efficient for text classification and other applications.
Random Forest	Random Forest is an ensemble machine learning algorithm that combines multiple decision trees to improve accuracy and prevent overfitting, commonly used for classification and regression tasks.
XGBoost	XGBoost (Extreme Gradient Boosting) is a powerful ensemble machine learning algorithm based on decision trees, designed for speed and performance in classification and regression tasks.
Long Short-Term Memory (LSTM)	LSTM is a type of recurrent neural network (RNN) architecture capable of learning and remembering patterns over long sequences of data, widely used in time-series analysis and NLP.

## 2 Related Studies

The intersection of social media analytics and mental health research has received increasing attention in recent years, leading to several important studies that highlight the potential for early detection and intervention. This section reviews key findings from various studies, emphasizing the relevance and applicability of social media data for identifying mental health disorders.

One of the seminal works in this domain is by Choudhury et al. (2013), who explored the predictive capabilities of social media content in identifying depression. They analyzed Twitter data and discovered that specific linguistic patterns, such as the use of negative emotion words, correlated strongly with self-reported depressive symptoms. This study demonstrated that social media could serve as a valuable resource for predicting mental health conditions, offering a potential tool for clinicians and researchers alike [2].

Similarly, Guntuku et al. (2017) conducted an integrative review that focused on detecting mental illness through social media. Their work synthesized various approaches and methodologies used in the field, providing insights into the effectiveness of different machine learning algorithms and sentiment analysis techniques. They found that social media platforms are rich sources of data that can reveal critical information about users' mental health, advocating for

the development of robust systems to analyze this data effectively [4].

A systematic review by Mathur et al. (2022) further emphasized the significance of mental health classification on social media. They examined various studies that utilized machine learning techniques for mental health detection, highlighting the success of these models in identifying depression and anxiety based on user-generated content. Their findings reinforced the notion that social media can be leveraged not only for individual assessments but also for broader epidemiological studies to understand population mental health trends [5].

In addition, Nadeem (2016) contributed to the discussion by investigating depression identification on Twitter. The study focused on developing algorithms that could discern emotional cues in tweets, indicating a user's mental state. The findings revealed that simple text analysis could lead to significant improvements in identifying at-risk individuals, further validating the potential of social media data in mental health monitoring [6].

Research by AlSagri and Ykhlef (2020) introduced a machine learning-based approach specifically for depression detection on Twitter. Their study incorporated both content and activity features, demonstrating that a combination of linguistic and behavioral analysis could enhance the accuracy of depression identification. This work illustrated the multifaceted nature of social media data and its ability to capture not just what users say but also how they interact online [1].

In a more recent study, Vaishnavi et al. (2022) investigated the application of various machine learning algorithms for predicting mental health illnesses. They found that certain algorithms outperformed others in classifying mental health conditions based on social media posts. This study provided a comparative analysis that could inform future research directions, emphasizing the importance of algorithm selection in the context of mental health detection [9].

Lastly, Safa et al. (2023) presented a roadmap for future development in predicting mental health using social media. Their work highlighted the ongoing challenges in the field, including ethical considerations and the need for improved data privacy measures. They emphasized that while social media offers rich data for mental health analysis, researchers must approach this opportunity with a strong ethical framework to ensure user safety and data security [7].

The paper titled "Single classifier vs. ensemble machine learning approaches for mental health prediction" (PMC9810771) explores the use of various machine learning techniques to predict mental health issues. Specifically, the study compares the performance of single classifiers (such as Logistic Regression, Gradient Boosting, Neural Networks, K-Nearest Neighbors, and

Support Vector Machine) with ensemble machine learning approaches, which combine multiple classifiers to improve prediction accuracy. The study is based on a dataset of survey responses collected by Open Sourcing Mental Illness (OSMI), and the goal is to classify mental health issues based on these responses. The authors also compare newer machine learning techniques like Extreme Gradient Boosting (XGBoost) and Deep Neural Networks (DNN). The results indicate that Gradient Boosting achieved the highest accuracy (88.80%), followed by Neural Networks (88.00%). The ensemble classifier, which combined multiple models, achieved an accuracy of 85.60%. Overall, the paper demonstrates that machine learning techniques, particularly ensemble models, show promise for automated prediction of mental health problems, providing valuable insights for early diagnosis and intervention in mental health care [3].

The paper titled "Ensemble of hybrid model based technique for early detecting of depression based on SVM and neural networks" (DOI: 10.1038/s41598-024-77193-0) presents a novel approach for early detection of depression using machine learning techniques. The study proposes an ensemble hybrid model combining Support Vector Machines (SVM) and Multilayer Perceptrons (MLP) to improve depression prediction accuracy. The DeprMVM hybrid model serves as a meta-learner, where the SVM and MLP networks act as level-0 learners. The model addresses class imbalance by applying the Synthetic Minority Over-sampling Technique (SMOTE) and cluster sampling, which improves detection accuracy and reduces the risk of overfitting. The study finds that the ensemble approach achieved an accuracy of 99.39% and an F1-score of 99.51%, outperforming previous models in depression detection. This paper highlights the potential of ensemble hybrid models for early detection of depression and their applicability in mental health care [8].

These studies collectively underscore the growing body of evidence supporting the integration of social media analytics and machine learning for mental health detection. They provide a solid foundation for the current project, which aims to enhance existing methodologies and develop a predictive model for identifying mental health disorders from social media posts.

## 3 Problem Definition and Preliminaries

### 3.1 Context and Background

Mental health disorders have become a significant public health concern worldwide. The World Health Organization (WHO) estimates that approximately 1 in 8 people globally experience mental health disorders, which encompass conditions such as depression, anxiety, bipolar disorder, and post-traumatic stress disorder (PTSD). The rise of social media platforms has changed how individuals express their mental health struggles, share experiences, and seek support. Posts on platforms like Reddit and Twitter provide a wealth of data reflecting real-time sentiments, issues, and conversations surrounding mental health. However, this vast amount of unstructured textual data presents challenges in effectively identifying and categorizing specific mental health disorders.

### 3.2 Objective

The primary objective of this project is to develop a robust system that can automatically analyze social media posts, specifically from Reddit and Twitter, to detect various mental health disorders. By leveraging Natural Language Processing (NLP) techniques and machine learning algorithms, this project aims to:

- **Classify Posts :** Accurately classify social media posts based on the type of mental health disorder mentioned, including but not limited to depression, anxiety, bipolar disorder, and PTSD.
- **Data Driven Insights :** Provide valuable insights into the prevalence and expression of mental health issues on social media, helping researchers, mental health professionals, and policymakers understand trends and patterns.

### 3.3 Challenges

- **Data Variability :** Social media posts can vary significantly in structure, style, and length. Users may employ slang, abbreviations, and informal language, making it difficult for algorithms to accurately interpret and classify posts.
- **Imbalanced Data :** Certain mental health issues may be underrepresented in social media discussions, leading to an imbalanced dataset. This imbalance can adversely affect model training and performance, making it harder to detect less frequent disorders.
- **Cultural and Contextual Nuances :** Mental health perceptions and discussions can vary across different cultures and contexts. The model needs to account for these nuances to avoid misclassification and provide accurate insights.

- **Privacy and Ethical Considerations :** Analyzing social media data raises ethical concerns regarding user privacy. It is crucial to handle sensitive information responsibly and comply with data protection regulations.

### 3.4 Scope

The scope of the project "Analyzing Social Media Posts for Mental Health Disorder Detection" delineates the specific aspects that will be covered, the methodologies employed, and the boundaries within which the research and analysis will occur. The project aims to harness the potential of machine learning and natural language processing (NLP) techniques to analyze social media sentiment and its correlation with mental health disorders, focusing specifically on a Reddit Dataset sourced using Python Reddit API Wrapper. This dataset contains user-generated content that reflects various emotional states, making it a valuable resource for this analysis.

- **Dataset Selection and Characteristics**

The primary data source for this project is the top textual posts from Reddit. This dataset includes posts that are labeled with mental health issues (Normal, Anxiety, Depression, Bipolar, PTSD). The selection of this dataset is pivotal, as it encapsulates a wide range of mental health-related discussions expressed by individuals on social media. Key characteristics of the dataset include:

- **User Anonymity :** To respect user privacy and adhere to ethical standards, the dataset does not contain personally identifiable information (PII) about the Twitter users. This ensures compliance with data protection regulations while allowing for robust analysis.

- **Analysis Objectives**

The project will focus on several key objectives:

- **Correlation with Mental Health Issues :** The analysis will explore the correlation between identified sentiments and specific mental health disorders, thereby contributing to the understanding of how social media discourse reflects mental health challenges.
- **Trend Analysis :** By analyzing sentiment trends over time, the project seeks to identify patterns in public discourse surrounding mental health, including any potential spikes in negative sentiments during particular events or crises.

- **Methodologies :** The project will employ various methodologies to achieve its objectives, including:

- **Data Processing** : Cleaning and preparing the dataset to ensure that it is suitable for analysis. This includes tasks such as removing noise (e.g., URLs, hashtags), tokenization, and normalization of text.
- **Feature Extraction** : Utilizing techniques like Term Frequency-Inverse Document Frequency (TF-IDF) to convert textual data into numerical representations suitable for machine learning algorithms.
- **Machine Learning Techniques** : Implementing various machine learning algorithms, including Logistic Regression and XGboost to classify posts and evaluate their performance based on accuracy, precision, recall, and F1 score.
- **Data Visualizations** : Employing visualization tools to present findings clearly, including heat maps for confusion matrices and ROC AUC Curve.

### 3.5 Exclusions

In delineating the boundaries of the project "Analyzing Social Media Posts for Mental Health Disorder Detection," it is crucial to specify what is excluded from the scope of this research to maintain a clear focus on the primary objectives. This project will not encompass the direct collection or real-time monitoring of Twitter data via the Twitter API, as it is solely reliant on the pre-existing Twitter sentiment dataset obtained from Kaggle. Therefore, any analysis involving the dynamic aspects of social media engagement, such as real-time sentiment shifts in response to current events or trending topics, will be outside the project's purview. Furthermore, the study will not address the technicalities of Reddit's platform-specific features, such as hashtags, user mentions, or reposts, subreddits in detail, as these elements are not central to the primary research focus on mental issue classification of individual users. While the project aims to analyze posts surrounding mental health, it will also include mapping the mental issue with mental wellbeing. Additionally, the research will not explore the ethical implications of data ownership or the responsibilities of social media platforms regarding user-generated content, as the focus will be primarily on data analysis techniques and outcomes rather than the broader ethical landscape.

### 3.6 Assumptions

In the context of the project "Analyzing Social Media Posts for Mental Health Disorder Detection," several key assumptions underpin the research framework and methodologies employed. Firstly, it is assumed that the Reddit dataset obtained using PRAW is representative of broader social media discourse regarding mental health issues, capturing a diverse range of sentiments expressed by users on the platform. This assumption is critical as it establishes the foundation

for analyzing sentiment trends and their potential correlations with various mental health disorders. Secondly, it is presumed that the posts recorded in the dataset accurately reflect the users' true emotions and perspectives at the time of posting, thereby providing valid data for analysis. Furthermore, it is assumed that the textual data within the dataset can be effectively processed and interpreted through natural language processing (NLP) techniques, allowing for the accurate classification of sentiments and identification of patterns. Another assumption is that the selected machine learning algorithms, including Logistic Regression and XGboost, will perform optimally with the provided data, leading to reliable and interpretable results regarding sentiment analysis and mental health correlations. Additionally, it is assumed that the sentiments expressed in social media posts can serve as a valid proxy for understanding public perceptions of mental health, enabling insights into societal attitudes and the potential stigmatization associated with these disorders. The project also assumes that the cleaning and preprocessing steps applied to the data will sufficiently prepare the dataset for analysis, minimizing noise and irrelevant information that could skew the results. Lastly, it is presumed that the ethical considerations surrounding the use of publicly available social media data have been adequately addressed, ensuring that the research adheres to relevant ethical standards and does not compromise user privacy or data integrity. These assumptions serve as the bedrock for the project's analytical framework, guiding the research processes and interpretations that follow.

## 4 Proposed Solution

The proposed work centers around "Analyzing Social Media Posts for Mental Health Disorder Detection," leveraging advanced data analytics and machine learning techniques to provide insights into the sentiments expressed in social media discussions related to mental health issues. This research is significant due to the increasing prevalence of mental health disorders globally and the role social media plays in shaping public perception and discourse around these issues. The core objective of this project is to develop a systematic approach to classify sentiments in tweets, thereby enabling better understanding and awareness of mental health conditions through social media analysis. Below, I outline the specific contributions and reusable components deployed in this project.

### 4.1 Special Contributions

- **Dataset Acquisition and Preparation :** The initial step involved sourcing a high-quality dataset from Kaggle, specifically the Twitter sentiment dataset. This dataset comprises user-generated tweets containing sentiments related to various mental health issues, serving as the primary data source for analysis. The preparation phase included extensive data

cleaning and preprocessing, wherein missing values were handled, duplicate entries removed, and text normalization performed. This crucial step ensured the dataset's integrity and suitability for subsequent analysis, allowing for a more accurate representation of sentiments.

- **Text Vectorization :** To enable machine learning models to interpret textual data, TF-IDF was implemented. This approach involved converting posts into a numerical format by creating a matrix representation of word frequencies across the dataset. The Scikit-learn library was instrumental in this process, offering functions for text vectorization and feature extraction. The reusable components for text preprocessing and vectorization were packaged into functions, allowing for easy application to future datasets or similar projects.
- **Implementation of Machine Learning Algorithms :** The project employed several machine learning algorithms, focusing primarily on Logistic Regression and XGboost for mental health classification. The Logistic Regression algorithm was chosen for its simplicity and effectiveness in classifying data points based on proximity in the feature space. In addition to Logistic Regression, XGboost was implemented due to its robustness in handling high-dimensional data and multi class classification tasks.
- **Model Evaluation :** Comprehensive evaluation metrics were employed to assess the performance of the machine learning models. Metrics such as accuracy, precision, recall, and F1-score were calculated to provide a holistic view of model effectiveness in classifying sentiments related to mental health. This evaluation process not only demonstrated the models' capabilities but also highlighted areas for improvement, providing a foundation for future iterations of the project. The evaluation framework, including metrics calculations and visualization, was designed as reusable components to streamline future model assessments.
- **Insights and Recommendations :** A critical aspect of this project is the generation of actionable insights based on the analysis of social media sentiments. The findings from the classification can inform mental health professionals, researchers, and policymakers about public sentiment trends, potential stigma associated with mental health issues, and the effectiveness of awareness campaigns. Recommendations for mental health awareness strategies can be derived from understanding how sentiments vary across different demographics and regions. This interpretative analysis, combined with quantitative results, contributes valuable knowledge to the ongoing conversation about mental health in society.

- **Documentation and Reproducibility :** To enhance the usability and impact of the project, thorough documentation was maintained throughout the research process. This documentation includes detailed explanations of the methodologies employed, code snippets, and instructions for reproducing the results. The aim is to ensure that the components developed in this project can be easily utilized by other researchers and practitioners in the field. By documenting the code and methodologies, I am contributing to the open-source community, allowing for collaborative improvements and innovations in mental health sentiment analysis.

## 4.2 Reusable Components

- **Data Collection Functions :** Modular functions designed for data collection, which can be reused across different platforms.
- **Data preprocessing Module :** A component that does data cleaning to remove the duplicates and empty rows and add a separate column for cleaned texts. This formatted dataset is then used further with TF-IDF to create a numerical matrix to be fed to the machine learning algorithms.
- **Machine and Deep Learning Model Functions :** Functions for implementing Logistic Regression, Naive Bayes, Support Vector Machine, Random Forest, XGboost, Long Short Term Memory algorithm allowing for easy retraining on varying datasets. These also features various evaluation metrics, making it easy to assess different models' performances.
- **Deployment Function :** A separate function that has the main python file for creating web based application on Streamlit Cloud. This also includes the requirements and package dependencies for deploying the application.

## 5 Project Planning

### 5.1 Software Life Cycle Model

In developing the project, I adopted an iterative approach to the Waterfall model, enabling a structured yet flexible framework for managing the various phases of the project. The project plan outlines specific tasks, dependencies, timelines, and milestones to ensure a systematic progression towards the final goal of analyzing social media posts for mental health disorder detection.

The Iterative Waterfall model was chosen for this project due to its structured approach while allowing for iterative revisions and refinements. Unlike the traditional Waterfall model, which emphasizes a linear progression through distinct phases, the iterative variant permits revisiting earlier stages based on findings and feedback. This flexibility is particularly beneficial in data-driven projects where insights gained during the analysis may necessitate adjustments to earlier stages, such as refining requirements or enhancing data preparation techniques.

In this project, the iterative nature of the Waterfall model facilitated ongoing improvement and adaptation throughout the development process. For example, initial results from the model evaluation phase may prompt a revisit to data preprocessing to enhance data quality or to explore alternative modeling techniques. This approach ultimately fosters a more robust final product, ensuring that the developed system meets the dynamic needs of mental health disorder detection in social media posts.

The key feature of the iterative approach is the feedback loop that exists between the phases. For instance, after completing the testing phase, if certain models do not meet performance expectations, the project can loop back to the implementation phase. This allows for modifications to the models, preprocessing techniques, or even revisiting the requirements to ensure alignment with the project's objectives.

The project is divided into distinct phases:

- **Requirement Gathering and Analysis :** This initial phase involved understanding the project's goals, objectives, and stakeholder expectations. It spanned approximately two weeks, culminating in a detailed requirements document that guided the subsequent stages.
- **Data Collection and Preparation :** Utilizing a Twitter sentiment dataset from Kaggle and Reddit API, the data collection phase was executed over a week. This included downloading the dataset, examining its structure, and performing data cleaning and prepossessing to ensure its suitability for analysis.
- **Model Development :** This phase, lasting about three weeks, included the creation of a Bag of Words model, splitting the dataset into training and test sets, and implementing various machine learning algorithms such as k-Nearest Neighbors (k-NN) and Support Vector Machines (SVM) for sentiment classification.
- **Model Evaluation :** Following model development, a week was allocated for rigorous testing and validation of the models, ensuring they met the required accuracy benchmarks. This phase involved using performance metrics such as accuracy, precision, recall, and F1-score to evaluate the models' effectiveness.

- **Final Deployment and Documentation :** The last phase, spanning two weeks, focused on deploying the best-performing model and creating comprehensive documentation. This included user manuals and technical documentation to facilitate future maintenance and enhancements.

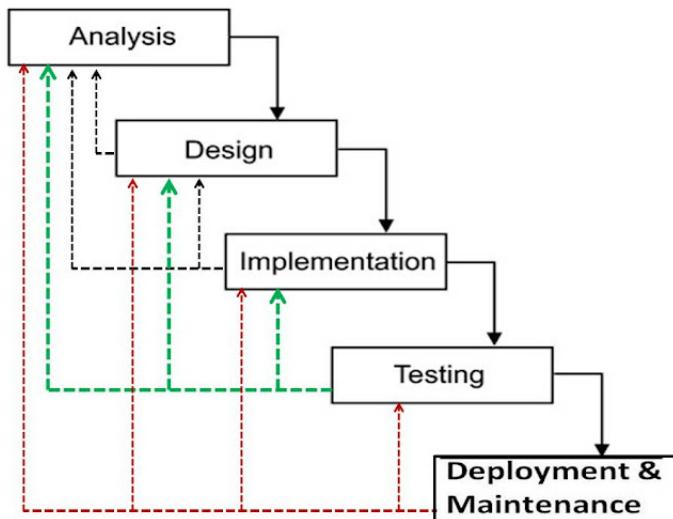


Figure 1: Iterative Waterfall Model

## 5.2 Dependencies and Milestones

Key dependencies were identified for successful project progression. For instance, completion of the data preparation phase was critical before proceeding to model development. Milestones were established at the end of each phase to ensure accountability and track progress. The successful completion of the requirement gathering phase marked the first milestone, followed by the data preparation phase, and so on.

## 5.3 Scheduling

Effective scheduling is crucial to the success of any project, as it establishes a clear timeline for tasks, milestones, and dependencies. In the context of our project on detecting mental health disorders through social media analysis, a detailed schedule has been developed to guide the project from inception to completion. This schedule includes specific tasks such as requirement gathering, data preprocessing, model implementation, testing, and deployment, each with clearly defined deadlines. The iterative nature of our chosen methodology allows for flexibility within the schedule, enabling adjustments based on testing outcomes and stakeholder feedback. Key milestones, such as the completion of data analysis, model validation, and user acceptance testing, have been identified to monitor progress and ensure timely delivery of the final product.

## ASMPFMHDD

By utilizing project management tools, such as Microsoft Project, we can visualize and track the progress of tasks, manage resources effectively, and maintain open communication among team members, ensuring that the project stays on schedule and meets its objectives. This proactive approach to scheduling enhances our ability to deliver a high-quality solution that aligns with our goals and stakeholder expectations.

ID	Task Mode	Task Name	Duration	Start	Finish	Predecessors	% Complete
1	✓	<b>GR29 ASMPFMHDD</b>	<b>243 days</b>	<b>Mon 01-07-24</b>	<b>Mon 02-06-25</b>		<b>100%</b>
2	✓	<b>Phase 1: 7th Semester Activities</b>	<b>150 days</b>	<b>Mon 01-07-24</b>	<b>Wed 22-01-25</b>		<b>100%</b>
3	✓	<b>Project Startup</b>	<b>20 days</b>	<b>Mon 01-07-24</b>	<b>Fri 26-07-24</b>		<b>100%</b>
4	✓	Team Building	2 days	Mon 01-07-24	Tue 02-07-24		100%
5	✓	Brainstorm on Project Topic	2 days	Wed 03-07-24	Thu 04-07-24	4	100%
6	✓	Project agreed with Guide	2 days	Fri 05-07-24	Mon 08-07-24	5	100%
7	✓	Related Study& Documentation	4 days	Mon 08-07-24	Thu 11-07-24	6	100%
8	✓	Deliver Project Synopsis for Guide's review	2 days	Fri 12-07-24	Mon 15-07-24	7	100%
9	✓	Close review feedbacks	9 days	Mon 15-07-24	Thu 25-07-24	8	100%
10	✓	<b>Project Synopsis Finalized</b>	<b>1 day</b>	<b>Fri 26-07-24</b>	<b>Fri 26-07-24</b>	<b>9</b>	<b>100%</b>
11	✓	<b>Requirement Analysis</b>	<b>17 days</b>	<b>Thu 01-08-24</b>	<b>Fri 23-08-24</b>		<b>100%</b>
12	✓	Gather Requirements	7 days	Thu 01-08-24	Fri 09-08-24	10	100%
13	✓	Prepare Draft Requirement Matrix	9 days	Mon 12-08-24	Thu 22-08-24	10	100%
14	✓	<b>Requirement Matrix Finalized</b>	<b>1 day</b>	<b>Fri 23-08-24</b>	<b>Fri 23-08-24</b>	<b>13</b>	<b>100%</b>
15	✓	<b>Design</b>	<b>46 days</b>	<b>Mon 26-08-24</b>	<b>Thu 24-10-24</b>	<b>14</b>	<b>100%</b>
16	✓	<b>Detailed Design</b>	<b>25 days</b>	<b>Mon 26-08-24</b>	<b>Thu 26-09-24</b>	<b>14</b>	<b>100%</b>
17	✓	Data Collection	3 days	Mon 26-08-24	Wed 28-08-24	14	100%
18	✓	Data Preprocessing	4 days	Thu 29-08-24	Mon 02-09-24	17	100%
19	✓	Model Training and Evaluation	18 days	Tue 03-09-24	Thu 26-09-24	18	100%
20	✓	<b>Test Plan Preparation</b>	<b>21 days</b>	<b>Fri 27-09-24</b>	<b>Thu 24-10-24</b>	<b>19</b>	<b>100%</b>
21	✓	Text Classification	4 days	Fri 27-09-24	Wed 02-10-24	19	100%
22	✓	Image Classification	9 days	Thu 03-10-24	Tue 15-10-24	21	100%
23	✓	Video Classification	4 days	Wed 16-10-24	Sat 19-10-24	22	100%
24	✓	Reddit and Twitter User Analysis	4 days	Mon 21-10-24	Thu 24-10-24	23	100%
25	✓	<b>Phase 1 Closure</b>	<b>59 days</b>	<b>Fri 01-11-24</b>	<b>Wed 22-01-25</b>	<b>3</b>	<b>100%</b>
26	✓	Prepare 7th Semester Project Rep	14 days	Fri 01-11-24	Wed 20-11-24	20	100%
27	✓	<b>Updated Requirement Matrix</b>	<b>2 days</b>	<b>Thu 21-11-24</b>	<b>Fri 22-11-24</b>	<b>26</b>	<b>100%</b>
28	✓	<b>Updated Project Plan</b>	<b>1 day</b>	<b>Fri 22-11-24</b>	<b>Fri 22-11-24</b>	<b>27</b>	<b>100%</b>
29	✓	Project Viva	2 days	Mon 20-01-25	Tue 21-01-25	28	100%
30	✓	<b>Approved Project Report - 7th Semester</b>	<b>1 day</b>	<b>Wed 22-01-25</b>	<b>Wed 22-01-25</b>	<b>29</b>	<b>100%</b>
31	✓	<b>Semester Gap</b>	<b>9 days</b>	<b>Thu 23-01-25</b>	<b>Tue 04-02-25</b>	<b>30</b>	<b>100%</b>
32	✓	<b>Phase 2: 8th Semester Activities</b>	<b>84 days</b>	<b>Wed 05-02-25</b>	<b>Mon 02-06-25</b>	<b>31</b>	<b>100%</b>
33	✓	<b>Coding &amp; Unit Testing</b>	<b>27 days</b>	<b>Wed 05-02-25</b>	<b>Thu 13-03-25</b>	<b>31</b>	<b>100%</b>
34	✓	Data Collection and Preprocessing	7 days	Wed 05-02-25	Thu 13-02-25	31	100%
35	✓	Model Training and Evaluation	6 days	Fri 14-02-25	Fri 21-02-25	34	100%
36	✓	Web Application Components	14 days	Mon 24-02-25	Thu 13-03-25	35	100%
37	✓	<b>System Integration Testing</b>	<b>45 days</b>	<b>Fri 14-03-25</b>	<b>Thu 15-05-25</b>	<b>36</b>	<b>100%</b>
38	✓	API Calls	23 days	Fri 14-03-25	Tue 15-04-25	36	100%
39	✓	Deployment	22 days	Wed 16-04-25	Thu 15-05-25	38	100%
40	✓	<b>Project Closure</b>	<b>14 days</b>	<b>Wed 14-05-25</b>	<b>Mon 02-06-25</b>	<b>39</b>	<b>100%</b>
41	✓	Prepare 8th Semester Project Report	5 days	Wed 14-05-25	Tue 20-05-25	39	100%
42	✓	<b>Updated Requirement Matrix</b>	<b>2 days</b>	<b>Wed 21-05-25</b>	<b>Thu 22-05-25</b>	<b>41</b>	<b>100%</b>
43	✓	<b>Updated Project Plan</b>	<b>5 days</b>	<b>Fri 23-05-25</b>	<b>Thu 29-05-25</b>	<b>42</b>	<b>100%</b>
44	✓	Review by Faculties	1 day	<b>Fri 30-05-25</b>	<b>Fri 30-05-25</b>	<b>43</b>	<b>100%</b>
45	✓	<b>Approved Project Report - 7th Semester</b>	<b>1 day</b>	<b>Mon 02-06-25</b>	<b>Mon 02-06-25</b>	<b>44</b>	<b>100%</b>

Figure 2: Project Plan

# ASMPFMHDD

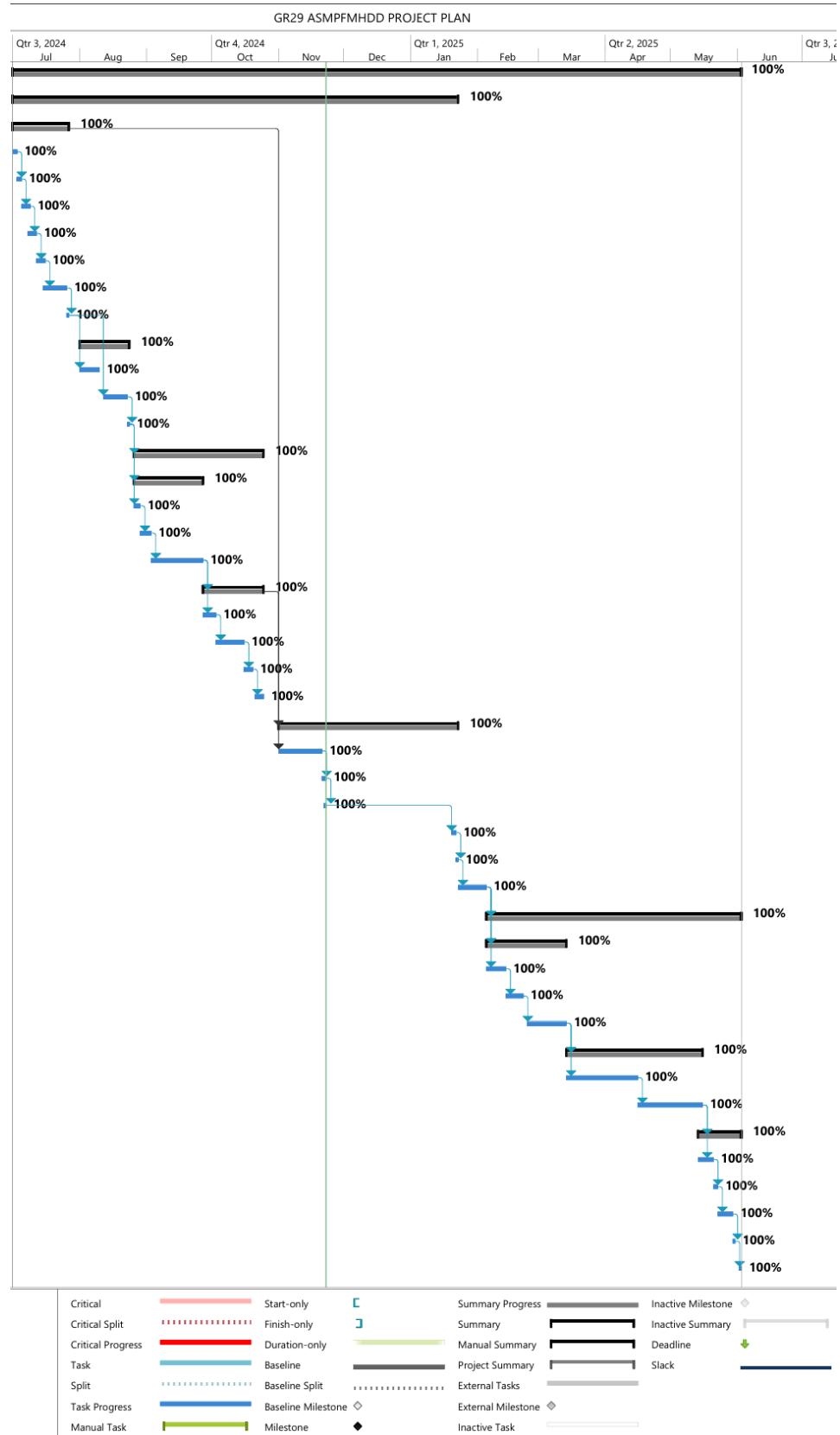


Figure 3: Gantt Chart

## 6 Requirement Analysis

### 6.1 Requirement Matrix

Rqmt ID	Requirement Item	Requirement Analysis Status	Design Module (As per Prototype folder structure)	Design Reference (section# under project Report)	Test Case Number (As per Prototype folder structure)	Technical Platform of Implementation	Prototype prepared?	Name of Program / Component
FR-001	Collect social media data from Reddit.	Completed ▾	D01	8.2.1	D04T04	PYTHON, COLAB	Yes ▾	Data Collection.ipynb
FR-002	Implement data cleaning and preprocessing.	Completed ▾	D02	8.2.2	D04T04	PYTHON, COLAB	Yes ▾	Data Preprocessing.ipynb
FR-003	Train machine learning and deep learning models.	Completed ▾	D03	8.2.3 - 8.2.10	D05DP01	PYTHON, COLAB	Yes ▾	WebAppV7.ipynb
FR-004	Evaluate models using performance metrics (accuracy, recall, F1 Score, Support).	Completed ▾	D03	9.2 - 9.9	D04T02, D04T03	PYTHON, COLAB	Yes ▾	[Algorithm_name.ipynb]
NFR-001	Testing different features of the web application	Completed ▾	D04	APPENDIX A - PROTOTYPE	D05DP02	PYTHON, STREAMLIT	Yes ▾	WebAppV8.ipynb
NFR-002	Final Web Application Deployment	Completed ▾	D05	APPENDIX A - PROTOTYPE	D05DP03	PYTHON, STREAMLIT	Yes ▾	WebAppV9.ipynb

Figure 4: Requirement Matrix

The Requirement Matrix is a comprehensive tool used to track and manage the key requirements of a project. It systematically organizes each requirement with a unique identifier, description, priority, and category (such as functional or non-functional). The matrix also records the source of the requirement, its current status (e.g., in progress, completed), any dependencies on other requirements, and the team or individual responsible for fulfilling it. Additionally, it includes a verification method to ensure the requirement is met, such as through testing or review. This structured format helps ensure that all requirements are clearly defined, prioritized, and tracked, enabling effective project management and ensuring alignment with stakeholder expectations.

### 6.2 Requirement Elaboration

#### 6.2.1 Functional Requirements

**Requirement ID: FR-001**

**Description:** Data Collection

**Priority:** High

**Category:** Functional

The system requires an ability to collect and ingest a large dataset of Reddit posts using Python Reddit API Wrapper. The data should include text content from tweets, associated sentiment labels, and other metadata such as timestamp and user details. This will serve as the primary source of information for mental health disorder detection. The system must ensure that the dataset is loaded correctly into the machine learning environment, and any discrepancies in the structure should be handled with pre-processing steps like cleaning, normalization.

***Requirement ID: FR-002******Description: Data Cleaning and Preprocessing******Priority: High******Category: Functional***

The system must include modules to clean the raw data, such as removing irrelevant characters, handling missing data, and tokenizing text. For the social media posts, it is essential to remove URLs, stopwords, and unnecessary punctuation. The pre-processing pipeline should also convert the text into a suitable numerical format using Term Frequency-Inverse Document Frequency (TF-IDF) for further analysis. Proper pre-processing ensures that the data is in a form that can be efficiently used by machine learning models.

***Requirement ID: FR-003******Description: Training Machine and Deep Learning Models******Priority: High******Category: Functional***

The system needs to implement machine learning algorithms such as Logistic Regression and XGBoost to classify social media posts based on text and detect potential signs of mental health disorders. The system must be able to train these models on historical data and then apply them to predict the sentiment and detect mental health-related issues in new posts.

***Requirement ID: FR-004******Description: Model Validation and Evaluation******Priority: High******Category: Functional***

The system must evaluate the performance of the trained models by splitting the dataset into training and test sets. Various performance metrics like accuracy, precision, recall, and F1-score should be computed to assess the model's effectiveness in detecting mental health disorders. Based on the evaluation, the system should allow for model fine-tuning, such as adjusting hyperparameters, to improve the overall performance.

## 6.2.2 Non Functional Requirements

***Requirement ID: NFR-001******Description: Testing******Priority: Medium******Category: Non Functional***

The addition of testing as a non-functional requirement ensures that the system is not only usable in the short term but also maintainable and extensible over time. This guarantees that future

updates and improvements to the system can be implemented without disrupting existing functionality or requiring a steep learning curve for new developers or users.

***Requirement ID: NFR-002***

***Description: Deployment***

***Priority: Medium***

***Category: Non Functional***

The system must be deployed considering the potential growth in the amount of social media posts that need to be analyzed. The data processing pipeline should be scalable to handle increasing data size without significant degradation in performance. This could involve implementing parallel processing techniques or leveraging cloud-based infrastructure to ensure that processing large datasets remains feasible even as the dataset scales.

The system is designed to support comprehensive mental health disorder detection from social media posts, with key functional and non-functional requirements. It requires the ability to collect and ingest large datasets of Reddit posts using the Python Reddit API Wrapper, capturing text content, sentiment labels, timestamps, and user metadata. This data forms the foundation for analysis, necessitating robust pre-processing steps to ensure quality and consistency. The pre-processing pipeline must clean raw data by removing irrelevant characters, URLs, and stopwords, handling missing values, and tokenizing text. Additionally, text should be converted into numerical formats using techniques like Term Frequency-Inverse Document Frequency (TF-IDF) for model training. The system incorporates machine learning algorithms such as Logistic Regression and XGBoost to classify social media posts and detect potential mental health disorders. It must support training these models on historical data while applying them to predict and classify new posts. To ensure model reliability, the system splits data into training and test sets, evaluating performance using metrics like accuracy, precision, recall, and F1-score. Model fine-tuning, including hyperparameter adjustments, is essential for optimizing performance. Non-functional requirements focus on testing to ensure the system remains maintainable and extensible, enabling future updates without significant disruption. Additionally, deployment considerations emphasize scalability to process increasing volumes of social media posts, leveraging parallel processing and cloud infrastructure to maintain efficiency. Together, these requirements provide a robust framework for effective and scalable mental health detection.

## 7 Design

### 7.1 Technical Environment

The technical environment for the project "Analyzing Social Media Posts for Mental Health Disorder Detection" comprises a combination of hardware, software, and tools that enable smooth data analysis, machine learning model training, and deployment. Below is a detailed overview of the minimum hardware configuration, software tools, and package details necessary to carry out this project effectively.

#### Minimum Hardware Configuration

Given the nature of the project, which involves processing textual data and training machine learning models, the hardware requirements are modest but significant enough to ensure optimal performance. The minimum configuration needed is:

- **Processor :** Intel Core i5 (or equivalent) with a base clock speed of at least 2.5 GHz. A multi-core processor is preferred as it helps in parallel processing, which is essential during model training and data preprocessing steps.
- **RAM :** 8 GB of RAM is recommended to handle the operations of data loading, cleaning, and transformation. Large datasets, like those used in this project, may require more memory to prevent memory overflow errors and reduce delays during processing. For larger datasets, 16 GB of RAM would be ideal.
- **Storage :** At least 256 GB of SSD storage is recommended. Faster storage access significantly impacts loading time for datasets and dependencies. SSD is preferred over traditional HDD because of its faster read/write speeds, which benefit large datasets like the Reddit-based social media posts used in this project.
- **Graphics Processing Unit (GPU) :** For basic machine learning tasks like Logistic Regression or SVM, a dedicated GPU is not necessary. However, if deep learning models or more complex neural networks were introduced later, a GPU like NVIDIA GTX 1060 with 4 GB VRAM or higher would be advantageous.
- **Operating System :** Windows 10 (64-bit) or higher, macOS 10.13 (High Sierra) or higher, or any stable Linux distribution (e.g., Ubuntu 18.04 or higher). The operating system should support all necessary machine learning libraries and be compatible with the tools required for the project.

## Software Tools and Packages

For the software stack, the project leverages a set of well-established tools, platforms, and programming libraries to ensure smooth execution from data preprocessing to model deployment:

- **Python** : The primary programming language used for data processing, model training, and evaluation. Python is chosen due to its rich ecosystem of libraries and frameworks tailored for machine learning and data science.
- **Pandas** : A powerful library for data manipulation and analysis, essential for data pre-processing tasks, such as handling missing values and restructuring datasets.
- **scikit-learn** : A comprehensive machine learning library in Python used for implementing and comparing algorithms in classification, regression, and clustering, along with various model evaluation tools.
- **Streamlit** : An open-source Python framework that facilitates the deployment of machine learning models and interactive web applications.
- **Pyngrok** : A Python wrapper for ngrok, used to create secure tunnels to locally deployed applications, which is particularly useful for sharing Streamlit applications over the web.
- **Google Colab** : A cloud-based platform used for writing, executing, and sharing Python code, with access to GPU and TPU resources, beneficial for model training. It integrates seamlessly with libraries like TensorFlow and PyTorch.
- **PRAW (Python Reddit API Wrapper)** : A Python library that allows for easy interaction with the Reddit API to access, retrieve, and analyze Reddit data, such as posts, comments, and user information.
- **pytesseract** : A Python wrapper for Tesseract OCR, used to extract text from images. It's essential for converting image-based text data into a format suitable for processing.
- **Pillow** : A Python Imaging Library that adds support for image processing, which aids in handling image files for text recognition tasks with pytesseract.
- **joblib** : A library for efficient serialization and deserialization of Python objects, especially useful for saving and loading machine learning models during deployment.
- **protobuf** : A protocol buffer library by Google used for serializing structured data, helpful in efficient data exchange between applications.
- **deep-translator** : A library that facilitates easy translation across different languages, enabling multilingual processing of text data.

- **Requests** : A user-friendly library for making HTTP requests, used to retrieve data from APIs or web resources as part of data collection.
- **google-generativeai** : A Python client library for Google's generative AI models, providing tools to integrate and utilize AI functionalities within the project.

## 7.2 Hierarchy of Modules

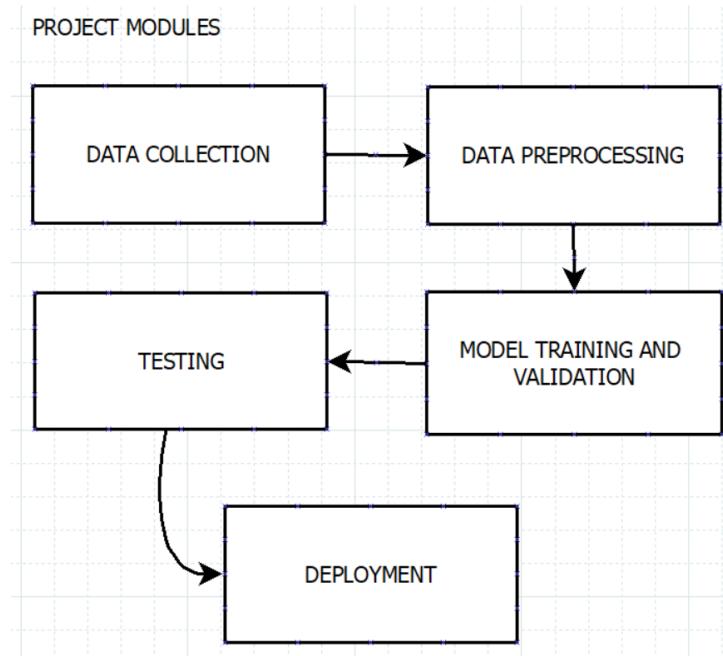


Figure 5: Project Modules

In this project, the system is structured into key modules to classify mental health issues based on text input effectively. The **Data Collection Module** gathers relevant text data, building a comprehensive dataset from sources like CSV files or platforms such as Reddit via PRAW. Next, the **Data Preprocessing Module** loads and cleans this data through tokenization, stop-word removal, and lemmatization, preparing it for analysis. Following this, the **Model Training and Validation Module** converts the text into numerical features using techniques like TF-IDF, splitting the data into training and validation sets to test various machine learning models, including Logistic Regression, Naive Bayes, SVM, Random Forest, XGboost and LSTM. Finally, the **Testing and Deployment Module** allows real-time predictions by deploying the model on platforms like Streamlit Cloud, providing an accessible solution for practical applications.

## ASMPFMHDD

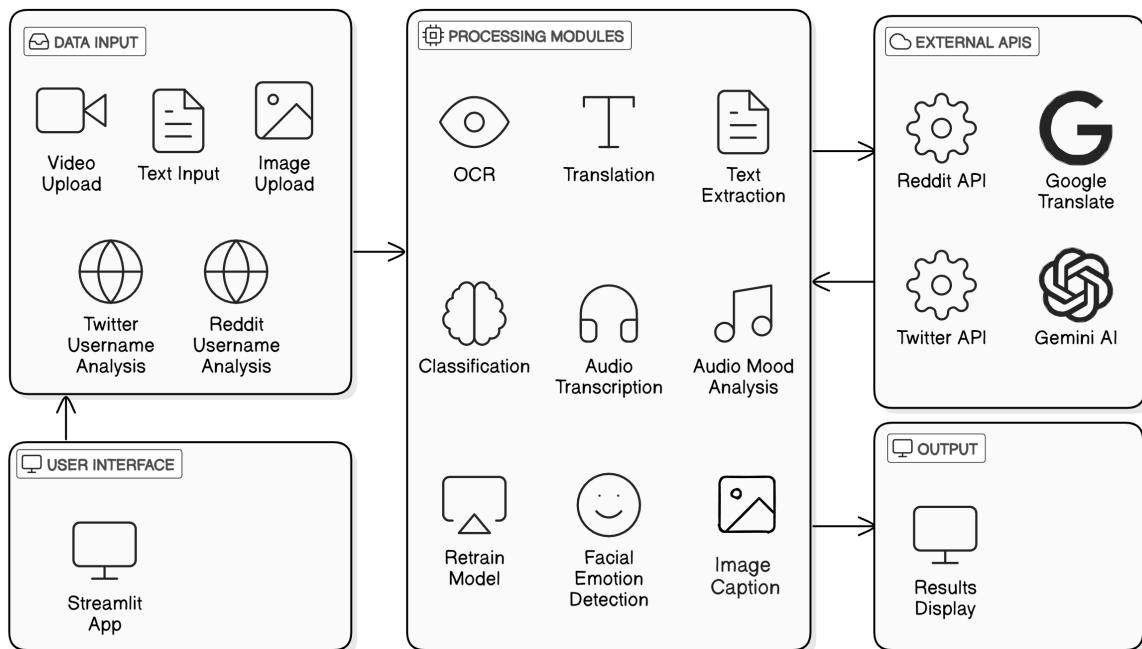


Figure 6: System Overview

### ML Model Workflow

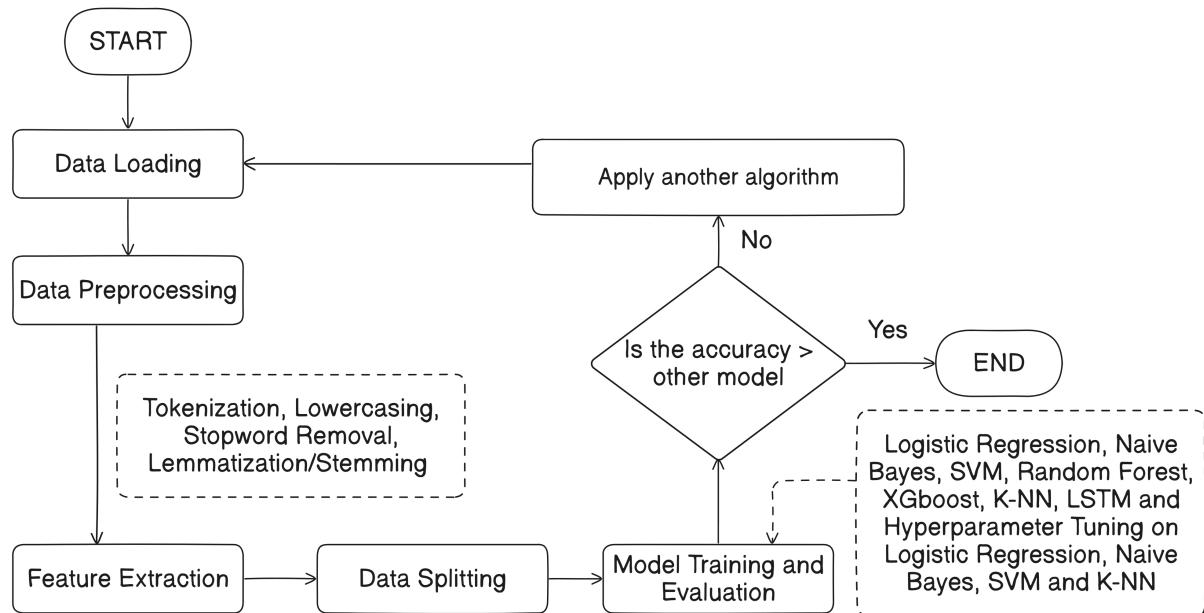


Figure 7: Model Workflow

## 7.3 Detailed Design

### 7.3.1 Data Loading and Preprocessing

The Data Loading and Preprocessing Module is the foundation of the system, responsible for ingesting and preparing the text data for analysis. This module begins by loading the dataset from the preprocessed\_mental\_health\_text.csv file, which contains various mental health-related text entries. Once the data is loaded, a series of preprocessing steps are conducted to ensure the text is clean and ready for feature extraction. This includes tokenization, where the text is split into individual words or tokens, and lowercasing to maintain uniformity across the dataset. Additionally, stop-word removal is performed to eliminate common words that do not contribute to the meaning, such as "and," "the," and "is." Finally, lemmatization or stemming is applied to reduce words to their base or root forms. These preprocessing techniques are crucial as they help improve the quality of the input data, ultimately leading to better model performance.

### 7.3.2 Feature Extraction

In the Feature Extraction Module, the preprocessed text data is transformed into a numerical format that machine learning algorithms can process. This module allows for the selection between two primary feature extraction methods: Term Frequency-Inverse Document Frequency (TF-IDF). The TF-IDF approach evaluates the importance of words in the dataset by considering their frequency in individual documents relative to their overall occurrence across all documents. This helps in highlighting the most informative words. By converting text into numerical features, this module prepares the data for the subsequent training and validation stages, ensuring that the classification models can effectively interpret the input.

### 7.3.3 Model Training and Validation

The Model Training and Validation Module is critical to developing a robust classification system. In this module, the dataset is split into training and testing sets to evaluate the performance of the models accurately. Various classification algorithms are employed, including Logistic Regression, Naive Bayes, Support Vector Machines, Random Forest, XGboost, K-Nearest-Neighbours and LSTM. Each model is trained on the training set, which involves adjusting the model parameters based on the input features and their corresponding labels. Following training, the models undergo validation to assess their performance using various metrics such as accuracy, precision, recall, and F1-score. A decision point is included to determine if the achieved accuracy meets the project requirements. If the accuracy is deemed acceptable, the model is accepted.

### 7.3.4 Prediction

The Prediction Module is designed to provide real-time classification of new input text related to mental health issues. Upon receiving user input, this module initiates a preprocessing workflow that mirrors the steps applied during the training phase, including tokenization, lowercasing, stop-word removal, and lemmatization or stemming. Once the input text is preprocessed, it is fed into the trained classification models to generate predictions. Each model may provide a classification result, allowing for a comprehensive analysis of the input. This module not only delivers the predicted mental health issue but also ensures that users receive an informative output that reflects the confidence level of each prediction, enabling them to understand the model's reasoning.

### 7.3.5 Testing and Deployment

The module focuses on making the trained models accessible for real-time predictions. Once the models have been validated and selected based on their performance, this module prepares them for deployment on suitable platforms like Streamlit Cloud. This involves packaging the models and creating a user interface where users can input text and receive predictions. The deployment process also includes considerations for scaling, ensuring that the system can handle multiple requests simultaneously while maintaining responsiveness. By providing a free and efficient deployment solution, this module enables users to access the mental health classification service easily. The deployment of the models ensures that the insights generated from the analysis can be utilized effectively in real-world applications.

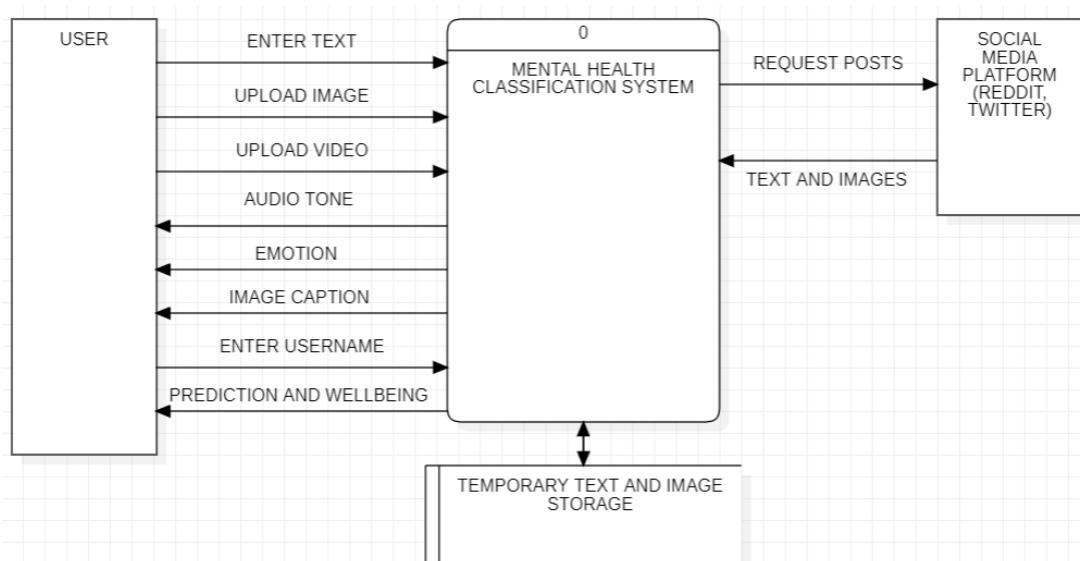


Figure 8: DFD Level 0 of the System

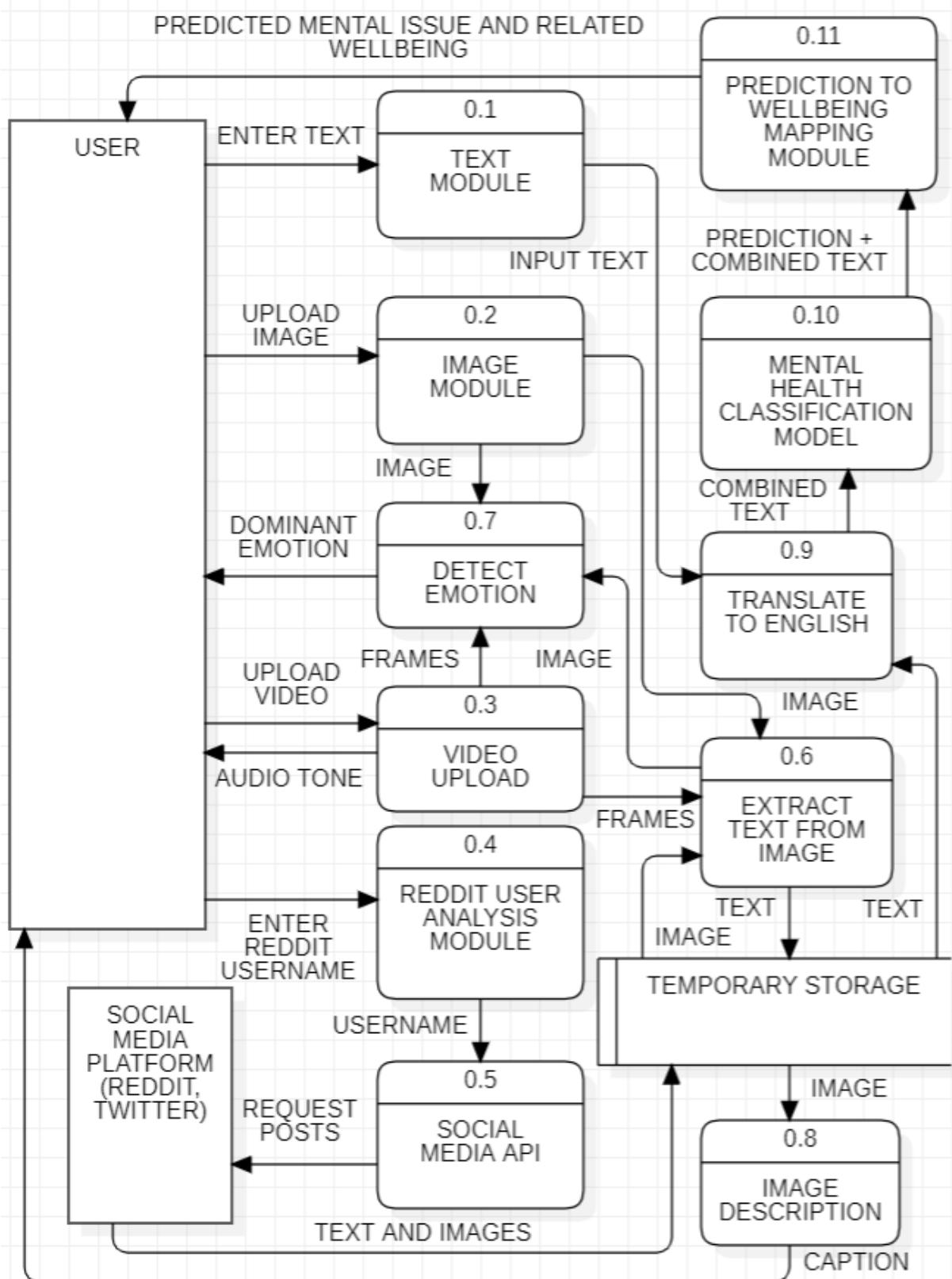


Figure 9: DFD Level 1 of the System

## 7.4 Emotion Detection Functionality

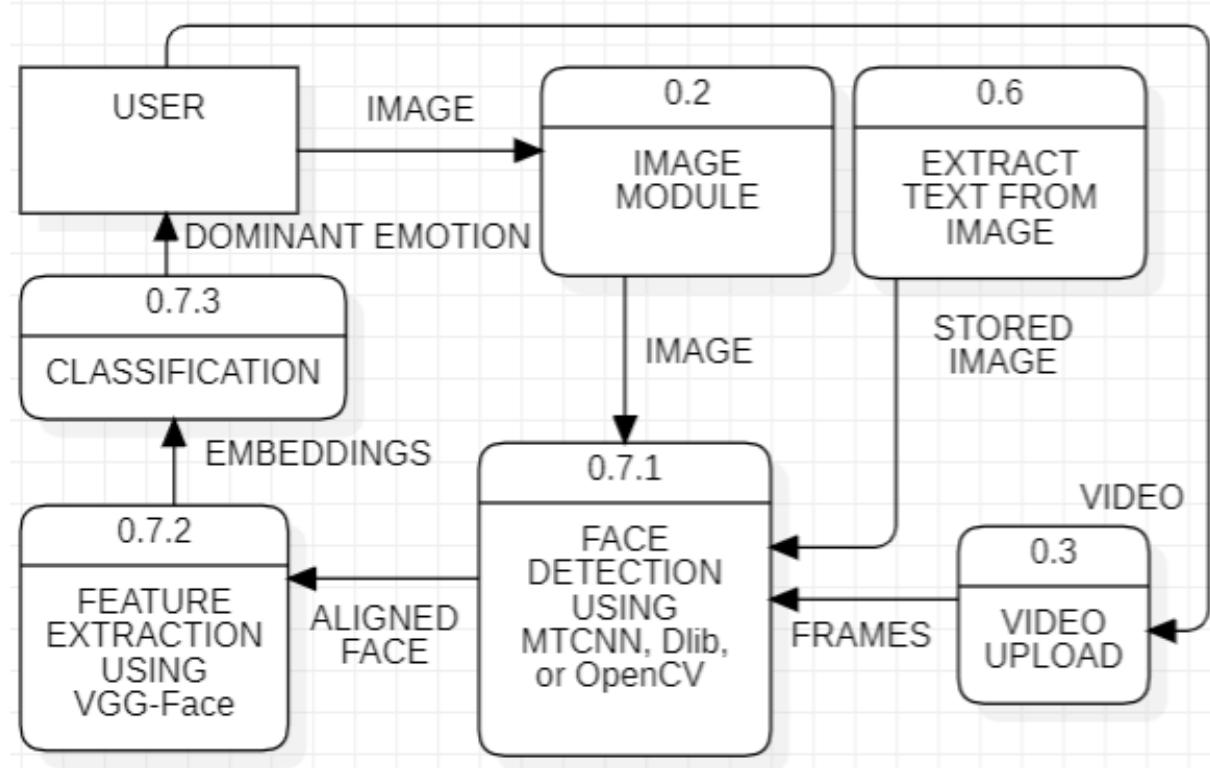


Figure 10: DFD Level 2 of Emotion Detection Functionality

DeepFace analysis involves several steps to process and analyze facial features from an uploaded image or video frame. The process begins with the user uploading the input, which is passed to DeepFace's pipeline. The first step is face detection, where models like MTCNN, Dlib, or OpenCV locate faces within the image. MTCNN (Multi-task Cascaded Convolutional Networks) detects faces using cascaded CNNs for high accuracy and alignment, while Dlib employs histogram-oriented gradients and machine learning techniques for robust detection. OpenCV, a widely-used computer vision library, uses Haar cascades or DNN-based approaches to detect faces. Once detected, the faces are cropped and aligned for consistency. Next is feature extraction, where pre-trained models like VGG-Face (a deep CNN trained specifically for face recognition) or Facenet (which uses triplet loss to generate highly discriminative face embeddings) convert the aligned face into numerical vectors called embeddings. These embeddings are compared using techniques like cosine similarity or Euclidean distance for tasks such as emotion detection (e.g., happiness, sadness), demographic analysis (e.g., age, gender), or face verification (matching two faces). Finally, the results, such as detected emotions or attributes, are displayed to the user, completing the analysis pipeline.

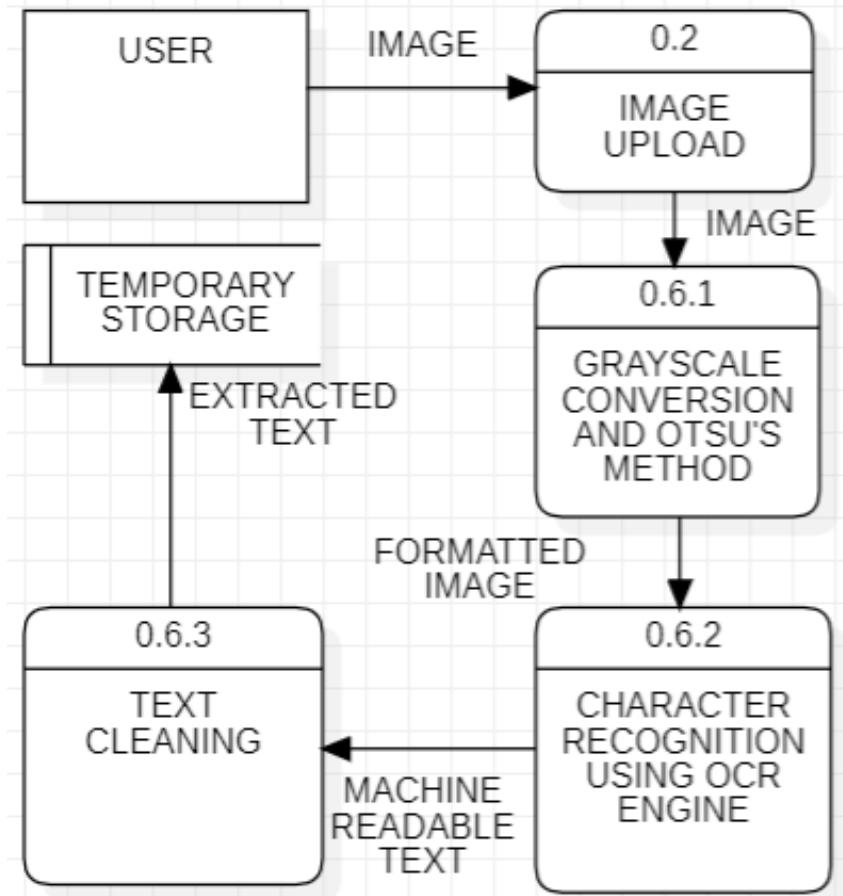


Figure 11: DFD Level 2 of Text Extraction from Image

## 7.5 Extract Text From Image

The process of extracting text from an image using Tesseract-OCR begins with the user uploading an image containing text, such as a scanned document or a photo of a sign. This raw input undergoes preprocessing to enhance the image for OCR recognition. The first step in preprocessing is grayscale conversion, where the image's RGB values are reduced to a single intensity value using a weighted formula. This simplifies the image by removing color information while retaining the text regions. Next, noise reduction is applied using techniques like Gaussian Blur or Median Blur, which smooth out artifacts while preserving the edges of text. After noise reduction, binarization is performed to convert the grayscale image into a binary format (black-and-white) using methods like Otsu's thresholding. This step isolates the text from the background. Region detection follows, where text-heavy areas are identified using contour detection or connected component analysis, filtering out non-text regions based on properties like aspect ratio and alignment. Once preprocessing is complete, the image is passed to Tesseract for character recognition. This involves segmenting the image into text lines and individual characters, which are then processed by Tesseract's LSTM-based neural network. The network

evaluates sequences of characters in context, improving the accuracy of word recognition. Recognized characters are combined into machine-readable text, leveraging language models and dictionary lookups to maintain the natural flow and correct structure of the text. The output is a string of recognized text, but it may still require postprocessing for better accuracy and readability. Postprocessing begins with error correction, where the recognized text is compared against dictionaries or spell-checking algorithms to identify and fix errors caused by OCR misrecognition. Rule-based corrections are also applied, such as replacing common misinterpreted characters (e.g., Òwith Òòr Ìwith Ì) depending on the context. After error correction, formatting is applied to structure the text into paragraphs, reconstruct line breaks, and retain alignments. This ensures that the extracted text is not only accurate but also well-organized for further use. Finally, the extracted text is either displayed to the user or saved to a file format such as .txt or .docx. This structured and cleaned output can then be utilized for various applications, from document analysis to content storage. The entire pipeline, from preprocessing to post-processing, ensures a high-quality extraction process by combining image enhancement, deep learning-based recognition, and advanced text correction methods.

## 7.6 Image Description

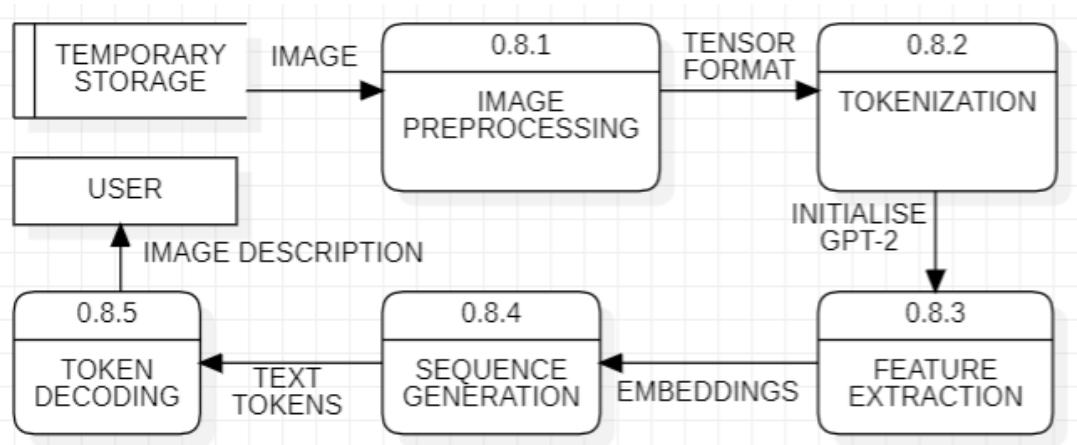


Figure 12: DFD Level 2 of Image Description

The process of generating an image description using Python's Transformers module and the ViT-GPT2 model begins with the user input, where the user uploads an image to be described. This image can be in any format (e.g., PNG, JPEG), and it is the raw data that will undergo a series of transformations before the description is generated. Once the image is received, it proceeds to the preprocessing stage. Preprocessing involves multiple steps: first, the image is resized to match the input dimensions required by the Vision Transformer (ViT) model, typically  $224 \times 224$  pixels. This ensures uniformity across all input data. The pixel values are then normalized to fall within a standard range (e.g., between 0 and 1), which improves the efficiency

and stability of the model’s computations. The image is subsequently transformed into a tensor format, which is the data structure required for processing by PyTorch or TensorFlow-based deep learning models. At this stage, the **“GPT-2 tokenizer”** is also initialized, preparing it to convert tokens into text during the sequence generation phase. Following preprocessing, the system moves to feature extraction, where the Vision Transformer (ViT) processes the image to extract meaningful features. ViT divides the image into smaller patches, typically  $16 \times 16$  pixels, and flattens each patch into a vector. These vectors are embedded using a linear projection layer, and positional embeddings are added to encode the spatial arrangement of the patches. The embedded patches are then passed through several transformer layers, where self-attention mechanisms analyze relationships between patches to capture both local and global image features. The final output of ViT is a high-dimensional embedding vector that represents the visual content of the image. This embedding vector is then passed to the GPT-2 decoder for the sequence generation phase. GPT-2 uses the image embeddings as input and applies its pre-trained language model to generate a sequence of text tokens. This is achieved through an iterative process where GPT-2 predicts the next token in the sequence based on the embeddings and previously generated tokens. The model continues this process until it reaches a stopping condition, such as an end-of-sequence (EOS) token. The generated tokens are numerical representations of words and phrases. Next comes the postprocessing stage, where the numerical tokens generated by GPT-2 are decoded back into human-readable text using the GPT-2 tokenizer. This step involves mapping the numerical tokens to their corresponding words in the vocabulary. The decoded text is then cleaned up for readability, ensuring proper grammar and structure. Any formatting issues are addressed during this stage to produce coherent and meaningful sentences. Finally, the system outputs the image description to the user. The description is a concise and accurate textual representation of the image content, such as “A group of people playing soccer on a field.” This description can be displayed on the application’s interface or saved to a file for further use. By leveraging the advanced capabilities of the Vision Transformer for feature extraction and GPT-2 for language modeling, this pipeline provides an efficient and accurate solution for automated image captioning.

## 7.7 Translation to English

The process of translating text to English using DeepTranslator begins with the user providing input, either by typing text in a non-English language or uploading a file containing text. This input is handled by the system, which prepares it for translation. The text is first preprocessed to ensure optimal translation results. This involves cleaning the input by removing unwanted characters such as extra whitespace, punctuation errors, or HTML tags that could interfere with accurate translation. If the source language is not specified, a language detection step is performed. This step uses machine learning models trained on multilingual datasets to analyze pat-

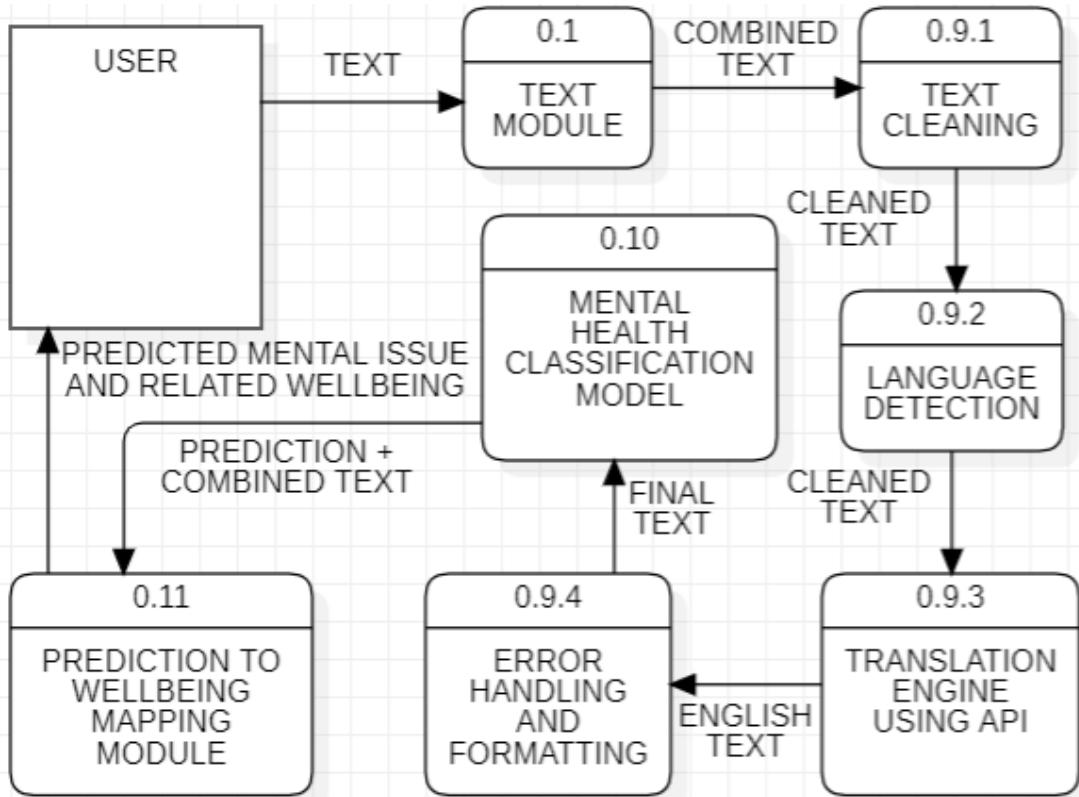


Figure 13: DFD Level 2 of Translation to English

terns in the text, such as character distribution and word frequency, to determine the language. Models like Google's Compact Language Detector employ statistical and supervised learning methods for this purpose. Once the text is prepared, it is passed to the DeepTranslator module, which interfaces with translation APIs like Google Translate or Microsoft Translator. These APIs rely on advanced neural machine translation (NMT) techniques to perform the translation. NMT models, which are based on deep learning, use sequence-to-sequence (seq2seq) architectures with encoder-decoder frameworks. The encoder processes the input text and converts it into a latent representation, a numerical vector capturing its meaning. This representation is then passed to the decoder, which generates the translated text in English. To ensure contextual accuracy, these models utilize attention mechanisms, which help focus on relevant parts of the input sentence during translation. Modern transformer-based architectures, like those used in Google Translate, enhance translation performance by enabling parallel processing of text and capturing long-range dependencies in sentences. After translation, the system postprocesses the output to ensure quality and readability. This includes validating the translated text for completeness, handling errors like untranslated phrases, and retrying failed requests if necessary. The output is formatted to resemble the original structure of the input, preserving elements like paragraphs, line breaks, and punctuation to maintain coherence. Finally, the translated text is displayed or saved, ensuring it is clean, accurate, and easy to understand. Throughout

this process, concepts from natural language processing (NLP), deep learning, and robust error handling are integrated to provide reliable and efficient translations.

## 7.8 Prediction to Wellbeing Mapping

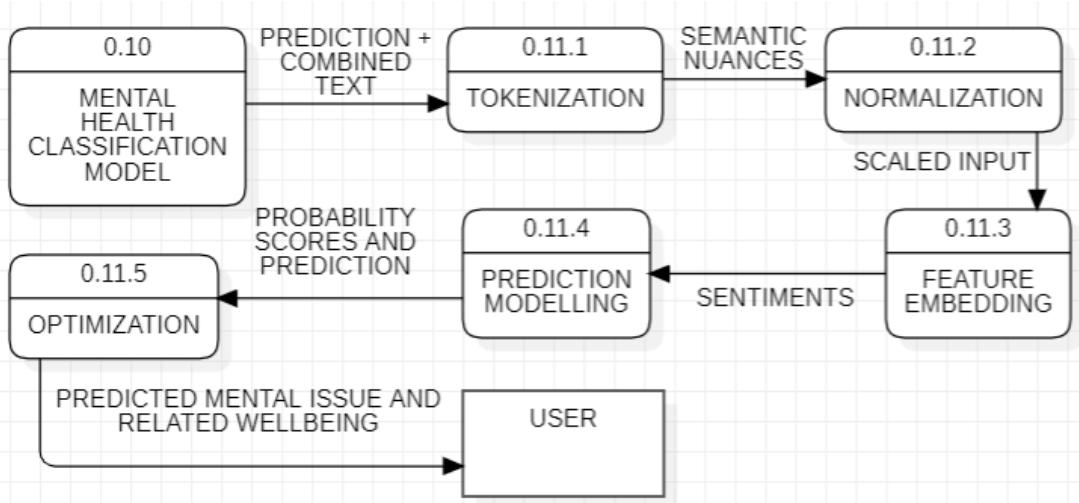


Figure 14: DFD Level 2 of Prediction to Wellbeing Mapping

The process begins when the user provides input data for prediction, which could be in the form of text, behavioral data, or health metrics. For instance, a user might input their mood and health conditions, share behavioral patterns, or upload relevant data such as physical activity logs or medical records. This data serves as the foundation for predicting the user's mental health or overall wellbeing. Once the data is provided, the next step is preprocessing, which involves cleaning and transforming the raw input. During the data cleaning phase, any noise, missing values, or redundant information are identified and removed. This ensures the dataset is clean and ready for processing. Additionally, feature engineering is performed to extract meaningful attributes or characteristics from the raw data. For example, from behavioral data, features like activity frequency or sleep patterns may be derived. This allows the model to work with relevant information that enhances the prediction process. The third step involves embedding generation, where the core model, GEMINI 1.5 FLASH, processes the preprocessed input to generate embeddings. Embeddings are high-dimensional vectors that capture the semantic and contextual properties of the data. For example, when analyzing text input, GEMINI 1.5 FLASH uses its deep learning algorithms to understand the context and meaning of the user's words, transforming them into embeddings that encapsulate both the surface-level meaning and the deeper nuances of the data. This representation of the input data is essential for subsequent prediction tasks. In the prediction model phase, the generated embeddings are passed to machine learning models, which use these embeddings to predict a user's mental health status or behavioral outcomes. These models may utilize various algorithms, such as decision trees,

support vector machines, or deep learning models, to make predictions. The system not only predicts the mental health condition but also maps these predictions to various wellbeing dimensions. These dimensions could include emotional wellbeing, such as happiness or stress; physical wellbeing, like energy levels or fatigue; and social wellbeing, including aspects like connectedness or isolation. By doing so, the system offers a comprehensive view of the user's overall wellbeing based on the predictions. Once predictions are made, postprocessing ensures the results are accurate and formatted properly. During the validation phase, the predictions are checked for any anomalies or inconsistencies, ensuring the output is reliable. This might involve techniques like consistency checks or outlier detection to verify that the predictions align with expected patterns. After validation, the results are formatted into a user-friendly structure. This might include organizing the wellbeing data into clear categories, such as emotional, physical, and social wellbeing, making it easy for the user to interpret. Finally, the system delivers the output, displaying or saving the prediction and associated wellbeing mapping. This output is presented in a way that helps the user understand the various aspects of their mental and physical health, giving them insights into their current state of wellbeing and providing actionable data for improvement.

## 7.9 Audio Mood Analysis

The analyze\_audio\_mood function begins with the user providing a path to a video file as input. This is the starting point of the process where the user supplies the video from which the audio will be extracted. In the preprocessing stage, the video file is passed to the extract\_audio\_from\_video function. This function extracts the audio content from the video, separating it for further analysis. Once the audio is extracted, the next step involves loading the audio file into memory using the librosa library, which prepares the audio data for feature extraction. The core of the audio processing happens during the MFCC (Mel-frequency cepstral coefficients) feature extraction. The librosa.feature.mfcc method is used to compute the MFCCs of the audio. These MFCCs represent different frequency bands of the audio signal and are crucial for analyzing the characteristics of the audio that correspond to various emotional tones. By capturing frequency patterns, MFCCs provide a robust feature set for mood classification. After the MFCCs are extracted, the next step is feature segmentation. The MFCC array is divided into four distinct frequency bands: low frequencies (MFCCs 0, 1, and 2), mid-low frequencies (MFCCs 3 and 4), mid-high frequencies (MFCCs 5, 6, and 7), and high frequencies (MFCCs 8, 9, 10, 11, and 12). For each of these frequency bands, the scalar mean of the MFCC values is computed. This scalar mean represents the overall intensity of the audio in each frequency range, simplifying the data for classification purposes. Finally, the mood classification process takes place. The scalar means for each of the frequency bands are compared to determine the dominant mood of the audio. Based on these comparisons, the function classifies the audio as

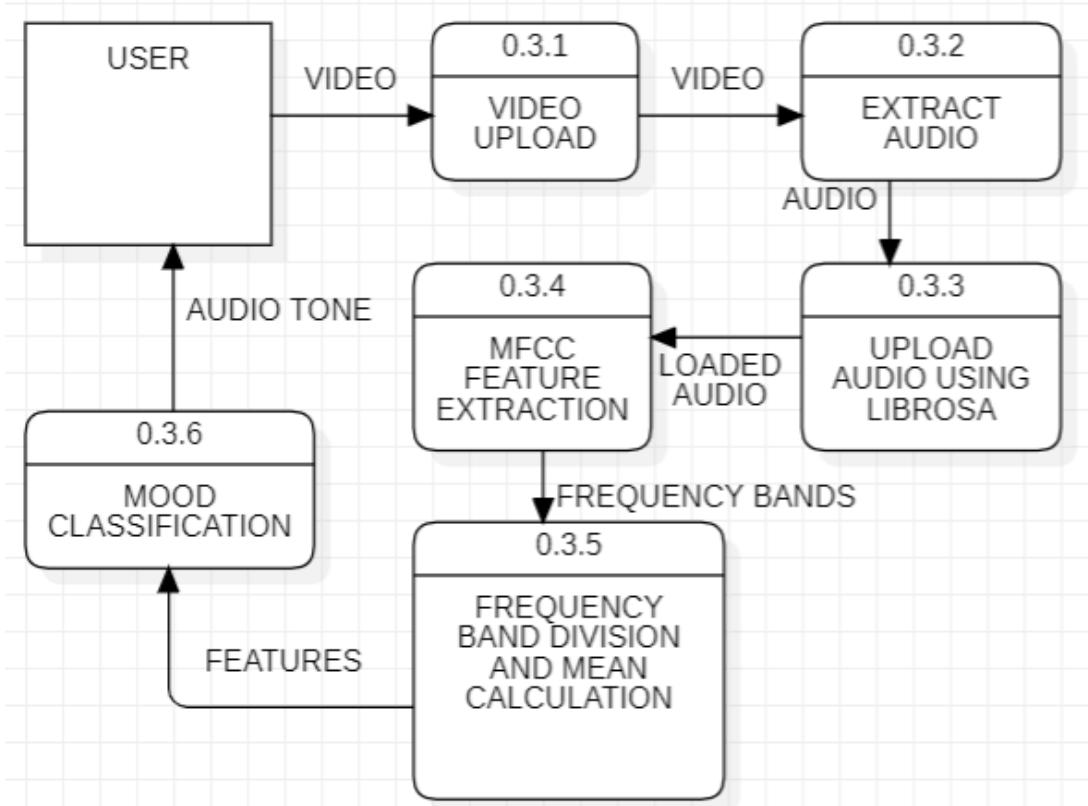


Figure 15: DFD Level 2 of Audio Mood Analysis

normal, neutral, melancholic, calm, anxious, or happy, depending on which frequency bands exhibit the strongest characteristics. This mood classification is then returned as the result of the analysis.

The various processes and technologies discussed throughout this conversation can play a pivotal role in enhancing the classification and prediction of mental health disorders. The use of machine learning (ML) and deep learning (DL) models enables the system to process and analyze large volumes of data, such as text, behavioral patterns, health metrics, and even multimedia inputs (e.g., images, audio, and video). By leveraging the DeepFace module for facial emotion recognition, the system can identify emotional cues from visual data, helping to assess a user's emotional state. This is complemented by text analysis models, such as those built with Logistic Regression, k-NN, SVM, and LSTM architectures, which are capable of classifying mental health issues from user-provided text, such as social media posts, personal narratives, or other forms of written communication.

Further enhancing the capability, the integration of DeepTranslator ensures that non-English input can be translated into English, allowing the system to support a wider variety of users and

ensuring the accuracy of the mental health analysis, even when the data comes from different languages. Meanwhile, behavioral data and health metrics are crucial inputs that can provide insights into a person's overall wellbeing. These data sources are analyzed using GEMINI 1.5 FLASH, which generates embeddings capturing the semantic and contextual properties of the input data, feeding into prediction models that classify users' mental health conditions and map them to various wellbeing dimensions. The classification of conditions like anxiety, depression, PTSD, and others can be made more accurate through these comprehensive data analyses.

By incorporating audio mood analysis through techniques such as MFCC for audio emotion detection and facial expression recognition for video inputs, the system further refines its ability to assess emotional and mental states. These audio and visual analyses provide additional layers of understanding that text alone might not fully capture. In conjunction with these features, streaming data from social media platforms like Reddit or Twitter allows for the monitoring of a user's public posts, providing real-time insights into their mental health as they interact with others online.

The combined use of these models and processes facilitates a holistic approach to mental health disorder classification. The system can classify mental health issues based on a range of factors, including written text, emotional tone from images, behavioral patterns, and audio cues. This multimodal approach helps create a more comprehensive, accurate, and nuanced prediction of mental health conditions, offering personalized insights into emotional, physical, and social wellbeing. As a result, users benefit from a system that not only detects potential mental health issues but also offers a deeper understanding of their overall wellbeing, which can inform targeted interventions, support, and wellbeing strategies.

## 8 Implementation

### 8.1 Features From RM

For the initial prototype development, a subset of the requirements from the Requirement Matrix (RM) was carefully selected to focus on implementing core functionalities and demonstrating proof of concept. The selection was based on a combination of high-priority functional requirements that form the backbone of the system, ensuring that critical features are built and validated before expanding the scope.

The requirements chosen for the prototype primarily involve data preprocessing, model training, and evaluation processes. These requirements were selected because they are fundamental to the project's success, ensuring that the data pipeline and model implementation work seamlessly together. This subset of features lays the groundwork for later integration with additional components and more complex functionality.

The filtered part of the RM focuses on the following high-priority requirements: data cleaning and feature extraction, training of machine learning models, and evaluation of model performance using standard metrics. These requirements were identified as crucial because they directly impact the system's ability to handle data, learn patterns, and provide meaningful outputs. Without successfully implementing these core features, the overall effectiveness of the solution would be significantly reduced.

Furthermore, these selected features align with the project's goals and provide a clear pathway for incremental development. By narrowing down the requirements to these foundational aspects, the development team can ensure that the prototype is not only functional but also extensible, providing a robust framework for future enhancements.

Rqmt ID	Requirement Item	Requirement Analysis Status
FR-001	Collect social media data from Reddit.	Completed ▾
FR-002	Implement data cleaning and preprocessing.	Completed ▾
FR-003	Train machine learning and deep learning models.	Completed ▾
FR-004	Evaluate models using performance metrics (accuracy, recall, F1 Score, Support).	Completed ▾
NFR-001	Testing different features of the web application	Completed ▾
NFR-002	Final Web Application Deployment	Completed ▾

Figure 16: Features from Requirement Matrix

## 8.2 Code Details and Output

### 8.2.1 Data Collection

#### Collecting Reddit Posts

```

import praw
import pandas as pd
import time
# Initialize Reddit API
reddit = praw.Reddit(client_id='YOUR_CLIENT_ID',
                      client_secret='YOUR_CLIENT_SECRET',
                      user_agent='Mental_Health_Classifier')
# Define subreddits and post types
subreddits = {'normal': ['news', 'AskReddit'],
              'depression': ['depression'],
              'ptsd': ['PTSD'],
              'anxiety': ['Anxiety'],
              'bipolar': ['BipolarReddit']}
post_types = ['hot', 'new', 'top']
posts_per_type = 100
# Collect and save posts
data = []
for label, subs in subreddits.items():
    for sub in subs:
        for post_type in post_types:
            posts = getattr(reddit.subreddit(sub), post_type)(
                limit=posts_per_type)
            for post in posts:
                data.append([post.title + "\n" + post.selftext,
                            label])
                time.sleep(1)
df = pd.DataFrame(data, columns=['text', 'label'])
df.to_csv(f'{label}_dataset.csv', index=False)

```

		text	mental_health_issue
0	2024 Election Due to the 2024 US Presidential ...		bipolar
1	Our most-broken and least-understood rules is ...		depression
2	Rules **UPDATED** 1. If you're here you must p...		normal
3	Happy Cakeday, r/Normal! Today you're 11 Let's...		normal
4	Happy Cakeday, r/Normal! Today you're 10 Let's...		normal
5	I am also a person I am very normal. Today I t...		normal
6	Happy Cakeday, r/Normal! Today you're 9 Let's ...		normal
7	I am a person I am a person		normal
8	Elections and Politics Hello friends!\n\nIt's ...		anxiety
9	You are more than just one emotion		ptsd

Figure 17: Obtained Dataset

## Combining Collected Datasets

```

import pandas as pd
from sklearn.utils import shuffle

# Load datasets
bipolar_df = pd.read_csv("bipolar_dataset.csv")
depression_df = pd.read_csv("depression_dataset.csv")
normal_df = pd.read_csv("normal_dataset.csv")
anxiety_df = pd.read_csv("anxiety_dataset.csv")
ptsd_df = pd.read_csv("ptsd_dataset.csv")

# Minimum length for balanced classes
min_length = min(len(bipolar_df), len(depression_df), len(
    normal_df) // 6, len(anxiety_df), len(ptsdf))

# Create balanced pattern
pattern_data = []
for i in range(min_length):
    pattern_data.extend([bipolar_df.iloc[i], depression_df.
        iloc[i]] +
        normal_df.iloc[i*6:(i+1)*6].to_dict(
            'records') +
        [anxiety_df.iloc[i], ptsd_df.iloc[i
            ]])

pattern_df = pd.DataFrame(pattern_data)
remaining_data = shuffle(pd.concat([bipolar_df[min_length:],

depression_df[min_length:],

normal_df[min_length
*6:], anxiety_df[
min_length:], ptsd_df
[min_length:]]))

final_df = pd.concat([pattern_df, remaining_data]).reset_index(drop=True)
final_df.to_csv("mental_health_combined.csv", index=False)

```

The first code snippet collects Reddit posts by connecting to specific subreddits associated with various mental health topics. Using the Reddit API, posts are retrieved from ‘hot’, ‘new’, and ‘top’ categories for each mental health type, including general subreddits (for the “normal” category). Each post’s title and content are combined, labeled, and saved into separate CSV files. The second code snippet merges these CSV files to form a combined dataset. It creates a balanced sample by selecting the minimum number of rows across each category and organizes the data in a structured pattern. Remaining rows are shuffled and appended, producing a final dataset suitable for machine learning applications in mental health classification.

## 8.2.2 Data Preprocessing

### Text Preprocessing

```

import pandas as pd
import re
from sklearn.feature_extraction.text import
    TfidfVectorizer
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
import nltk

# Download stopwords (if you haven't already)
nltk.download('stopwords')
nltk.download('punkt')

# Load the dataset
df = pd.read_csv('mental_health.csv')

# 1. Handling Missing Values
df.dropna(subset=['text'], inplace=True)

# 2. Removing duplicates (if any)
df.drop_duplicates(subset=['text'], inplace=True)

# 3. Text Preprocessing
negative_words = {"not", "no", "nor", "never", "nothing",
                  "nowhere", "neither", "cannot", "n't", "without", "barely",
                  "hardly", "scarcely"}

def clean_text(text):
    text = re.sub(r'http\S+', '', text) # Remove URLs
    text = re.sub(r@\w+, '', text) # Remove mentions
        (@username)
    text = re.sub(r'[^a-zA-Z\s]', '', text) # Remove
        special characters, numbers, and punctuations
    text = text.lower() # Convert text to lowercase
    tokens = word_tokenize(text) # Tokenize the text
    tokens = [word for word in tokens if word not in
              stopwords.words('english') or word in
              negative_words] # Remove stopwords, but keep
              negative words
    clean_text = '_'.join(tokens) # Join the tokens
        back into a single string
    return clean_text

df['cleaned_text'] = df['text'].apply(clean_text) #
    Apply the cleaning function to the 'text' column

```

## Text Preprocessing

```

# 4. Feature Extraction using TF-IDF Vectorization
vectorizer = TfidfVectorizer(max_features=10000) #
    Adjust the max_features
X = vectorizer.fit_transform(df['cleaned_text']) # Fit
    and transform the cleaned text data
X_df = pd.DataFrame(X.toarray(), columns=vectorizer.
    get_feature_names_out()) # Convert the result to a
    DataFrame for easier understanding

# Save the preprocessed dataset (optional)
df.to_csv('preprocessed_mental_health.csv', index=False)

```

The code snippet above demonstrates the preprocessing steps for a mental health dataset. It handles missing values and duplicates in the text data, performs text cleaning by removing URLs, mentions, and special characters, and applies tokenization. The cleaned text is then vectorized using TF-IDF to convert the textual data into a format suitable for machine learning models. Finally, the preprocessed dataset is saved as a new CSV file for future use in classification tasks.

```

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!
    aaaaaaaaaaaaaaaaaaaaaaaaaa ab abandon abandoned \
0 0.0                      0.0 0.0      0.0      0.0
1 0.0                      0.0 0.0      0.0      0.0
2 0.0                      0.0 0.0      0.0      0.0
3 0.0                      0.0 0.0      0.0      0.0
4 0.0                      0.0 0.0      0.0      0.0

    abandoning abandonment abc abilities ability ... zemeckis zemlja \
0      0.0        0.0 0.0      0.0 0.00000 ...      0.0      0.0
1      0.0        0.0 0.0      0.0 0.03726 ...      0.0      0.0
2      0.0        0.0 0.0      0.0 0.00000 ...      0.0      0.0
3      0.0        0.0 0.0      0.0 0.00000 ...      0.0      0.0
4      0.0        0.0 0.0      0.0 0.00000 ...      0.0      0.0

    zero zoloft zombie zombieland zombies zone zoom zuckerberg
0  0.0   0.0   0.0     0.0     0.0   0.0   0.0   0.0
1  0.0   0.0   0.0     0.0     0.0   0.0   0.0   0.0
2  0.0   0.0   0.0     0.0     0.0   0.0   0.0   0.0
3  0.0   0.0   0.0     0.0     0.0   0.0   0.0   0.0
4  0.0   0.0   0.0     0.0     0.0   0.0   0.0   0.0

[5 rows x 10000 columns]

```

Figure 18: Data Collection and Preprocessing

### 8.2.3 Logistic Regression Model for Classification

#### Logistic Regression for Mental Health Classification

```

import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score,
    classification_report, confusion_matrix

# Load the preprocessed dataset
dataset = pd.read_csv('preprocessed_mental_health.csv')

# Check if 'cleaned_text' and 'mental_health_issue' columns
# exist
if 'cleaned_text' not in dataset.columns or \
    'mental_health_issue' not in dataset.columns:
    raise ValueError("The dataset must have 'cleaned_text' and 'mental_health_issue' columns.")

# Remove rows with missing values in 'cleaned_text' column
dataset.dropna(subset=['cleaned_text'], inplace=True)

# Initialize the CountVectorizer and fit/transform the
# cleaned text
LRvectorizer = CountVectorizer()
X = LRvectorizer.fit_transform(dataset['cleaned_text'])

# Prepare the target variable
y = dataset['mental_health_issue']

# Split the dataset into Training and Test Sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
    test_size=0.2, random_state=42)

# Initialize the Logistic Regression model
LRmodel = LogisticRegression(multi_class='ovr', max_iter
    =10000)

# Train the model
LRmodel.fit(X_train, y_train)

# Make predictions on the test set
y_pred = LRmodel.predict(X_test)

```

## Logistic Regression for Mental Health Classification

```

# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy:{accuracy*100:.2f}%')

# Print classification report
print("Classification_Report:\n", classification_report(
    y_test, y_pred))

# Print confusion matrix
print("Confusion_Matrix:\n", confusion_matrix(y_test, y_pred))

```

The provided code demonstrates how to apply Logistic Regression for mental health classification using a preprocessed dataset. First, the dataset is loaded, and a check is performed to ensure that it contains the necessary columns, specifically `cleaned_text` and `mental_health_issue`. If these columns are missing, an error is raised. The dataset is then cleaned by removing rows that have missing values in the `cleaned_text` column using the `dropna()` function. Next, the `CountVectorizer` is initialized and applied to the `cleaned_text` column to convert the text data into a numerical format suitable for machine learning. This is accomplished by transforming the text into a document-term matrix where each row represents a document (post) and each column represents a unique term (word). The target variable, `mental_health_issue`, is also extracted from the dataset. The dataset is then split into training and testing sets using `train_test_split()`, where 80% of the data is used for training, and 20% is reserved for testing. The `random_state` is set to ensure reproducibility of the results. A Logistic Regression model is initialized with a multi-class strategy (`multi_class='ovr'`) and a high number of iterations (`max_iter=10000`) to allow the model to converge. The model is then trained on the training data using the `fit()` method. After training, predictions are made on the test set using the `predict()` method. The model's performance is evaluated by calculating the accuracy score, which is printed as a percentage. Additionally, a detailed classification report is generated, which includes metrics such as precision, recall, and F1-score for each class. Lastly, a confusion matrix is printed to visualize the model's classification performance and to understand how well the model distinguishes between different mental health issues. This end-to-end pipeline is crucial for applying Logistic Regression to textual data, allowing for effective classification of mental health issues based on Reddit posts.

### 8.2.4 Naive Bayes for Classification

#### Naive Bayes for Mental Health Classification

```

import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import
    CountVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score,
    classification_report, confusion_matrix

# Load the preprocessed dataset
dataset = pd.read_csv('preprocessed_mental_health.csv')

# Check if 'cleaned_text' and 'mental_health_issue'
# columns exist
if 'cleaned_text' not in dataset.columns or 'mental_health_issue' not in dataset.columns:
    raise ValueError("The dataset must have 'cleaned_text' and 'mental_health_issue' columns.")

# Remove rows with missing values in 'cleaned_text' column
dataset.dropna(subset=['cleaned_text'], inplace=True)

# Initialize the CountVectorizer and fit/transform the cleaned text
NBvectorizer = CountVectorizer()
X = NBvectorizer.fit_transform(dataset['cleaned_text'])

# Prepare the target variable
y = dataset['mental_health_issue']

# Split the dataset into Training and Test Sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
    test_size=0.2, random_state=42)

# Initialize the Naive Bayes classifier
NBmodel = MultinomialNB()

# Fit the model
NBmodel.fit(X_train, y_train)

```

## Naive Bayes for Mental Health Classification

```

# Make predictions
y_pred = NBmodel.predict(X_test)

# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy * 100:.2f}%')

# Print classification report
print("Classification_Report:\n", classification_report(
    y_test, y_pred))

# Print confusion matrix
print("Confusion_Matrix:\n", confusion_matrix(y_test,
    y_pred))

```

The provided code demonstrates how to apply the Naive Bayes classifier (MultinomialNB) for mental health classification using a preprocessed dataset. First, the dataset is loaded, and a check is performed to ensure that it contains the necessary columns, specifically `cleaned_text` and `mental_health_issue`. If these columns are missing, an error is raised. The dataset is then cleaned by removing rows that have missing values in the `cleaned_text` column using the `dropna()` function. Next, the `CountVectorizer` is initialized and applied to the `cleaned_text` column to convert the text data into a numerical format suitable for machine learning. This is accomplished by transforming the text into a document-term matrix where each row represents a document (post) and each column represents a unique term (word). The target variable, `mental_health_issue`, is also extracted from the dataset. The dataset is then split into training and testing sets using `train_test_split()`, where 80% of the data is used for training, and 20% is reserved for testing. The `random_state` is set to ensure reproducibility of the results. A Naive Bayes model (`MultinomialNB`) is initialized and trained on the training data using the `fit()` method. After training, predictions are made on the test set using the `predict()` method. The model's performance is evaluated by calculating the accuracy score, which is printed as a percentage. Additionally, a detailed classification report is generated, which includes metrics such as precision, recall, and F1-score for each class. Lastly, a confusion matrix is printed to visualize the model's classification performance and to understand how well the model distinguishes between different mental health issues. This end-to-end pipeline is crucial for applying the Naive Bayes algorithm to textual data, allowing for effective classification of mental health issues based on Reddit posts.

### 8.2.5 Support Vector Machine for Classification

#### Support Vector Classifier Implementation

```

import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score,
    classification_report, confusion_matrix

# Load the preprocessed dataset
dataset = pd.read_csv('preprocessed_mental_health.csv')

# Check if 'cleaned_text' and 'mental_health_issue' columns
# exist
if 'cleaned_text' not in dataset.columns or \
    'mental_health_issue' not in dataset.columns:
    raise ValueError("The dataset must have 'cleaned_text' and 'mental_health_issue' columns.")

# Remove rows with missing values in 'cleaned_text' column
dataset.dropna(subset=['cleaned_text'], inplace=True)

# Initialize the CountVectorizer and fit/transform the
# cleaned text
SVMvectorizer = CountVectorizer()
X = SVMvectorizer.fit_transform(dataset['cleaned_text'])

# Prepare the target variable
y = dataset['mental_health_issue']

# Split the dataset into Training and Test Sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
    test_size=0.2, random_state=42)

# Initialize the Support Vector Classifier
SVMmodel = SVC(kernel='linear', C=1, random_state=42,
    probability=True) # You can adjust parameters as needed

# Train the model
SVMmodel.fit(X_train, y_train)

# Make predictions on the test set
y_pred = SVMmodel.predict(X_test)

```

## Support Vector Classifier Implementation

```

# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy:{accuracy*100:.2f}%')

# Print classification report
print("Classification_Report:\n", classification_report(
    y_test, y_pred))

# Print confusion matrix
print("Confusion_Matrix:\n", confusion_matrix(y_test, y_pred))

```

The provided code demonstrates how to apply the Support Vector Classifier (SVC) for mental health classification using a preprocessed dataset. First, the dataset is loaded, and a check is performed to ensure that it contains the necessary columns, specifically `cleaned_text` and `mental_health_issue`. If these columns are missing, an error is raised. The dataset is then cleaned by removing rows that have missing values in the `cleaned_text` column using the `dropna()` function. Next, the `CountVectorizer` is initialized and applied to the `cleaned_text` column to convert the text data into a numerical format suitable for machine learning. This is accomplished by transforming the text into a document-term matrix where each row represents a document (post) and each column represents a unique term (word). The target variable, `mental_health_issue`, is also extracted from the dataset. The dataset is then split into training and testing sets using `train_test_split()`, where 80% of the data is used for training, and 20% is reserved for testing. The `random_state` is set to ensure reproducibility of the results. A Support Vector Classifier model (SVC) is initialized with a linear kernel (`kernel='linear'`), regularization parameter `C=1`, and the probability flag set to `True` to enable probability estimates. The model is then trained on the training data using the `fit()` method. After training, predictions are made on the test set using the `predict()` method. The model's performance is evaluated by calculating the accuracy score, which is printed as a percentage. Additionally, a detailed classification report is generated, which includes metrics such as precision, recall, and F1-score for each class. Lastly, a confusion matrix is printed to visualize the model's classification performance and to understand how well the model distinguishes between different mental health issues. This end-to-end pipeline is crucial for applying the Support Vector Machine (SVM) algorithm to textual data, allowing for effective classification of mental health issues based on Reddit posts.

### 8.2.6 Random Forest for Classification

#### Random Forest Classifier Implementation

```

import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score,
    classification_report, confusion_matrix

# Load the preprocessed dataset
dataset = pd.read_csv('preprocessed_mental_health.csv')

# Check if 'cleaned_text' and 'mental_health_issue' columns
# exist
if 'cleaned_text' not in dataset.columns or \
    'mental_health_issue' not in dataset.columns:
    raise ValueError("The dataset must have 'cleaned_text' and 'mental_health_issue' columns.")

# Remove rows with missing values in 'cleaned_text' column
dataset.dropna(subset=['cleaned_text'], inplace=True)

# Initialize the CountVectorizer and fit/transform the
# cleaned text
RFvectorizer = CountVectorizer()
X = RFvectorizer.fit_transform(dataset['cleaned_text'])

# Prepare the target variable
y = dataset['mental_health_issue']

# Split the dataset into Training and Test Sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
    test_size=0.2, random_state=42)

# Initialize the Random Forest Classifier with added
# parameters
RFmodel = RandomForestClassifier(
    n_estimators=3000,           # Number of trees
    max_depth=None,             # Maximum depth of each tree
    min_samples_split=20,        # Minimum number of samples
    to split a node
)

```

## Random Forest Classifier Implementation

```

        min_samples_leaf=1,           # Minimum number of samples
        in a leaf node
        max_features='sqrt',         # Number of features to
        consider at each split
        bootstrap=False,             # Whether to use
        bootstrapping

        random_state=42             # For reproducibility
    )

# Train the model
RFmodel.fit(X_train, y_train)

# Make predictions on the test set
y_pred = RFmodel.predict(X_test)

# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy:{accuracy*100:.2f}%')

# Print classification report
print("Classification_Report:\n", classification_report(
    y_test, y_pred))

# Print confusion matrix
print("Confusion_Matrix:\n", confusion_matrix(y_test, y_pred))

```

The provided code demonstrates how to apply the Random Forest Classifier for mental health classification using a preprocessed dataset. First, the dataset is loaded, and a check is performed to ensure that it contains the necessary columns, specifically `cleaned_text` and `mental_health_issue`. If these columns are missing, an error is raised. The dataset is then cleaned by removing rows that have missing values in the `cleaned_text` column using the `dropna()` function. Next, the `CountVectorizer` is initialized and applied to the `cleaned_text` column to convert the text data into a numerical format suitable for machine learning. This is accomplished by transforming the text into a document-term matrix where each row represents a document (post) and each column represents a unique term (word). The target variable, `mental_health_issue`, is also extracted from the dataset. The dataset is then split into training and testing sets using `train_test_split()`, where 80% of the data is used for training, and 20% is reserved for testing. The `random_state` is set to ensure reproducibility of the results. A Random Forest Classifier model is initialized with several hyperparameters. Specifically, `n_estimators=3000` defines the number of decision trees in the forest. The

`max_depth=None` means the trees are allowed to grow until all leaves are pure or contain fewer than the minimum samples required to split a node. The `min_samples_split=20` and `min_samples_leaf=1` specify the minimum number of samples required to split an internal node and the minimum number of samples required to be at a leaf node, respectively. The `max_features='sqrt'` sets the maximum number of features to consider for the best split at each node to be the square root of the total number of features. `bootstrap=False` disables bootstrapping, meaning the entire dataset is used to build each tree. After initialization, the model is trained using the `fit()` method on the training data. After training, predictions are made on the test set using the `predict()` method. The model's performance is evaluated by calculating the accuracy score, which is printed as a percentage. Additionally, a detailed classification report is generated, which includes metrics such as precision, recall, and F1-score for each class. Lastly, a confusion matrix is printed to visualize the model's classification performance and to understand how well the model distinguishes between different mental health issues. This end-to-end pipeline is essential for applying the Random Forest algorithm to textual data, allowing for effective classification of mental health issues based on Reddit posts.

### 8.2.7 XGBoost for Classification

#### XGBoost Classifier Implementation

```

import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import accuracy_score,
    classification_report, confusion_matrix
import xgboost as xgb

# Load the dataset
data = pd.read_csv('preprocessed_mental_health.csv')

# Separate features and target
X = data['text']
y = data['mental_health_issue']

# Encode target labels
label_encoder = LabelEncoder()
y = label_encoder.fit_transform(y)

# Split dataset into training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
    test_size=0.2, random_state=42)

```

## XGBoost Classifier Implementation

```

# Convert text data to numerical data using TF-IDF
Vectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
tfidf = TfidfVectorizer(max_features=5000)
X_train = tfidf.fit_transform(X_train)
X_test = tfidf.transform(X_test)

# Define the XGBoost classifier
xgb_clf = xgb.XGBClassifier(objective='multi:softmax',
    num_class=5, eval_metric='mlogloss', use_label_encoder=
    False)

# Train the model
xgb_clf.fit(X_train, y_train)

# Predict on the test set
y_pred = xgb_clf.predict(X_test)

# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
report = classification_report(y_test, y_pred, target_names=
    label_encoder.classes_)

print(f"Accuracy:{accuracy*100:.2f}%")
print("Classification_Report:\n", report)

# Print confusion matrix
print("Confusion_Matrix:\n", confusion_matrix(y_test, y_pred
    ))

```

The provided code demonstrates how to apply XGBoost for mental health classification using a preprocessed dataset. First, the dataset is loaded, and the target variable (`mental_health_issue`) is separated from the feature variable (`text`). The target variable is then encoded using `LabelEncoder()` to convert the categorical labels into numerical values, which are required for machine learning algorithms. Next, the dataset is split into training and testing sets using `train_test_split()`, where 80% of the data is used for training, and 20% is reserved for testing. The `random_state` is set to ensure reproducibility of the results. The text data is then transformed into numerical features using the `TfidfVectorizer()`. This vectorizer converts the raw text into a matrix of TF-IDF (Term Frequency-Inverse Document Frequency) features, which are more suitable for training a machine learning model. The `max_features=5000` parameter limits the number of features to the 5000 most frequent terms in the dataset. The XGBoost classifier is then initialized with several hyperparameters.

The `objective='multi:softmax'` indicates that the task is a multi-class classification problem. The `num_class=5` parameter specifies the number of classes for the classification task, and `eval_metric='mlogloss'` is set to use the log loss metric for evaluation. `use_label_encoder=False` disables the automatic label encoding of target labels, as we have already manually encoded the labels. After initialization, the model is trained on the training data using the `fit()` method. After training, predictions are made on the test set using the `predict()` method. The model's performance is evaluated by calculating the accuracy score, which is printed as a percentage. Additionally, a detailed classification report is generated, which includes metrics such as precision, recall, and F1-score for each class. Lastly, a confusion matrix is printed to visualize the model's classification performance and to understand how well the model distinguishes between different mental health issues. This pipeline demonstrates the use of XGBoost in combination with TF-IDF vectorization for text classification, enabling effective classification of mental health issues based on Reddit posts.

### 8.2.8 K Nearest Neighbours for Classification

```
k-NN Classifier Implementation for Mental Health Classification

import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score,
    classification_report, confusion_matrix

# Load the preprocessed dataset
dataset = pd.read_csv('preprocessed_mental_health.csv')

# Check if 'cleaned_text' and 'mental_health_issue' columns
# exist
if 'cleaned_text' not in dataset.columns or 'mental_health_issue'
    not in dataset.columns:
    raise ValueError("The dataset must have 'cleaned_text' and 'mental_health_issue' columns.")

# Remove rows with missing values in 'cleaned_text' column
dataset.dropna(subset=['cleaned_text'], inplace=True)

# Initialize the CountVectorizer and fit/transform the cleaned
# text
KNNvectorizer = CountVectorizer()
X = KNNvectorizer.fit_transform(dataset['cleaned_text'])
```

## k-NN Classifier Implementation for Mental Health Classification

```

# Prepare the target variable
y = dataset['mental_health_issue']

# Split the dataset into Training and Test Sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
    test_size=0.2, random_state=42)

# Initialize the k-NN classifier
KNNmodel = KNeighborsClassifier(n_neighbors=5)    # You can adjust
    the number of neighbors

# Fit the model
KNNmodel.fit(X_train, y_train)

# Make predictions
y_pred = KNNmodel.predict(X_test)

# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy * 100:.2f}%')

# Print classification report
print("Classification_Report:\n", classification_report(y_test,
    y_pred))

# Print confusion matrix
print("Confusion_Matrix:\n", confusion_matrix(y_test, y_pred))

```

The code provided demonstrates how to implement a k-Nearest Neighbors (k-NN) classifier for mental health issue classification based on text data. Initially, the dataset is loaded from a CSV file, and a check is performed to ensure the presence of the necessary columns, namely `cleaned_text` and `mental_health_issue`. If any of these columns are missing, the program raises an error. The next step involves removing rows with missing text data from the `cleaned_text` column using `dropna()`. The `CountVectorizer` from `sklearn` is used to transform the textual data into a matrix of token counts, which is suitable for machine learning. This transformation converts each post (document) into a vector of word occurrences, with each unique word in the dataset being represented by a column. The target variable, `mental_health_issue`, is extracted, and the dataset is split into training and test sets using `train_test_split()`. The training set consists of 80% of the data, and the remaining 20% is used for testing. A k-NN classifier is initialized with 5 neighbors (`n_neighbors=5`), though this number can be adjusted for tuning the model. The classifier is trained on the training data using the `fit()` method. Once the model is trained, it is used to predict mental health

issues for the test set, and the accuracy of the predictions is calculated and printed. In addition to the accuracy score, a detailed classification report is generated, which includes metrics like precision, recall, and F1-score for each class. Finally, a confusion matrix is printed to help visualize the model's performance in terms of how well it classifies different mental health issues. This end-to-end pipeline demonstrates how to apply a k-NN classifier to a text classification task, specifically identifying mental health issues based on user-generated content such as social media posts.

### 8.2.9 Long Short Term Memory based Classification

#### LSTM Model Implementation

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import confusion_matrix,
    classification_report
from sklearn.preprocessing import LabelEncoder
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import
    pad_sequences
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Embedding, LSTM, Dense,
    Dropout
from tensorflow.keras.utils import to_categorical
from sklearn.model_selection import train_test_split

# Load the dataset
data = pd.read_csv('preprocessed_mental_health.csv')

# Separate features and target
X = data['text']
y = data['mental_health_issue']

# Encode target labels
label_encoder = LabelEncoder()
y = label_encoder.fit_transform(y)
y = to_categorical(y) # Convert labels to one-hot encoded
# format for multi-class classification

# Split dataset into training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
    test_size=0.2, random_state=42)

```

## LSTM Model Implementation

```

# Tokenize and pad the text sequences
vocab_size = 10000 # Set a vocabulary size
max_length = 100 # Set a max length for padding

tokenizer = Tokenizer(num_words=vocab_size, oov_token=<OOV>
    ")
tokenizer.fit_on_texts(X_train)

X_train_seq = tokenizer.texts_to_sequences(X_train)
X_test_seq = tokenizer.texts_to_sequences(X_test)

X_train_padded = pad_sequences(X_train_seq, maxlen=
    max_length, padding='post', truncating='post')
X_test_padded = pad_sequences(X_test_seq, maxlen=max_length,
    padding='post', truncating='post')

# Build the LSTM model
model = Sequential([
    Embedding(vocab_size, 128, input_length=max_length),
    LSTM(128, return_sequences=True),
    Dropout(0.2),
    LSTM(64),
    Dropout(0.2),
    Dense(64, activation='relu'),
    Dense(y.shape[1], activation='softmax') # Output layer
        with softmax for multi-class classification
])

model.compile(optimizer='adam', loss='
    categorical_crossentropy', metrics=['accuracy'])

# Train the model
history = model.fit(X_train_padded, y_train, epochs=10,
    batch_size=32, validation_data=(X_test_padded, y_test))

# Plot training history
plt.figure(figsize=(12, 4))
plt.subplot(1, 2, 1)
plt.plot(history.history['accuracy'], label='Training_
    Accuracy')
plt.plot(history.history['val_accuracy'], label='Validation_
    Accuracy')
plt.title('Model_Accuracy')
plt.xlabel('Epochs')
plt.ylabel('Accuracy')
plt.legend()

```

## LSTM Model Implementation

```

plt.subplot(1, 2, 2)
plt.plot(history.history['loss'], label='Training_Loss')
plt.plot(history.history['val_loss'], label='Validation_Loss')
plt.title('Model_Loss')
plt.xlabel('Epochs')
plt.ylabel('Loss')
plt.legend()
plt.show()

# Evaluate the model on test data
test_loss, test_accuracy = model.evaluate(X_test_padded,
                                           y_test)
print(f"Test_Accuracy:{test_accuracy*100:.2f}%")

# Generate predictions and convert back from one-hot
encoding
y_pred = model.predict(X_test_padded)
y_pred_classes = np.argmax(y_pred, axis=1)
y_test_classes = np.argmax(y_test, axis=1)

# Output classification report
print("Classification_Report:\n", classification_report(
    y_test_classes, y_pred_classes, target_names=
    label_encoder.classes_))

# Plot Confusion Matrix
cm = confusion_matrix(y_test_classes, y_pred_classes)
plt.figure(figsize=(8, 6))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues',
            xticklabels=label_encoder.classes_, yticklabels=
            label_encoder.classes_)

plt.xlabel('Predicted_Labels')
plt.ylabel('True_Labels')
plt.title('Confusion_Matrix')
plt.show()

```

This code demonstrates how to apply a Long Short-Term Memory (LSTM) neural network for text classification, specifically for predicting mental health issues based on preprocessed Reddit data. The dataset is loaded and the target variable (`mental_health_issue`) is encoded into a one-hot format for multi-class classification. The dataset is split into training and testing sets. Text data is tokenized and padded to ensure uniform input lengths for the LSTM model. The model consists of embedding layers, LSTM layers with dropout for regularization, and dense

## ASMPFMHDD

layers for classification. The model is trained on the training data, and performance is monitored using validation data. The training and validation accuracy and loss are plotted. The model is evaluated on the test data, and performance metrics such as accuracy and classification report are displayed. A confusion matrix is plotted to visualize model performance.

```
Epoch 1/10
/usr/local/lib/python3.10/dist-packages/keras/src/layers/core/embedding.py:90: UserWarning: Argument `input_length` is deprecated. Just remove it.
  warnings.warn(
465/465 ━━━━━━━━ 135s 280ms/step - accuracy: 0.6046 - loss: 1.1142 - val_accuracy: 0.6841 - val_loss: 0.7390
Epoch 2/10
465/465 ━━━━━━━━ 131s 281ms/step - accuracy: 0.6681 - loss: 0.7686 - val_accuracy: 0.6634 - val_loss: 0.8092
Epoch 3/10
465/465 ━━━━━━━━ 144s 286ms/step - accuracy: 0.6499 - loss: 0.8130 - val_accuracy: 0.6680 - val_loss: 0.7538
Epoch 4/10
465/465 ━━━━━━━━ 139s 280ms/step - accuracy: 0.6368 - loss: 0.8657 - val_accuracy: 0.6605 - val_loss: 0.7481
Epoch 5/10
465/465 ━━━━━━━━ 130s 280ms/step - accuracy: 0.6949 - loss: 0.6574 - val_accuracy: 0.6922 - val_loss: 0.6711
Epoch 6/10
465/465 ━━━━━━━━ 141s 278ms/step - accuracy: 0.7132 - loss: 0.6056 - val_accuracy: 0.6933 - val_loss: 0.6697
Epoch 7/10
465/465 ━━━━━━━━ 147s 289ms/step - accuracy: 0.7304 - loss: 0.5676 - val_accuracy: 0.7081 - val_loss: 0.6606
Epoch 8/10
465/465 ━━━━━━━━ 131s 282ms/step - accuracy: 0.7483 - loss: 0.5433 - val_accuracy: 0.7516 - val_loss: 0.6319
Epoch 9/10
465/465 ━━━━━━━━ 142s 282ms/step - accuracy: 0.8079 - loss: 0.4523 - val_accuracy: 0.7484 - val_loss: 0.6037
Epoch 10/10
465/465 ━━━━━━━━ 142s 282ms/step - accuracy: 0.8564 - loss: 0.3823 - val_accuracy: 0.8355 - val_loss: 0.5368
```

Figure 19: Output for LSTM Epochs

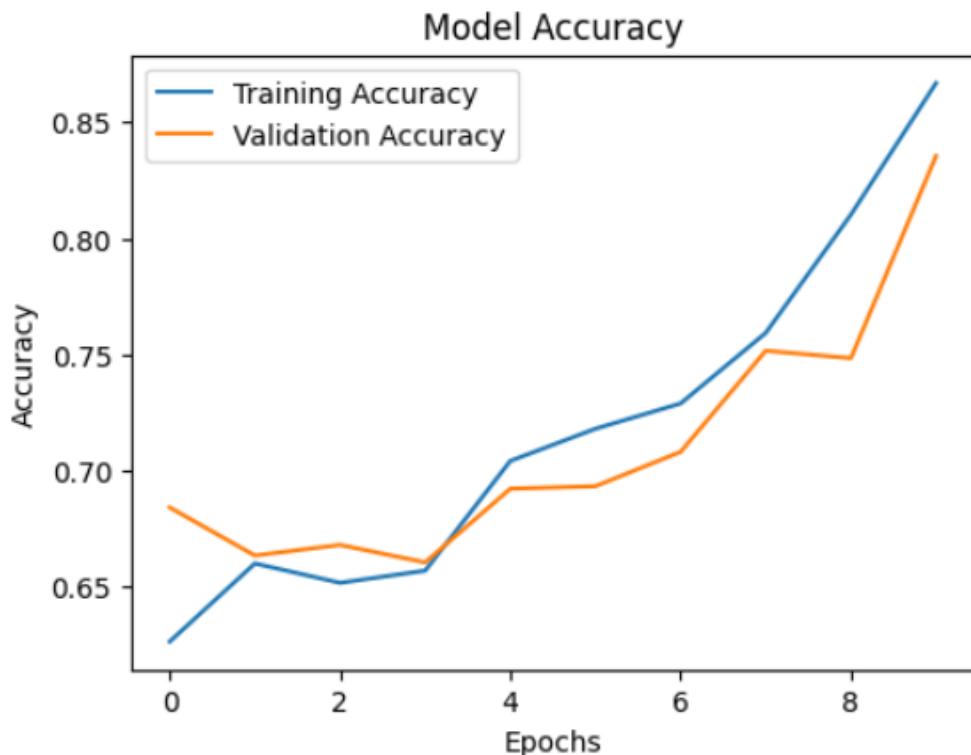


Figure 20: Accuracy v/s Epoch

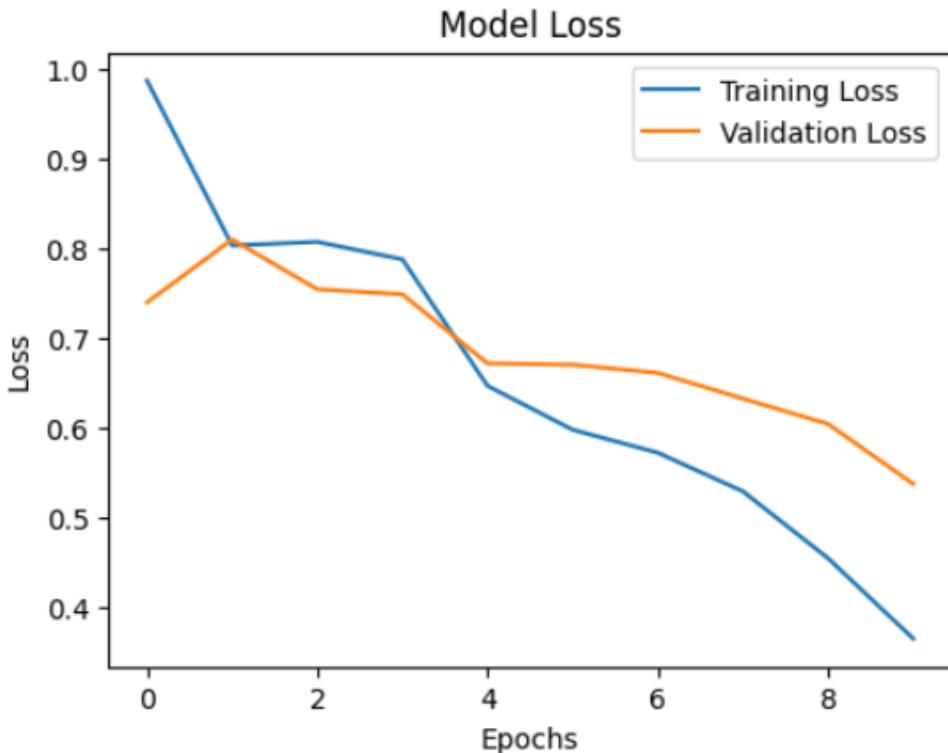


Figure 21: Loss v/s Epoch

### 8.2.10 Hyperparameter Tuning Using RandomizedSearchCV

#### Logistic Regression

```

import pandas as pd
from sklearn.model_selection import train_test_split,
    RandomizedSearchCV
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score,
    classification_report

# Load the preprocessed dataset
dataset = pd.read_csv('preprocessed_mental_health.csv')

# Check if 'cleaned_text' and 'mental_health_issue' columns
# exist
if 'cleaned_text' not in dataset.columns or 'mental_health_issue'
    not in dataset.columns:
    raise ValueError("The dataset must have 'cleaned_text' and 'mental_health_issue' columns.")

```

## Logistic Regression

```

# Remove rows with missing values in 'cleaned_text' column
dataset.dropna(subset=['cleaned_text'], inplace=True)

# Initialize the CountVectorizer and fit/transform the cleaned
text
HPLRvectorizer = CountVectorizer()
X = HPLRvectorizer.fit_transform(dataset['cleaned_text'])

# Prepare the target variable
y = dataset['mental_health_issue']

# Split the dataset into Training and Test Sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
    test_size=0.2, random_state=42)

# Initialize the Logistic Regression model
HPLRmodel = LogisticRegression(max_iter=200)

# Define the hyperparameter grid for Randomized Search
param_distributions = {
    'C': [0.001, 0.01, 0.1, 1, 10, 100],           # Inverse of
    regularization strength
    'penalty': ['l1', 'l2', 'elasticnet', 'none'],   #
    Regularization types
    'solver': ['liblinear', 'saga']                  # Solvers that
    support l1, elasticnet
}

# Initialize RandomizedSearchCV
random_search = RandomizedSearchCV(estimator=HPLRmodel,
    param_distributions=param_distributions,
    n_iter=10, scoring='accuracy',
    cv=5, n_jobs=-1,
    random_state=42)

# Fit RandomizedSearchCV
random_search.fit(X_train, y_train)

# Best hyperparameters from Random Search
print("Best_Hyperparameters:", random_search.best_params_)

# Best model from Random Search
best_modellR = random_search.best_estimator_

# Make predictions using the best model
y_pred = best_modellR.predict(X_test)

```

## Logistic Regression

```
# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy * 100:.2f}%')

# Print classification report
print("Classification_Report:\n", classification_report(y_test,
y_pred))
```

The above code demonstrates a machine learning workflow for classifying mental health issues using LogisticRegression with hyperparameter tuning. The dataset is loaded using pandas, and a check ensures the presence of 'cleaned\_text' and 'mental\_health\_issue' columns. Missing values in 'cleaned\_text' are dropped to prevent processing errors. The textual data is transformed into a numerical format using CountVectorizer, which converts text into a bag-of-words representation. The dataset is then split into training and testing sets using train\_test\_split. A logistic regression model is initialized, and the hyperparameter grid is defined, including parameters like C (inverse of regularization strength), penalty, and solver. Hyperparameter tuning is performed using RandomizedSearchCV, which searches for the best model configuration by sampling parameter combinations. The best model is used to make predictions on the test set, and evaluation metrics such as accuracy and a classification report are printed. This approach ensures optimal model performance and generalization to new data.

## K Nearest Neighbours

```
import pandas as pd
from sklearn.model_selection import train_test_split,
RandomizedSearchCV
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score,
classification_report, confusion_matrix

# Load the preprocessed dataset
dataset = pd.read_csv('preprocessed_mental_health.csv')
# Check if 'cleaned_text' and 'mental_health_issue' columns
# exist
if 'cleaned_text' not in dataset.columns or 'mental_health_issue'
    not in dataset.columns:
    raise ValueError("The dataset must have 'cleaned_text' and "
'mental_health_issue' columns.)
```

## K Nearest Neighbours

```

# Remove rows with missing values in 'cleaned_text' column
dataset.dropna(subset=['cleaned_text'], inplace=True)

# Initialize the CountVectorizer and fit/transform the cleaned
text
HPTKNNvectorizer = CountVectorizer()
X = HPTKNNvectorizer.fit_transform(dataset['cleaned_text'])

# Prepare the target variable
y = dataset['mental_health_issue']

# Split the dataset into Training and Test Sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
    test_size=0.2, random_state=42)

# Initialize the k-NN classifier
knn = KNeighborsClassifier()

# Define the hyperparameter grid for Randomized Search
param_distributions = {
    'n_neighbors': [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 14,
        15, 16, 17, 18, 19, 20],
        # Different values for number of neighbors
    'metric': ['euclidean', 'manhattan', 'chebyshev', 'minkowski
        '], # Different distance metrics
    'weights': ['uniform', 'distance'] # Weighing options for
        neighbors
}

# Initialize RandomizedSearchCV
random_search = RandomizedSearchCV(estimator=knn,
    param_distributions=param_distributions,
        n_iter=10, scoring='accuracy
            ', cv=5, n_jobs=-1,
            random_state=42)

# Fit RandomizedSearchCV
random_search.fit(X_train, y_train)

# Best hyperparameters from Random Search
print("Best_Hyperparameters:", random_search.best_params_)

# Best model from Random Search
best_knn = random_search.best_estimator_

# Make predictions using the best model
y_pred = best_knn.predict(X_test)

```

## K Nearest Neighbours

```
# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy * 100:.2f}%')

# Print classification report
print("Classification_Report:\n", classification_report(y_test,
y_pred))

# Print confusion matrix
print("Confusion_Matrix:\n", confusion_matrix(y_test, y_pred))
```

This code implements a k-Nearest Neighbors (k-NN) classifier for classifying mental health issues from text data using hyperparameter tuning via RandomizedSearchCV. First, it loads a preprocessed dataset, ensuring the existence of critical columns, `cleaned_text` (features) and `mental_health_issue` (target). Rows with missing values in the `cleaned_text` column are removed to maintain data integrity. The text is vectorized into numerical format using CountVectorizer, converting text into a bag-of-words model. The features ( $X$ ) and target ( $y$ ) are split into training (80%) and test (20%) datasets to ensure proper model evaluation. The k-NN model is initialized without predefined parameters, and a hyperparameter grid is defined for tuning. This grid explores various values for the number of neighbors (`n_neighbors`), distance metrics (`metric`), and neighbor weighting options (`weights`). RandomizedSearchCV is then used to optimize these parameters by training multiple k-NN models with combinations from the grid, evaluating them with 5-fold cross-validation, and scoring based on accuracy. The best combination of hyperparameters is identified and used to configure the final k-NN model. The model is then tested on unseen data (test set), and its performance is evaluated by calculating the accuracy, a classification report detailing precision, recall, F1-score for each class, and a confusion matrix showing true vs. predicted classifications. This approach balances computational efficiency with robustness, enabling the selection of an optimized k-NN model for mental health classification.

## Support Vector Machine

```
import pandas as pd
from sklearn.model_selection import train_test_split,
RandomizedSearchCV
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score,
classification_report
```

## Support Vector Machine

```

# Load the preprocessed dataset
dataset = pd.read_csv('preprocessed_mental_health.csv')

# Check if 'cleaned_text' and 'mental_health_issue' columns
# exist
if 'cleaned_text' not in dataset.columns or 'mental_health_issue'
    not in dataset.columns:
    raise ValueError("The dataset must have 'cleaned_text' and 'mental_health_issue' columns.")

# Remove rows with missing values in 'cleaned_text' column
dataset.dropna(subset=['cleaned_text'], inplace=True)

# Initialize the CountVectorizer and fit/transform the cleaned
# text
vectorizer = CountVectorizer()
X = vectorizer.fit_transform(dataset['cleaned_text'])

# Prepare the target variable
y = dataset['mental_health_issue']

# Split the dataset into Training and Test Sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Initialize the SVC model
model = SVC()

# Define the hyperparameter grid for Randomized Search
param_distributions = {
    'C': [0.1, 1, 10, 100],                      # Regularization
    'kernel': ['linear', 'rbf', 'poly'],           # Kernel types
    'gamma': ['scale', 'auto', 0.1, 1],            # Kernel coefficient
    'for rbf', 'poly', and 'sigmoid'
}

# Initialize RandomizedSearchCV
random_search = RandomizedSearchCV(estimator=model,
    param_distributions=param_distributions,
    n_iter=10, scoring='accuracy',
    cv=5, n_jobs=-1,
    random_state=42)

# Fit RandomizedSearchCV
random_search.fit(X_train, y_train)

```

## Support Vector Machine

```

# Best hyperparameters from Random Search
print("Best_Hyperparameters:", random_search.best_params_)

# Best model from Random Search
best_model = random_search.best_estimator_

# Make predictions using the best model
y_pred = best_model.predict(X_test)

# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy:{accuracy*100:.2f}%')

# Print classification report
print("Classification_Report:\n", classification_report(y_test,
    y_pred))

```

This code implements a Support Vector Classifier (SVC) for classifying mental health issues from text data using hyperparameter tuning via RandomizedSearchCV. It begins by loading a preprocessed dataset and ensuring that essential columns, `cleaned_text` (features) and `mental_health_issue` (target), are present. Any rows with missing values in the `cleaned_text` column are removed to ensure clean data. The text is then vectorized into a numerical format using CountVectorizer, transforming it into a bag-of-words representation. The features ( $X$ ) and target ( $y$ ) are split into training (80%) and test (20%) sets for model validation. The SVC model is initialized without predefined hyperparameters, and a hyperparameter grid is defined for tuning. This grid includes different values for the regularization parameter ( $C$ ), kernel types (`kernel`), and the kernel coefficient (`gamma`). RandomizedSearchCV is employed to find the best combination of these parameters by evaluating multiple models with various hyperparameter combinations through 5-fold cross-validation, using accuracy as the scoring metric. The best set of hyperparameters is selected to configure the final SVC model. The model is then evaluated on the test set by making predictions, and its performance is measured by calculating the accuracy, as well as generating a classification report that includes precision, recall, and F1-score for each class. This approach ensures the selection of the most optimized SVC model for the classification of mental health issues from text data.

## Naive Bayes

```

import pandas as pd
from sklearn.model_selection import train_test_split,
    RandomizedSearchCV
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score,
    classification_report, confusion_matrix
from scipy.stats import uniform

# Load the preprocessed dataset
dataset = pd.read_csv('preprocessed_mental_health.csv')

# Check if 'cleaned_text' and 'mental_health_issue' columns
# exist
if 'cleaned_text' not in dataset.columns or 'mental_health_issue'
    not in dataset.columns:
    raise ValueError("The dataset must have 'cleaned_text' and 'mental_health_issue' columns.")

# Remove rows with missing values in 'cleaned_text' column
dataset.dropna(subset=['cleaned_text'], inplace=True)

# Initialize the CountVectorizer and fit/transform the cleaned
# text
vectorizer = CountVectorizer()
X = vectorizer.fit_transform(dataset['cleaned_text'])

# Prepare the target variable
y = dataset['mental_health_issue']

# Split the dataset into Training and Test Sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
    test_size=0.2, random_state=42)

# Initialize the Naive Bayes model
naive_bayes_model = MultinomialNB()

# Define the parameter distribution for Randomized Search
param_distributions = {
    'alpha': uniform(0.001, 5.0) # Sampling alpha from a
        uniform distribution
}

```

## Naive Bayes

```

# Initialize RandomizedSearchCV
random_search = RandomizedSearchCV(estimator=naive_bayes_model,
    param_distributions=param_distributions,
        n_iter=10, scoring='accuracy',
        cv=5, n_jobs=-1,
        random_state=42)

# Fit RandomizedSearchCV
random_search.fit(X_train, y_train)

# Best hyperparameters from Random Search
print("Best_Hyperparameters:", random_search.best_params_)

# Best model from Random Search
best_model = random_search.best_estimator_

# Make predictions using the best model
y_pred = best_model.predict(X_test)

# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy * 100:.2f}%')

# Print classification report
print("Classification_Report:\n", classification_report(y_test,
    y_pred))

# Print confusion matrix
print("Confusion_Matrix:\n", confusion_matrix(y_test, y_pred))

```

This code implements a Naive Bayes classifier using the MultinomialNB algorithm to classify mental health issues from text data, with hyperparameter tuning via RandomizedSearchCV. The dataset is first loaded, ensuring the required columns, `cleaned_text` (features) and `mental_health_issue` (target), exist. Rows with missing values in the `cleaned_text` column are removed to maintain data quality. The text is vectorized using CountVectorizer, converting it into a bag-of-words representation. The features ( $X$ ) and target ( $y$ ) are then split into training (80%) and test (20%) sets to facilitate proper evaluation. The Naive Bayes model is initialized, and a parameter distribution for `alpha` is defined using a uniform distribution, sampling values between 0.001 and 5.0. RandomizedSearchCV is utilized to explore different values for `alpha` over 10 iterations, using 5-fold cross-validation and accuracy as the scoring metric. The best hyperparameters are identified and used to configure the final Naive Bayes model. The model is then tested on the unseen data (test set).

### 8.2.11 Ensemble Model

#### Using Logistic Regression (meta learner) and XGBoost

##### Model Loading and Preprocessing

```

import pickle
from sklearn.ensemble import StackingClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score,
    classification_report
import xgboost as xgb
import pandas as pd

# Load the Logistic Regression model and vectorizer
with open('LRmodel.pkl', 'rb') as file:
    lr_model = pickle.load(file)
with open('LRvectorizer.pkl', 'rb') as file:
    lr_vectorizer = pickle.load(file)
# Load the XGBoost model, label encoder, and TF-IDF vectorizer
with open('xgb_model.pkl', 'rb') as file:
    xgb_model = pickle.load(file)
with open('label_encoder.pkl', 'rb') as file:
    label_encoder = pickle.load(file)
with open('tfidf_vectorizer.pkl', 'rb') as file:
    tfidf_vectorizer = pickle.load(file)

```

This Python code is centered around preparing a machine learning workflow by loading pre-trained models and their corresponding preprocessing tools. It begins by importing essential libraries for model evaluation and stacking. The `pickle` library is specifically used to load serialized Python objects, enabling the restoration of previously trained models and utilities saved as `.pkl` files. The first part of the code loads a Logistic Regression model along with its associated vectorizer. The Logistic Regression model is deserialized from the `LRmodel.pkl` file, which represents a trained model capable of predicting outcomes based on vectorized text data. The corresponding vectorizer, stored in `LRvectorizer.pkl`, is also loaded. This vectorizer, likely a `CountVectorizer`, converts raw text into numerical feature vectors, which serve as the input for the Logistic Regression model. By loading these components, the workflow ensures consistency with the preprocessing and modeling steps used during training. In the next section, the code loads an XGBoost model alongside its associated tools. The XGBoost model is deserialized from `xgb_model.pkl` and represents another trained model designed to classify text data. To facilitate this, the code also loads a label encoder from `label_encoder.pkl`. The label encoder is responsible for mapping textual class labels (e.g., "depression" or "normal") to numerical values required by machine learning models.

Furthermore, the `tfidf_vectorizer.pkl` is loaded to preprocess text data using the Term Frequency-Inverse Document Frequency (TF-IDF) approach, which captures the importance of words in a document relative to a corpus. This ensures that the input text is transformed in a manner consistent with the original training setup of the XGBoost model. The design of this code emphasizes efficiency and consistency by leveraging pre-trained components. Instead of retraining models, it restores them from their serialized forms, enabling immediate use for predictions or integration into ensemble learning setups. This approach also ensures that text preprocessing remains identical to the training phase, avoiding discrepancies in input data handling. By combining these tools, the foundation is laid for complex tasks such as making predictions or building advanced models like stacked classifiers.

### Data Preparation

```
# Load the test dataset
data = pd.read_csv('preprocessed_mental_health.csv')

# Check if 'cleaned_text' column exists
if 'cleaned_text' not in data.columns:
    raise ValueError("The dataset must have a 'cleaned_text' column.")

# Remove rows with missing values in 'cleaned_text'
data.dropna(subset=['cleaned_text'], inplace=True)

# Split features and target
X_test = data['cleaned_text']
y_test = data['mental_health_issue']

# Encode target labels
y_test = label_encoder.transform(y_test)

# Transform the text using the respective vectorizers
X_test_lr = lr_vectorizer.transform(X_test) # Logistic Regression vectorizer
X_test_xgb = tfidf_vectorizer.transform(X_test) # XGBoost vectorizer
```

This code snippet focuses on preparing the test dataset for evaluation with pre-trained machine learning models, specifically Logistic Regression and XGBoost. First, the test dataset is loaded from a CSV file named `preprocessed_mental_health.csv` using `pd.read_csv`. The dataset is expected to have been preprocessed earlier, containing features and corresponding labels. The code then checks for the existence of a critical column named `cleaned_text`. If this column is missing, a `ValueError` is raised, ensuring that the dataset meets the re-

quired format for subsequent steps. Next, the code removes rows with missing values in the `cleaned_text` column using `data.dropna`. This step ensures the dataset is clean and free of incomplete rows that could potentially cause errors during text transformation or model evaluation. After this cleaning process, the dataset is split into features (`X_test`) and target labels (`y_test`). The features contain the preprocessed text, while the target labels represent the mental health issues corresponding to each text entry. The target labels in `y_test` are then encoded into numerical format using a preloaded `label_encoder`. This encoding is crucial because machine learning models require numerical representations of categorical labels for classification tasks. By transforming the labels to integers, the encoded labels are consistent with the classes used during the model's training phase. Finally, the text data in `X_test` is vectorized using two pre-trained vectorizers. The `lr_vectorizer`, likely a `CountVectorizer` or similar, is applied to produce feature vectors suitable for the Logistic Regression model. Simultaneously, the `tfidf_vectorizer` is applied to transform the same text data into feature vectors optimized for the XGBoost model. These transformations ensure that the test data aligns with the feature representations learned by the respective models during training. This setup allows for seamless evaluation of the Logistic Regression and XGBoost models on the same dataset, maintaining consistency with their training configurations.

### Combine Base Model Predictions

```
import numpy as np

# Get predictions from the base models
lr_predictions_proba = lr_model.predict_proba(X_test_lr)
xgb_predictions_proba = xgb_model.predict_proba(X_test_xgb)

# Combine predictions as new features
stacked_features = np.hstack((lr_predictions_proba,
    xgb_predictions_proba))
```

This Python code demonstrates a fundamental step in stacking ensemble learning by combining the prediction probabilities of base models into a single feature set. The process begins with obtaining predictions from two pre-trained models: a Logistic Regression model (`lr_model`) and an XGBoost model (`xgb_model`). These models generate class probability distributions for the test dataset, which are then used as input features for a meta-learner in an ensemble framework. First, the code uses the `predict_proba` method of each model to calculate the probabilities associated with each class for every test instance. For the Logistic Regression model, the feature set (`X_test_lr`) is passed to `lr_model.predict_proba`, producing a 2D array (`lr_predictions_proba`) where each row contains the class probabilities for a corresponding instance. Similarly, the XGBoost model processes its input fea-

tures (`X_test_xgb`) to generate a comparable 2D array (`xgb_predictions_proba`) of class probabilities. Next, the code combines these probability arrays to form a new feature set, referred to as `stacked_features`. This is accomplished using the `np.hstack` function from the NumPy library, which horizontally stacks the arrays column-wise. The resulting `stacked_features` array contains concatenated probability distributions from both models for each instance. For example, if there are three classes and 100 instances, the combined feature set will have 100 rows and six columns (three from each model). This stacked feature set serves as input to a meta-learner, enabling it to make final predictions by leveraging the strengths of both base models. By integrating predictions in this manner, the stacking approach aims to improve classification performance by reducing errors associated with individual models. The meta-learner effectively learns how to optimally combine these predictions to produce more accurate outcomes.

### Train Meta-Learner

```
# Train meta-learner using combined features (optional step if
# not pre-trained)
X_train_meta, y_train_meta = stacked_features, y_test # Example
# using test data as meta-training data

meta_learner = LogisticRegression(max_iter=5000)
meta_learner.fit(X_train_meta, y_train_meta)

# Save the trained meta-learner
with open('meta_learner.pkl', 'wb') as file:
    pickle.dump(meta_learner, file)
```

This Python code focuses on training a meta-learner in the context of stacking ensemble learning. A meta-learner is a secondary machine learning model trained on the outputs of base models, with the goal of combining their predictions to achieve higher accuracy and robustness. The first step involves defining the training data for the meta-learner. The variable `X_train_meta` is set to the combined feature set (`stacked_features`), which contains the concatenated prediction probabilities from base models like Logistic Regression and XG-Boost. These features represent the likelihood of each class for every instance, as estimated by the base models. The variable `y_train_meta` holds the true labels (`y_test`), which are used as the target variable for training the meta-learner. The code then initializes a Logistic Regression model as the meta-learner. The parameter `max_iter=5000` ensures that the optimization process has sufficient iterations to converge, particularly for complex datasets or large feature sets. The `fit` method is called to train the meta-learner using the stacked features and corresponding labels. During this training process, the meta-learner learns how to optimally

combine the prediction probabilities from the base models to produce more accurate final predictions. After training, the code saves the meta-learner to a file named `meta_learner.pkl` using the `pickle.dump` function. This step serializes the trained model, enabling its reuse for future predictions without needing to retrain it. By saving the meta-learner, the workflow becomes more efficient, as the combined expertise of the base models and the meta-learner can be leveraged in subsequent tasks. This approach exemplifies the flexibility and power of stacking ensemble learning, where a meta-learner synthesizes the strengths of multiple base models to enhance predictive performance. The ability to save the meta-learner ensures that this improved model can be deployed in practical applications with minimal computational overhead.

#### Evaluate Meta-Learner and Ensemble Model

```
from sklearn.metrics import confusion_matrix

# Load the pre-trained meta-learner
with open('meta_learner.pkl', 'rb') as file:
    meta_learner = pickle.load(file)

# Predict using the meta-learner
final_predictions = meta_learner.predict(stacked_features)

# Evaluate the ensemble model
accuracy = accuracy_score(y_test, final_predictions)
report = classification_report(y_test, final_predictions,
                                target_names=label_encoder.classes_)

print(f"Accuracy: {accuracy * 100:.2f}%")
print("Classification Report:\n", report)
# Print confusion matrix
print("Confusion Matrix:\n", confusion_matrix(y_test,
                                              final_predictions))
```

This Python code evaluates the performance of a stacking ensemble model, specifically focusing on the meta-learner. The workflow begins by loading the pre-trained meta-learner from a file named `meta_learner.pkl`. Using the `pickle.load` function, the serialized meta-learner model is deserialized and prepared for making predictions. The next step involves generating predictions with the meta-learner. The model uses the stacked features, which are the combined prediction probabilities from the base models (e.g., Logistic Regression and XGBoost). The `predict` method of the meta-learner outputs the final predicted labels for the test data. To evaluate the ensemble model's effectiveness, the code computes its accuracy using the `accuracy_score` function. Accuracy is expressed as the percentage of correctly classified instances relative to the total instances in the test set. Additionally, the

`classification_report` function is used to generate a detailed performance summary. This report includes precision, recall, F1-score, and support for each mental health class. The `target_names` parameter ensures that the report associates the predicted numerical labels with their corresponding class names, as encoded by the label encoder. The evaluation also involves generating a confusion matrix, which provides an in-depth view of the classification performance. The `confusion_matrix` function compares the true labels (`y_test`) with the predicted labels (`final_predictions`), resulting in a matrix that shows the counts of true positives, true negatives, false positives, and false negatives for each class. Finally, the accuracy, classification report, and confusion matrix are printed to the console. This comprehensive evaluation allows the user to assess the ensemble model's overall performance, identify any weaknesses (e.g., misclassifications for specific classes), and confirm the effectiveness of the meta-learner in combining predictions from base models. This structured approach to evaluation underscores the advantages of stacking ensemble learning, demonstrating how the combined insights from multiple base models and a meta-learner can lead to robust predictions and meaningful performance metrics.

In ensemble learning, combining different models like Logistic Regression and XGBoost often leads to improved results due to their complementary strengths. Logistic Regression, being a linear model, excels in problems where the relationship between features is linear. It is simple, interpretable, and computationally efficient, making it effective for text-based classification tasks when the relationship between the features (such as word counts or TF-IDF scores) and the target variable is linear. However, Logistic Regression has some limitations: it struggles with capturing complex non-linear relationships and can perform poorly when the data is highly imbalanced or when there are interactions between features that the model cannot capture. On the other hand, XGBoost (Extreme Gradient Boosting) is a powerful tree-based model that can capture complex, non-linear relationships and interactions between features, making it particularly effective for handling large, noisy, and high-dimensional datasets like text. It builds an ensemble of decision trees and iteratively improves the model by minimizing errors. However, XGBoost also has its drawbacks. It requires careful tuning of hyperparameters (e.g., learning rate, number of trees, and tree depth) and can be computationally expensive, especially for large datasets. Additionally, while it can model non-linear relationships well, it lacks the interpretability of simpler models like Logistic Regression. By combining Logistic Regression and XGBoost in an ensemble learning setup, the two models can complement each other. Logistic Regression can provide a robust baseline when the relationships between features are linear, while XGBoost can enhance the model's ability to capture more complex patterns. This hybrid approach leverages the strengths of both models—Logistic Regression's simplicity and speed, alongside XGBoost's power in modeling complex interactions.

## Using Logistic Regression, Naive Bayes, Support Vector Machine, LSTM and XGBoost (meta learner)

### Evaluate Meta-Learner and Ensemble Model

```

import pickle
import numpy as np
import pandas as pd
from sklearn.metrics import accuracy_score,
    classification_report, confusion_matrix
from tensorflow.keras.models import load_model
from tensorflow.keras.utils import pad_sequences
from xgboost import XGBClassifier

# Load Logistic Regression model and vectorizer
with open('LRmodel.pkl', 'rb') as file:
    lr_model = pickle.load(file)
with open('LRvectorizer.pkl', 'rb') as file:
    lr_vectorizer = pickle.load(file)
# Load SVM model and vectorizer
with open('SVMmodel.pkl', 'rb') as file:
    svm_model = pickle.load(file)
with open('SVMvectorizer.pkl', 'rb') as file:
    svm_vectorizer = pickle.load(file)
# Load XGBoost model, vectorizer, and label encoder
with open('xgb_model.pkl', 'rb') as file:
    xgb_model = pickle.load(file)
with open('tfidf_vectorizer.pkl', 'rb') as file:
    tfidf_vectorizer = pickle.load(file)
with open('label_encoder.pkl', 'rb') as file:
    label_encoder = pickle.load(file)
# Load LSTM model, tokenizer, and label encoder
lstm_model = load_model('lstm_model.h5')
with open('LSTM_tokenizer.pkl', 'rb') as file:
    lstm_tokenizer = pickle.load(file)
# Load Naive Bayes model and vectorizer
with open('NBmodel.pkl', 'rb') as file:
    nb_model = pickle.load(file)
with open('NBvectorizer.pkl', 'rb') as file:
    nb_vectorizer = pickle.load(file)
# Load the test dataset
data = pd.read_csv('preprocessed_mental_health.csv')

# Check if 'cleaned_text' column exists
if 'cleaned_text' not in data.columns:
    raise ValueError("The dataset must have a 'cleaned_text' column.")

```

## Evaluate Meta-Learner and Ensemble Model

```

# Remove rows with missing values in 'cleaned_text'
data.dropna(subset=['cleaned_text'], inplace=True)
# Split features and target
X_test = data['cleaned_text']
y_test = data['mental_health_issue']

# Encode target labels
y_test = label_encoder.transform(y_test)

# Process the text for each model
X_test_lr = lr_vectorizer.transform(X_test) # Logistic
    Regression vectorizer
X_test_svm = svm_vectorizer.transform(X_test) # SVM vectorizer
X_test_xgb = tfidf_vectorizer.transform(X_test) # XGBoost
    vectorizer
X_test_nb = nb_vectorizer.transform(X_test) # Naive Bayes
    vectorizer
X_test_lstm = lstm_tokenizer.texts_to_sequences(X_test) # LSTM
    tokenizer

# Pad sequences for LSTM
X_test_lstm = pad_sequences(X_test_lstm, maxlen=100, padding='post',
    truncating='post')

# Get predictions from the base models
lr_predictions_proba = lr_model.predict_proba(X_test_lr) # Logistic
    Regression probabilities
svm_predictions_proba = svm_model.predict_proba(X_test_svm) # SVM
    probabilities
xgb_predictions_proba = xgb_model.predict_proba(X_test_xgb) # XGBoost
    probabilities
nb_predictions_proba = nb_model.predict_proba(X_test_nb) # Naive
    Bayes probabilities
lstm_predictions_proba = lstm_model.predict(X_test_lstm) # LSTM
    probabilities

# Stack the predictions of all models to create the feature
matrix for the meta-learner
stacked_features = np.hstack((
    lr_predictions_proba,
    svm_predictions_proba,
    xgb_predictions_proba,
    nb_predictions_proba,
    lstm_predictions_proba
))

```

## Evaluate Meta-Learner and Ensemble Model

```

# Train the meta-learner (XGBoost as the meta-learner)
meta_learner_xgb = XGBClassifier(use_label_encoder=False,
    eval_metric='mlogloss')
meta_learner_xgb.fit(stacked_features, y_test)

# Save the trained XGBoost meta-learner
with open('meta_learner_xgb.pkl', 'wb') as file:
    pickle.dump(meta_learner_xgb, file)

# Predict using the XGBoost meta-learner
final_predictions_xgb = meta_learner_xgb.predict(
    stacked_features)

# Evaluate the XGBoost ensemble model
accuracy_xgb = accuracy_score(y_test, final_predictions_xgb)
report_xgb = classification_report(y_test, final_predictions_xgb,
    target_names=label_encoder.classes_)

print(f"XGBoost_Meta-Learner_Accuracy:{accuracy_xgb*100:.2f}%")
print("Classification_Report_(XGBoost):\n", report_xgb)
print("Confusion_Matrix_(XGBoost):\n", confusion_matrix(y_test,
    final_predictions_xgb))

```

This script is designed to load several pre-trained machine learning models, apply them to a dataset of mental health-related text data, combine their predictions using a meta-learner approach, and evaluate the performance of the meta-learner model. The models used in this pipeline include logistic regression, support vector machine (SVM), XGBoost, Naive Bayes, and a deep learning-based Long Short-Term Memory (LSTM) model. The purpose of this setup is to leverage the individual strengths of different models to improve classification accuracy for predicting mental health issues based on text data. First, the script imports various necessary libraries like `pickle`, `numpy`, `pandas`, and machine learning tools from `sklearn` and `xgboost`. These libraries help load, transform, and evaluate models and datasets. It then proceeds to load each pre-trained model and associated preprocessing objects (like vectorizers and tokenizers) from disk using `pickle.load()`. Specifically, models for logistic regression, SVM, Naive Bayes, XGBoost, and LSTM are loaded along with their respective vectorizers or tokenizers. These models are already trained and saved previously, allowing the script to reuse them for making predictions on new data. Next, the script loads the test dataset, which contains preprocessed mental health-related text in a column called `cleaned_text`. It ensures that no rows with missing values are present in this column by dropping such rows. The target variable, `mental_health_issue`, which indicates whether a mental health issue is present, is also

extracted. This target variable is then encoded using a label encoder that was previously saved. Encoding is essential because machine learning models require numerical labels rather than categorical ones. For each model, the script prepares the test data differently. The text data is processed using the corresponding vectorizer or tokenizer to convert the text into numerical features that the models can understand. For models like logistic regression, SVM, Naive Bayes, and XGBoost, the test data is transformed into feature matrices using their respective vectorizers, such as `lr_vectorizer` for logistic regression and `svm_vectorizer` for the SVM model. In contrast, the LSTM model requires tokenized sequences, so the `lstm_tokenizer` is used to convert the text into sequences of integers. Since the LSTM model requires fixed-length input, the sequences are padded to a maximum length of 100 tokens using `pad_sequences`.

After transforming the data, the script proceeds to generate predictions from each model. The predictions for each model are given in the form of probabilities (instead of class labels), using methods like `predict_proba()`. These probabilities represent the model's confidence in each class for a given input. For the LSTM model, the prediction probabilities are generated using `model.predict()`, which outputs the probability distribution across possible classes. With these prediction probabilities in hand, the script constructs a feature matrix for the meta-learner, which is an ensemble learning method that combines the predictions of the base models. This meta-learner is an XGBoost classifier, which is a powerful machine learning algorithm known for its efficiency and predictive accuracy. The probabilities predicted by each base model are stacked together using `np.hstack()`, creating a combined feature vector for each sample in the dataset. These stacked features are then used to train the XGBoost meta-learner model. The trained meta-learner, now stored in `meta_learner_xgb`, is used to make final predictions. These predictions represent the ensemble decision, which ideally benefits from the strengths of all the base models. The script then evaluates the performance of the meta-learner using accuracy, classification report, and confusion matrix. Accuracy gives an overall percentage of correctly classified instances, while the classification report provides detailed metrics like precision, recall, and F1-score for each class. The confusion matrix helps visualize the performance of the classifier, showing how many samples were correctly or incorrectly classified for each class. Finally, the trained meta-learner is saved to disk using `pickle.dump()` for future use, so it can be reused to make predictions on new data without retraining. This entire process ensures that the ensemble model, which combines the strengths of different classifiers, is capable of making robust predictions on the mental health issue dataset.

## Using Logistic Regression (meta learner), Naive Bayes, Support Vector Machine, LSTM and XGBoost

### Evaluate Meta-Learner and Ensemble Model

```

import pickle
import numpy as np
import pandas as pd
from sklearn.metrics import accuracy_score,
    classification_report, confusion_matrix
from tensorflow.keras.models import load_model
from tensorflow.keras.utils import pad_sequences
from sklearn.linear_model import LogisticRegression # Import
    Logistic Regression
from sklearn.model_selection import cross_val_score,
    StratifiedKFold # For cross-validation

# Load Logistic Regression model and vectorizer
with open('LRmodel.pkl', 'rb') as file:
    lr_model = pickle.load(file)
with open('LRvectorizer.pkl', 'rb') as file:
    lr_vectorizer = pickle.load(file)

# Load SVM model and vectorizer
with open('SVMmodel.pkl', 'rb') as file:
    svm_model = pickle.load(file)
with open('SVMvectorizer.pkl', 'rb') as file:
    svm_vectorizer = pickle.load(file)

# Load XGBoost model, vectorizer, and label encoder
with open('xgb_model.pkl', 'rb') as file:
    xgb_model = pickle.load(file)
with open('tfidf_vectorizer.pkl', 'rb') as file:
    tfidf_vectorizer = pickle.load(file)
with open('label_encoder.pkl', 'rb') as file:
    label_encoder = pickle.load(file)

# Load LSTM model, tokenizer, and label encoder
lstm_model = load_model('lstm_model.h5')
with open('LSTM_tokenizer.pkl', 'rb') as file:
    lstm_tokenizer = pickle.load(file)

# Load Naive Bayes model and vectorizer
with open('NBmodel.pkl', 'rb') as file:
    nb_model = pickle.load(file)
with open('NBvectorizer.pkl', 'rb') as file:
    nb_vectorizer = pickle.load(file)

```

## Evaluate Meta-Learner and Ensemble Model

```

# Load the test dataset
data = pd.read_csv('preprocessed_mental_health.csv')

# Check if 'cleaned_text' column exists
if 'cleaned_text' not in data.columns:
    raise ValueError("The dataset must have a 'cleaned_text' column.")

# Remove rows with missing values in 'cleaned_text'
data.dropna(subset=['cleaned_text'], inplace=True)

# Split features and target
X_test = data['cleaned_text']
y_test = data['mental_health_issue']

# Encode target labels
y_test = label_encoder.transform(y_test)

# Process the text for each model
X_test_lr = lr_vectorizer.transform(X_test) # Logistic Regression vectorizer
X_test_svm = svm_vectorizer.transform(X_test) # SVM vectorizer
X_test_xgb = tfidf_vectorizer.transform(X_test) # XGBoost vectorizer
X_test_nb = nb_vectorizer.transform(X_test) # Naive Bayes vectorizer
X_test_lstm = lstm_tokenizer.texts_to_sequences(X_test) # LSTM tokenizer

# Pad sequences for LSTM
X_test_lstm = pad_sequences(X_test_lstm, maxlen=100, padding='post', truncating='post')

# Get predictions from the base models
lr_predictions_proba = lr_model.predict_proba(X_test_lr) # Logistic Regression probabilities
svm_predictions_proba = svm_model.predict_proba(X_test_svm) # SVM probabilities
xgb_predictions_proba = xgb_model.predict_proba(X_test_xgb) # XGBoost probabilities
nb_predictions_proba = nb_model.predict_proba(X_test_nb) # Naive Bayes probabilities
lstm_predictions_proba = lstm_model.predict(X_test_lstm) # LSTM probabilities

```

## Evaluate Meta-Learner and Ensemble Model

```

# Stack the predictions of all models to create the feature
# matrix for the meta-learner
stacked_features = np.hstack((
    lr_predictions_proba,
    svm_predictions_proba,
    xgb_predictions_proba,
    nb_predictions_proba,
    lstm_predictions_proba
))

# Define the Logistic Regression model for the meta-learner
meta_learner_lr = LogisticRegression(max_iter=5000)

# Cross-validation setup (Stratified K-Fold for class balance)
cv = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)

# Perform cross-validation to evaluate the meta-learner model
cross_val_scores = cross_val_score(meta_learner_lr,
                                     stacked_features, y_test, cv=cv, scoring='accuracy')

# Print the cross-validation results
print(f"Cross-Validation_Accuracy_Scores:{cross_val_scores}")
print(f"Mean_Cross-Validation_Accuracy:{np.mean(
    cross_val_scores)*100:.2f}%")

# Train the Logistic Regression meta-learner on the full dataset
meta_learner_lr.fit(stacked_features, y_test)

# Save the trained Logistic Regression meta-learner
with open('meta_learner_lr.pkl', 'wb') as file:
    pickle.dump(meta_learner_lr, file)

# Predict using the Logistic Regression meta-learner
final_predictions_lr = meta_learner_lr.predict(stacked_features)

# Evaluate the Logistic Regression ensemble model
accuracy_lr = accuracy_score(y_test, final_predictions_lr)
report_lr = classification_report(y_test, final_predictions_lr,
                                   target_names=label_encoder.classes_)

print(f"Logistic_Regression_Meta-Learner_Accuracy:{accuracy_lr*100:.2f}%")
print("Classification_Report_(Logistic_Regression):\n",
      report_lr)
print("Confusion_Matrix_(Logistic_Regression):\n",
      confusion_matrix(y_test, final_predictions_lr))

```

The code begins by importing the necessary libraries, which include pickle for loading the saved models and vectorizers, numpy and pandas for data manipulation, tensorflow for the LSTM model, and various machine learning tools from sklearn like LogisticRegression and StratifiedKFold for cross-validation. The first part of the code loads the pre-trained models and their corresponding preprocessing components from disk using pickle for models such as logistic regression, support vector machine, Naive Bayes, and XGBoost, and tensorflow for the LSTM model. The models include their respective vectorizers, tokenizers, and the label encoder, which were previously trained and saved. These models are designed to predict mental health issues based on text data, and each model uses different techniques to process the input data. After loading the models, the code proceeds to load the test dataset, which is stored in a CSV file containing preprocessed mental health-related text in the column `cleaned_text`. It ensures that any missing values in the `cleaned_text` column are removed before proceeding with the analysis. The target variable, `mental_health_issue`, is then encoded into numerical labels using the label encoder to prepare the data for model prediction. The test text data is then transformed using the appropriate vectorizer for each model. For instance, the logistic regression and SVM models use their respective vectorizers to convert the text into feature matrices, while the LSTM model tokenizes the text into integer sequences and pads the sequences to a fixed length of 100 tokens to ensure uniform input size. Once the text data is processed for each model, the code proceeds to generate prediction probabilities for each base model using methods like `predict_proba` for logistic regression, SVM, XGBoost, and Naive Bayes, while the LSTM model uses the `predict` method to output class probabilities. These probabilities represent the models' confidence in predicting each possible class for each input. The predictions from each model are then stacked together using numpy to create a combined feature matrix, which serves as input for the meta-learner model. The meta-learner, in this case, is a logistic regression model that combines the strengths of the base models to improve the overall prediction accuracy. Before training the meta-learner, a cross-validation process is set up using StratifiedKFold, which ensures that each fold of the cross-validation maintains the class distribution of the target variable. This step is important to assess the stability and generalization of the meta-learner model across different subsets of the data. Cross-validation is performed using the `cross_val_score` method from sklearn, which calculates the accuracy for each fold and prints the mean cross-validation accuracy score. After evaluating the model through cross-validation, the meta-learner is trained on the full dataset, and the model is saved to disk using pickle for future use. The trained meta-learner is then used to make final predictions by predicting the class labels based on the stacked features. The model's performance is evaluated using `accuracy_score`, `classification_report`, and `confusion_matrix` from sklearn to provide a comprehensive evaluation of the model's performance, including precision, recall, and F1-score for each class, along with a confusion matrix.

## 9 Results and Analysis

The metrics used for evaluating the performance of the classification models include Precision, Recall, F1-Score, and Support, along with the Confusion Matrix. These metrics are crucial for assessing how well the models are able to differentiate between various classes, providing insight into their accuracy, ability to capture relevant instances, and the overall balance between precision and recall. By analyzing these metrics, a comprehensive understanding of the model's performance can be obtained, enabling informed decisions for further optimization and tuning.

### 9.1 Classification Metrics and Confusion Matrix

- **Precision:** Precision is the ratio of correctly predicted positive observations to the total predicted positives. It can be calculated as:

$$\text{Precision} = \frac{TP}{TP + FP}$$

where  $TP$  represents True Positives, and  $FP$  represents False Positives.

- **Recall:** Recall is the ratio of correctly predicted positive observations to all observations in the actual class:

$$\text{Recall} = \frac{TP}{TP + FN}$$

where  $TP$  is True Positives, and  $FN$  is False Negatives.

- **F1-Score:** F1-Score is the harmonic mean of Precision and Recall, calculated as:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- **Support:** Support refers to the number of actual occurrences of each class in the dataset:

$$\text{Support} = \text{Number of samples in the true class}$$

- **Confusion Matrix:** A confusion matrix is used to evaluate the performance of a classification model. It is structured as follows:

$$\begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix}$$

where:

- $TP$  = True Positives
- $FP$  = False Positives
- $FN$  = False Negatives
- $TN$  = True Negatives

## 9.2 Results of Logistic Regression

**Logistic Regression Classification Report**

Class	Precision	Recall	F1-Score	Support
Anxiety	0.83	0.77	0.80	379
Bipolar	0.74	0.55	0.63	384
Depression	0.76	0.76	0.76	373
Normal	0.92	0.99	0.95	2183
PTSD	0.87	0.77	0.82	394
<b>Accuracy</b>	87.66%			
<b>Macro Avg</b>	0.82	0.77	0.79	3713
<b>Weighted Avg</b>	0.87	0.88	0.87	3713

**Confusion Matrix for Logistic Regression Model**

	Anxiety	Bipolar	Depression	Normal	PTSD
Anxiety	291	18	27	24	19
Bipolar	7	213	29	125	10
Depression	26	31	283	18	15
Normal	3	10	4	2165	1
PTSD	24	16	29	22	303

**ROC Curve Areas for Each Class**

Class	ROC AUC
Anxiety	0.95
Bipolar	0.92
Depression	0.96
Normal	0.99
PTSD	0.95

The Logistic Regression model performed well with an overall accuracy of 87.66%, indicating that the model correctly classified the majority of the instances. The classification report shows high precision, recall, and F1-scores for the 'Normal' class, which was expected due to its large number of instances. However, the 'Bipolar' and 'Anxiety' classes have lower recall and F1-scores, suggesting that the model struggles more with these classes. The confusion matrix highlights the misclassifications. For example, 'Anxiety' is often confused with 'Depression' and 'PTSD,' while the 'Normal' class is well-separated from the other classes. The large number of instances in the 'Normal' class could have contributed to the high accuracy but also to the imbalance in performance across other classes. The ROC curve AUC scores indicate that the model performs well in distinguishing between the classes. The 'Normal' class has the highest AUC (0.99), which is expected due to the large proportion of 'Normal' instances. Other classes, like 'Anxiety' and 'Depression,' also have high AUC scores (0.95 and 0.96, respectively), indicating that the model is capable of distinguishing them effectively.

## ASMPFMHDD

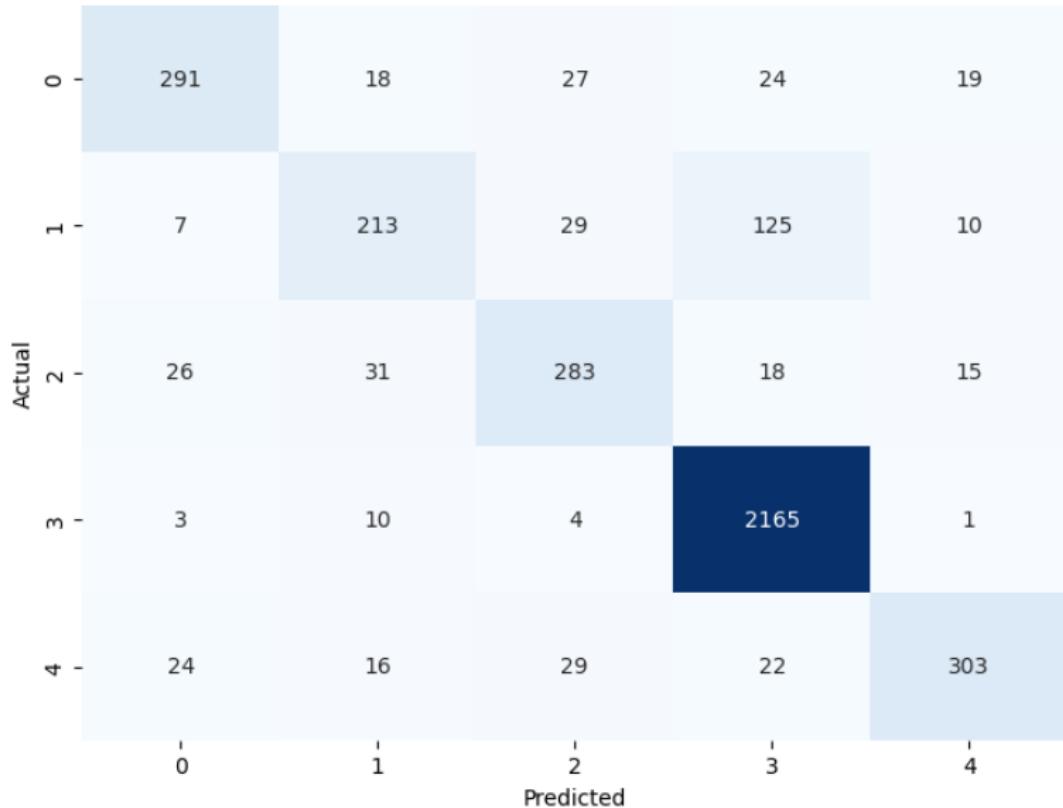


Figure 22: Confusion Matrix (Logistic Regression)

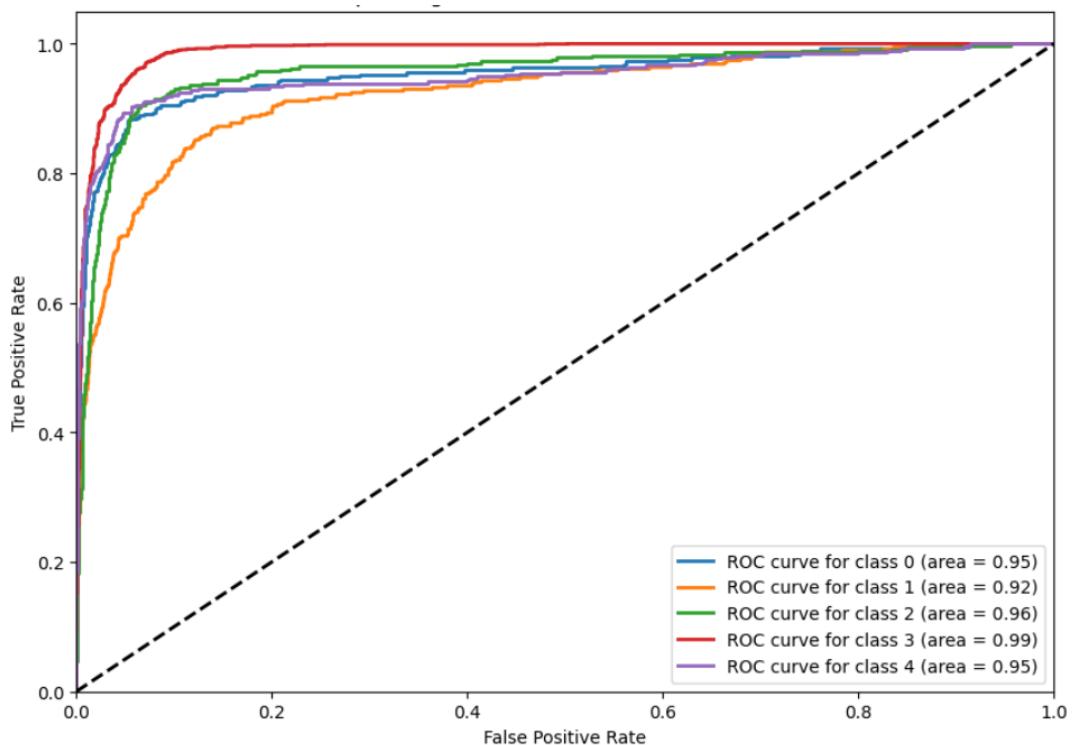


Figure 23: ROC AUC (Logistic Regression)

### 9.3 Results of Naive Bayes

**Naive Bayes Classification Report**

<b>Class</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>	<b>Support</b>
Anxiety	0.70	0.73	0.72	379
Bipolar	0.83	0.45	0.58	384
Depression	0.59	0.87	0.70	373
Normal	0.96	0.92	0.94	2183
PTSD	0.71	0.83	0.76	394
<b>Accuracy</b>	83.63%			
<b>Macro Avg</b>	0.76	0.76	0.74	3713
<b>Weighted Avg</b>	0.85	0.84	0.84	3713

**Confusion Matrix for Naive Bayes Model**

	<b>Anxiety</b>	<b>Bipolar</b>	<b>Depression</b>	<b>Normal</b>	<b>PTSD</b>
<b>Anxiety</b>	278	4	63	3	31
<b>Bipolar</b>	30	171	62	86	35
<b>Depression</b>	26	6	323	0	18
<b>Normal</b>	38	21	68	2006	50
<b>PTSD</b>	24	5	34	4	327

**ROC Curve Areas for Each Class**

<b>Class</b>	<b>ROC AUC</b>
Anxiety	0.92
Bipolar	0.89
Depression	0.94
Normal	0.99
PTSD	0.94

The Naive Bayes model achieved an overall accuracy of 83.63%, indicating a reasonable performance. The classification report shows strong results for the Normal class, with a precision of 0.96 and a recall of 0.92, resulting in an F1-score of 0.94. However, the Bipolar class exhibits much lower recall (0.45) and F1-score (0.58), suggesting that the model struggles to accurately identify Bipolar instances. Misclassifications are more frequent for Bipolar, often being confused with Depression and PTSD. The Anxiety class, although having decent precision (0.70), shows a lower recall (0.73), indicating that it also faces challenges in classification. The Confusion Matrix highlights that the model has difficulty distinguishing Bipolar and Depression from each other, with a large number of misclassifications across these classes. On the other hand, the Normal class is well-separated and correctly classified, with 2006 true positives, which likely contributes to the high overall accuracy. PTSD also shows relatively good classification results, with misclassifications being less frequent. The ROC AUC Curve Areas show that the model

## ASMPFMHDD

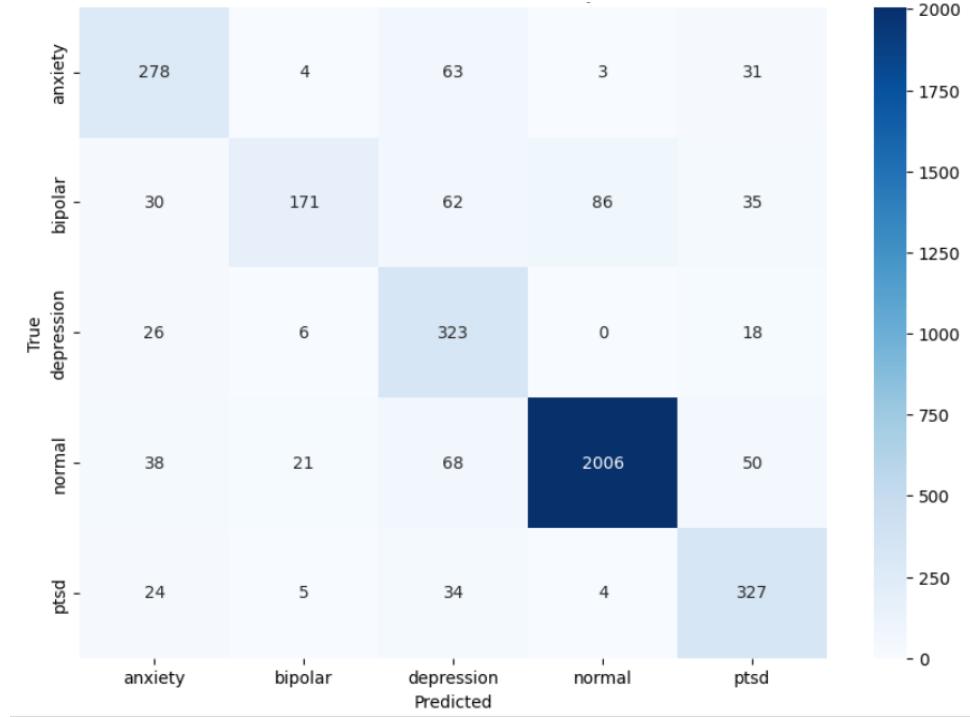


Figure 24: Confusion Matrix (Naive Bayes)

performs well for most classes, with the Normal class having the highest AUC score (0.99), followed by Depression and PTSD with AUCs of 0.94. While the model performs well for distinguishing between classes like Normal and Depression, the lower AUC for Bipolar (0.89) indicates that there may still be room for improvement in distinguishing this class from others.

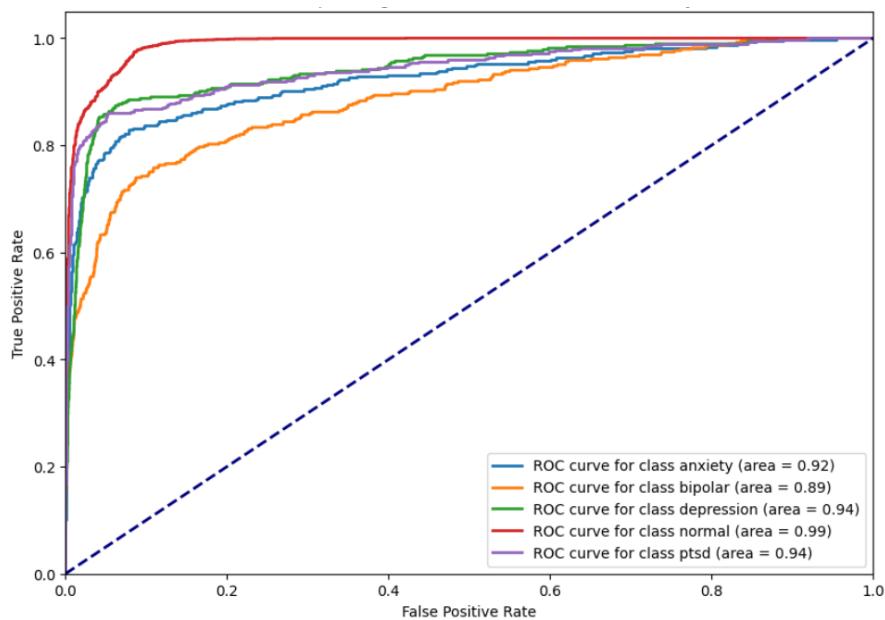


Figure 25: ROC AUC (Naive Bayes)

## 9.4 Results of Support Vector Machine

**SVM Classification Report**

<b>Class</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>	<b>Support</b>
Anxiety	0.72	0.76	0.74	379
Bipolar	0.62	0.61	0.61	384
Depression	0.74	0.71	0.72	373
Normal	0.94	0.95	0.95	2183
PTSD	0.78	0.74	0.76	394
<b>Accuracy</b>	85.13%			
<b>Macro Avg</b>	0.76	0.75	0.76	3713
<b>Weighted Avg</b>	0.85	0.85	0.85	3713

**Confusion Matrix for SVM Model**

	<b>Anxiety</b>	<b>Bipolar</b>	<b>Depression</b>	<b>Normal</b>	<b>PTSD</b>
<b>Anxiety</b>	287	22	29	13	28
<b>Bipolar</b>	20	233	29	88	14
<b>Depression</b>	37	31	265	11	29
<b>Normal</b>	20	62	8	2084	9
<b>PTSD</b>	34	26	29	13	292

**ROC Curve Areas for Each Class**

<b>Class</b>	<b>ROC AUC</b>
Anxiety	0.96
Bipolar	0.90
Depression	0.96
Normal	0.98
PTSD	0.96

The Support Vector Machine (SVM) model achieved an accuracy of 85.13%, demonstrating strong performance in classifying the various mental health conditions. The classification report indicates that the Normal class has the highest precision (0.94) and recall (0.95), resulting in an F1-score of 0.95, showing that the model is particularly good at identifying instances of normal behavior. However, the Bipolar class shows lower performance, with a precision of 0.62 and recall of 0.61, suggesting that the model struggles more with identifying bipolar disorder instances. The recall for the Anxiety class is relatively high (0.76), though precision is somewhat lower (0.72), indicating a balanced classification performance for this condition. The confusion matrix reveals that the model is generally accurate, with the Normal class being correctly classified most of the time (2084 true positives). However, some misclassifications occur for classes like Bipolar, Depression, and PTSD, with notable misclassifications of Bipolar as Depression and Normal. These misclassifications are likely affecting the overall recall for certain

## ASMPFMHDD



Figure 26: Confusion Matrix (SVM)

classes. The ROC AUC curve areas demonstrate that the model has excellent performance in distinguishing between most classes. The Normal class has the highest AUC of 0.98, followed by Anxiety, Depression, and PTSD with AUCs of 0.96. Bipolar, although still high at 0.90, lags behind the other classes, reflecting the model's challenge in distinguishing Bipolar from other conditions.

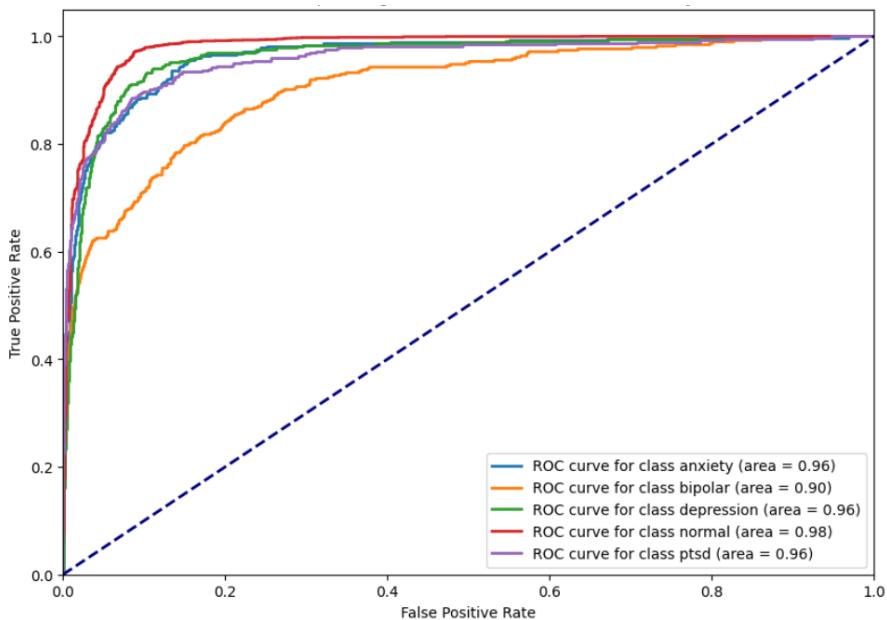


Figure 27: ROC AUC (SVM)

## 9.5 Results of Random Forest

**Random Forest Classification Report**

Class	Precision	Recall	F1-Score	Support
Anxiety	0.81	0.70	0.75	379
Bipolar	0.93	0.47	0.62	384
Depression	0.72	0.77	0.74	373
Normal	0.88	1.00	0.93	2183
PTSD	0.92	0.74	0.82	394
<b>Accuracy</b>	86.00%			
<b>Macro Avg</b>	0.85	0.73	0.77	3713
<b>Weighted Avg</b>	0.86	0.86	0.85	3713

**Confusion Matrix for Random Forest Model**

	Anxiety	Bipolar	Depression	Normal	PTSD
Anxiety	264	3	34	68	10
Bipolar	12	180	46	141	5
Depression	26	3	286	48	10
Normal	3	6	1	2173	0
PTSD	19	2	30	53	290

**ROC Curve Areas for Each Class**

Class	ROC AUC
Anxiety	0.96
Bipolar	0.89
Depression	0.97
Normal	0.97
PTSD	0.97

The Random Forest model achieved an accuracy of 86.00%, indicating a strong performance in classifying the mental health conditions. The classification report shows that the Normal class achieved the highest recall (1.00) and a precision of 0.88, resulting in a high F1-score of 0.93. This reflects the model's ability to correctly classify the majority of the Normal instances. The Bipolar class, however, shows a much lower recall (0.47), which indicates that the model has difficulty identifying instances of Bipolar disorder, as reflected by its F1-score of 0.62. The Anxiety and PTSD classes have relatively balanced performance, with moderate precision and recall values. The confusion matrix illustrates the distribution of misclassifications. The Normal class is correctly classified almost entirely (2173 true positives), while other classes like Bipolar and PTSD exhibit significant misclassifications, particularly Bipolar, which is often misclassified as Depression and Normal. The ROC AUC scores for all classes are quite high, indicating

## ASMPFMHDD

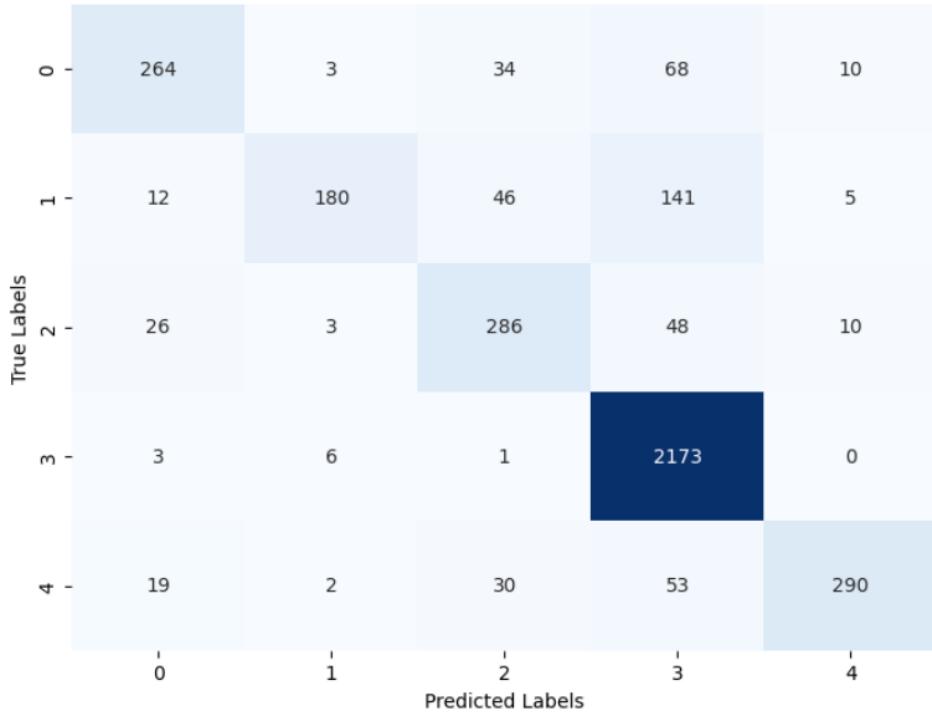


Figure 28: Confusion Matrix (Random Forest)

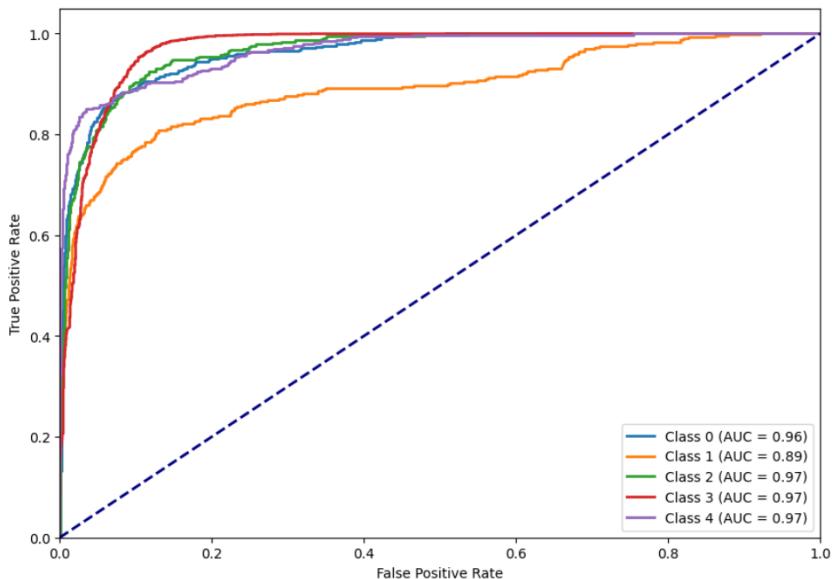


Figure 29: ROC AUC (Random Forest)

that the model is effective at distinguishing between the classes. The Anxiety, Depression, Normal, and PTSD classes each have AUC values above 0.96, with the Normal and Depression classes achieving 0.97. Bipolar has the lowest AUC at 0.89, which corresponds to its lower classification performance. These AUC values suggest that the Random Forest model is proficient in distinguishing between most of the classes, though it faces challenges in identifying Bipolar disorder.

## 9.6 Results of XGBoost

**XGBoost Classification Report**

Class	Precision	Recall	F1-Score	Support
Anxiety	0.81	0.74	0.77	403
Bipolar	0.77	0.62	0.69	397
Depression	0.72	0.81	0.76	387
Normal	0.93	0.98	0.95	2137
PTSD	0.86	0.75	0.80	396
<b>Accuracy</b>	87.39%			
<b>Macro Avg</b>	0.82	0.78	0.80	3720
<b>Weighted Avg</b>	0.87	0.87	0.87	3720

**Confusion Matrix for XGBoost Model**

	Anxiety	Bipolar	Depression	Normal	PTSD
Anxiety	297	20	39	24	23
Bipolar	13	248	36	92	8
Depression	30	13	313	17	14
Normal	3	28	8	2096	2
PTSD	22	12	36	29	297

**ROC Curve Areas for Each Class**

Class	ROC AUC
Anxiety	0.97
Bipolar	0.95
Depression	0.97
Normal	0.99
PTSD	0.97

The XGBoost model achieved an accuracy of 87.39%, demonstrating strong overall performance. The classification report indicates high precision and recall for the 'Normal' class, which is likely due to its substantial representation in the dataset. The 'Anxiety' and 'PTSD' classes also showed reasonable results, with the 'Anxiety' class achieving a precision of 0.81 and a recall of 0.74. However, the 'Bipolar' class exhibited a lower recall (0.62), suggesting that the model struggles more with this class. The confusion matrix highlights a well-separated 'Normal' class, while the 'Bipolar' and 'PTSD' classes experience more confusion with other categories. The ROC AUC scores further emphasize the model's good discriminatory capability, with AUC values of 0.97 for 'Anxiety,' 'Depression,' and 'PTSD,' and 0.99 for 'Normal.' These scores indicate that the model is proficient at distinguishing between different mental health conditions in the dataset.

## ASMPFMHDD

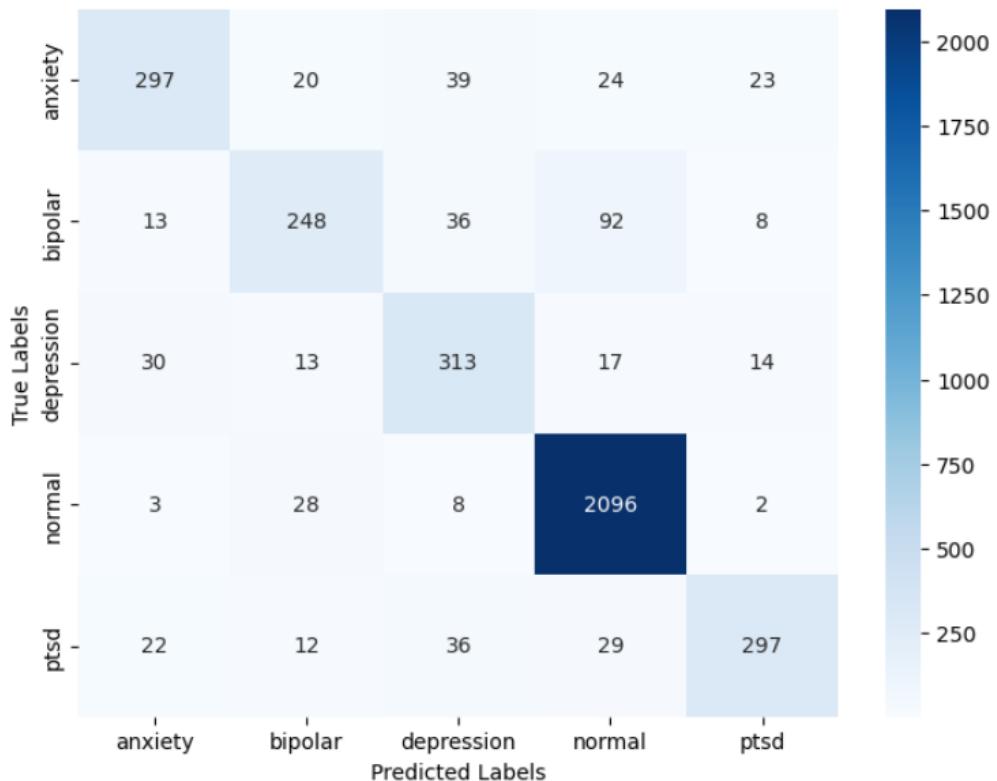


Figure 30: Confusion Matrix (XGBoost)

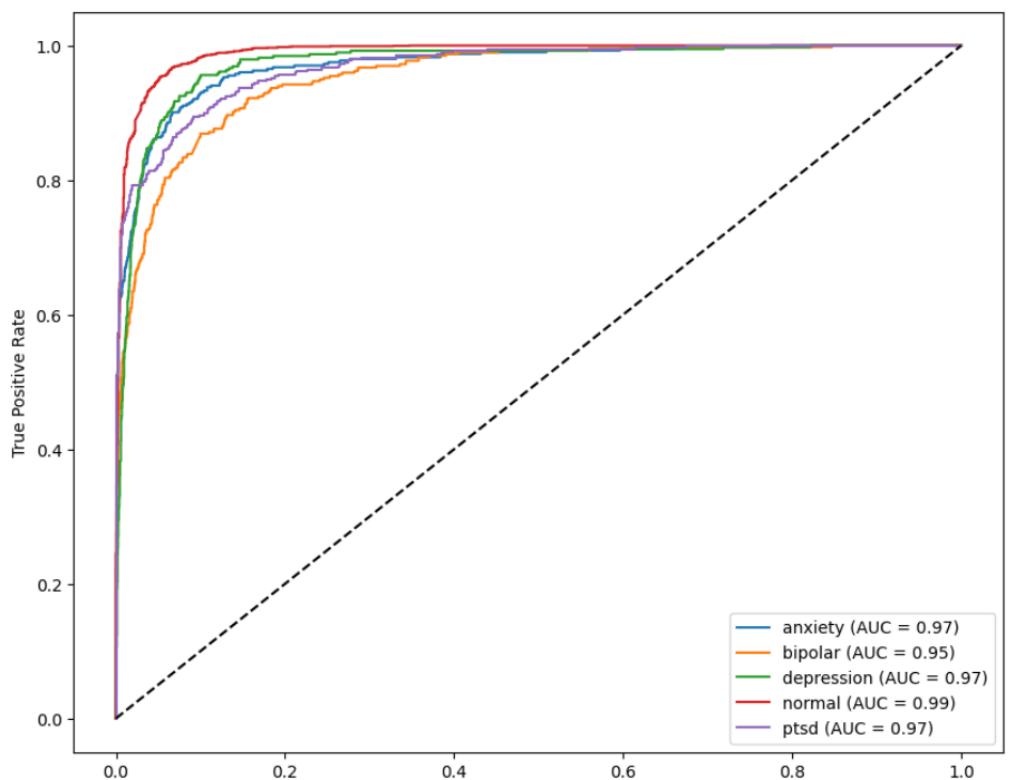


Figure 31: ROC AUC (XGBoost)

## 9.7 Results of KNN

**KNN Classification Report**

Class	Precision	Recall	F1-Score	Support
Anxiety	0.58	0.31	0.40	379
Bipolar	0.18	0.59	0.28	384
Depression	0.47	0.39	0.43	373
Normal	0.79	0.69	0.73	2183
PTSD	0.80	0.09	0.16	394
<b>Accuracy</b>	54.46%			
<b>Macro Avg</b>	0.56	0.41	0.40	3713
<b>Weighted Avg</b>	0.67	0.54	0.56	3713

**Confusion Matrix for KNN Model**

	Anxiety	Bipolar	Depression	Normal	PTSD
Anxiety	117	122	35	101	4
Bipolar	7	226	40	108	3
Depression	24	129	145	74	1
Normal	11	657	15	1499	1
PTSD	41	120	74	124	35

**ROC Curve Areas for Each Class**

Class	ROC AUC
Anxiety	0.73
Bipolar	0.67
Depression	0.77
Normal	0.80
PTSD	0.67

The KNN model achieved an accuracy of 54.46%, which is considerably lower than the XGBoost model. The classification report reveals a significant imbalance in performance across classes. The 'Normal' class achieved relatively high precision (0.79) and recall (0.69), reflecting the model's tendency to perform well with dominant classes. However, the 'Anxiety' and 'PTSD' classes performed poorly, especially with PTSD having a very low recall of 0.09, indicating substantial misclassification. The confusion matrix suggests that the model struggles with distinguishing between some classes, especially 'Anxiety' and 'Bipolar,' where confusion with other categories is high. The ROC AUC scores are relatively lower compared to the XGBoost model, with 'Normal' having the highest AUC of 0.80. These results indicate that while the KNN model offers a lower overall performance, it still provides a reasonable level of discrimination for the 'Normal' class.

## ASMPFMHDD

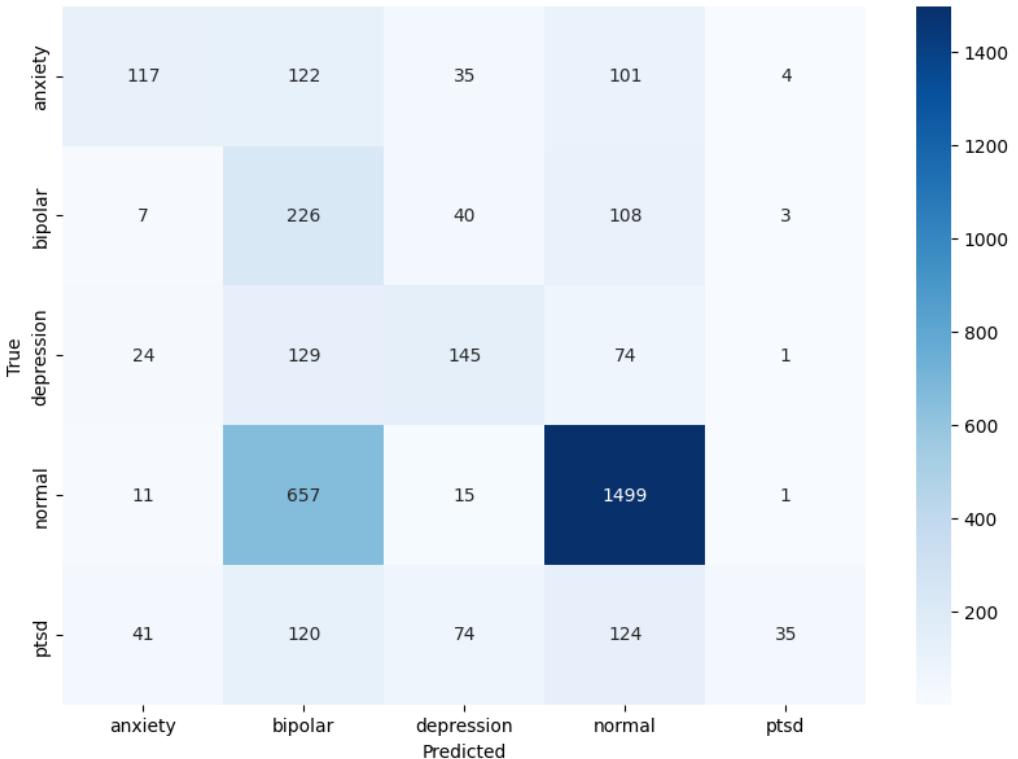


Figure 32: Confusion Matrix (KNN)

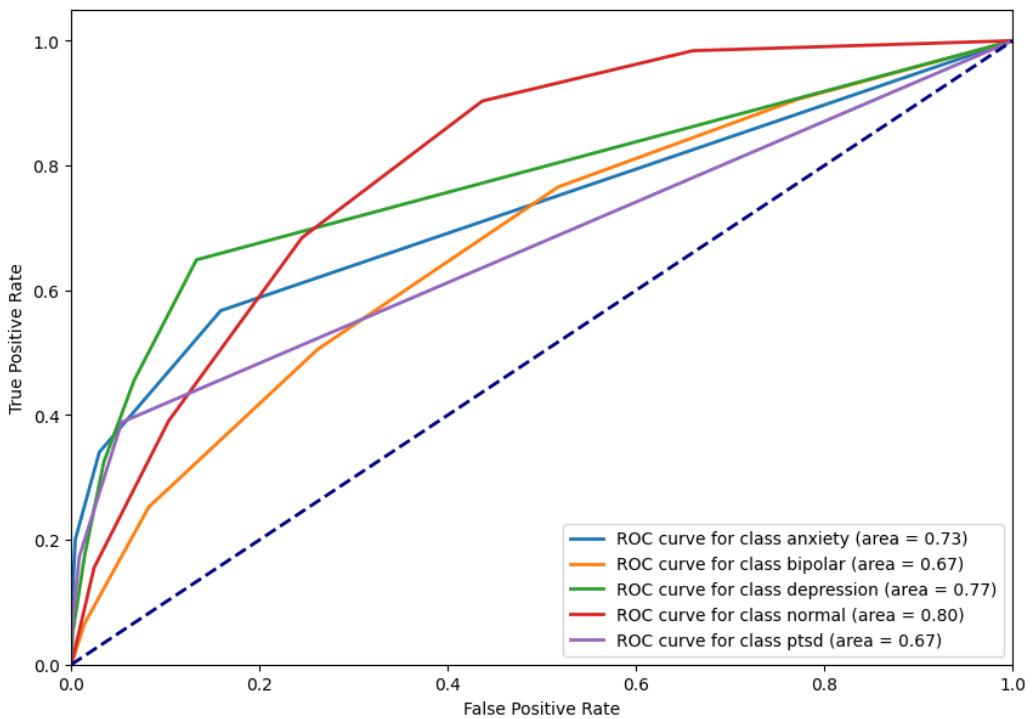


Figure 33: ROC AUC (KNN)

## 9.8 Results of LSTM

**LSTM Classification Report**

Class	Precision	Recall	F1-Score	Support
Anxiety	0.81	0.68	0.74	403
Bipolar	0.60	0.59	0.60	397
Depression	0.56	0.77	0.65	387
Normal	0.96	0.96	0.96	2137
PTSD	0.76	0.61	0.68	396
<b>Accuracy</b>	83.55%			
<b>Macro Avg</b>	0.74	0.72	0.73	3720
<b>Weighted Avg</b>	0.84	0.84	0.84	3720

**Confusion Matrix for LSTM Model**

	Anxiety	Bipolar	Depression	Normal	PTSD
Anxiety	274	36	32	28	33
Bipolar	26	233	42	85	11
Depression	25	31	299	19	13
Normal	6	18	11	2043	9
PTSD	31	33	33	35	264

**ROC Curve Areas for Each Class**

Class	ROC AUC
Anxiety	0.95
Bipolar	0.92
Depression	0.94
Normal	0.99
PTSD	0.95

The LSTM model achieved an accuracy of 83.55%, indicating strong performance, although it lags slightly behind other models such as XGBoost and SVM. The classification report shows high precision, recall, and F1-scores for the 'Normal' class, which dominates the dataset. The 'Anxiety' class has reasonable performance, with a precision of 0.81 and recall of 0.68, but the 'Bipolar' and 'PTSD' classes have lower precision and recall, especially for 'Bipolar' (0.60 in both precision and recall). The confusion matrix reflects this, with 'Bipolar' and 'PTSD' misclassified with other classes, while 'Normal' is well-separated. The ROC AUC scores highlight the model's ability to distinguish between classes effectively, with 'Normal' scoring the highest AUC of 0.99, followed by other classes such as 'Anxiety' and 'Depression,' both with AUCs above 0.90.

## ASMPFMHDD

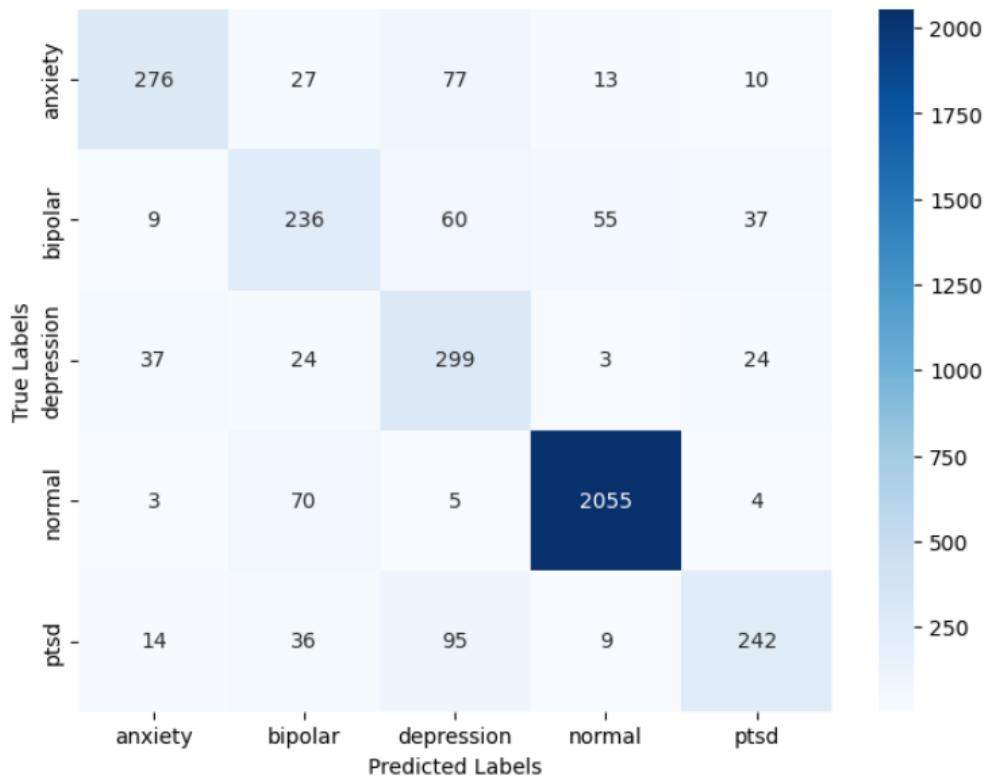


Figure 34: Confusion Matrix (LSTM)

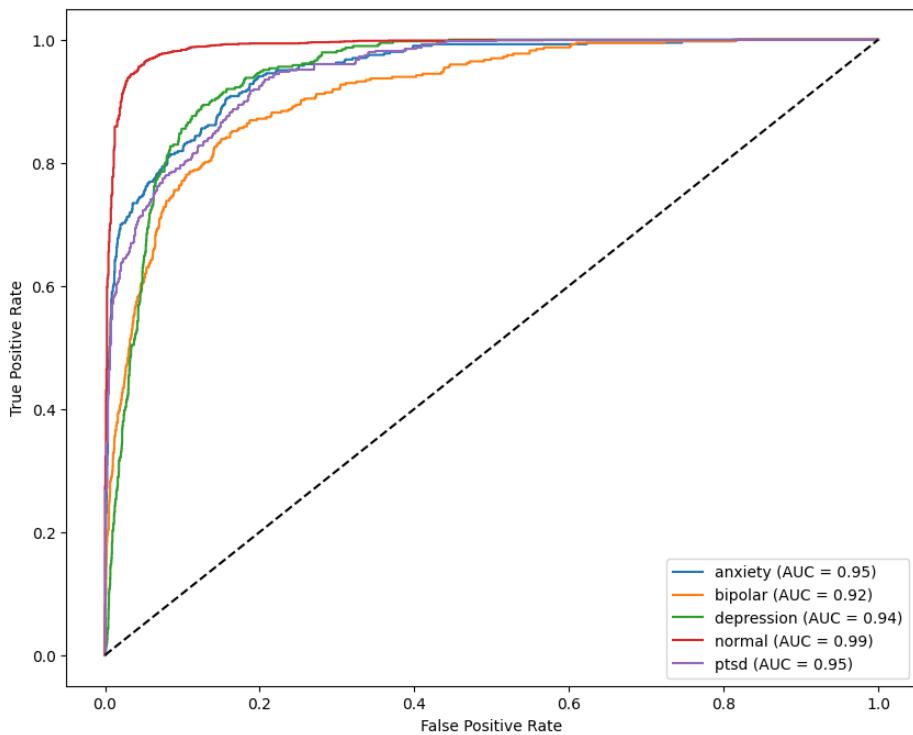


Figure 35: ROC AUC (LSTM)

Among the various machine learning algorithms tested, Logistic Regression emerged as the best performing model. With an accuracy of 87.66%, it outperformed the other models in terms of overall accuracy, precision, recall, and F1-score. The classification report for Logistic Regression shows strong performance across all classes, especially the 'Normal' class, which had high precision (0.92) and recall (0.99). Although the model faced challenges with the 'Bipolar' class, which had lower recall and F1-score, its overall ability to differentiate between other mental health conditions was impressive. The confusion matrix and ROC AUC scores further highlight the model's robustness, with an AUC of 0.99 for the 'Normal' class and strong values for other classes as well. This indicates that Logistic Regression not only achieves high accuracy but also performs well in distinguishing between different mental health categories, making it the most reliable choice for this classification task.

## 9.9 Results of Hyperparameter Tuning

**Hyperparameter Tuning on Logistic Regression**

Best Hyperparameters	Value
Solver	liblinear
Penalty	l2
C	1

**Classification Report**

Class	Precision	Recall	F1-Score	Support
Anxiety	0.83	0.77	0.80	379
Bipolar	0.74	0.55	0.64	384
Depression	0.76	0.76	0.76	373
Normal	0.92	0.99	0.95	2183
PTSD	0.87	0.77	0.82	394
Accuracy	0.88			3713
Macro avg	0.83	0.77	0.79	3713
Weighted avg	0.87	0.88	0.87	3713

**ROC Curve Areas for Each Class**

Class	ROC AUC
Anxiety	0.95
Bipolar	0.92
Depression	0.96
Normal	0.99
PTSD	0.95

The results of hyperparameter tuning on the Logistic Regression model show that the best hyperparameters were a solver of `liblinear`, a penalty of `l2`, and a regularization parameter (`C`) of 1. The model achieved an overall accuracy of 87.72%, with high performance across

## ASMPFMHDD

most classes. The classification report reveals that the model performs well on normal class with an accuracy of 92% in predicting this class, whereas performance on bipolar is slightly lower 64%  $F1 - score$ . The ROC AUC scores indicate that the model is highly effective in distinguishing between different classes, with normal achieving a near-perfect AUC of 0.99, followed closely by depression at 0.96. These results suggest that the model is well-tuned, particularly for classes like normal and anxiety.

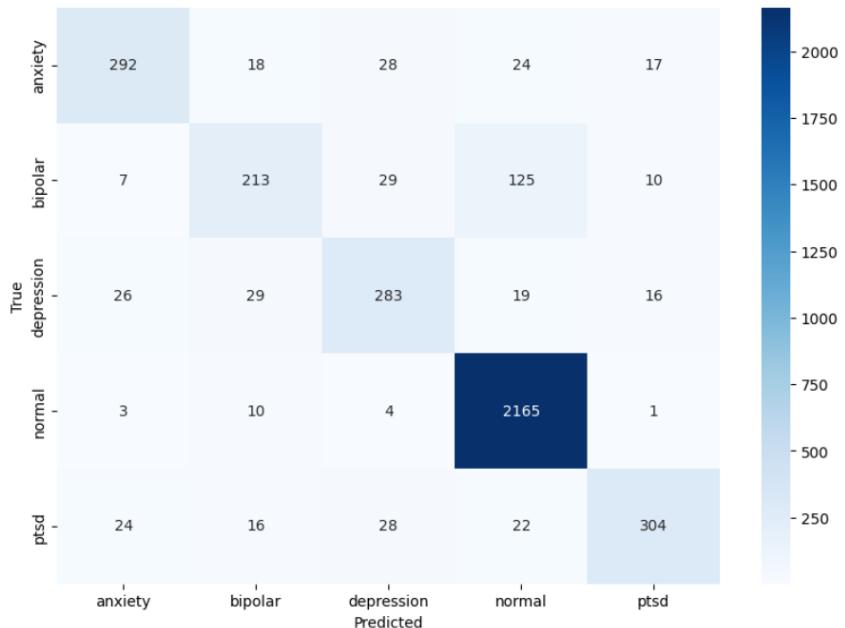


Figure 36: Classification Matrix after Hyperparameter Tuning (Logistic Regression)

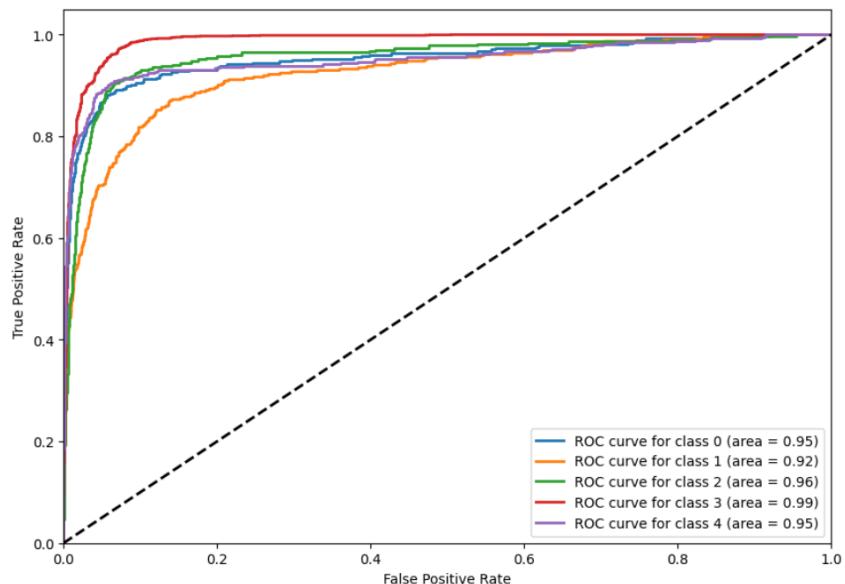


Figure 37: ROC AUC after Hyperparameter Tuning (Logistic Regression)

**Hyperparameter Tuning on k-NN**

Best Hyperparameters	Value
Weights	distance
n_neighbors	10
Metric	euclidean

**Classification Report**

Class	Precision	Recall	F1-Score	Support
Anxiety	0.73	0.23	0.35	379
Bipolar	0.18	0.60	0.27	384
Depression	0.47	0.40	0.43	373
Normal	0.74	0.65	0.69	2183
PTSD	0.83	0.10	0.18	394
Accuracy	0.52			3713
Macro avg	0.59	0.40	0.39	3713
Weighted avg	0.66	0.52	0.53	3713

**ROC Curve Areas for Each Class**

Class	ROC AUC
Anxiety	0.77
Bipolar	0.67
Depression	0.79
Normal	0.79
PTSD	0.71

The results of hyperparameter tuning on the k-Nearest Neighbors (k-NN) model show that the best hyperparameters were weights set to distance, n\_neighbors set to 10, and metric set to euclidean. Despite these optimizations, the model achieved an overall accuracy of 52.03%, which is relatively low. The classification report indicates that the model struggled to effectively predict the anxiety and PTSD classes, with particularly low recall values (0.23 and 0.10, respectively). However, it performed reasonably well on the normal class with an F1-score of 0.69. The ROC AUC scores reflect this by showing relatively poor performance for bipolar (0.67) and PTSD (0.71), while normal and depression had slightly better values at 0.79 each. These results suggest that k-NN is not the best model for this dataset, as it fails to consistently classify the minority classes effectively.

## ASMPFMHDD

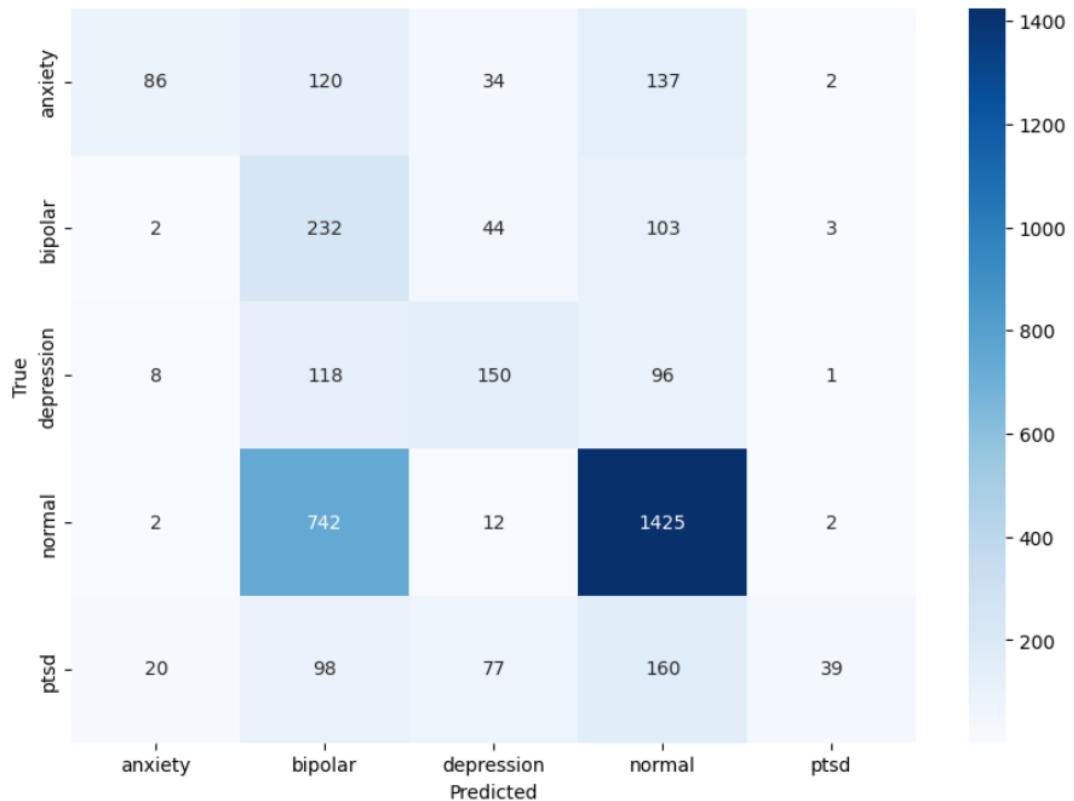


Figure 38: Classification Matrix after Hyperparameter Tuning (KNN)

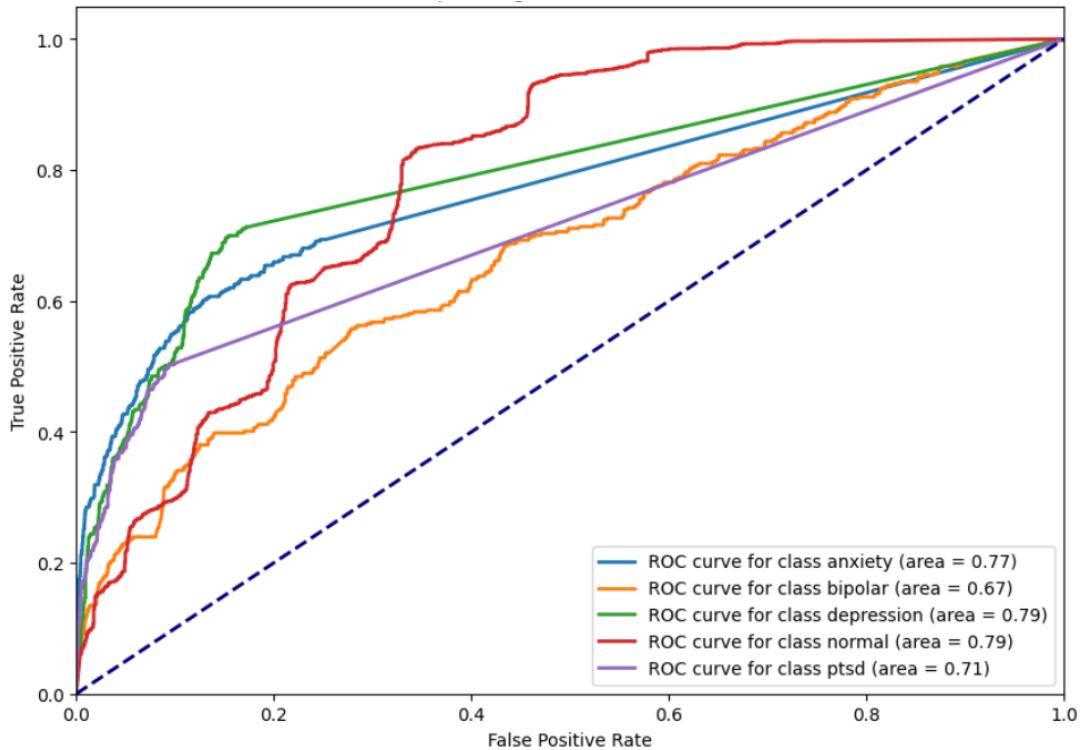


Figure 39: ROC AUC after Hyperparameter Tuning (KNN)

**Hyperparameter Tuning on SVM**

Best Hyperparameters	Value
Kernel	linear
Gamma	scale
C	1

**Classification Report**

Class	Precision	Recall	F1-Score	Support
Anxiety	0.72	0.76	0.74	379
Bipolar	0.62	0.61	0.61	384
Depression	0.74	0.71	0.72	373
Normal	0.94	0.95	0.95	2183
PTSD	0.78	0.74	0.76	394
Accuracy	0.85			3713
Macro avg	0.76	0.75	0.76	3713
Weighted avg	0.85	0.85	0.85	3713

**ROC Curve Areas for Each Class**

Class	ROC AUC
Anxiety	0.92
Bipolar	0.88
Depression	0.95
Normal	0.98
PTSD	0.94

The results of hyperparameter tuning on the Support Vector Machine (SVM) model reveal that the best hyperparameters were kernel set to linear, gamma set to scale, and C set to 1. With these settings, the model achieved an overall accuracy of 85.13%. The classification report indicates solid performance across most classes, with normal achieving the highest F1-score of 0.95, followed by depression at 0.72 and anxiety at 0.74. The bipolar class had a slightly lower F1-score of 0.61, indicating that the model was less effective in distinguishing this class. The ROC AUC values were strong for most classes, with normal reaching 0.98 and depression, PTSD, and anxiety all scoring above 0.90, showing that the SVM model effectively discriminates between the classes. Overall, the SVM model performed well on this dataset, particularly for the normal and anxiety classes, making it a strong contender for mental health issue classification.

## ASMPFMHDD



Figure 40: Classification Matrix after Hyperparameter Tuning (SVM)

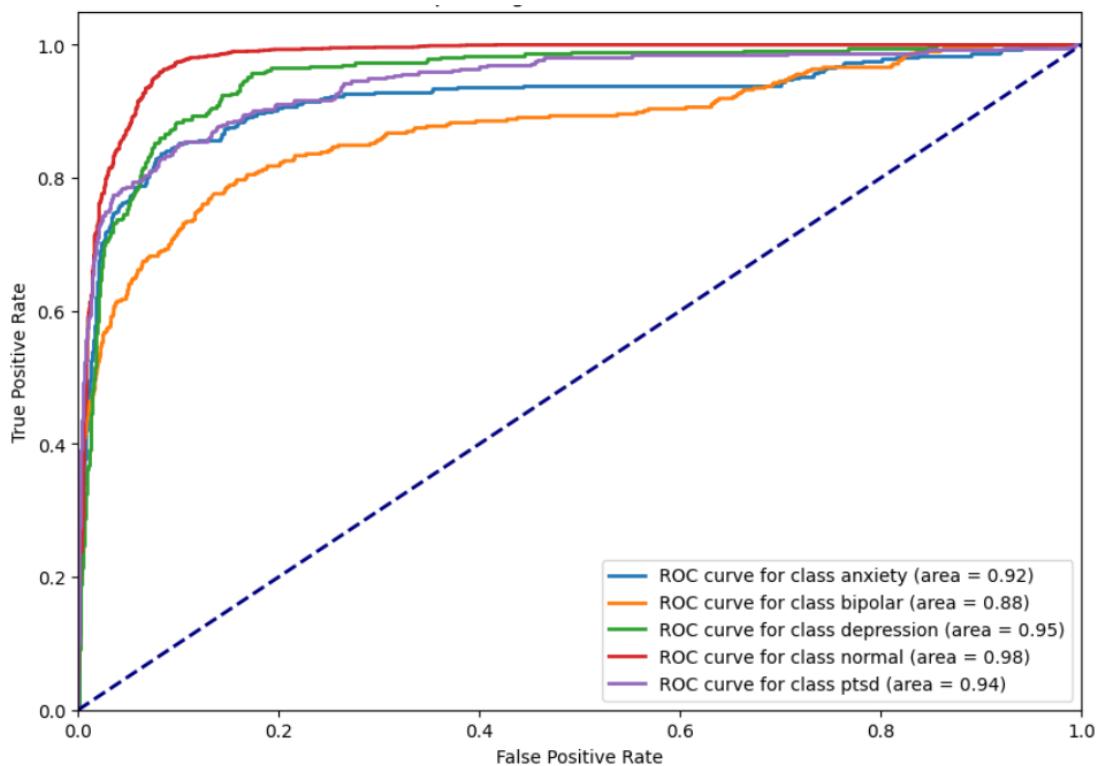


Figure 41: ROC AUC after Hyperparameter Tuning (SVM)

**Hyperparameter Tuning on Naive Bayes**

<b>Best Hyperparameters</b>	<b>Value</b>
Alpha	0.2914180608409973

**Classification Report**

<b>Class</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>	<b>Support</b>
Anxiety	0.69	0.76	0.72	379
Bipolar	0.75	0.55	0.64	384
Depression	0.60	0.83	0.70	373
Normal	0.96	0.91	0.94	2183
PTSD	0.73	0.79	0.76	394
Accuracy	0.84			3713
Macro avg	0.75	0.77	0.75	3713
Weighted avg	0.85	0.84	0.84	3713

**ROC Curve Areas for Each Class**

<b>Class</b>	<b>ROC AUC</b>
Anxiety	0.93
Bipolar	0.93
Depression	0.93
Normal	0.99
PTSD	0.94

The results of hyperparameter tuning on the Naive Bayes model show that the best hyperparameter was alpha set to 0.2914. With this optimal value, the model achieved an accuracy of 83.95%. The classification report reveals that the normal class performed the best, with a high F1-score of 0.94, followed by PTSD at 0.76. The anxiety class achieved a solid F1-score of 0.72, while bipolar and depression had lower scores. The ROC AUC curve areas demonstrate strong performance across all classes, with normal reaching a perfect 0.99 and the other classes scoring above 0.90. Overall, the Naive Bayes model performed well, especially in identifying normal and PTSD, making it a good candidate for mental health issue classification.

## ASMPFMHDD

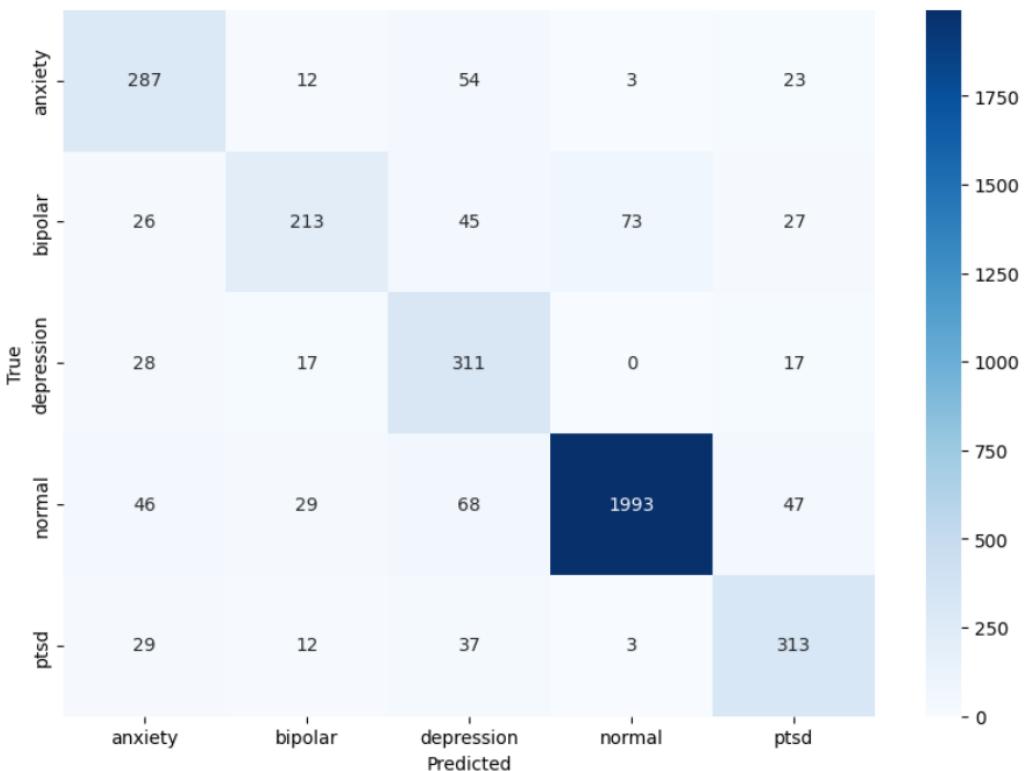


Figure 42: Classification Matrix after Hyperparameter Tuning (Naive Bayes)

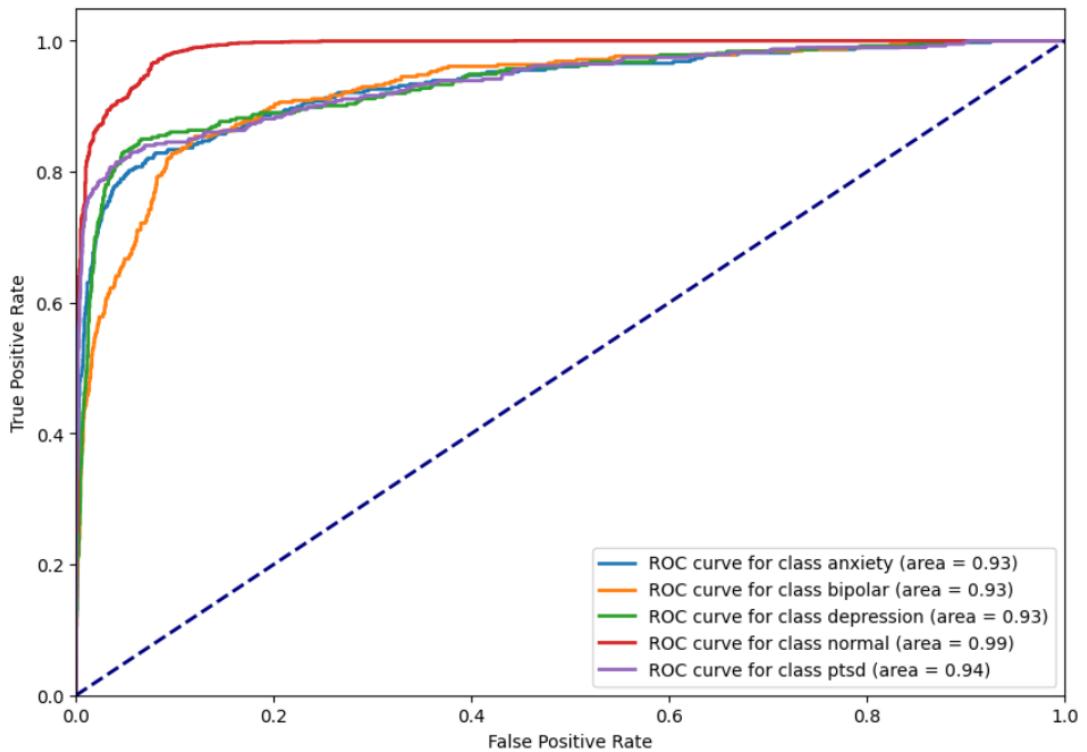


Figure 43: ROC AUC after Hyperparameter Tuning (Naive Bayes)

After performing hyperparameter tuning using RandomizedSearchCV, Logistic Regression emerged as the best-performing model for mental health classification. The model achieved an accuracy of 87.72%, demonstrating its strong ability to correctly classify mental health issues based on text data. The best hyperparameters found during tuning were 'solver': 'liblinear', 'penalty': 'l2', 'C': 1, which contributed to the model's robustness and efficiency. In terms of classification performance, Logistic Regression showed a high recall and precision, especially for classes like *normal*, with a recall of 0.99, indicating its proficiency in identifying the most prevalent class in the dataset. Furthermore, the ROC AUC scores for each class were impressive, with values close to or exceeding 0.90, which suggests excellent model discrimination between classes. Overall, Logistic Regression's combination of high accuracy, solid precision-recall balance, and optimal hyperparameters makes it the most suitable model for this task.

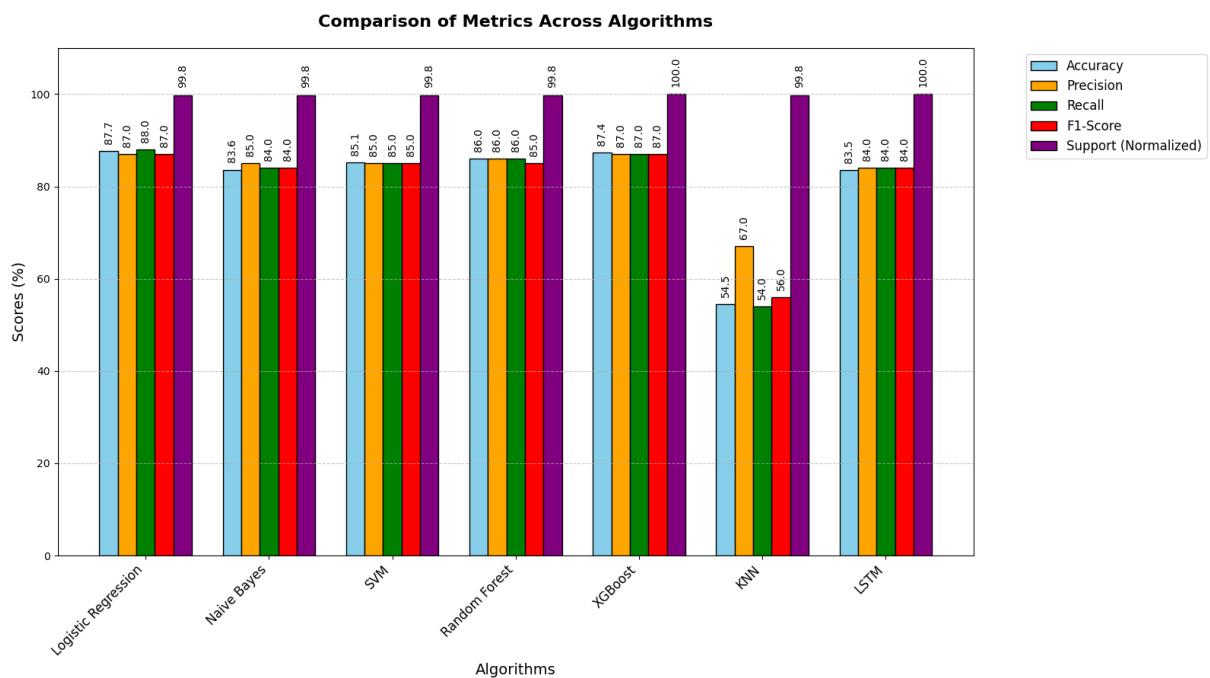


Figure 44: Result Comparison of the Algorithms

K-Nearest Neighbors (KNN) performs significantly poorly on this dataset compared to other algorithms, both before and after hyperparameter tuning. The dataset used, which includes Reddit posts with corresponding mental health labels, presents challenges for KNN due to its reliance on distance-based metrics. Specifically, KNN uses the distance between data points to classify them, but this approach is sensitive to the scale and distribution of features. In the case of text data, the feature space can be sparse and high-dimensional, making it difficult for KNN to find meaningful relationships between the posts and their associated mental health labels. After hyperparameter tuning, KNN was optimized with parameters such as **weights='distance'**,

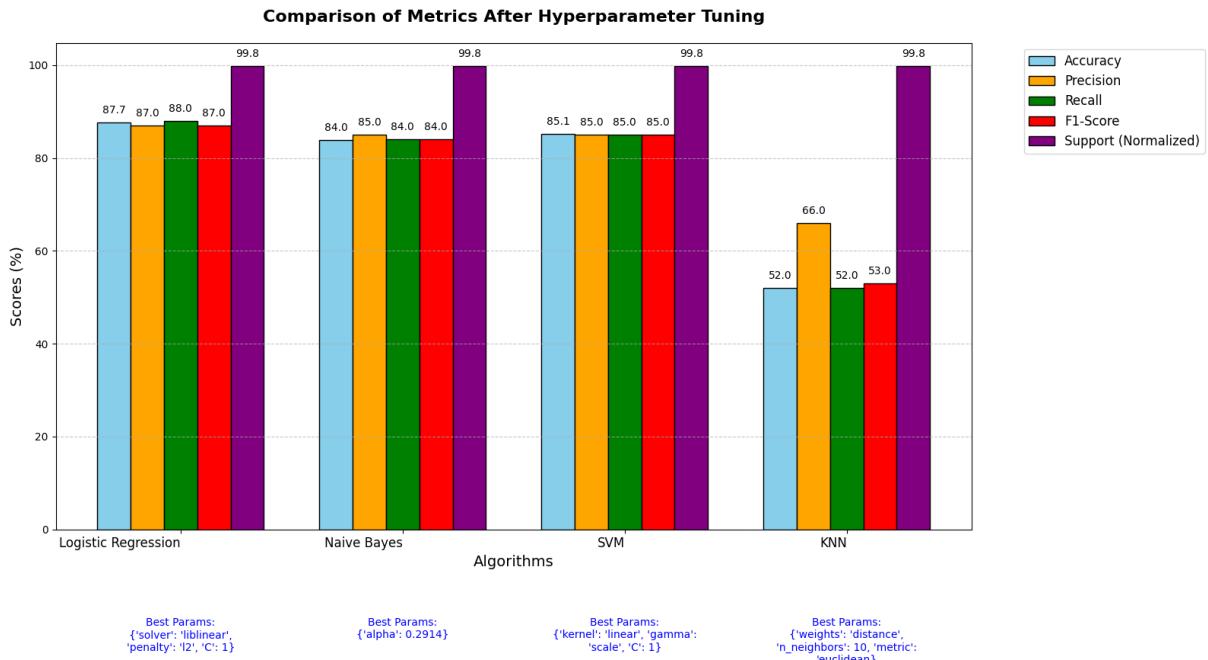


Figure 45: Result Comparison after Hyperparameter Tuning

**n\_neighbors=10**, and **metric='euclidean'**, yet its performance remains suboptimal. This suggests that KNN struggles with the complexities inherent in text classification tasks, where the relationships between features (such as words or phrases in a Reddit post) are often non-linear and context-dependent. Despite tuning, the algorithm's performance metrics—accuracy, precision, recall, and F1-score—remain notably lower than those of other algorithms like Logistic Regression, Naive Bayes, or SVM. Additionally, KNN's inability to capture these subtleties in language means it performs poorly, particularly in distinguishing between categories like anxiety, bipolar disorder, or PTSD, where the text patterns may be more nuanced. Overall, KNN's reliance on proximity without considering the richer context of textual data likely contributes to its lower performance on this mental health classification task. Other algorithms, particularly those that are more suited for high-dimensional and sparse data (like Logistic Regression or SVM), show much better results as they can better account for the underlying patterns in the Reddit posts.

## 9.10 Results from Ensemble Model Training and Testing

### 9.10.1 Logistic Regression and XGBoost

**Classification Report for Ensemble Model**

	Precision	Recall	F1-Score	Support
<b>Anxiety</b>	0.96	0.93	0.95	1999
<b>Bipolar</b>	0.90	0.86	0.88	1941
<b>Depression</b>	0.95	0.94	0.94	1959
<b>Normal</b>	0.97	0.99	0.98	10679
<b>PTSD</b>	0.98	0.95	0.96	1987
<b>Accuracy</b>	<b>96.09%</b>			
<b>Macro avg</b>	0.95	0.93	0.94	18565
<b>Weighted avg</b>	0.96	0.96	0.96	18565

**ROC AUC Scores for Each Class**

Class	ROC AUC
Anxiety	1.0
Bipolar	0.99
Depression	0.99
Normal	1.0
PTSD	1.0

The classification report provides an insightful summary of the ensemble model's performance, which combines Logistic Regression and XGBoost to classify mental health issues. Precision is a key metric that measures the proportion of correctly predicted positive cases out of all the positive predictions made by the model. For example, a precision of 0.96 for anxiety means that when the model predicts anxiety, it is correct 96% of the time. Recall, on the other hand, reflects the model's ability to identify all actual positive cases in the dataset. A recall of 0.93 for anxiety suggests that the model correctly identifies 93% of all true anxiety cases. The F1-score, which balances precision and recall, is also an important metric. With an F1-score of 0.95 for anxiety, the model strikes a good balance between precision and recall, indicating that it performs well across both metrics. Support represents the number of actual instances of a class in the dataset, and for anxiety, there are 1999 instances in the test set. The model's overall accuracy is 96.09%, meaning that it correctly predicts the mental health condition for 96.09% of the test instances, reflecting the overall effectiveness of the model. The macro average, which takes the unweighted mean of precision, recall, and F1-scores across all classes, shows an average precision of 0.95, recall of 0.93, and F1-score of 0.94. This provides a general idea of the model's performance across all classes without accounting for class imbalance. The weighted average, however, adjusts for the class distribution by giving more weight to classes with more instances. The weighted averages for precision, recall, and F1-score are all 0.96,

## ASMPFMHDD

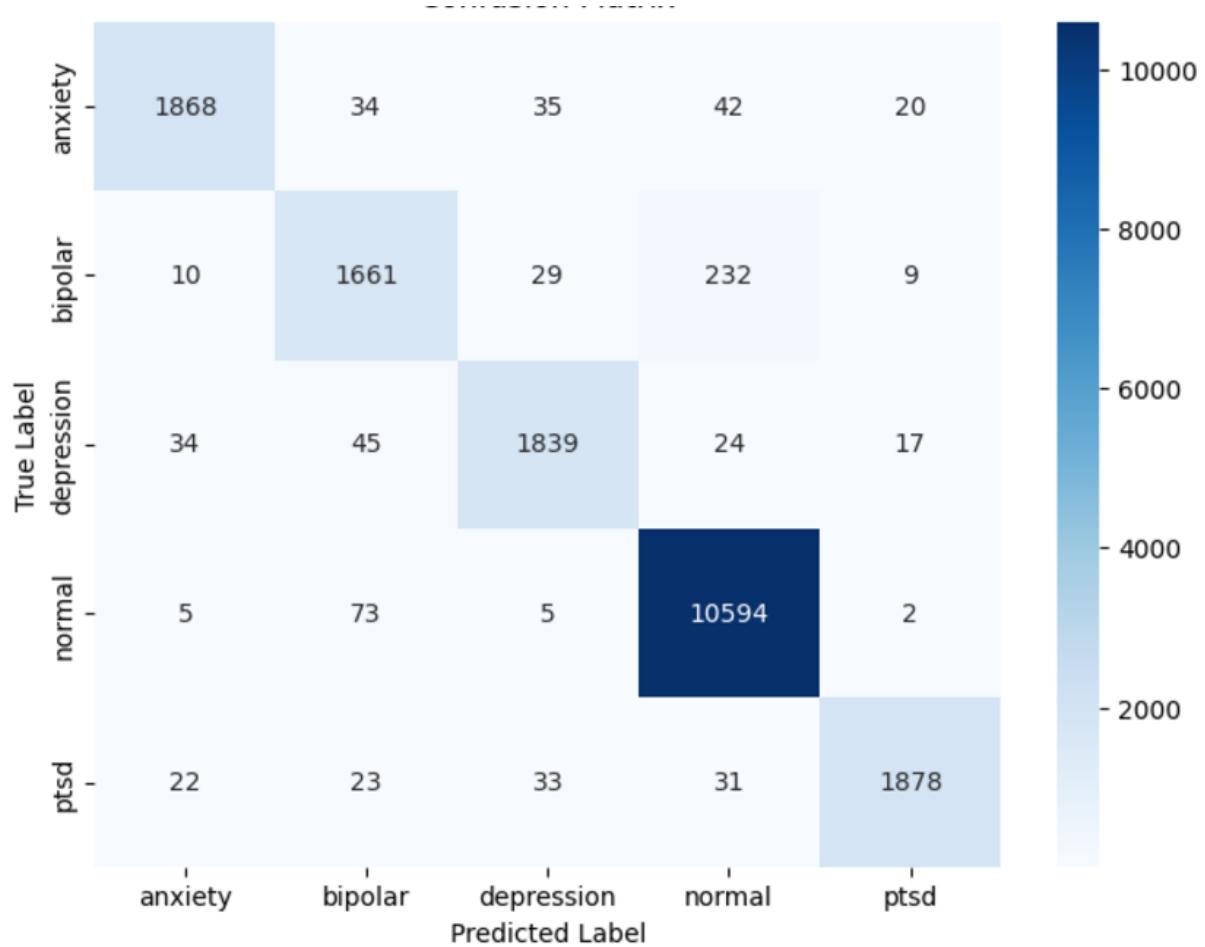


Figure 46: Confusion Matrix for Ensemble Model(LR + XGB)

reflecting strong overall performance, especially when considering the differing class sizes. The ROC AUC scores are another key performance indicator that evaluates the model's ability to distinguish between different classes. The AUC score ranges from 0 to 1, with 1 representing perfect classification. A perfect AUC score of 1.0 for anxiety, normal, and PTSD indicates that the model can perfectly differentiate between these classes and others. The AUC scores for bipolar and depression are 0.99, which is almost perfect, further demonstrating the model's strong performance across all classes. Overall, these metrics indicate that the ensemble model is highly effective at classifying mental health issues, with robust precision, recall, and the ability to distinguish between the different conditions. The confusion matrix will provide insight into the number of false positives, false negatives, true positives, and true negatives for each class. By looking at the confusion matrix, you can better understand where the model might be misclassifying certain classes, such as when it confuses anxiety with depression or bipolar with PTSD. For example, a high number of false positives in the anxiety class would mean the model is often predicting anxiety when it's actually a different mental health condition. Conversely, a high number of false negatives would mean the model is missing anxiety cases.

## ASMPFMHDD

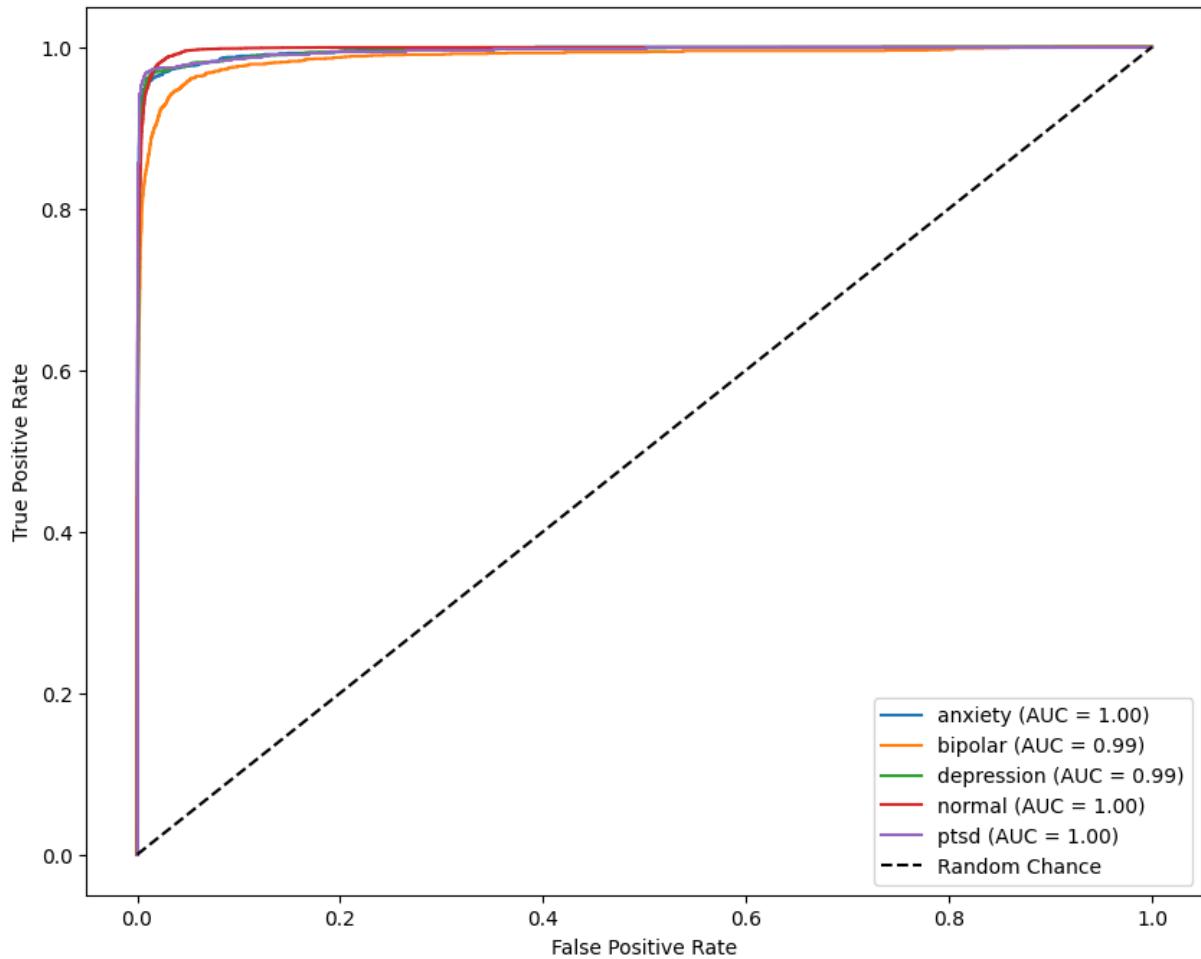


Figure 47: ROC AUC Curve for Ensemble Model(LR + XGB)

The combination of logistic regression and XGBoost allows the ensemble model to leverage the strengths of both algorithms. While logistic regression is simple and interpretable, it might struggle with non-linearities in the data. XGBoost, on the other hand, is a powerful gradient boosting algorithm that can capture complex patterns in the data. By stacking their predictions, the ensemble model benefits from both simplicity and complexity, achieving higher accuracy and robustness in text-based classification tasks.

This ensemble model, combining Logistic Regression and XGBoost, is highly effective for the mental health classification web application due to its ability to leverage the strengths of both algorithms. Logistic Regression provides a simple yet powerful linear approach to classification, offering interpretability and efficiency, while XGBoost excels at handling complex, non-linear relationships with its gradient boosting framework, enabling better accuracy and robustness. By stacking these models, the ensemble approach mitigates the individual weaknesses of each algorithm—Logistic Regression's limitations in capturing non-linear patterns and XGBoost's susceptibility to overfitting on small datasets—resulting in a more balanced, accurate, and gen-

eralizable model. This synergy ensures that the mental health classification web application delivers reliable, high-precision predictions across diverse text-based inputs.

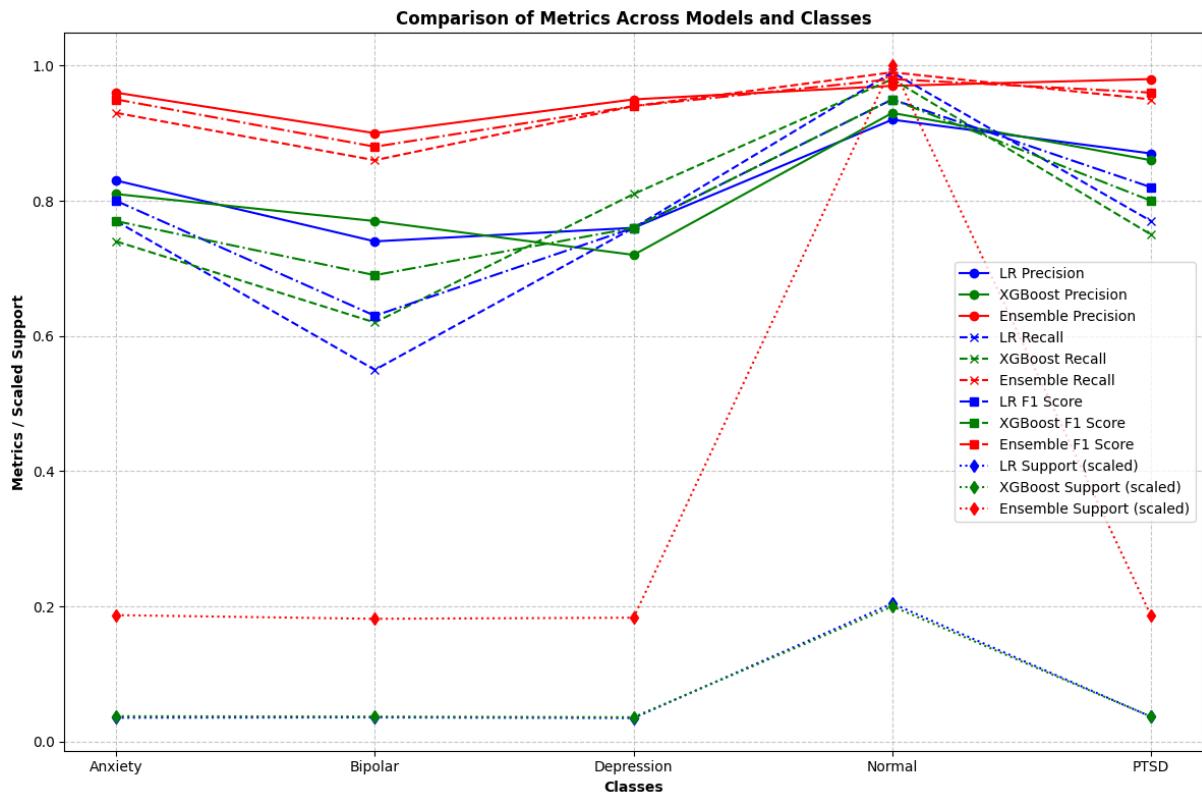


Figure 48: Comparison of results from LR, XGB and Ensemble Model

### 9.10.2 Logistic Regression, SVM, Naive Bayes, XGBoost(meta-learner), LSTM

#### Classification Report

	Precision	Recall	F1-Score	Support
<b>Anxiety</b>	1.00	1.00	1.00	1999
<b>Bipolar</b>	1.00	1.00	1.00	1941
<b>Depression</b>	1.00	1.00	1.00	1959
<b>Normal</b>	1.00	1.00	1.00	10679
<b>PTSD</b>	1.00	1.00	1.00	1987
<b>Accuracy</b>	<b>99.59%</b>			
<b>Macro avg</b>	1.00	1.00	1.00	18565
<b>Weighted avg</b>	1.00	1.00	1.00	18565

### ROC AUC Scores for Each Class

Class	ROC AUC
Anxiety	1.0
Bipolar	1.0
Depression	1.0
Normal	1.0
PTSD	1.0

Each row of the matrix represents the true class, while each column represents the predicted class. The diagonal elements (1997, 1936, 1959, 10671, 1987) show the number of correct predictions for each class (anxiety, bipolar, depression, normal, PTSD), while the off-diagonal elements represent misclassifications. For example, the model misclassified 2 samples from the anxiety class as depression and 1 sample from the bipolar class as normal, but overall, the model made very few misclassifications. An ROC AUC score of 1.0 indicates perfect classification performance for each class, meaning that the model has perfectly separated the classes based on the probability predictions. This reinforces the outstanding performance of the XGBoost model.

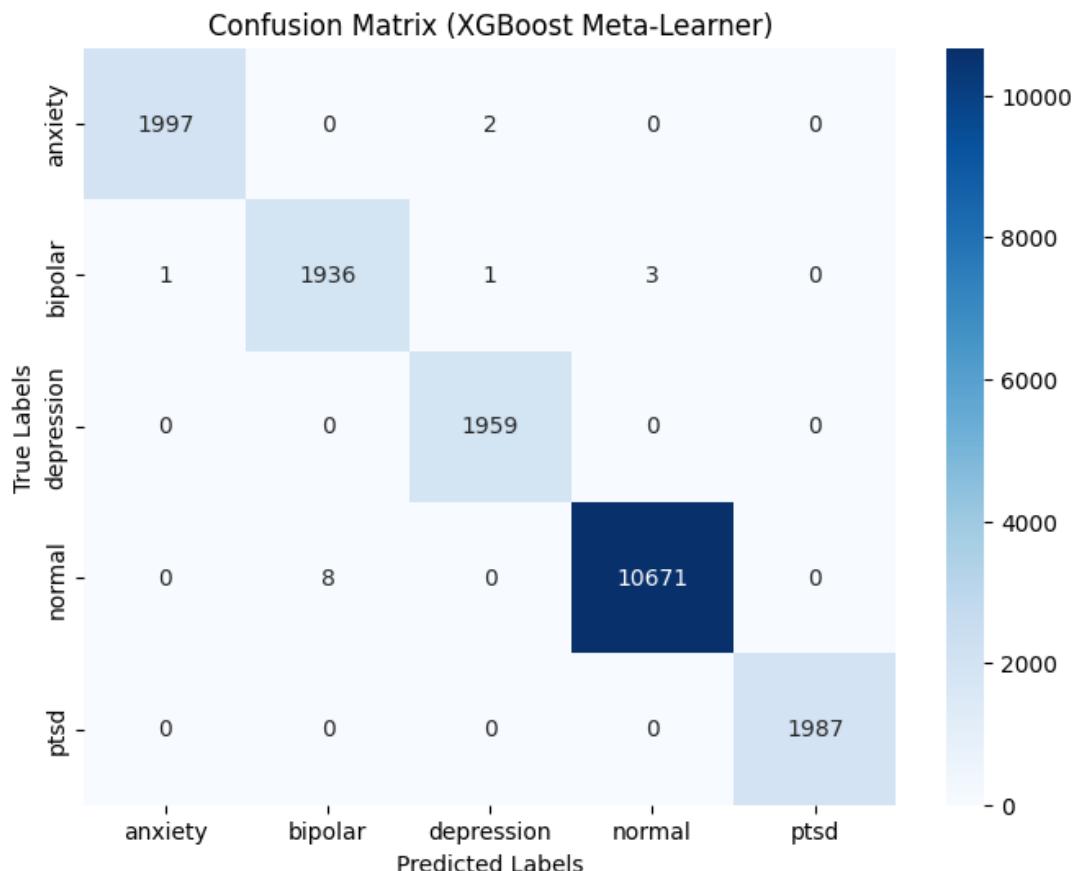


Figure 49: Confusion Matrix for ensemble model (XGB - meta learner)

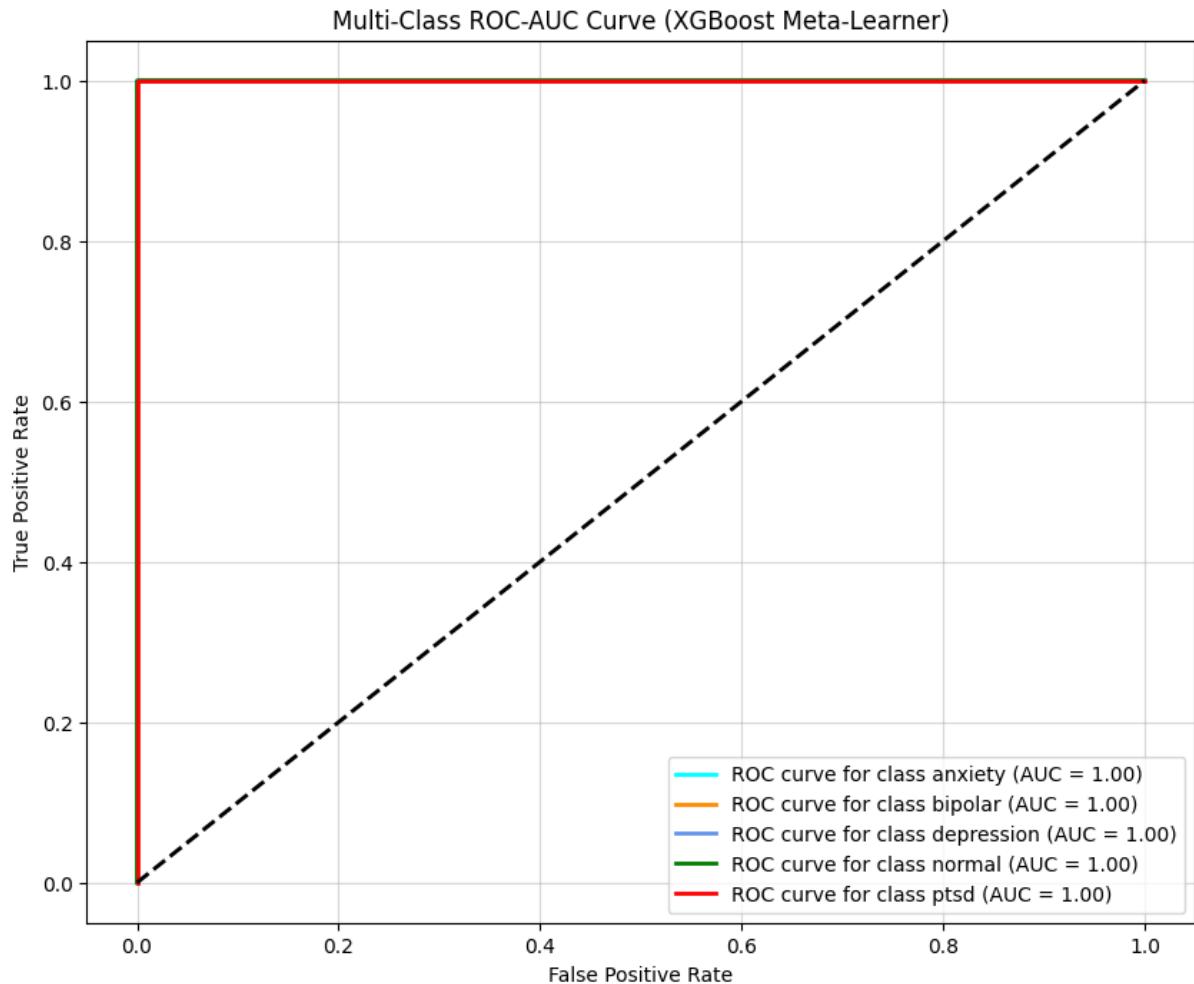


Figure 50: ROC AUC Curve for multiple models in ensemble model (XGB - meta learner)

Using XGBoost as a meta-learner in an ensemble model is a strategic decision that brings several advantages, both in terms of performance and interpretability. To understand why XGBoost is an excellent choice, it is important to first appreciate the concept of a meta-learner and how ensemble learning works. Ensemble learning involves combining multiple base models to improve prediction accuracy by leveraging their individual strengths. In a typical ensemble method, the predictions of several base models (e.g., logistic regression, SVM, Naive Bayes, LSTM) are used as input features for a meta-learner, which then learns how to best combine these predictions to make a final decision. The meta-learner's role is to determine which base model's predictions are more reliable for each sample, thus improving the overall performance of the ensemble. XGBoost, a gradient boosting algorithm, has several features that make it an ideal candidate for a meta-learner in this context. First, XGBoost is known for its ability to handle complex relationships in data, especially in terms of interactions between features. Unlike simpler models like logistic regression or Naive Bayes, XGBoost builds decision trees iteratively, where each new tree is focused on correcting the errors made by the previous trees.

## ASMPFMHDD

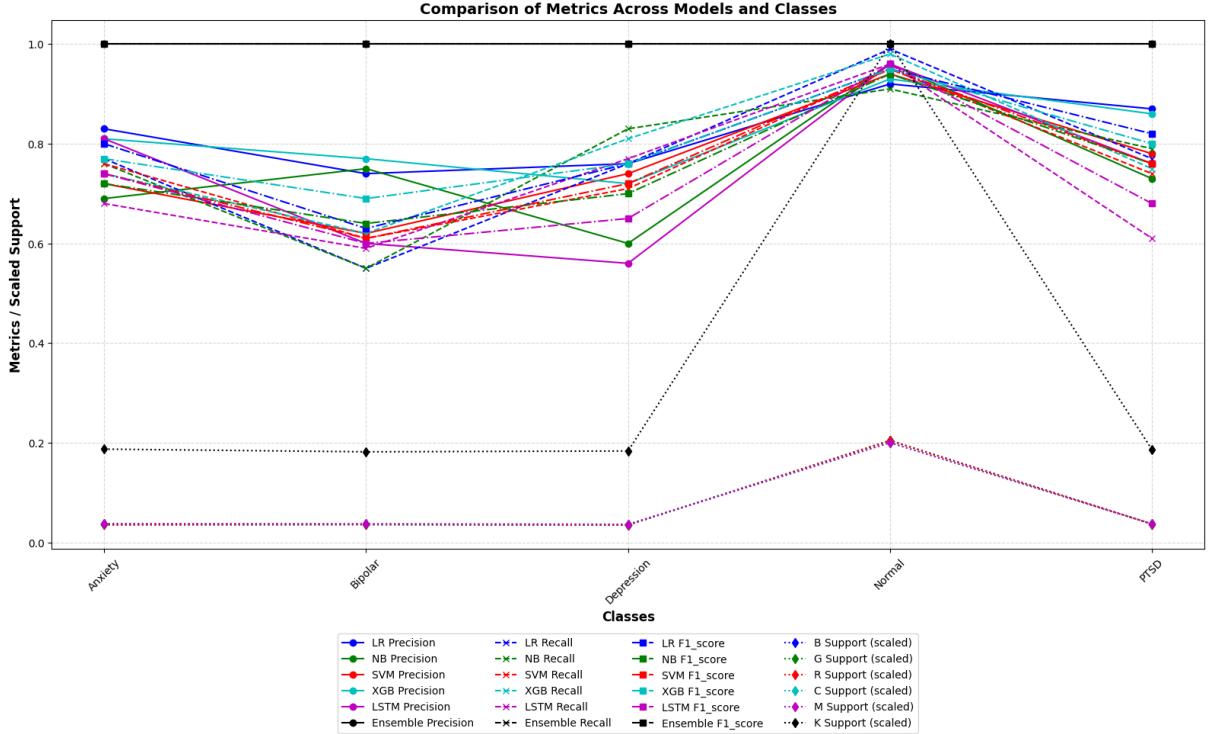


Figure 51: Comparison of results with Ensemble Model (XGB - meta learner)

This iterative approach allows XGBoost to capture non-linear patterns and subtle interactions that might be missed by individual base models. When used as a meta-learner, XGBoost can intelligently combine the predictions from the base models by learning the optimal way to weigh the output of each base model. This is especially important in a multi-model setting where different classifiers might have complementary strengths—XGBoost can effectively "learn" when to trust each model's prediction more, depending on the characteristics of the input data. Another advantage of using XGBoost as a meta-learner is its regularization capabilities. XGBoost includes both L1 and L2 regularization, which helps prevent overfitting. In the context of ensemble learning, this is particularly useful because, with multiple models contributing to the final prediction, the meta-learner needs to ensure it doesn't overly rely on any one base model, especially if that model is prone to overfitting. By applying regularization, XGBoost encourages the model to generalize well, ensuring that the ensemble is robust and does not become overly sensitive to the training data or to noise.

Additionally, XGBoost is highly efficient in terms of both computation and memory usage. It is designed to work well with large datasets, and its ability to handle sparse data effectively makes it a good fit for a wide variety of tasks, including those involving textual data or other high-dimensional inputs. In ensemble models, where multiple base models generate a large number of features (e.g., prediction probabilities from each base model), XGBoost's efficiency in processing these features allows it to scale well even as the number of base models or the size

of the dataset grows. Another key benefit is XGBoost’s ability to handle imbalanced datasets. Many classification tasks, especially in medical or mental health data, suffer from class imbalances, where some classes (e.g., “Normal”) are more frequent than others (e.g., “Bipolar” or “PTSD”). XGBoost has built-in features to adjust for class imbalances, such as using class weights or employing custom loss functions. This is important in a meta-learner context because the base models might not be equally effective across all classes. XGBoost can ensure that the ensemble model doesn’t bias its predictions toward the majority class, which is crucial for improving the overall classification performance, especially for underrepresented classes. Furthermore, XGBoost provides excellent interpretability through feature importance scores. In an ensemble model, where several base models contribute to the final prediction, understanding how each base model influences the outcome is valuable for both improving the model and for explaining its behavior. XGBoost offers built-in tools to analyze the importance of individual features (in this case, the predictions of the base models), allowing practitioners to understand how the meta-learner is making decisions and which base models are contributing the most to its predictions. Lastly, XGBoost’s performance in various machine learning competitions has demonstrated its robustness. It is one of the top-performing algorithms in tasks such as classification and regression, often outperforming other methods, including random forests and neural networks. Its popularity and track record make it a safe and reliable choice for a meta-learner.

While using models like XGBoost, SVM, Naive Bayes, and LSTM as meta-learners in ensemble modeling can provide high accuracy, these models can also be prone to **overfitting**, especially when they are trained on the predictions of multiple base models. Overfitting occurs when the model learns the noise or patterns specific to the training data that do not generalize well to unseen data. For XGBoost, although it is a powerful and highly flexible model, its ability to capture complex relationships between features can lead to overfitting, especially when the base models are already strong. When combined with a large number of features from multiple base models, the XGBoost meta-learner may excessively adjust to small fluctuations in the training data, leading to high performance on the training set but poor generalization to new data. This risk is heightened if the base models already provide overlapping or correlated predictions, which can reinforce irrelevant patterns for the meta-learner.

Similarly, SVM as a meta-learner can overfit in ensemble setups. SVM, especially when using a non-linear kernel, is very sensitive to the scale of features and may become highly specialized to the training data if not properly tuned. When stacking the output of multiple base models, each with its own biases and prediction tendencies, the SVM meta-learner may overemphasize specific instances that are not representative of the general data distribution. This can lead to poor performance on new, unseen data despite strong accuracy on the training set.

Naive Bayes is another model that can overfit when used as a meta-learner. Although it assumes independence between features, stacking it with predictions from base models may violate this assumption, leading to overly optimistic performance during training. Naive Bayes models are particularly sensitive to correlated features, and in an ensemble, where the base models might produce similar predictions, the meta-learner might end up fitting to these correlations too closely. This results in high training accuracy but a significant drop in performance when applied to unseen data.

Finally, LSTM models, although excellent at capturing long-range dependencies in sequential data, can also overfit when used as a meta-learner. LSTMs are prone to memorizing the training data, especially when the number of parameters in the model is large relative to the size of the training set. In ensemble settings, the LSTM meta-learner might overly adapt to the specific features of the training data, particularly when the base models' predictions are very similar, reinforcing the same patterns in the data. This can cause the LSTM meta-learner to perform well on training data but fail to generalize effectively to new data.

In conclusion, while these models can provide high accuracy in ensemble setups, they are also susceptible to overfitting, which can undermine their ability to generalize to new, unseen data. Proper regularization, cross-validation, and tuning of hyperparameters are necessary to mitigate these risks and improve the model's robustness and generalization capabilities.

### **9.10.3 Logistic Regression(meta-learner), SVM, Naive Bayes, XGBoost, LSTM**

**Cross-Validation Accuracy Scores**

<b>Cross-Validation Accuracy Scores</b>	<b>Value</b>
1	0.97333692
2	0.96768112
3	0.97091301
4	0.96606518
5	0.97037436
<b>Mean Cross-Validation Accuracy</b>	<b>96.97%</b>

**Classification Report**

	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>	<b>Support</b>
<b>Anxiety</b>	0.96	0.95	0.95	1999
<b>Bipolar</b>	0.93	0.90	0.92	1941
<b>Depression</b>	0.95	0.95	0.95	1959
<b>Normal</b>	0.98	0.99	0.99	10679
<b>PTSD</b>	0.97	0.96	0.96	1987
<b>Accuracy</b>	97.05%			
<b>Macro avg</b>	0.96	0.95	0.95	18565
<b>Weighted avg</b>	0.97	0.97	0.97	18565

**ROC AUC Scores**

<b>Anxiety</b>	1.00
<b>Bipolar</b>	0.99
<b>Depression</b>	1.00
<b>Normal</b>	1.00
<b>PTSD</b>	1.00

The confusion matrix for the Logistic Regression model shows the performance of the classifier for each of the five mental health issues. For anxiety, 1900 true positives, 24 false positives, 33 false negatives, and 18 true negatives were recorded. This means the model was very accurate in detecting anxiety with minimal errors. Similarly, the bipolar disorder class had 1754 true positives and relatively few misclassifications, indicating high precision and recall. Depression also had a high number of correct predictions with 1857 true positives. The normal class had 10608 true positives and only minor misclassifications, highlighting the model's ability to classify the majority class accurately. The PTSD class showed similar performance with 1899 true positives. Overall, the confusion matrix suggests that the Logistic Regression model is highly effective in correctly classifying mental health conditions with very few errors across all classes. The ROC AUC scores for each class are all exceptionally high, with the anxiety, depression, normal, and PTSD classes achieving a perfect score of 1.00. The bipolar class has a slightly lower but still impressive score of 0.99. These high ROC AUC values indicate that the model is able to distinguish between the different classes with very high accuracy and minimal overlap between the predicted and actual values, ensuring robust model performance.

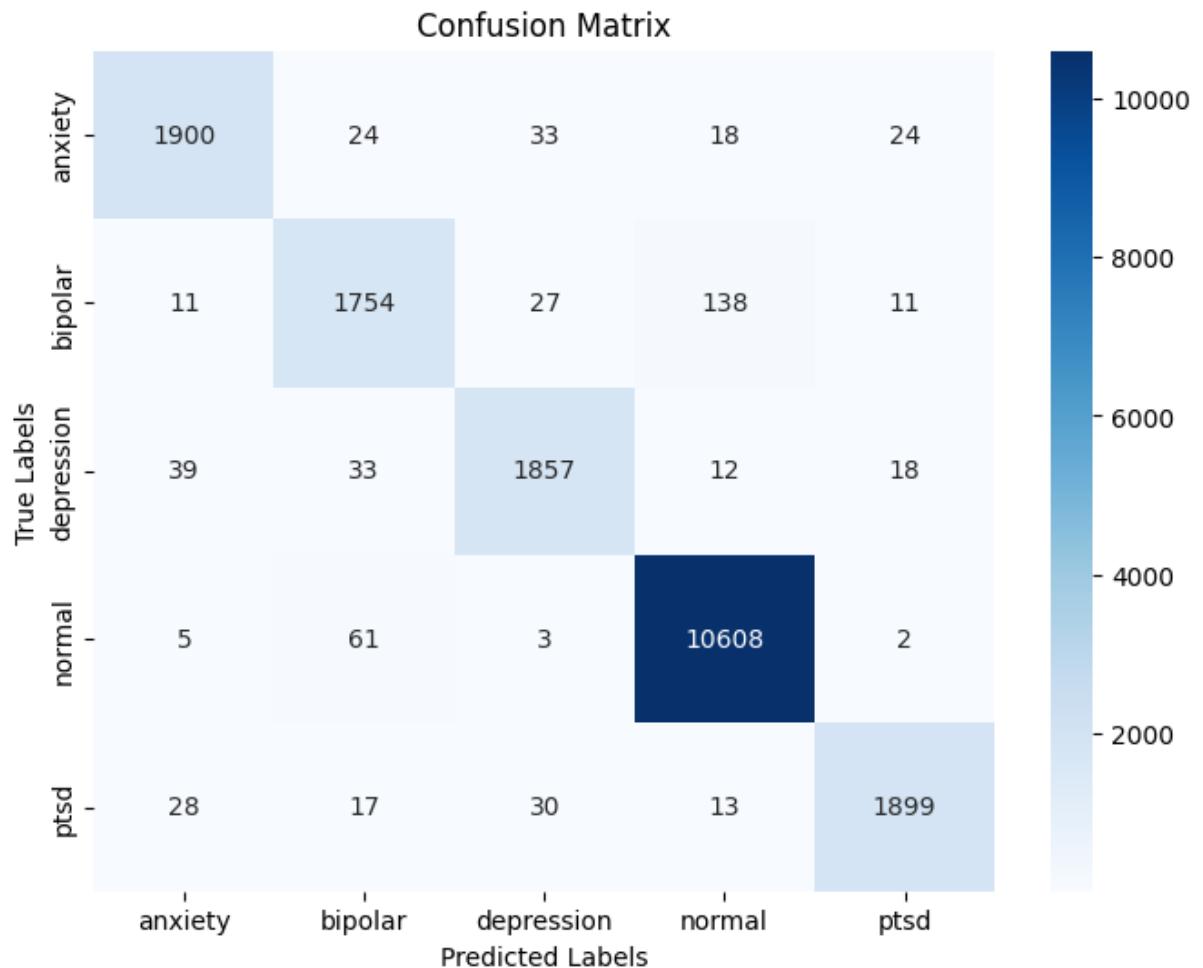


Figure 52: Confusion Matrix for Ensemble Model (LR - meta learner)

Logistic Regression is also highly effective at combining probabilistic outputs from base models. Since ensemble methods often rely on predictions in the form of probabilities, Logistic Regression is able to leverage these to produce more accurate final predictions. It can weigh the predictions from each base model appropriately, allowing the ensemble to take advantage of the strengths of each model. This ability to combine probabilities is critical when base models have different levels of confidence in their predictions.

Additionally, Logistic Regression is computationally efficient, making it a great choice for large ensembles. Training and inference times for Logistic Regression are relatively fast compared to more complex models like deep neural networks or gradient boosting machines. This efficiency is especially useful when dealing with large numbers of base models in the ensemble, as it ensures that the meta-learner can be trained quickly and can make predictions in real-time.

The regularization capabilities of Logistic Regression are another advantage when used as a meta-learner. Regularization techniques such as L1 and L2 regularization can help prevent

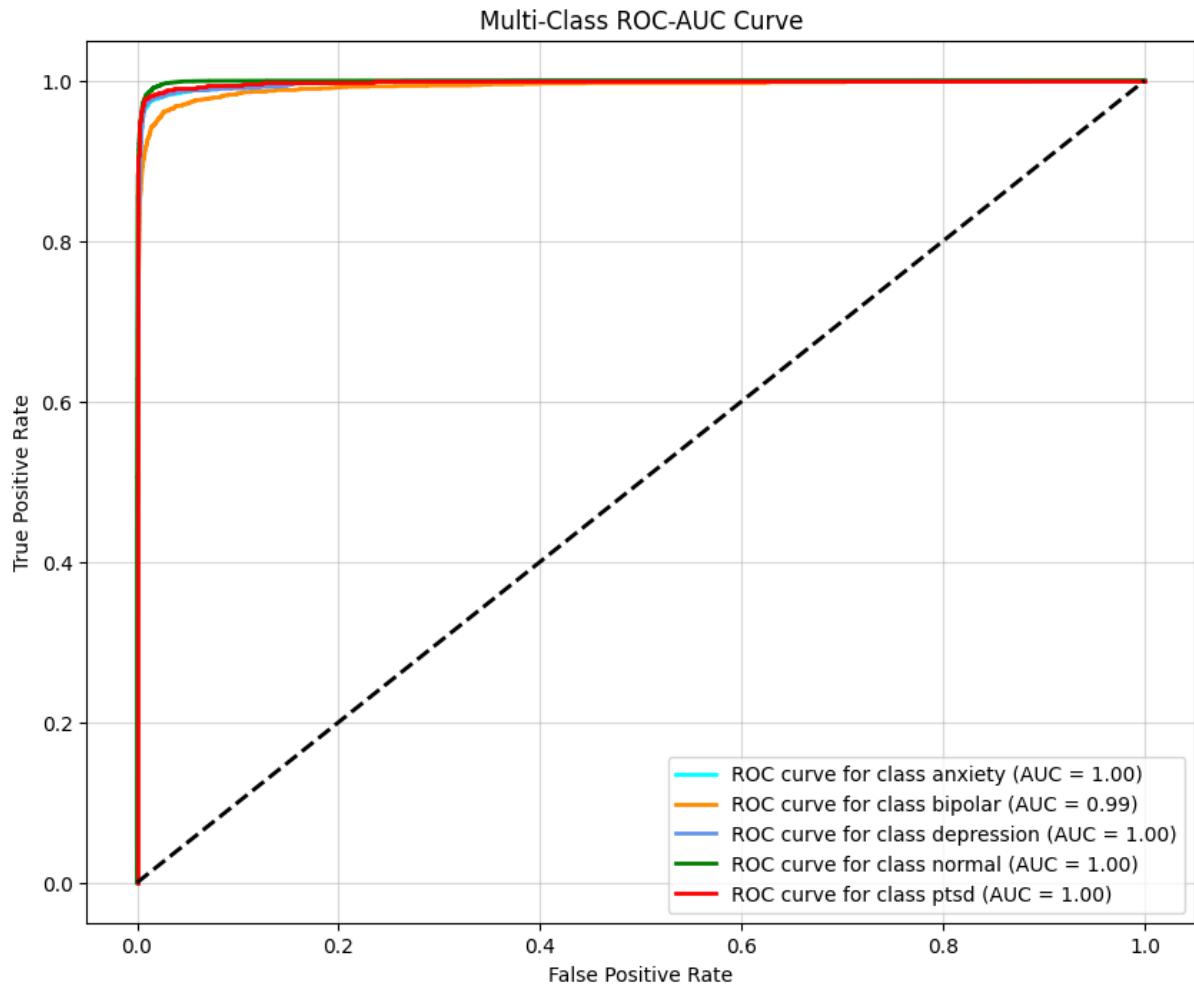


Figure 53: ROC AUC for Ensemble Model (LR - meta learner)

overfitting, which is a common issue when combining multiple models. By applying regularization, Logistic Regression ensures that the ensemble does not become too complex or overly tuned to the training data, improving the generalization performance on unseen data.

Moreover, Logistic Regression handles imbalanced data well, which is often a challenge in classification tasks. In ensemble settings, some base models might be more sensitive to the majority class, but Logistic Regression can adjust for class imbalances by applying class weights, ensuring that the ensemble model remains fair and does not favor the dominant class.

Another advantage of Logistic Regression is its ability to handle a large number of input features. In the case of an ensemble model, each base model's prediction can be treated as a feature, and Logistic Regression can effectively combine these features to make the final decision. This makes Logistic Regression versatile and adaptable when working with ensembles composed of diverse models, whether they are based on different algorithms or different data representations.

## ASMPFMHDD

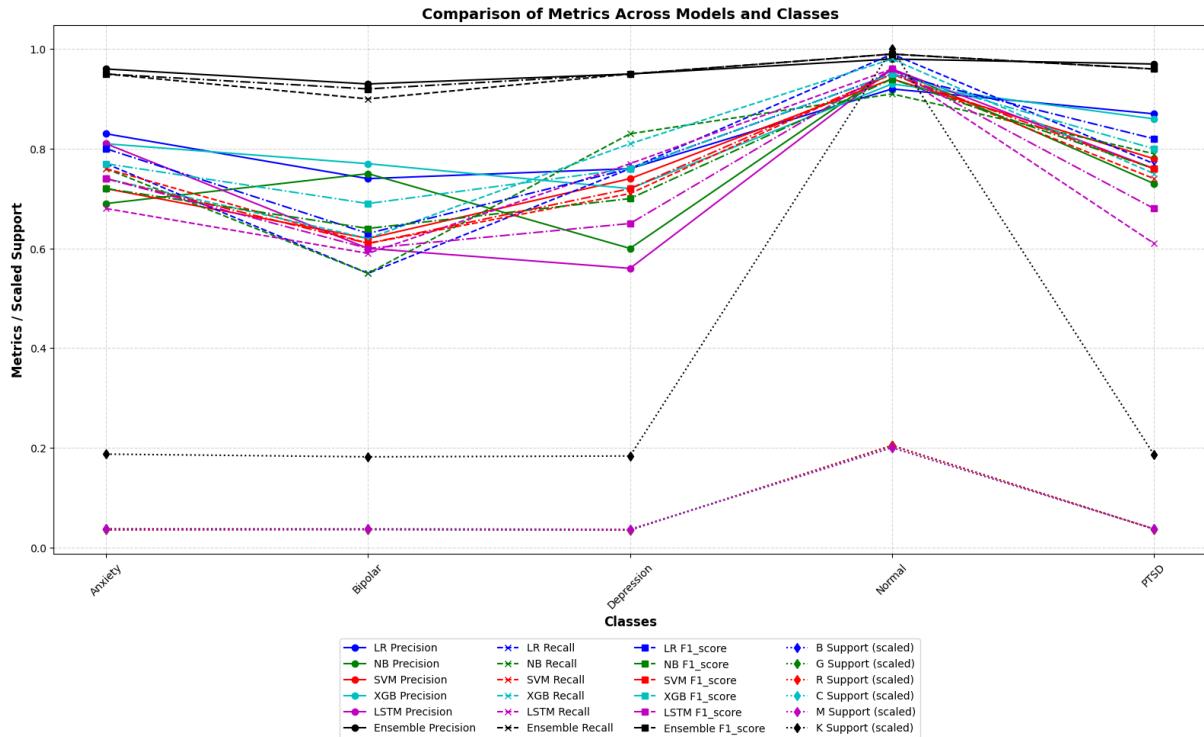


Figure 54: Comparison of results with Ensemble Model (LR - meta learner)

In summary, Logistic Regression is an ideal choice as a meta-learner in ensemble models because of its simplicity, interpretability, efficiency, and strong performance in combining probabilistic outputs. Its ability to apply regularization, handle imbalanced data, and process multiple features makes it a reliable and effective meta-learner that enhances the performance of the overall ensemble model.

## 10 Conclusion

### 10.1 Project Benefits

The project on detecting mental health disorders through social media analysis offers a wide array of significant benefits, both immediate and long-term, across multiple dimensions. First and foremost, it addresses a critical issue in mental health care—early detection and intervention. Social media has become a ubiquitous platform where people express their thoughts, feelings, and emotional states, often unconsciously. By leveraging the vast amounts of data available on social media platforms, our project seeks to tap into this resource to identify early signs of mental health disorders such as anxiety, depression, and more severe conditions like bipolar disorder or schizophrenia. The ability to detect mental health issues through real-time social media data is a game-changer for public health systems, mental health practitioners, and even individuals who may not realize they are at risk. Early detection enables timely intervention, reducing the overall burden of mental health disorders on society by preventing escalation into more severe conditions that often lead to hospitalization, self-harm, or even suicide. In this sense, the project aligns with global health initiatives that emphasize early diagnosis and preventive care.

Moreover, this project holds significant potential for improving the accuracy and efficiency of mental health diagnostics. Traditional diagnostic methods are often time-consuming, subjective, and reliant on self-reporting, which can lead to underdiagnosis or misdiagnosis. By utilizing machine learning algorithms and natural language processing techniques, our project automates the process of sentiment and behavioral analysis on social media platforms, offering a more objective and data-driven approach. This automated system can process large volumes of data much faster than human professionals, providing insights that would be impossible to glean from manual analysis. The algorithms developed as part of this project can be easily scaled to analyze millions of social media posts, enabling a broader reach in monitoring public mental health trends. Additionally, the project offers practical benefits for mental health professionals, allowing them to focus on treatment and intervention rather than diagnosis. It provides a tool that can be integrated into telehealth systems, offering mental health screening at scale, which is particularly valuable in underserved or rural areas where access to mental health professionals is limited.

From a technological standpoint, the project offers a host of reusable components and methodologies. The machine learning models developed, the sentiment analysis tools, and the overall data pipeline are designed to be scalable and modular. These components can be adapted and extended to other domains beyond mental health, such as market sentiment analysis, public opinion monitoring, or even detecting harmful behavior like cyberbullying and harassment online. By advancing the state of the art in social media analytics, this project contributes to the

growing field of AI-driven health care solutions. Furthermore, it provides a blueprint for future interdisciplinary work that integrates data science, psychology, and public health.

## 10.2 Future Scope for Improvements

While this project offers numerous immediate benefits, there is substantial room for future enhancements that can broaden its applicability, accuracy, and effectiveness. Currently, the project focuses on analyzing Reddit data using PRAW, which is limited to a specific social media platform and dataset. In the future, incorporating data from other platforms like Facebook, Instagram, Twitter, and even niche forums could provide a more comprehensive understanding of an individual's mental health status. Different platforms cater to different demographics and social behaviors, and expanding the dataset will allow for a more holistic analysis of mental health indicators across various user bases. Additionally, expanding the dataset to include multilingual posts or integrating language translation capabilities could make the system applicable to a global audience, helping to identify mental health issues in non-English speaking populations.

Moreover, future improvements could focus on integrating real-time data analysis capabilities. Currently, our project is based on batch processing of historical data. However, in future iterations, the system could be developed to perform real-time monitoring, offering immediate feedback and potentially alerting health professionals or loved ones when someone shows signs of mental distress. This real-time capability would be invaluable in emergency situations, allowing for immediate intervention. Developing a mobile application or a web-based interface where users can voluntarily connect their social media accounts to monitor their mental health status could also increase user engagement and provide individuals with direct feedback on their well-being.

Another significant future enhancement could involve incorporating ethical considerations and improving user privacy. As mental health is a sensitive subject, ensuring that the system is designed with robust privacy protections is critical. Future work could focus on using differential privacy or other anonymization techniques to ensure that user data remains confidential while still allowing for effective analysis. Moreover, collaborating with psychologists, ethicists, and legal experts could help refine the system to ensure it adheres to ethical guidelines and avoids potential harm, such as misdiagnosis or privacy violations.

Lastly, the future scope of this project could include expanding its use in clinical settings. While the current system is primarily designed as a research tool, future iterations could be developed in collaboration with mental health professionals to ensure that it meets clinical standards.

## 11 References

- [1] Hatoon S AlSagri and Mourad Ykhlef. Machine learning-based approach for depression detection in twitter using content and activity features. *IEICE Transactions on Information and Systems*, 103(8):1825–1832, 2020.
- [2] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. Predicting depression via social media. *Proceedings of the International AAAI Conference on Web and Social Media*, 2013.
- [3] Jetli Chung and Jason Teo. Single classifier vs. ensemble machine learning approaches for mental health prediction. *Brain Informatics*, 10(1):1, jan 2023.
- [4] Sharath Chandra Guntuku, David Bryce Yaden, Margaret L. Kern, Lyle H. Ungar, and Johannes C. Eichstaedt. Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences*, 18:43–49, 2017.
- [5] Priya Mathur, Amit Kumar Gupta, and Abhishek Dadhich. Mental health classification on social-media: Systematic review. *Proceedings of the 4th International Conference on Information Management & Machine Intelligence*, 2022.
- [6] Moin Nadeem. Identifying depression on twitter. *arXiv preprint arXiv:1607.07384*, 2016.
- [7] Ramin Safa, S. A. Edalatpanah, and Ali Sorourkhah. Predicting mental health using social media: A roadmap for future development, 2023.
- [8] Dip Kumar Saha, Tuhin Hossain, Mejdl Safran, Sultan Alfarhood, M. F. Mridha, and Dunren Che. Ensemble of hybrid model based technique for early detecting of depression based on svm and neural networks. *Scientific Reports*, 14(1):25470, oct 2024.
- [9] Konda Vaishnavi, U Nikhitha Kamath, B Ashwath Rao, and N V Subba Reddy. Predicting mental health illness using machine learning algorithms. *Journal of Physics: Conference Series*, 2161(1):012021, jan 2022.

## APPENDIX A - Prototype

### System Dependencies Installation

```
!apt-get install -y ffmpeg libsm6 libxext6
!apt-get install -y tesseract-ocr
!apt-get install -y portaudio19-dev
```

This code installs essential system-level dependencies required for multimedia processing, Optical Character Recognition (OCR), and audio manipulation tasks. The ffmpeg library provides robust capabilities for processing video and audio data, supporting operations like format conversion, compression, and extraction. The libsm6 and libxext6 libraries are X Window System components needed for graphical processing and enabling compatibility with multimedia frameworks like OpenCV. The tesseract-ocr package installs Tesseract, a powerful OCR engine used for extracting text from images and documents. Lastly, portaudio19-dev is a development package for the PortAudio library, which enables cross-platform audio processing and is often required for speech recognition and audio streaming applications. These installations ensure that the environment is properly configured to handle complex data processing workflows.

### Import Libraries

```
import streamlit as st
import joblib
import pandas as pd
import praw
from PIL import Image
from deep_translator import GoogleTranslator
import requests
from io import BytesIO
from collections import Counter
import google.generativeai as genai
import cv2
import numpy as np
import whisper
import tempfile
import os
from pydub import AudioSegment
import subprocess
import re
import librosa
import librosa.display
import tensorflow as tf
import pytesseract
```

This code snippet includes various library imports essential for building a multi-functional application that processes and analyzes multimedia data, text, and machine learning tasks. Streamlit is used for building interactive web apps. Joblib assists with object serialization, and pandas handles structured data manipulation. PRAW enables interaction with Reddit's API, while Pillow facilitates image processing. GoogleTranslator supports text translation, and requests fetches data from web sources. BytesIO handles in-memory byte streams, and collections.Counter aids in counting elements in datasets. Google Generative AI tools are also incorporated for advanced AI tasks. For media processing, OpenCV and NumPy handle image and video manipulations. Whisper is an ASR (Automatic Speech Recognition) tool, and tempfile manages temporary files. Pydub processes audio, and subprocess executes shell commands. Re provides regular expressions for text parsing, Librosa supports audio feature extraction and visualization, and TensorFlow enables building machine learning models. Lastly, pytesseract extracts text from images using OCR (Optical Character Recognition), completing a robust toolset for diverse computational tasks.

### Configuration and Model Initialization

```

# Configure Tesseract and FFMPEG
pytesseract.pytesseract.tesseract_cmd = '/usr/bin/tesseract'
os.environ["FFMPEG_BINARY"] = "/usr/bin/ffmpeg"
# Load Whisper model for audio transcription
whisper_model = whisper.load_model("base")
# Load the saved logistic regression model and vectorizer
model = joblib.load('LRmodel.pkl')
vectorizer = joblib.load('LRvectorizer.pkl')
# Initialize Reddit API
reddit = praw.Reddit(client_id='<CLIENT_ID>',
client_secret='<CLIENT_SECRET_KEY>', user_agent='Mental_Health')
# Configure Gemini API for wellbeing insights
genai.configure(api_key="<GEMINI_API_KEY>")
generation_config = {
    "temperature": 1, "top_p": 0.95, "top_k": 40,           "
    "max_output_tokens": 8192, "response_mime_type": "text
    /plain",
}
gemini_model = genai.GenerativeModel(
    model_name="gemini-1.5-flash",           generation_config=
    generation_config,
)

```

This code initializes and configures various tools and models for a complex data-processing pipeline. The Tesseract OCR engine is configured by specifying its executable path to enable text extraction from images. The FFMPEG binary is similarly set to facilitate multimedia

processing tasks. The Whisper model, a speech-to-text solution, is loaded with its base configuration for audio transcription. Saved machine learning artifacts, a logistic regression model, and a vectorizer are loaded using joblib, providing a pre-trained setup for text classification tasks. The Reddit API is initialized using PRAW, with credentials to interact with Reddit's platform for data retrieval. For generative AI tasks, the Gemini API is configured with an API key and generation parameters such as temperature, top-p sampling, and maximum output token limit. A generative model is instantiated using these configurations, designed to provide wellbeing insights. This setup creates a cohesive framework for multimedia processing, natural language processing, and machine learning applications.

#### Fetching Reddit User Text Posts

```
# Function to fetch text-based posts from Reddit
def fetch_user_text_posts(username):
    try:
        user = reddit.redditor(username)
        posts = [post.title + "\n" + post.selftext for post
                 in user.submissions.new(limit=20)]
        return posts
    except Exception as e:
        st.write(f"Error fetching text posts: {e}")
        return []
```

This function, `fetch_user_text_posts`, is designed to fetch text-based posts from a specified Reddit user. The function takes a `username` as an input and attempts to retrieve the most recent 20 text-based posts from that user's Reddit submissions. Using the PRAW library, the `reddit.redditor(username)` method is called to access the user's posts, and the function iterates over these posts, concatenating the title and the content (`selftext`) of each post into a single string. These concatenated post details are then stored in a list. If an error occurs during this process (such as network issues or invalid user input), the function catches the exception and displays an error message using Streamlit's `st.write()`. If an exception is caught, an empty list is returned. This function is useful for collecting Reddit text data for further analysis or processing.

## Fetching Image-Based Posts from Reddit and Performing OCR

```

# Function to fetch image-based posts from Reddit and
perform OCR

def fetch_user_images_and_extract_text(username):
    try:
        user = reddit.redditor(username)
        images = [post.url for post in user.submissions.new(
            limit=20) if post.url.endswith('.jpg', '.jpeg',
            '.png', '.webp', '.bmp', '.tiff'))]

        extracted_texts = []
        for image_url in images:
            try:
                response = requests.get(image_url)
                image = Image.open(BytesIO(response.content))
                st.image(image, caption="Fetched_Image",
                         use_column_width=True)

                extracted_text = extract_text_from_image(
                    image)
                if extracted_text.strip():
                    translated_text = GoogleTranslator(
                        source='auto', target='en').translate(
                        extracted_text)
                    extracted_texts.append(translated_text)
                    st.write("Extracted_and_Translated_Text_
                             from_Image:")
                    st.text(translated_text)
            except Exception as e:
                st.write(f"Error_processing_image_{image_url}
                         :_{e}")

        return extracted_texts
    except Exception as e:
        st.write(f"Error_fetching_images:_{e}")
        return []

```

This function, `fetch_user_images_and_extract_text`, is designed to fetch image-based posts from a specified Reddit user and perform Optical Character Recognition (OCR) to extract text from the images. The function first attempts to retrieve the most recent 20 submissions from the specified Reddit user using the PRAW library. It filters the posts to include only those with image URLs that match common image file formats (e.g., .jpg, .jpeg, .png, etc.). For each valid image URL, the function fetches the image using the `requests` library and then processes it by opening the image with Pillow's `Image.open()` method. The image is dis-

played on the Streamlit app using `st.image()` with the option to show it in the appropriate column width. After displaying the image, the function calls `extract_text_from_image`, which performs OCR using Tesseract (assumed to be defined elsewhere in the code). If any text is successfully extracted, it is translated into English using the `GoogleTranslator` from the `deep_translator` library, and the translated text is displayed on the app using `st.write()` and `st.text()`. In case of errors (e.g., issues fetching the image or processing the OCR), exceptions are caught and an error message is displayed via Streamlit's `st.write()`. The function returns a list of extracted and translated texts from the images, or an empty list if any errors occur during the process. This function is helpful for gathering, processing, and translating text from image posts on Reddit.

### Classifying Text and Displaying Results

```
# Function to classify text and display result
def classify_text(text):
    input_vectorized = vectorizer.transform([text])
    prediction_proba = model.predict_proba(input_vectorized)

    issue_labels = model.classes_
    proba_df = pd.DataFrame(prediction_proba, columns=
        issue_labels).T
    proba_df.columns = ['Probability']

    top_issue = proba_df['Probability'].idxmax()
    top_probability = proba_df['Probability'].max()

    st.write(f"The most likely mental health concern is:{top_issue} with a probability of {top_probability:.2%}")

get_wellbeing_insight(text, top_issue)
```

This function, `classify_text`, is used to classify a given text and display the most likely mental health concern along with its probability. The function first vectorizes the input `text` using the `vectorizer.transform()` method, which converts the text into a format suitable for the machine learning model. The `model.predict_proba()` method is then called to get the probabilities for each possible class label, which represents different mental health issues. The class labels (issues) are retrieved using `model.classes_`, and a Pandas DataFrame (`proba_df`) is created to display the predicted probabilities for each issue. The DataFrame is transposed and renamed to give a clearer view, with a column for the probability values. The function then identifies the issue with the highest probability using `idxmax()` and retrieves the maximum probability value using `max()`. Finally, the most likely issue and its associated

probability are displayed on the Streamlit app using `st.write()`. Additionally, the function calls `get_wellbeing_insight`, passing the text and the top issue, likely to generate further insights into the mental health concern identified. This function is integral to classifying text data for mental health analysis and providing actionable insights based on the results.

### Getting Wellbeing Insights from Gemini Model

```
# Function to get wellbeing insights from Gemini model
def get_wellbeing_insight(text, top_issue):
    try:
        chat_session = gemini_model.start_chat(history[])
        prompt = f"<Prompt_to_get_the_well_being_based_on_Ryff_Scale_Six_Factor_Model>"

        response = chat_session.send_message(prompt)

        st.write("###_Wellbeing_Insight:")
        st.write(response.text)
    except Exception as e:
        st.write(f"Error_retrieving_wellbeing_insights:{e}")
    )
```

The function `get_wellbeing_insight` interacts with the Gemini AI model to generate insights related to a mental health issue based on the Ryff Scale of Psychological Well-Being. The function starts by initializing a chat session with the Gemini model using `gemini_model.start_chat()`. It then constructs a detailed prompt that outlines the six factors of well-being: autonomy, environmental mastery, personal growth, positive relations with others, purpose in life, and self-acceptance. These factors are crucial for evaluating an individual's psychological well-being. The prompt includes specific example statements related to each factor, offering context for how the mental health issue (`top_issue`) might affect the individual in each area. The function sends this prompt to the Gemini model and retrieves the response, which provides detailed advice and reflections on how the issue impacts each well-being factor. The response includes short paragraphs for each of the six factors, analyzing the potential effects of the issue on the individual's ability to function in these areas. After receiving the response, the function uses Streamlit's `st.write()` method to display the wellbeing insights. In case of any errors, such as issues with the Gemini model or the chat session, an error message is displayed using `st.write()`. This function is valuable for generating personalized psychological insights and offering practical advice based on the impact of a specific mental health issue.

### Extract Text from Image Using Tesseract

```
# Function to extract text from image using Tesseract
def extract_text_from_image(image):
    extracted_text = pytesseract.image_to_string(image)
    return extracted_text.splitlines()

# Function to extract text from an image using Tesseract
def extract_text_from_image_video(image):
    extracted_text = pytesseract.image_to_string(image)
    return extracted_text if extracted_text else "" # Return empty string if no text is found
```

The provided code consists of two functions designed to extract text from an image using the Tesseract OCR (Optical Character Recognition) engine.

The first function, `extract_text_from_image(image)`, takes an `image` object as input, applies the `pytesseract.image_to_string(image)` method to extract text from the image, and then splits the resulting text into individual lines using the `splitlines()` method. This allows the function to return a list where each element corresponds to a line of extracted text, making it easier to handle multiline results. The second function, `extract_text_from_image_video(image)`, operates similarly by extracting text from the image, but it includes an additional check to determine whether any text was successfully extracted. If no text is found, the function returns an empty string (" "), ensuring that it always returns a consistent type rather than `None`, which could potentially cause issues in other parts of the code. Both functions rely on the `pytesseract` library, which is a Python wrapper for the open-source Tesseract OCR engine, to perform the text extraction. These functions are useful for extracting text from images, which can be helpful in various use cases such as document scanning, analyzing images with embedded text, or processing video frames to capture text content.

### Extract Frames from Video File

```
# Function to extract 20 frames from a video file
def extract_frames(video_path, num_frames=20):
    cap = cv2.VideoCapture(video_path)
    total_frames = int(cap.get(cv2.CAP_PROP_FRAME_COUNT))

    frames = []
    frame_interval = total_frames // num_frames # Calculate frame interval
```

## ASMPFMHDD

### Extract Frames from Video File

```
for i in range(num_frames):
    cap.set(cv2.CAP_PROP_POS_FRAMES, i * frame_interval)

    ret, frame = cap.read()

    if ret:
        frame = cv2.cvtColor(frame, cv2.COLOR_BGR2RGB)
        frames.append(frame)

cap.release()
return frames
```

The `extract_frames(video_path, num_frames=20)` function extracts a specified number of frames from a given video file. The function uses the OpenCV library, specifically the `cv2.VideoCapture()` method, to load the video from the file path provided by the `video_path` argument. First, it calculates the total number of frames in the video using `cap.get(cv2.CAP_PROP_FRAME_COUNT)`. Next, it calculates the frame interval by dividing the total number of frames by the desired number of frames to be extracted (`num_frames`). For each of the frames, the function sets the position of the video playback to a specific frame using `cap.set(cv2.CAP_PROP_POS_FRAMES, i * frame_interval)`. It then reads the frame, converts the color from BGR (Blue-Green-Red) to RGB using `cv2.cvtColor(frame, cv2.COLOR_BGR2RGB)`, and appends the frame to a list. This process is repeated for the number of frames specified. Finally, the video capture object is released with `cap.release()`, and the list of frames is returned. This function can be useful in scenarios where frame extraction is needed from videos for further processing, such as image analysis, video summarization, or even object detection tasks.

### Transcribe Audio from Video

```
def transcribe_audio_from_video(video_file):
    try:
        # Save the uploaded video file to a temporary file
        with tempfile.NamedTemporaryFile(delete=False, suffix=".mp4") as temp_video_file:
            temp_video_file.write(video_file.read())
            temp_video_path = temp_video_file.name

        audio_path = tempfile.NamedTemporaryFile(suffix=".wav",
                                                delete=False).name
```

## Transcribe Audio from Video

```

# Extract audio from video using subprocess
subprocess.run(["ffmpeg", "-i", temp_video_path, "-q:a",
    "0", "-map", "a", audio_path, "-y"])
audio = AudioSegment.from_file(audio_path)

# Use Whisper to transcribe the audio
result = whisper_model.transcribe(audio_path)

# Get the transcribed text and translate if necessary
transcribed_text = result["text"]
translated_text = GoogleTranslator(source="auto", target
    ="en").translate(transcribed_text)

# Clean up temporary files
os.remove(temp_video_path)
os.remove(audio_path)

return translated_text

except Exception as e:
    # Display a user-friendly message if the video is too
    long or another error occurs
    if "duration" in str(e).lower() or "length" in str(e).
        lower():
        return "The_video_is_too_long_to_process._Please_
            upload_a_shorter_video."
    else:
        return f"An_error_occurred:_{e}"

```

The `transcribe_audio_from_video(video_file)` function processes a video file by first saving it to a temporary file on disk. It then extracts the audio using the `ffmpeg` command-line tool, saving the extracted audio as a WAV file. The extracted audio is processed using the Whisper model for transcription, which outputs the transcribed text. Additionally, the function utilizes the `GoogleTranslator` to translate the text into English if necessary. Temporary files created during processing are deleted to manage resources efficiently. In case of errors, the function returns a user-friendly message, especially handling scenarios where the video duration exceeds allowable limits. This function is useful for converting spoken content in videos to text for further processing.

### Translate Text Using DeepL

```
# Function to translate text using DeepL
def translate_text(text, target_lang="en"):
    try:
        if text:
            translated_text = GoogleTranslator(source="auto",
                                              target=target_lang).translate(text)
            return translated_text
        return "" # Return empty string if text is empty or None
    except Exception as e:
        return f"Error_translating_text:{str(e)}"
```

The `translate_text(text, target_lang="en")` function provides an interface to translate a given string of text into a specified target language, with English ("en") set as the default. Using the `GoogleTranslator` library, the function automatically detects the source language (`source="auto"`) and translates the text to the desired target language. If the input text is empty or `None`, the function gracefully returns an empty string. In case of any errors during the translation process, such as network issues or invalid input, an error message containing the exception details is returned. This function is particularly useful for applications requiring multilingual support, enabling seamless translation of text between languages.

### Extract Audio from Video File

```
# Function to extract audio from a video file
def extract_audio_from_video(video_path):
    try:
        # Generate a temporary audio file path
        audio_path = tempfile.NamedTemporaryFile(delete=False,
                                                suffix=".wav").name

        # Use FFmpeg to extract audio from video
        subprocess.run(["ffmpeg", "-i", video_path, "-q:a", "0",
                      "-map", "a", audio_path, "-y"])

        # Return the path of the extracted audio
        return audio_path

    except Exception as e:
        return f"Error_extracting_audio:{str(e)}"
```

The `extract_audio_from_video(video_path)` function facilitates the extraction of audio content from a video file. It first generates a temporary file path with a `.wav` suffix

to store the extracted audio. The function uses the `ffmpeg` command-line tool to process the video file specified by `video_path`, extracting the audio stream with high quality (`-q:a 0`) and saving it to the temporary file. The path to the extracted audio file is returned for further use. In the event of an error, such as invalid file paths or processing issues, the function catches the exception and returns a descriptive error message. This function is useful in multimedia applications where the separation of audio from video content is needed, such as transcription or audio analysis workflows.

### Analyze Audio Mood Based on Extracted Audio

```
# Function to analyze audio mood based on extracted audio
def analyze_audio_mood(video_path):
    try:
        # Extract audio from the video (assuming
        # extract_audio_from_video is implemented)
        audio_path = extract_audio_from_video(video_path)

        # Load the audio file using librosa
        y, sr = librosa.load(audio_path)

        # Extract MFCCs (Mel-frequency cepstral coefficients)
        # from the audio signal
        mfcc = librosa.feature.mfcc(y=y, sr=sr, n_mfcc=13)

        # Divide the MFCC array into 4 frequency bands and
        # calculate scalar mean for each band

        # Low Frequencies: MFCC 0, 1, 2
        low_freq_mfcc = np.mean(mfcc[0:3], axis=1)
        mean_low = np.mean(low_freq_mfcc) # Scalar mean for low
                                         # frequencies

        # Mid-Low Frequencies: MFCC 3, 4
        mid_low_freq_mfcc = np.mean(mfcc[3:5], axis=1)
        mean_mid_low = np.mean(mid_low_freq_mfcc) # Scalar mean
                                                # for mid-low frequencies

        # Mid-High Frequencies: MFCC 5, 6, 7
        mid_high_freq_mfcc = np.mean(mfcc[5:8], axis=1)
        mean_mid_high = np.mean(mid_high_freq_mfcc) # Scalar
                                                    # mean for mid-high frequencies

        # High Frequencies: MFCC 8, 9, 10, 11, 12
        high_freq_mfcc = np.mean(mfcc[8:13], axis=1)
        mean_high = np.mean(high_freq_mfcc) # Scalar mean for
                                         # high frequencies
```

## Analyze Audio Mood Based on Extracted Audio

```

# Now use these scalar means for classification

if mean_high <= mean_low and mean_high <= mean_mid_low
and mean_high <= mean_mid_high:
    return "Audio_sounds_normal,_with_no_dominant_
emotion_detected"

elif mean_mid_high <= mean_low and mean_mid_high <=
mean_mid_low and mean_mid_high <= mean_high:
    return "Audio_sounds_neutral,_calm,_or_peaceful"

elif mean_mid_low <= mean_low and mean_mid_low <=
mean_mid_high and mean_mid_low <= mean_high:
    return "Audio_sounds_slightly_melancholic,_or_neutral
"

elif mean_low <= mean_mid_low and mean_low <=
mean_mid_high and mean_low <= mean_high:
    return "Audio_sounds_calm,_or_melancholic,_with_less_
intensity"

elif mean_high > mean_low and mean_high > mean_mid_low
and mean_high <= mean_mid_high:
    return "Audio_sounds_depressive,_or_anxious,in_nature
"

else :
    return "Audio_sounds_upbeat,_and_energetic_(Happy)"

except Exception as e:
    return f"Error_analyzing_audio_mood:{str(e)}"
```

The `analyze_audio_mood(video_path)` function determines the mood of an audio segment extracted from a given video file. It begins by extracting audio using the `extract_audio_from_video` function and loading the audio file with `librosa`. The function computes Mel-frequency cepstral coefficients (MFCCs), which are features widely used in audio signal processing for mood or emotion analysis. The MFCCs are divided into four frequency bands (low, mid-low, mid-high, high), and scalar means are computed for each band. Based on these scalar values, a series of conditions classify the mood as normal, neutral, melancholic, calm, depressive, or upbeat. In case of an exception, an error message with details is returned. This function can be utilized in applications such as multimedia content analysis, emotion recognition, or mood-based music recommendations.

## Twitter API call and post extraction

```

# Initialize Twitter API
BEARER_TOKEN = "<TWITTER-BEARER-TOKEN>"
client = tweepy.Client(bearer_token=BEARER_TOKEN)

# Twitter
def fetch_image_content(image_url):
    """Fetch_and_process_an_image_from_a_URL."""
    try:
        response = requests.get(image_url, timeout=10)
        response.raise_for_status() # Ensure the request was
                                    # successful
        return Image.open(BytesIO(response.content))
    except Exception as e:
        st.write(f"Error_fetching_image:{e}")
        return None

def get_latest_tweets_with_images(username, max_items=10):
    """Fetch_latest_tweets_with_text_and_associated_images."""
    # Fetch user details to get user ID
    user = client.get_user(username=username)
    if not user.data:
        return [], []

    user_id = user.data.id

    # Fetch the latest tweets (exclude retweets and replies)
    response = client.get_users_tweets(
        id=user_id,
        tweet_fields=["attachments"],
        expansions=["attachments.media_keys"],
        media_fields=["url"],
        exclude=["retweets", "replies"],
        max_results=max_items
    )

    tweet_data = []

    if response.data:
        for tweet in response.data:
            # Extract text
            text = tweet.text

```

## Twitter API call and post extraction

```

# Extract images if available
images = []
if hasattr(tweet, "attachments") and tweet.attachments is not None:
    if "media_keys" in tweet.attachments:
        for media_key in tweet.attachments["media_keys"]:
            media = next(
                (media for media in response.includes.get("media", []) if
                 media["media_key"] == media_key),
                None
            )
            if media and media.type == "photo":
                images.append(media.url)

# Append tweet data
tweet_data.append({"text": text, "images": images})

return tweet_data

```

The above code snippet initializes the Twitter API using the `tweepy` library, specifically with a BEARER\_TOKEN for authentication. It defines two functions: `fetch_image_content` and `get_latest_tweets_with_images`. The `fetch_image_content` function retrieves an image from a provided URL using the `requests` library. It ensures successful HTTP requests via `response.raise_for_status()` and opens the image using the `Pillow` library's `Image.open`, handling errors gracefully by returning `None` if an exception occurs. The `get_latest_tweets_with_images` function fetches the latest tweets from a specified username, extracting text and associated image URLs. It first retrieves the user's unique Twitter ID using `client.get_user`. It then fetches tweets using `client.get_users_tweets`, excluding retweets and replies, and includes `attachments.media_keys` to identify images. For each tweet, it extracts the text and resolves media URLs by matching `media_keys` with `response.includes.media` data, appending only those of type `photo`. The final result is a list of dictionaries, each containing tweet text and a list of image URLs.

## Streamlit Mental Health Disorder Detection App

```

# Define the Streamlit app
def run_app():
    st.title("Mental_Health_Disorder_Detection")

    option = st.sidebar.selectbox(
        "Choose_an_option",
        ["Text_Input", "Image_Upload", "Video_Upload", "Reddit_Username_Analysis"]
    )

# Text Input
if option == "Text_Input":
    st.subheader("Enter_Text_to_Classify_Mental_Health_Issue")
    input_text = st.text_area("Enter_your_text_here:")

    if st.button("Classify_Text"):
        if input_text.strip() == "":
            st.write("Please_enter_some_text_to_classify.")
        else:
            translated_text = GoogleTranslator(source='auto',
                                              target='en').translate(input_text)
            st.write("Translated_Text_(to_English):")
            st.write(translated_text)
            classify_text(translated_text)

# Image Upload
elif option == "Image_Upload":
    st.subheader("Upload_an_Image_to_Extract_and_Classify_Text")
    uploaded_image = st.file_uploader("Upload_an_Image",
                                       type=["jpg", "jpeg", "png", "webp", "bmp", "tiff"])

    if uploaded_image is not None:
        image = Image.open(uploaded_image)
        st.image(image, caption="Uploaded_Image",
                  use_column_width=True)

        extracted_text = extract_text_from_image(image)
        translated_text = GoogleTranslator(source='auto',
                                           target='en').translate("\n".join(extracted_text))

        st.subheader("Translated_Text_(to_English) ")
        st.text(translated_text)

```

## ASMPFMHDD

### Streamlit Mental Health Disorder Detection App

```
if st.button("Classify_Extracted_Text"):
    if not translated_text or translated_text.strip() == "":
        st.write("It_is_normal_with_probability_100%")
    else:
        classify_text(translated_text)

# Video Upload
elif option == "Video_Upload":
    st.subheader("Upload_a_Video_to_Extract_and_Classify_Text")
    video_file = st.file_uploader("Choose_a_video_file",
                                  type=["mp4", "mov", "avi"])

    if video_file:
        video_path = "/tmp/uploaded_video.mp4"
        with open(video_path, "wb") as f:
            f.write(video_file.getbuffer())

        st.video(video_file)

        frames = extract_frames(video_path)
        combined_text = ""

        st.write("Extracting_frames_from_video...")
        for idx, frame in enumerate(frames):
            st.image(frame, caption=f"Frame_{idx+1}",
                     use_column_width=True)
            text_from_frame = extract_text_from_image_video(
                frame)
            if text_from_frame and text_from_frame not in combined_text:
                combined_text += text_from_frame + " "

        st.write("Text_Extracted_from_Video_Frames:")
        st.text(combined_text)

        transcribed_audio_text = transcribe_audio_from_video(
            video_file)
        st.write("Transcribed_Audio_Text:")
        st.text(transcribed_audio_text)

        full_combined_text = combined_text + " " +
            transcribed_audio_text
        translated_combined_text = translate_text(
            full_combined_text)
```

## ASMPFMHDD

### Streamlit Mental Health Disorder Detection App

```
st.write("Translated_Combined_Text_(Frames+_Audio):")
st.text(translated_combined_text)

st.write("Analyzing_Audio_Mood...")
mood_result = analyze_audio_mood(video_path)
st.write(mood_result)

if st.button("Classify_Extracted_Text"):
    classify_text(translated_combined_text)

# Reddit Username Analysis
elif option == "Reddit_Username_Analysis":
    st.subheader("Enter_Reddit_Username_for_Analysis")
    username = st.text_input("Enter_Reddit_username:")

if st.button("Analyze"):
    text_posts = fetch_user_text_posts(username)
    image_texts = fetch_user_images_and_extract_text(
        username)

    all_text = text_posts + image_texts
    if all_text:
        predictions = [model.predict(vectorizer.
            transform([text]))[0] for text in all_text]
        issue_counts = Counter(predictions)
        top_issue, top_count = issue_counts.most_common
            (1)[0]
        st.write(f"The_most_frequent_issue:_{top_issue}_"
            f"({(top_count/_len(predictions))_*100:.2f}%")
        issue_distribution = pd.DataFrame(issue_counts.
            items(), columns=['Mental_Health_Issue', 'Count'])
        st.write("Mental_health_issue_distribution_"
            "across_posts:")
        st.write(issue_distribution)

    # Call the Gemini model to get well-being
    # insights
    get_wellbeing_insight("".join(all_text),
        top_issue)
else:
    st.write("No_valid_text_found_for_analysis.")
```

## ASMPFMHDD

### Streamlit Mental Health Disorder Detection App

```
# Twitter Username Analysis
elif option == "Twitter_Username_Analysis":
    st.subheader("Enter_Twitter_Username_for_Analysis")
    username = st.text_input("Enter_Twitter_username:")

    if st.button("Analyze"):
        if username.strip() == "":
            st.write("Please_enter_a_Twitter_username.")
        else:
            # Fetch the latest tweets with associated images
            tweets_with_images =
                get_latest_tweets_with_images(username)

            # Extract text content from tweets
            text_posts = [tweet['text'] for tweet in
                tweets_with_images if tweet['text']]
            st.write("Recent_Text_Posts_from_Tweets:")
            st.write(text_posts[:3]) # Display a few posts
            for review

            # Extract and process text from associated
            # images
            image_texts = []
            for tweet in tweets_with_images:
                for image_url in tweet['images']:
                    image = fetch_image_content(image_url)
                    if image:
                        st.image(image, caption=f"Image_from
                            _Tweet", use_column_width=True)
                    if image:
                        extracted_text =
                            extract_text_from_image(image) # Assuming a text extraction
                        function is defined
                    if extracted_text:
                        image_texts.append(
                            extracted_text)

            # Combine text from both tweet text and
            # extracted image text
            all_text = text_posts + image_texts

            # Ensure all entries in all_text are strings
            all_text = [str(text) for text in all_text if
                text]
```

## ASMPFMHDD

### Streamlit Mental Health Disorder Detection App

```
if all_text:
    predictions = []
    for text in all_text:
        try:
            # Vectorize and classify each text
            input_vectorized = vectorizer.
                transform([text])
            prediction = model.predict(
                input_vectorized)
            predictions.append(prediction[0])
        except Exception as e:
            st.write(f"Error_processing_text:{text[:50]}...-{e}")
            continue

    # Count the most common mental health issue
    issue_counts = Counter(predictions)
    top_issue, top_count = issue_counts.
        most_common(1)[0]
    top_percentage = (top_count / len(
        predictions)) * 100

    st.write(f"The_most_frequently_detected_
        mental_health_concern_is:{top_issue},_
        appearing_in_{top_percentage:.2f}%_of_
        analyzed_text.")
    issue_distribution = pd.DataFrame(
        issue_counts.items(), columns=['Mental_
            Health_Issue', 'Count'])
    st.write("Mental_health_issue_distribution_
        across_posts:")
    st.write(issue_distribution)

    # Call the Gemini model to get well-being
    # insights
    get_wellbeing_insight("".join(all_text),
        top_issue)
else:
    st.write("No_valid_text_found_for_analysis."
        )

# Run the app
if __name__ == '__main__':
    run_app()
```

The `run_app()` function defines a Streamlit application to detect mental health disorders using multiple input methods: text input, image upload, video upload, and Reddit username analysis. Users can interact with a sidebar to choose their preferred mode of input. Each mode processes the data accordingly, utilizing pre-defined helper functions for text translation, text extraction from images, audio transcription, and Reddit user data retrieval. Depending on the mode, the app combines extracted and processed data to classify mental health concerns or display insights. It leverages libraries like `GoogleTranslator`, `librosa`, and `FFmpeg` for processing and analysis. The results include mood analysis, classification probabilities, and potential mental health issues. This app provides a comprehensive tool for mental health-related data analysis. For the Twitter analysis mode, the app allows users to input a Twitter username, fetches the latest tweets from the user, and extracts both text content and any associated images. It uses the `tweepy` library to interact with the Twitter API and retrieve tweets that are not retweets or replies. For each tweet, the app extracts the text and checks for any media attachments, specifically images. If images are found, it fetches and processes them using the `fetch_image_content` function, which downloads the image and uses OCR (Optical Character Recognition) to extract text from the image. Both the tweet text and any extracted text from the images are combined to create a comprehensive dataset for analysis. This data is then passed through a pre-trained machine learning model to classify the mental health concerns present in the posts. The model generates predictions, and the app displays the most frequently detected mental health issue along with its percentage occurrence. It also shows the distribution of mental health issues across the posts and provides insights using a separate well-being model. This feature enables the app to offer a detailed analysis of the mental health state inferred from Twitter posts, based on the extracted text and images. Additionally, the app integrates the Gemini API for advanced mood analysis and mental well-being insights. The Gemini API, accessed through a Flask server, provides an external service that can process large volumes of text data for sentiment analysis, mood classification, and mental health predictions. The Gemini API is a machine learning-powered API that allows users to analyze textual data by passing in relevant text input. In this case, the app sends the combined text from Twitter posts (including extracted text from images) to the Gemini API for analysis. The API returns insights such as sentiment scores, emotional tones, and relevant mental health classifications based on the input. The Flask server acts as an intermediary between the Streamlit app and the Gemini API. It handles HTTP requests from the Streamlit app, passes the text data to the Gemini API, and processes the responses before sending them back to the Streamlit interface. When a user submits their data, the app makes an HTTP request to the Flask server, which in turn queries the Gemini API. The Flask server ensures that the data is properly formatted and can handle various types of inputs, whether they are text, audio, or other formats. The Gemini API processes this data and returns a structured response that contains mood-related insights and mental health predictions.

## ASMPFMHDD

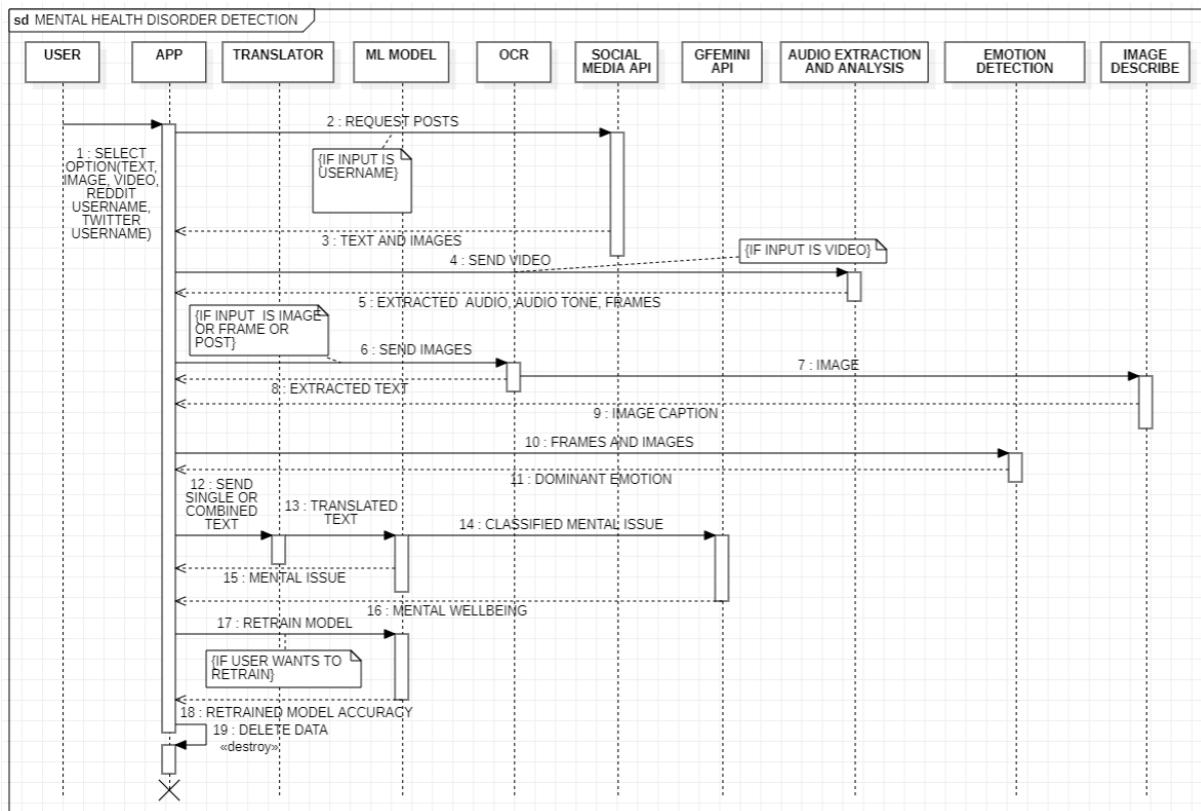


Figure 55: Sequence Diagram of the Application

Below are some screenshots from the web application.

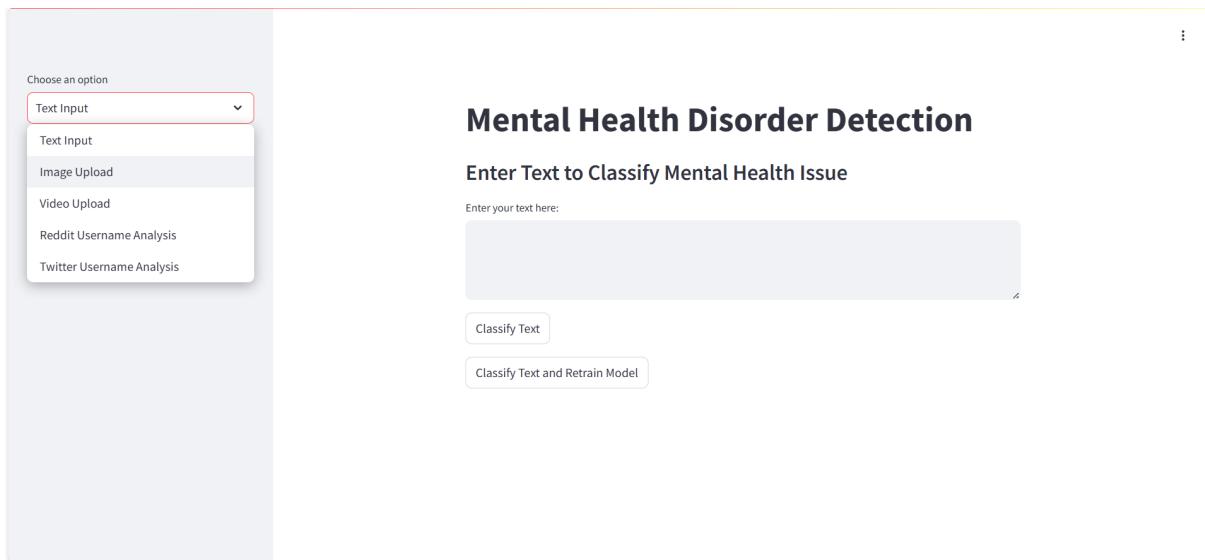


Figure 56: Website with all options

# ASMPFMHDD

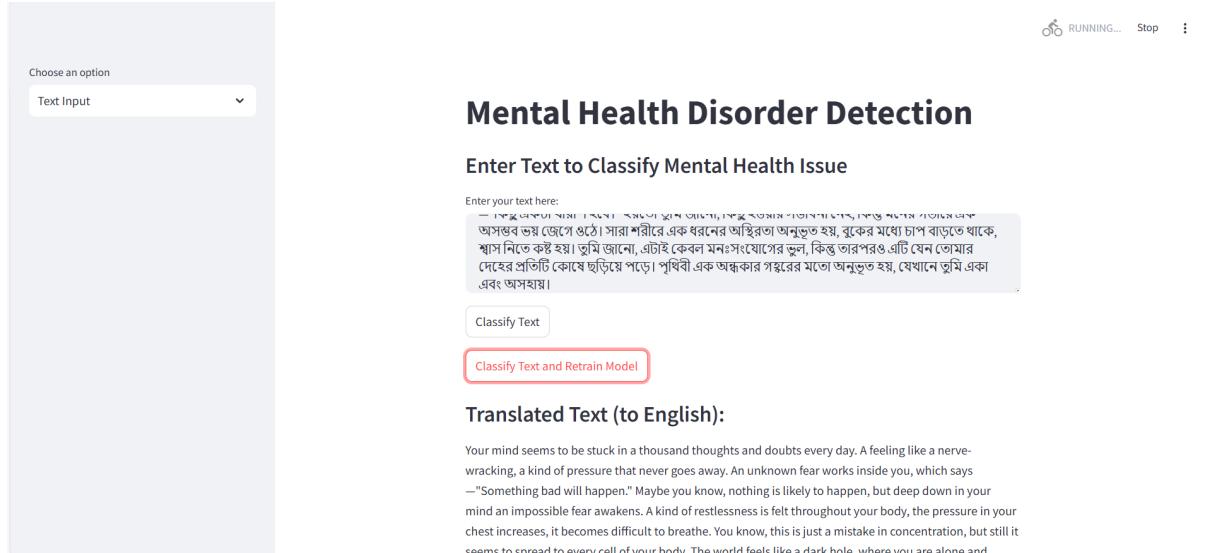


Figure 57: Entering Text for classification

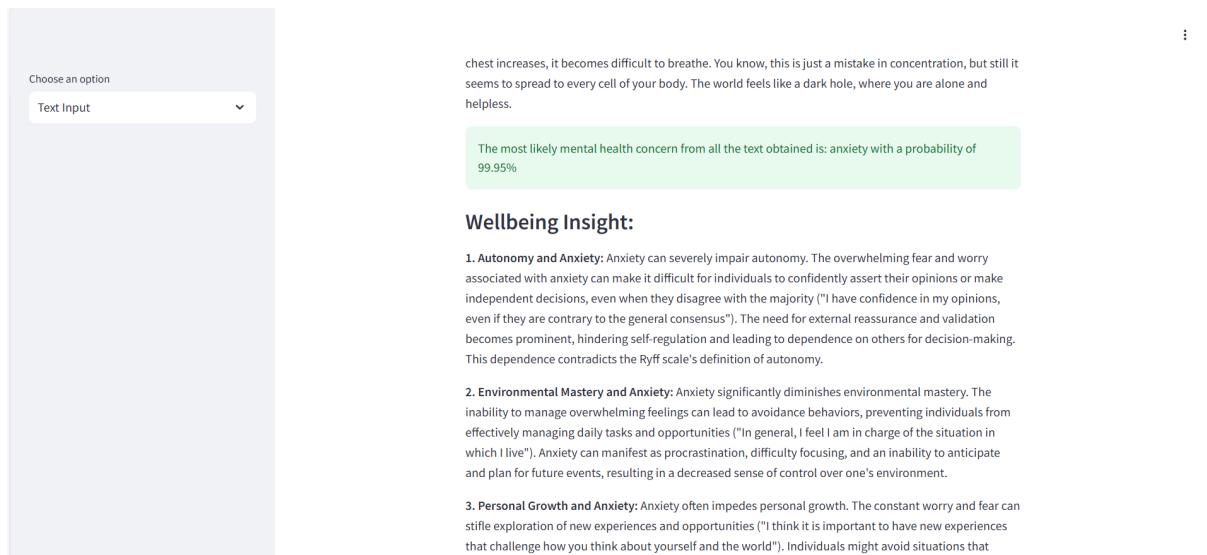


Figure 58: Text Classification Result

# ASMPFMHDD

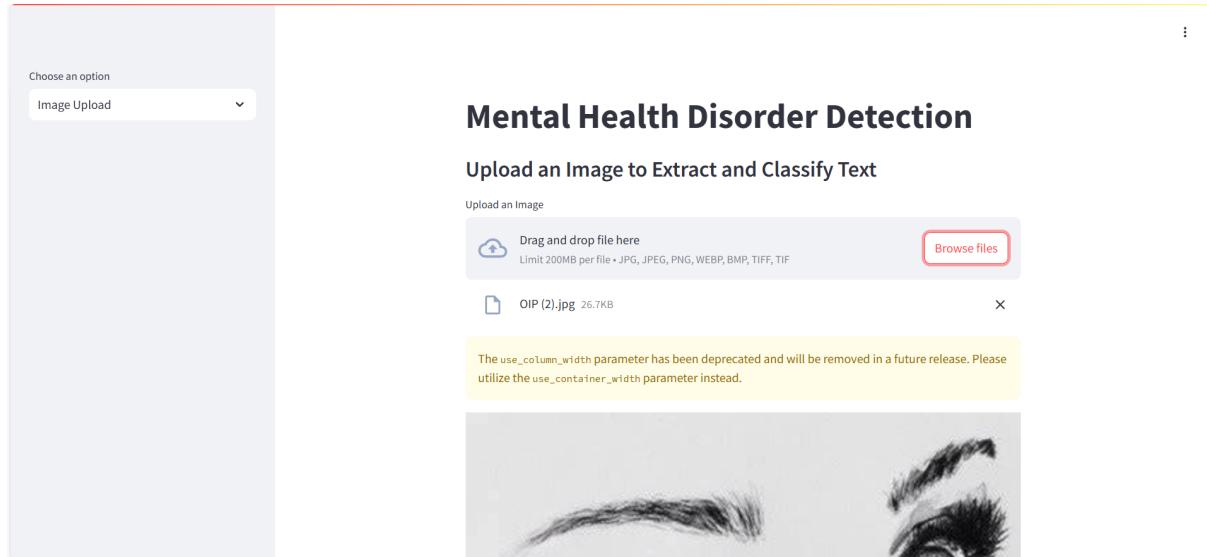


Figure 59: Upload Image

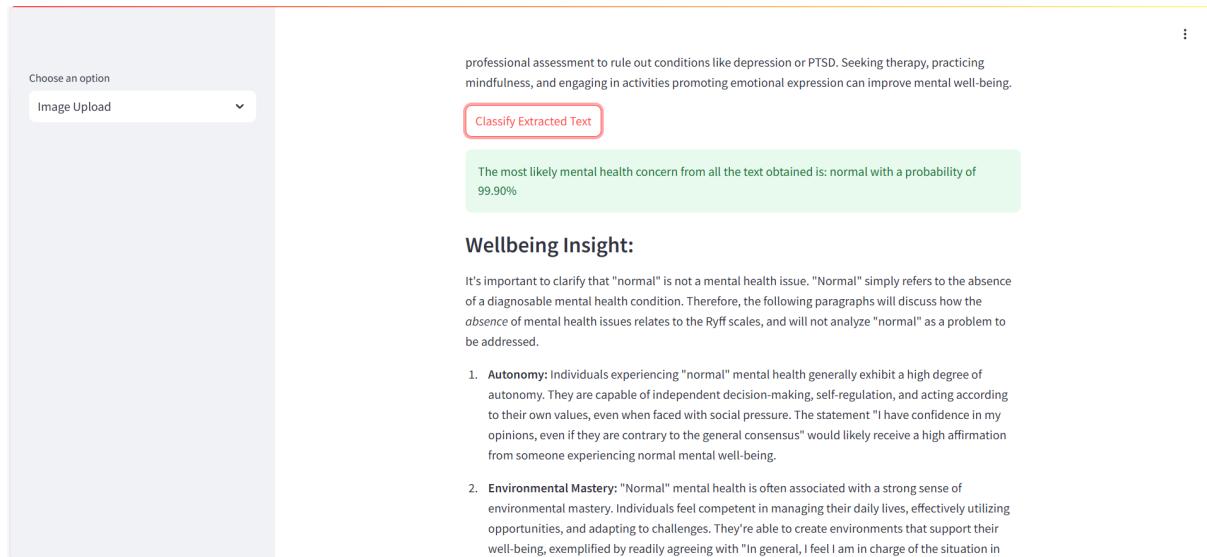


Figure 60: Image Classification Result

# ASMPFMHDD

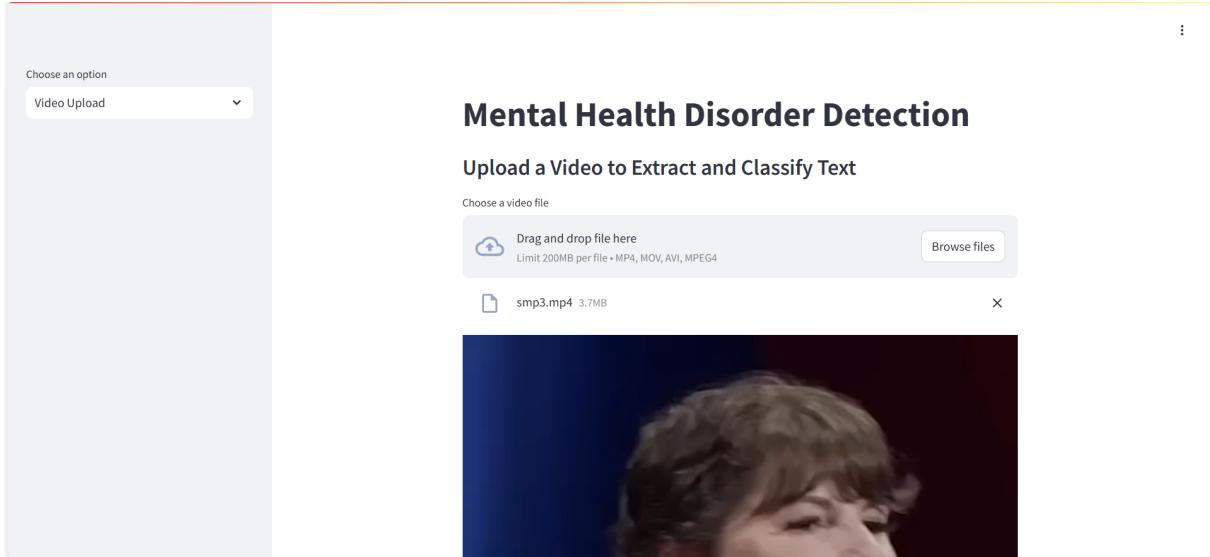


Figure 61: Upload Video

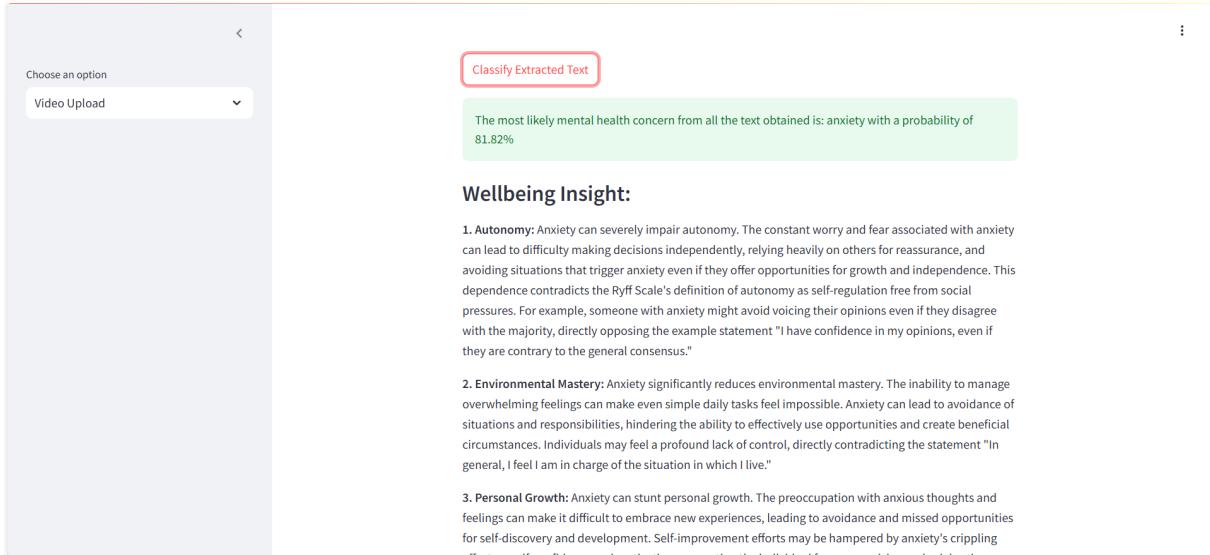


Figure 62: Video Classification Result

# ASMPFMHDD

The screenshot shows the ASMPFMHDD application interface. On the left, a sidebar has a dropdown menu with "Choose an option" and "Reddit Username Analysis" selected. At the top right, there's a status indicator "RUNNING..." with a stop button. The main area is titled "Mental Health Disorder Detection" and "Enter Reddit Username for Analysis". A text input field contains "flowerpower0601" and a red "Analyze" button is below it. Under "Recent Text Posts:", a list of posts from the user "flowerpower0601" is shown:

- 0 : "Taylor Swift's 'Eras' show. What's ACTUALLY going on? What do you guys think of this?"
- 1 : "Taylor Swift's 'Eras' show. What's ACTUALLY going on? [removed]"
- 2 : "Taylor Swift und die Eras Tour

Below the posts, a message states: "Ich habe letztens dieses Video gefunden, welches behauptet das die Eras Tour von Taylor Swift überhaupt nicht live ist und das selbst, dass was sich "live" anhört einfach pre-recorded ist und die Band gar nicht live spielt sondern die ganze Show ein Backing Track ist, der abgespielt wird."

Figure 63: Reddit User Analysis

The screenshot shows the ASMPFMHDD application interface. On the left, a sidebar has a dropdown menu with "Choose an option" and "Reddit Username Analysis" selected. At the top right, there's a status indicator "RUNNING..." with a stop button. The main area displays a green box stating: "The most frequently detected mental health concern from all the text obtained is: normal appearing in 55.00% of analyzed text." Below this, a section titled "Mental health issue distribution across posts:" shows a table:

	Mental Health Issue	Count
0	normal	11
1	depression	4
2	ptsd	3
3	anxiety	2

Below the table, a section titled "Wellbeing Insight:" provides information about mental health issues. It clarifies that "normal" is a baseline descriptor and not a mental health issue. It discusses the six factors of Ryff's Psychological Well-being scale and the typical spectrum of mental health. A numbered list follows:

1. **Autonomy:** Individuals within the "normal" range generally demonstrate a good degree of autonomy. They are capable of making independent decisions and regulating their behavior, even when facing social pressures. They likely agree with statements like "I have confidence in my opinions, even if they are contrary to the general consensus." However, the level of autonomy can vary widely within this

Figure 64: Result from Reddit Posts Analysis

# ASMPFMHDD

Choose an option

Twitter Username Analysis

## Mental Health Disorder Detection

### Enter Twitter Username for Analysis

Enter Twitter username:

Analyze

Recent Text Posts from Tweets:

0 :  
"Well said. It is good that this truth is coming out, and that too in a way common people can see it.

A fake narrative can persist only for a limited period of time. Eventually, the facts will always come out! <https://t.co/8XXo5hQe2y>"

1 :  
"महान बाधासाहेब ठाकरे जी यांच्या पुण्यतिमोळिमित मी खंगा आदराजवी अर्पण करते. महाराष्ट्राचा विकास आणि मराठी लोकाच्या समस्ता यांची अवाही असे ते एक शैष व्यक्तिमत्त होते. भारतीय संस्कृती आणि मूर्खांचे सर्वोन करून लाविष्याचा अस्थिमान वृद्धित करण्यावर याचा दर्श... <https://t.co/oHRV3DUFKj>"

2 :

The most frequently detected mental health concern is: normal, appearing in 100.00% of analyzed text.

Mental health issue distribution across posts:

Mental Health Issue	Count
normal	24

Figure 65: Twitter User Analysis

Choose an option

Twitter Username Analysis

The most frequently detected mental health concern is: normal, appearing in 100.00% of analyzed text.

Mental health issue distribution across posts:

Mental Health Issue	Count
normal	24

### Wellbeing Insight:

It's important to clarify that "normal" in the context of mental health is a relative and complex term. It doesn't represent a specific clinical diagnosis but rather a general state of mental well-being without significant impairment. Therefore, the following analysis explores how a lack of significant mental health challenges might relate to the Ryff scales, acknowledging that variations within the "normal" range are expected.

- 1. Autonomy:** Individuals experiencing "normal" mental well-being typically demonstrate a high degree of autonomy. They are able to make independent choices, regulate their behavior effectively, and resist undue social pressure. They confidently express their opinions, even if they differ from the majority. A lack of autonomy, even within the "normal" range, might manifest as occasional indecisiveness or a tendency to seek excessive external validation.
- 2. Environmental Mastery:** "Normal" mental health often correlates with a strong sense of environmental mastery. Individuals can effectively manage their daily lives, overcome challenges, and create environments conducive to their well-being. Those on the lower end of the "normal" spectrum might experience occasional feelings of being overwhelmed or lacking control in specific areas of their life.

Wellbeing Insight:

It's important to clarify that "normal" in the context of mental health is a relative and complex term. It doesn't represent a specific clinical diagnosis but rather a general state of mental well-being without significant impairment. Therefore, the following analysis explores how a lack of significant mental health challenges might relate to the Ryff scales, acknowledging that variations within the "normal" range are expected.

- 1. Autonomy:** Individuals experiencing "normal" mental well-being typically demonstrate a high degree of autonomy. They are able to make independent choices, regulate their behavior effectively, and resist undue social pressure. They confidently express their opinions, even if they differ from the majority. A lack of autonomy, even within the "normal" range, might manifest as occasional indecisiveness or a tendency to seek excessive external validation.
- 2. Environmental Mastery:** "Normal" mental health often correlates with a strong sense of environmental mastery. Individuals can effectively manage their daily lives, overcome challenges, and create environments conducive to their well-being. Those on the lower end of the "normal" spectrum might experience occasional feelings of being overwhelmed or lacking control in specific areas of their life.

Figure 66: Result from Twitter Posts Analysis

## Model Retraining Functionality

```

# File path for dataset
dataset_path = 'preprocessed_mental_health.csv'

def update_dataset(new_text, mental_health_issue):
    if os.path.exists(dataset_path):
        dataset = pd.read_csv(dataset_path)
    else:
        # If the dataset doesn't exist, create a new one
        dataset = pd.DataFrame(columns=['text', 'mental_health_issue', 'cleaned_text'])

    # Clean the text (adjust cleaning based on your preprocessing)
    cleaned_text = new_text.lower()

    # Create a new DataFrame for the new row
    new_row = pd.DataFrame({
        'text': [new_text],
        'mental_health_issue': [mental_health_issue],
        'cleaned_text': [cleaned_text]
    })

    # Use pd.concat to append the new row
    dataset = pd.concat([dataset, new_row], ignore_index=True)

    # Save the updated dataset
    dataset.to_csv(dataset_path, index=False)

def retrain_model():
    """Retrain the Logistic Regression model with the updated dataset."""
    # Load the updated dataset
    dataset = pd.read_csv(dataset_path)

    # Ensure required columns exist
    if 'cleaned_text' not in dataset.columns or 'mental_health_issue' not in dataset.columns:
        st.error("Dataset must have 'cleaned_text' and 'mental_health_issue' columns.")
    return None, None

    # Preprocessing
    dataset.dropna(subset=['cleaned_text'], inplace=True)
    vectorizer = CountVectorizer()
    X = vectorizer.fit_transform(dataset['cleaned_text'])
    y = dataset['mental_health_issue']

```

## Model Retraining Functionality

```

# Split dataset
X_train, X_test, y_train, y_test = train_test_split(X, y,
    test_size=0.2, random_state=42)

# Train Logistic Regression
model = LogisticRegression(multi_class='ovr', max_iter
    =10000)
model.fit(X_train, y_train)

# Save the updated model and vectorizer
joblib.dump(model, 'LRmodel.pkl')
joblib.dump(vectorizer, 'LRvectorizer.pkl')

# Evaluate
accuracy = accuracy_score(y_test, model.predict(X_test))
return model, accuracy

def update_and_retrain(example_text, example_issue):
    try:
        # Update dataset
        update_dataset(example_text, example_issue)

        # Display the last three rows of the updated dataset
        dataset = pd.read_csv(dataset_path)
        st.write("Updated_Dataset_(Last_3_Rows):")
        st.write(dataset.tail(3))

        # Notify user about retraining process
        st.info("Model_is_being_retrained...")

        # Retrain model
        model, accuracy = retrain_model()

        # Display retraining success message
        if model:
            st.success(f"Model_retrained_successfully!_Accuracy:
                {_accuracy_*_100:.2f}%")
        except Exception as e:
            st.error(f"An_error_occurred:_{e}")

# Function to classify text, display result and retrain model
def classify_text_retrain_model(text):
    input_vectorized = vectorizer.transform([text])
    prediction_proba = model.predict_proba(input_vectorized)

```

## Model Retraining Functionality

```

issue_labels = model.classes_
proba_df = pd.DataFrame(prediction_proba, columns=
    issue_labels).T
proba_df.columns = ['Probability']

top_issue = proba_df['Probability'].idxmax()
top_probability = proba_df['Probability'].max()

st.write(f"The most likely mental health concern is: {top_issue} with a probability of {top_probability:.2%}")

get_wellbeing_insight(text, top_issue)

# Adding Model Retraining Functionality
update_and_retrain(text, top_issue)

```

This code provides a system for updating a dataset with new text data, retraining a Logistic Regression model based on the updated data, and classifying new input text while also retraining the model. The `update_dataset` function appends new text and its associated mental health issue to an existing CSV file. If the dataset doesn't already exist, it creates a new one with columns for the text, the mental health issue, and the cleaned version of the text. The text is then cleaned (in this case, by converting it to lowercase). After updating the dataset, the function saves it back to the CSV file. The `retrain_model` function reads the dataset, ensures it contains the necessary columns, cleans the data by removing any rows with missing values, and preprocesses the text using a `CountVectorizer`. It splits the dataset into training and testing sets, then trains a Logistic Regression model. The model and vectorizer are saved to disk using `joblib`, and the function returns the trained model and its accuracy on the test set. The `update_and_retrain` function combines these operations, first calling `update_dataset` to add new data, then displaying the last three rows of the updated dataset using Streamlit's `st.write`, and finally retraining the model. If successful, it shows the retrained model's accuracy. The `classify_text_retrain_model` function takes an input text, vectorizes it using the pre-trained vectorizer, and uses the trained model to predict the probability of different mental health issues. It converts these probabilities into a DataFrame and identifies the mental health issue with the highest probability. The function displays this prediction along with its associated probability and calls another function, `get_wellbeing_insight`, to provide additional insights. Finally, it calls `update_and_retrain` to update the dataset and retrain the model with the new prediction.

## Retraining the Ensemble Model

```

def retrain_model():
    # Load Logistic Regression model and vectorizer
    with open('LRmodel.pkl', 'rb') as file:
        lr_model = pickle.load(file)
    with open('LRvectorizer.pkl', 'rb') as file:
        lr_vectorizer = pickle.load(file)

    # Load SVM model and vectorizer
    with open('SVMmodel.pkl', 'rb') as file:
        svm_model = pickle.load(file)
    with open('SVMvectorizer.pkl', 'rb') as file:
        svm_vectorizer = pickle.load(file)

    # Load XGBoost model, vectorizer, and label encoder
    with open('xgb_model.pkl', 'rb') as file:
        xgb_model = pickle.load(file)
    with open('tfidf_vectorizer.pkl', 'rb') as file:
        tfidf_vectorizer = pickle.load(file)
    with open('label_encoder.pkl', 'rb') as file:
        label_encoder = pickle.load(file)

    # Load LSTM model, tokenizer, and label encoder
    lstm_model = load_model('lstm_model.h5')
    with open('LSTM_tokenizer.pkl', 'rb') as file:
        lstm_tokenizer = pickle.load(file)

    # Load Naive Bayes model and vectorizer
    with open('NBmodel.pkl', 'rb') as file:
        nb_model = pickle.load(file)
    with open('NBvectorizer.pkl', 'rb') as file:
        nb_vectorizer = pickle.load(file)

    # Load the test dataset
    data = pd.read_csv('preprocessed_mental_health.csv')

    # Check if 'cleaned_text' column exists
    if 'cleaned_text' not in data.columns:
        raise ValueError("The dataset must have a 'cleaned_text' column.")

```

## Retraining the Ensemble Model

```

# Remove rows with missing values in 'cleaned_text'
data.dropna(subset=['cleaned_text'], inplace=True)

# Split features and target
X_test = data['cleaned_text']
y_test = data['mental_health_issue']

# Encode target labels
y_test = label_encoder.transform(y_test)

# Process the text for each model
X_test_lr = lr_vectorizer.transform(X_test)    # Logistic
                                                Regression vectorizer
X_test_svm = svm_vectorizer.transform(X_test)   # SVM
                                                vectorizer
X_test_xgb = tfidf_vectorizer.transform(X_test)  # XGBoost
                                                vectorizer
X_test_nb = nb_vectorizer.transform(X_test)      # Naive Bayes
                                                vectorizer
X_test_lstm = lstm_tokenizer.texts_to_sequences(X_test)  # LSTM
                                                tokenizer

# Pad sequences for LSTM
X_test_lstm = pad_sequences(X_test_lstm, maxlen=100, padding
                            ='post', truncating='post')

# Get predictions from the base models
lr_predictions_proba = lr_model.predict_proba(X_test_lr)    #
                                                Logistic Regression probabilities
svm_predictions_proba = svm_model.predict_proba(X_test_svm)
                        # SVM probabilities
xgb_predictions_proba = xgb_model.predict_proba(X_test_xgb)
                        # XGBoost probabilities
nb_predictions_proba = nb_model.predict_proba(X_test_nb)    #
                                                Naive Bayes probabilities
lstm_predictions_proba = lstm_model.predict(X_test_lstm)    #
                                                LSTM probabilities

# Stack the predictions of all models to create the feature
matrix for the meta-learner
stacked_features = np.hstack((
    lr_predictions_proba,
    svm_predictions_proba,
    xgb_predictions_proba,
    nb_predictions_proba,
    lstm_predictions_proba
))

```

### Retraining the Ensemble Model

```

# Define the Logistic Regression model for the meta-learner
meta_learner_lr = LogisticRegression(max_iter=5000)

# Train the Logistic Regression meta-learner on the full
dataset
meta_learner_lr.fit(stacked_features, y_test)

# Save the trained Logistic Regression meta-learner
with open('meta_learner_lr.pkl', 'wb') as file:
    pickle.dump(meta_learner_lr, file)

# Predict using the Logistic Regression ensemble model
final_predictions_lr = meta_learner_lr.predict(
    stacked_features)

# Evaluate the Logistic Regression ensemble model
accuracy_lr = accuracy_score(y_test, final_predictions_lr)

return meta_learner_lr, accuracy_lr

```

The above function, `retrain_model`, is designed to retrain an ensemble learning model using an updated dataset. Initially, it loads multiple base models (Logistic Regression, SVM, XGBoost, Naive Bayes, and LSTM) along with their respective vectorizers, tokenizers, and label encoder from previously saved pickle files. The function then loads and preprocesses the updated dataset, ensuring it contains the necessary text column (`cleaned_text`) and splitting it into features (`X_test`) and target labels (`y_test`). The target labels are encoded using the label encoder, while the text features are transformed using the corresponding vectorizers or tokenizers for each model. Predictions from each model are then collected, stacked together, and used to train the meta-learner, a `LogisticRegression` classifier. Once trained, the meta-learner is saved for future use, and its performance is evaluated using accuracy. The function returns the trained XGBoost meta-learner and its accuracy score, allowing for the dynamic adaptation of the ensemble model to any updates in the dataset.

## ASMPFMHDD

Choose an option

Text Input

### Mental Health Disorder Detection

Enter Text to Classify Mental Health Issue

Enter your text here:

তোমার মন কখনো চরম উচ্চতায় পৌছায়, এক ধরনের অদম্য শক্তি অনুভব হয়...এটি মনে হয় জীবনের সবচেয়ে চমকপ্রদ মুহূর্ত। তোমার ভাবনায়, এমন এক তরঙ্গ, যা সারা পৃথিবীকে জয় করার মতো। কিন্তু ঠিক প্রের মুহূর্তে, সবচিক বদলে যায়, মনে হয় পৃথিবী ঢেতে পড়ছে। এক মুহূর্তে ভূমি একেবারে শক্তিশালী। আবার প্রক্ষণণেই, পৃথিবী দোষকে অসহায় মনে হতে থাকে। একদিনকে এক

Classify Text

Classify Text and Retrain Model

Translated Text (to English):

Sometimes your mind reaches extreme heights, you feel a kind of invincible power - it seems like the most amazing moment in life. There is a wave in your thoughts that is like conquering the whole world. But the next moment, everything changes, it seems like the world is falling apart. One moment you are absolutely strong, and the next moment, the world seems helpless to you. On one side there is a frenzy of feelings, on the other side there is darkness. Sometimes it feels like you are floating on clouds, and the next moment your feet start falling to the ground. Sometimes you feel like you are living in a different world, where

Figure 67: Text Classification and Retrain Model

Choose an option

Text Input

- Education:** Learning about bipolar disorder and its management strategies can empower individuals and their loved ones.
- Self-compassion:** Practicing self-compassion and self-care is essential for navigating the challenges of the illness.

It's important to note that these are general observations, and individual experiences with bipolar disorder can vary greatly. Professional guidance from a psychiatrist and therapist is crucial for personalized treatment and support.

Updated Dataset (Last 3 Rows):

	text	mental_health_issue
18,595	Does anybody else get really anxious from hearing a phone ring or even vibrate? I've	anxiety
18,596	I was born trying to kill myself Came out looking like fucking megamind. Some bable	depression
18,597	Sometimes your mind reaches extreme heights, you feel a kind of invincible power - i	bipolar

Model is being retrained...

Model retrained successfully! Accuracy: 97.05%

Figure 68: Result from Text Classification and Model Retraining

# ASMPFMHDD

Choose an option

Reddit Username Analysis

## Mental Health Disorder Detection

### Enter Reddit Username for Analysis

Enter Reddit username:

Analyze

Analyze and Retrain Model

Recent Text Posts:

① : "Unpredictability is the suckiest part of life. Some people never drink or smoke, but they still get cancer and die young. Some people drink and smoke still gets to live a long fulfilling life.

Some people study so hard to get their dream job, have a promising future, but on his first day to work, he got killed by a drunk driver.

Some people work so hard and accumulate a lot of fortune for enjoying retirement

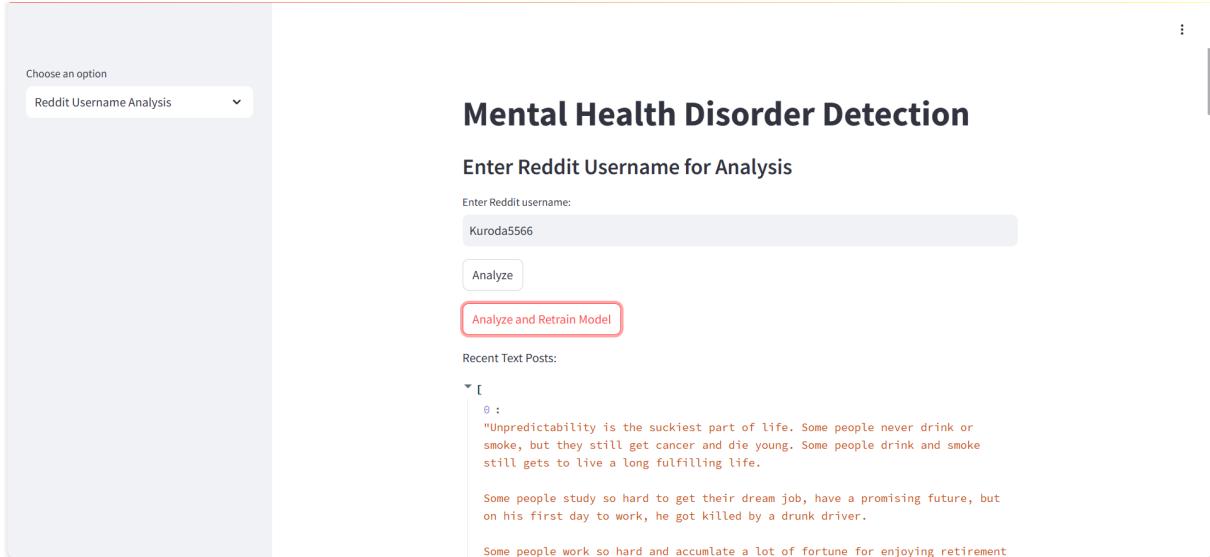


Figure 69: Reddit Analysis with Model Retraining

Choose an option

Reddit Username Analysis

Practical Advice to Enhance Well-being (even within a "normal" range):

To improve scores across these areas, individuals could consider engaging in activities that promote mindfulness, stress management techniques (like yoga or meditation), pursuing hobbies and interests, fostering strong social connections, setting achievable goals, and practicing self-compassion. Regular exercise, a balanced diet, and sufficient sleep are also crucial for overall well-being. If significant challenges persist in any of these areas despite self-help efforts, seeking professional guidance from a therapist or counselor can be beneficial.

Updated Dataset (Last 3 Rows):

	text	mental_health_issue
18,596	I was born trying to kill myself Came out looking like fucking megamind. Some babies	depression
18,597	Sometimes your mind reaches extreme heights, you feel a kind of invincible power - i	bipolar
18,598	I can accept rdr2 made with rdr1 graphics. It might sound weird, but I really love the	normal

Model is being retrained...

Model retrained successfully! Accuracy: 97.05%

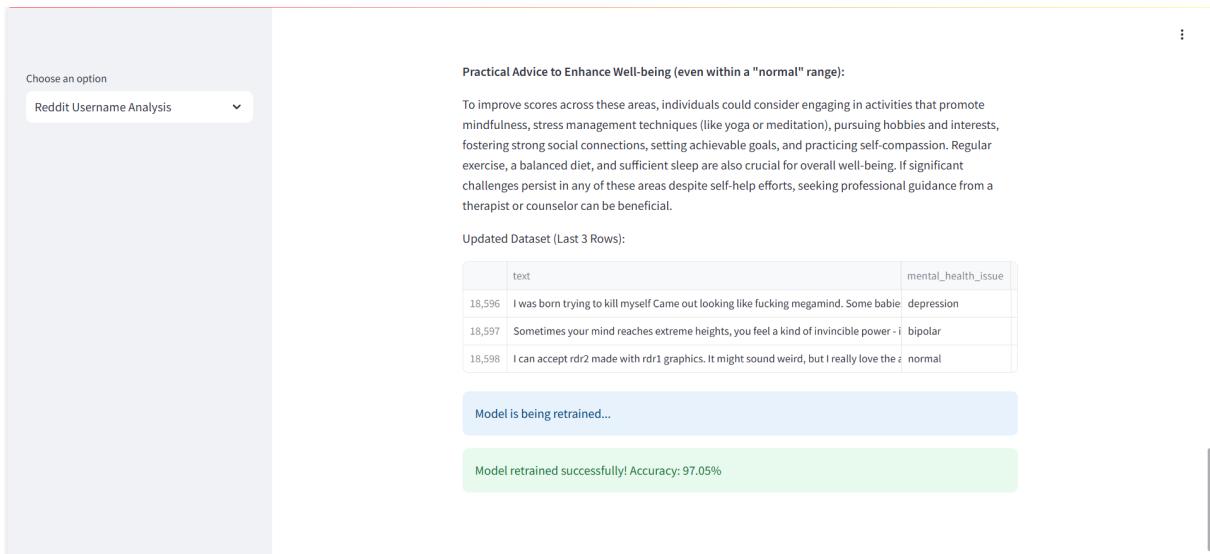


Figure 70: Result from Reddit Analysis and Model Retraining

## Emotion Analysis and Mental Health Insights

```

def detect_emotions_from_frame(frame):
    try:
        # Use DeepFace to analyze emotions
        result = DeepFace.analyze(frame, actions=['emotion'],
                                  enforce_detection=False)
        return result[0]['dominant_emotion']
    except Exception as e:
        print(f"No_expression_or_error_detecting_emotion:{e}")
        return None

def analyze_emotions_from_frames(frames):
    # Initialize an emotion counts dictionary
    emotion_counts = {'happy': 0, 'sad': 0, 'angry': 0, 'disgust':
                      0, 'fear': 0, 'surprise': 0, 'neutral': 0}
    frame_emotions = []

    for idx, frame in enumerate(frames):
        # Detect emotions from the current frame
        emotion = detect_emotions_from_frame(frame)
        if emotion:
            frame_emotions.append(emotion)
            if emotion in emotion_counts:
                emotion_counts[emotion] += 1

    return emotion_counts, frame_emotions

def display_emotion_summary(emotion_counts):
    # Convert the emotion counts to a DataFrame for display
    emotion_df = pd.DataFrame(list(emotion_counts.items()),
                              columns=['Emotion', 'Count'])
    st.write("Emotion_Analysis_Summary:")
    st.table(emotion_df)
    return max(emotion_counts, key=emotion_counts.get)

def analyze_with_gemini(dominant_emotion, emotion_counts):
    try:
        # Start a chat session with the Gemini API
        chat_session = gemini_model.start_chat(history=[])

        # Summarize emotion counts for the prompt
        emotion_summary = ", ".join([f"{emotion}:{count}" for
                                    emotion, count in emotion_counts.items()])

```

## Emotion Analysis and Mental Health Insights

```

# Create a prompt for the Gemini API
prompt = (
    f"The_detected_dominant_emotion_is_{dominant_emotion}'."
    f"{emotion_summary}.Based_on_this_information,_"
    analyze_the_potential_implications_for_mental_"
    health_"
    f"conditions_such_as_depression,_anxiety,_PTSD,_or_"
    bipolar_disorder._Provide_actionable_insights_in_"
    three_lines."
)

# Send the prompt and get a response
response = chat_session.send_message(prompt)
st.write(response.text)
except Exception as e:
    print(f"Error_in_Gemini_API_call:{e}")
    return "An_error_occurred_while.communicating_with_the_"
        Gemini_API._Please_try_again_later."

```

The functions in the code work together to analyze emotions from video frames. The `detect_emotions_from_frame(frame)` function utilizes the DeepFace library to detect the dominant emotion in a given video frame, returning `None` if no emotion is detected or an error occurs. The `analyze_emotions_from_frames(frames)` function iterates over multiple video frames, applying `detect_emotions_from_frame` to each and counting the frequency of detected emotions in a dictionary. The `display_emotion_summary(emotion_counts)` function uses Streamlit to display a summary of detected emotions in a table format and identifies the dominant emotion with the highest count. Lastly, `analyze_with_gemini(dominant_emotion, emotion_counts)` communicates with the Gemini API by preparing a prompt based on the emotion analysis, sending it to the API, and displaying the response in Streamlit, while also handling any potential errors during the API interaction.

The DeepFace Python module is a comprehensive framework for facial recognition and facial attribute analysis using deep learning. It provides a high-level interface to process and analyze facial data through a variety of pre-trained deep learning models. The module is designed to make facial recognition, emotion detection, age estimation, and gender detection straightforward by abstracting the complexities of deep learning. DeepFace supports multiple state-of-the-art face recognition models, including VGG-Face, Google FaceNet, OpenFace, Facebook DeepFace, and Dlib. Users can choose any of these models depending on their specific requirements. It also integrates with various facial attribute models for emotion analysis, enabling it

to classify emotions such as happiness, sadness, anger, and more. The module uses TensorFlow and Keras as its backend frameworks to run deep learning operations efficiently. At its core, DeepFace operates by detecting and aligning faces in an image using a pre-trained facial detector, ensuring consistent positioning before feeding the face into recognition or analysis pipelines. This preprocessing step is critical because it eliminates variability caused by head tilt, lighting, and facial orientation. For emotion detection specifically, DeepFace uses a softmax classifier at the final layer of its models to predict probabilities for different emotions. DeepFace is also known for its flexibility and simplicity. Users can invoke facial analysis or recognition with a single line of code. Additionally, the module includes an **analyze** function for emotion, age, and gender detection, as well as a **verify** function to compare two images for identity matching. It also allows users to toggle between facial detectors, including OpenCV, SSD, and MTCNN, depending on their hardware and processing needs. The module emphasizes usability with optional parameters for tuning accuracy, speed, and error handling. For example, users can enable or disable facial detection enforcement, which is useful in scenarios where non-human faces or partial occlusions are present. The ability to batch process multiple images and integrate seamlessly with large datasets makes DeepFace a preferred choice for production-level applications. Overall, DeepFace abstracts the complexities of implementing deep learning models for facial analysis, providing a user-friendly and powerful tool for a variety of facial recognition and emotion detection tasks.

In a mental health disorder detection application, DeepFace plays a pivotal role in extracting facial expressions from images and videos, which can be critical for assessing emotional states and providing insights into potential mental health concerns. The module leverages deep learning models trained to recognize a wide array of facial expressions, such as happiness, sadness, anger, surprise, fear, and disgust. These expressions are vital indicators of a person's emotional state, which, when analyzed over time, can offer valuable clues about their mental health. DeepFace first processes the input image or video by detecting faces through a facial recognition model, ensuring that only human faces are analyzed. Once faces are detected, the module aligns and normalizes the images to account for variations in lighting, angles, and head positions, ensuring that the analysis remains accurate regardless of how the subject is positioned or oriented. DeepFace's emotion detection model then works by analyzing the subtle changes in facial muscle movements that correspond to different emotional expressions. For instance, a furrowed brow and tight lips may indicate anger, while a smiling face may suggest happiness. These expressions are captured and classified using convolutional neural networks (CNNs), which are adept at identifying patterns in visual data. DeepFace outputs these emotions as probability scores, allowing the application to interpret the likelihood of each emotion being present in a given frame. In a video context, the module processes each frame sequentially, enabling the tracking of emotional transitions over time, which is particularly valuable in detecting

mood fluctuations or episodes commonly associated with mental health conditions like depression, anxiety, or bipolar disorder. The ability to extract and analyze these facial expressions in real-time or from pre-recorded content empowers mental health disorder detection systems to better understand the emotional well-being of individuals. By assessing facial expressions in conjunction with other data points, such as speech or text analysis, DeepFace helps build a comprehensive profile of a person's emotional state. This can aid in detecting mental health issues early, identifying stress triggers, or tracking the effectiveness of therapeutic interventions. In this way, DeepFace's emotion detection capabilities become an essential tool in a mental health disorder detection application, providing critical insights that can inform diagnosis and personalized care plans.

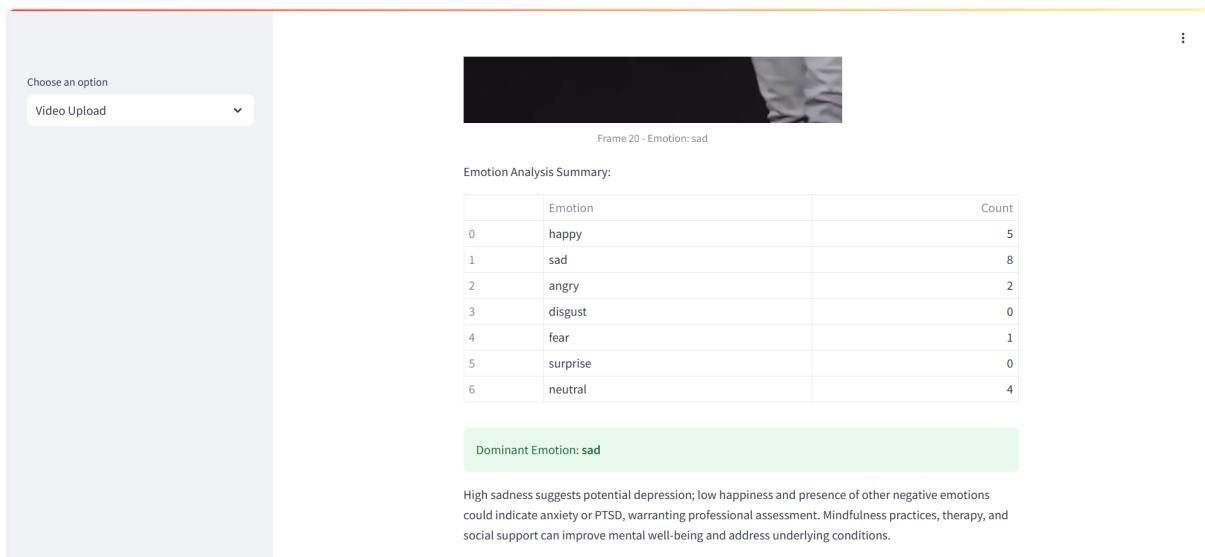


Figure 71: Result from the emotion analysis of facial expression

## Image Captioning

```

# Function to load the model (cached for efficiency)
@st.cache_resource
def load_model():
    model_name = "nlpconnect/vit-gpt2-image-captioning"
    model = VisionEncoderDecoderModel.from_pretrained(model_name)
    )
    feature_extractor = ViTImageProcessor.from_pretrained(
        model_name)
    tokenizer = AutoTokenizer.from_pretrained(model_name)
    return model, feature_extractor, tokenizer

# Load the model
IDmodel, IDfeature_extractor, IDtokenizer = load_model()

# Set device to GPU if available
device = torch.device("cuda" if torch.cuda.is_available() else "
    cpu")
IDmodel.to(device)

# Function to generate caption
def generate_caption(image):
    # Preprocess the image
    if image.mode != "RGB":
        image = image.convert(mode="RGB")
    pixel_values = IDfeature_extractor(images=image,
        return_tensors="pt").pixel_values
    pixel_values = pixel_values.to(device)

    # Generate caption (you can adjust max_length and num_beams as needed)
    with torch.no_grad():
        output_ids = IDmodel.generate(pixel_values, max_length
            =16, num_beams=4)
    caption = IDtokenizer.decode(output_ids[0],
        skip_special_tokens=True)
    return caption

```

The provided code implements two key functionalities: image captioning and mental health text classification. For image captioning, it loads a pre-trained VisionEncoderDecoder model (vit-gpt2-image-captioning) using Hugging Face tools, processes input images, and generates captions. The **vit-gpt2-image-captioning** model combines two advanced architectures, the Vision Transformer (ViT) and GPT-2, to generate descriptive captions for images. It functions as an encoder-decoder system where ViT acts as the encoder to process and extract meaningful features from the image, and GPT-2 acts as the decoder to generate coherent natural language

captions based on those features. The Vision Transformer (ViT) is a transformer-based model for image recognition. Unlike traditional convolutional neural networks (CNNs), which process images in a grid-like fashion, ViT splits the image into smaller patches (e.g., 16x16 pixels). These patches are flattened into vectors and passed through a transformer encoder. This encoder treats each patch as a token, similar to how words are tokens in natural language processing. By leveraging self-attention mechanisms, ViT captures both local and global dependencies in the image, resulting in a comprehensive feature representation. These features are passed to the decoder for further processing. GPT-2, a pre-trained generative language model, serves as the decoder. It uses the features extracted by ViT to generate captions word by word. Transformers like GPT-2 rely heavily on self-attention and positional encodings to understand the relationships between words in a sequence. In this context, GPT-2 uses the visual feature embeddings as the initial context and generates text sequentially. At each step, it predicts the next word in the caption based on the previous words and the image features, ensuring that the captions are contextually relevant and grammatically accurate. Transformers are crucial here because they allow the model to handle complex dependencies and interactions in both image and text data. ViT uses transformers to learn rich visual representations by modeling relationships between patches in the image, while GPT-2 leverages transformers to generate fluent and coherent text based on the encoded image features.



Figure 72: Generate Image Caption