**Multimodal AI Framework for Social Media Based Mental Disorder Detection and Personalized Wellbeing Insights**

Submitted by

| **SOUMYADEEP NANDY** | **(13000121033)** |
| --- | --- |
| **PRITHWISH SARKAR** | **(13000121037)** |
| **SAGNIK MUKHOPADHYAY** | **(13000121040)** |
| **ARKAPRATIM GHOSH** | **(13000121058)** |

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, SaltLake, Sector V, Kolkata– 700 091**  **CERTIFICATE**  This is to certify that the project entitled “Multimodal AI Framework for Social Media Based Mental Disorder Detection and Personalized Wellbeing Insights” prepared by SOUMYADEEP NANDY (13000121033), PRITHWISH SARKAR (13000121037), SAGNIK MUKHOPADHYAY (13000121040) and ARKAPRATIM GHOSH (13000121058) of B.Tech (Computer Science & Engineering), Final Year, has been done according to the regulations of the Degree of Bachelor of Technology in Computer Science & Engineering. The candidates have fulfilled the requirements for the submission of the project report.  It is to be understood that, the undersigned does not necessarily endorse 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**

**. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .**

**ACKNOWLEDGEMENTS**

We Would like to express our sincere gratitude to our project guide in the Computer Science and Engineering department. We are extremely thankful for the keen interest he / she took in advising us, for the books and reference materials provided for the moral 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 staff for the gracious hospitality they offered us.

**SOUMYADEEP NANDY (13000121033)**

**PRITHWISH SARKAR (13000121037)**

**SAGNIK MUKHOPADHYAY (13000121040)**

**ARKAPRATIM GHOSH (13000121058)**

Techno Main Salt Lake

Date : 11th June, 2025

**Contents**

[**Abstract**](#_4qj5edj044l6) **5**

[**1. Introduction**](#_6mlsomhnw4ms) **5**

[1.1 Project Overview](#_wih918htxgi3) 5

[1.2 Project Purpose](#_srvudj2hxiq7) 5

[1.3 Technical Domain Specifications](#_29lb4eb3jfrc) 6

[1.4 Business Domain Specifications](#_hqqgaruclr0u) 6

[1.5 Glossary/Keywords](#_9edlj5dcltj3) 7

[**2. Related Studies**](#_mowvav2zns2w) **9**

[**3. Problem Definition and Preliminaries**](#_fuf6dwepozpd) **11**

[3.1 Context and Background](#_gre1yjvzl724) 11

[3.2 Objective and Challenges](#_5hxix498wd83) 11

[3.3 Scope, Exclusions and Assumptions](#_bjpmb243b4lg) 11

[**4. Proposed Solution**](#_60dh97x02gns) **12**

[**5. Project Planning**](#_u7q5efmlhjnu) **13**

[5.1 Software Life Cycle Model](#_h5ax7amt505r) 13

[5.2 Dependencies, Milestones and Scheduling](#_l74k6y38ywt5) 14

[**6. Requirement Analysis**](#_amh4ckhq8qpy) **16**

[6.1 Requirement Matrix](#_gdaxj73zxh3n) 16

[6.2 Requirement Elaboration](#_15xwurzgu5b8) 17

[**7. Design**](#_6n96m4skocxe) **18**

[7.1 Technical Environment](#_bbn554fiycg0) 18

[7.2 Hierarchy of Modules](#_yakb9pxef82m) 20

[7.3 Detailed Design](#_gfs2sdy6xlun) 21

[**8. Implementation**](#_te337oircnsj) **31**

[8.1 Data Collection and Dataset Preparation](#_kthndp30tv20) 31

[8.2 Data Cleaning and Feature Extraction](#_2sqt8ubyfri) 33

[8.3 Machine Learning Models](#_lucjt2kqnf4s) 35

[8.4 Deep Learning Models](#_dvznlgrmmn8q) 36

[8.5 Ensemble Model](#_pup8jckl00kq) 41

[8.6 Wellbeing Survey and Association Matrix](#_40lg2cqw2qp4) 42

[8.7 RAG for Wellbeing Insights](#_qw2nos7ub5ng) 46

[**9. Test Plans, Results and Analysis**](#_b8kpm2vajoht) **50**

[9.1 Results from Base Models](#_s1ypme6zm2o4) 53

[9.2 Comparison of different tokenizations](#_nn78qyp37oc0) 63

[9.3 Results from Ensemble Model Training and Testing](#_egya36r4gny6) 64

[9.4 Results from Hierarchical Ensemble Models](#_blkrtidhcpxh) 70

[9.5 User Interface of the Application](#_btnu66fu2ea5) 73

[**10. Conclusion**](#_q8pb7t9oqu0d) **77**

[**11. References**](#_h177baaou774) **78**

# **Abstract**

This project presents a scalable, multimodal AI framework for the early detection of mental health issues that combines an ensemble of diverse classifiers with a hierarchical model to achieve up to 98.03% accuracy on standard benchmarks and 96.25% on larger datasets. Packaged as a web application, the system offers continuous model retraining and dynamic knowledge-base updates, interactive well-being assessments, and retrieval-augmented generation to deliver personalized, evidence-based insights, complemented by rich visual analytics and actionable recommendations to enable timely interventions. Extensive evaluations demonstrate the platform’s effectiveness and extensibility for proactive mental health monitoring.

# **1. Introduction**

## **1.1 Project Overview**

Mental health disorders—including depression, anxiety, bipolar disorder, and PTSD—affect millions worldwide and often go undetected until they manifest in crises. Meanwhile, people increasingly share their thoughts, feelings, and experiences on social media platforms (Reddit, Twitter etc) and in digital documents, leaving behind rich clues about their emotional state. In this work, we develop a multimodal AI framework that ingests text, images, video, and document feeds, uses OCR and deep-learning emotion analysis, and aligns user responses to established well-being scales. By fusing these signals through an ensemble of machine-learning and neural models, our system aims to flag early warning signs of distress and guide users toward timely, personalized support.

## **1.2 Project Purpose**

Early identification and intervention are critical for mitigating the severity of mental health crises, yet current screening methods often rely on self-report or occasional clinical encoun ters. This project aims to fill that gap by delivering a continuously learning, data-driven monitoring tool that passively and proactively analyzes digital footprints—from social posts to uploaded documents and survey responses—to surface warning signs long before a crisis point. By putting actionable insights directly into the hands of individuals, caregivers, and healthcare providers, it seeks to enable truly preventative mental-health care at scale.

## **1.3 Technical Domain Specifications**

| **Domain** | **Specification** |
| --- | --- |
| **Hardware** | Standard Machine With≥8GB RAM and multi-coreCPU. (Optional:GPUforlargerdatasetsorcomplexmodeltraining.) |
| **Operating System** | Cross-platform support:macOS,Windows10/11,Linux distributions(e.g.Ubuntu,LinuxMint). |
| **Programming Languages** | Python 3.x(primary language for,data analysis,NLP). |
| **Libraries/ Frameworks** | •Data Processing: Pandas, NumPy  •Machine Learning: Scikit-learn, XGBoost, TensorFlow, Transformers  • Image/text analysis: OpenCV, Tesseract, Pytesseract, DeepFace •Audioprocessing: Librosa, PyDub, SpeechRecognition  • Social Media Integration: PRAW, Tweepy  •Visualization: Plotly, Matplotlib  •Additionaltools: Streamlit, NLTK, GoogleGenerative AI |
| **Development Environment** | Google Colab (cloud execution with optionalGP for large datasets or model training). |

## **1.4 Business Domain Specifications**

| **Stakeholder** | **Role / Use case** |
| --- | --- |
| **Mental Health Services** | Mental Health Providers, including hospitals and therapy centers,canleveragemachinelearningtodetectearlysigns ofmentaldisordersfromsocial mediadata.This Proactive approach complements traditional self-reporting and clinical assessments,enabling earlier intervention and support for patients. |
| **SocialMedia Platforms** | Social media platforms like Twitter and Reddit are key spaces for expressing thoughts and emotions, including mental health struggles. This project’s machine learning models can help these platforms safeguard user well-being by identifying concerns early, while maintaining ethical standards. |
| **PublicHealth Organizations** | Public health organizations can use real-time social media data to monitor mental well-being, identify trends, and design data-driven interventions. By analyzing language patterns, they can create targeted awareness campaigns that better engage individuals facing mental health challenges. |

## **1.5 Glossary/Keywords**

| **Term** | **Definition** |
| --- | --- |
| 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. |
| Retrieval-Augmented Generation (RAG) | A hybrid NLP framework that combines information retrieval and text generation by fetching relevant context from a knowledge base before generating responses, improving factual accuracy and relevance. |
| 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 for accessing Reddit data, such as posts, comments, and user information. |
| Tesseract OCR | Tesseract OCR 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 digitising printed text. |
| Depression | There is a difference between depression and mood swings or short-lived emotional reactions to daily experiences; it is a mental state causing painful symptoms that adversely disrupt normal activities (e.g., sleeping). |

| **Term** | **Definition** |
| --- | --- |
| Anxiety | Several behavioral disturbances are associated with anxiety disorders, including excessive fear and worry. Severe symptoms cause significant impairment in functioning and considerable distress. Anxiety disorders come in many forms, such as social anxiety, generalized anxiety, panic, etc. |
| BipolarDisorder | 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. |
| 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. |
| 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. |
| TransformersModule | 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 2.0 Flash | Gemini 2.0 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. |
| HyperparameterTuning | Hyperparameter tuning involves selecting the best parameters for a machine learning model to optimize its performance on a given task using grid search or random search. |
| Embedding  Model | A neural network that transforms individual text inputs into fixed-length vector representations in a continuous semantic space, enabling efficient similarity search and downstream tasks like clustering or retrieval. |
| Cross-Encoder | A model that jointly processes a pair of inputs (e.g., query and document) through a shared encoder and directly produces a relevance score or classification, allowing for richer interaction at the cost of higher compute per pair. |

# **2. Related Studies**

|  |
| --- |

|  |
| --- |

# **3. Problem Definition and Preliminaries**

## **3.1 Context and Background**

Mental health disorders affect approximately 1 in 8 people globally. Social media platforms like Reddit and Twitter offer vast amounts of real-time data reflecting mental health struggles, but the unstructured nature of this data poses challenges for effective identification and categorization of specific disorders.

## **3.2 Objective and Challenges**

The primary objective is to develop a system that uses NLP and machine learning to analyze Reddit and Twitter posts for detecting mental health disorders like depression, anxiety, bipolar disorder, and PTSD. It seeks to classify posts accurately and provide data-driven insights into mental health trends for researchers, professionals, and policymakers.

• **Data Variability:** Social media posts vary in structure, style, and language.

• **Imbalanced Data**: Uneven distribution of mental issues can impact model training.

• **Cultural Nuances**: Mental health discussions differ across cultures.

• **Privacy and Ethics:** Analyzing social media data raises concerns about user privacy.

## **3.3 Scope, Exclusions and Assumptions**

The scope of the project is to develop a multimodal AI framework to detect mental disorders from social media inputs using ML and NLP. It analyzes text, images, videos, PDFs, dynamic responses to images shown and Reddit/Twitter data via APIs, classifying inputs into Normal, Anxiety, Depression, Bipolar, or PTSD using a Reddit dataset. Finally, an association is created between mental health disorder and mental wellbeing parameters from Ryff’s Psychological Well-being Scale. This project excludes real-time sentiment analysis, platform-specific features like hashtags or subreddits, and ethical implications of data ownership. It also does not analyze comments, metadata, or Reddit/Twitter-specific elements, focusing solely on detecting mental health disorders from user profiles and mapping them to wellbeing insights, though the dataset has been made solely from Reddit posts from various subreddits. The project assumes that the Reddit dataset obtained via PRAW represents diverse mental health discussions and that user posts accurately reflect emotions. NLP techniques are presumed effective for sentiment classification, and selected ML models (Logistic Regression, SVM, Naïve Bayes, LSTM, Transformer, XGBoost) are expected to perform optimally. Social media sentiments are considered valid proxies for public mental health perceptions. Data preprocessing is assumed sufficient to reduce noise, and ethical standards are maintained to protect user privacy. The users’ responses to the well-being surveys are considered valid for updating the association matrix between mental health disorders and well-being parameters.

# **4. Proposed Solution**

|  |
| --- |

# **5. Project Planning**

## **5.1 Software Life Cycle Model**

| **Fig. 5.1 Iterative Waterfall Model** |
| --- |

|  |
| --- |

## **5.2 Dependencies, Milestones and Scheduling**

Key dependencies were identified for successful project progression. For instance, com pletion 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. Effective scheduling is vital for project success. A detailed timeline with tasks like requirement gathering, data prepro cessing, model implementation, and testing was created, with flexibility for adjustments based on feedback. Key milestones, including data analysis, model validation, and user acceptance testing, ensure progress tracking. Using tools like Microsoft Project, we monitor tasks, manage resources, and maintain communication to deliver a high-quality solution on time.

| **Fig. 5.2 Project Plan** |
| --- |

| **Fig. 5.3 Gantt Chart** |
| --- |

# **6. Requirement Analysis**

## **6.1 Requirement Matrix**

| **Fig. 6.1 Requirement Matrix** |
| --- |

## **6.2 Requirement Elaboration**

|  |
| --- |

# **7. Design**

## **7.1 Technical Environment**

The technical environment for the project “Multimodal AI Framework for Social Media-Based Mental Disorder Detection and Personalized Well-being Insights” comprises a combination of hardware, software, and tools that enable smooth data analysis, machine learning model training, and deployment. Below is an overview of the minimum hardware configuration, software tools, and package details necessary to carry out this project.

|  |
| --- |

|  |
| --- |

|  |
| --- |

## **7.2 Hierarchy of Modules**

| **Fig. 7.1 Project Modules** |
| --- |

|  |
| --- |

## **7.3 Detailed Design**

|  |
| --- |

**Fig. 7.2 System Overview**

| **Fig. 7.3 DFD Level 0** |
| --- |

| **Fig. 7.4 DFD Level 1 of the System** |
| --- |

**Image Description**

| **Fig. 7.5 DFD Level 2 of Image Description** |
| --- |

The image captioning process using Python’s Transformers module and the ViT-GPT2 model involves several key steps. The user uploads an image, which is 224×224 pixels, normalizing

pixel values, and converting it into a tensor format. The Vision Transformer (ViT) divides the image into 16×16 patches, embeds them, and processes them through transformer layers to extract visual features. The resulting embedding is passed to GPT-2, which generates a sequence of text tokens iteratively using its language model. Tokens are decoded back into human-readable text using the GPT-2 tokenizer, cleaned for readability, and output as a concise description. This pipeline leverages ViT for feature extraction and GPT-2 for language generation, enabling efficient and almost accurate image captioning.

**Emotion Detection Functionality**

| **Fig. 7.6 DFD Level 2 for emotion detection** |
| --- |

DeepFace analysis processes facial features from uploaded images or video frames through a structured pipeline. It starts with face detection using models like MTCNN, Dlib, or OpenCV to locate and align faces. The detected faces are cropped and passed to feature extraction, where pre-trained models like VGG-Face or Facenet generate numerical embeddings. These embeddings are compared using cosine similarity or Euclidean distance for tasks like emotion detection (e.g., happiness, sadness), demographic analysis (e.g., age, gender), or face verification. The results, such as detected emotions or attributes, are then displayed to the user. This process enables real-time analysis of facial expressions and emotions, providing valuable insights for mental health monitoring and wellbeing assessment.

**Extract Text From Image**

| **Fig. 7.7 DFD Level 2 for text extraction from image** |
| --- |

The process of extracting text from an image using Tesseract-OCR begins with preprocessing the uploaded image. The image is converted to grayscale, noise is reduced using Gaussian or Median Blur, and binarization (e.g., Otsu’s thresholding) isolates text from the background. Text regions are detected using contour analysis or connected components. The processed image is then passed to Tesseract, which uses an LSTM-based neural network to recognize characters and generate machine-readable text, enhanced by language models for accuracy. Postprocessing corrects errors via spell-checking and rule-based replacements (e.g., 0 → O), and formats the text into paragraphs or lines. The final output, displayed or saved as .txt or .docx, is ready for applications like document analysis. This pipeline combines image enhancement, deep learning, and text correction for high-quality results.

**Translation to English**

The process of translating text to English using DeepTranslator begins with preprocessing the input to clean unwanted characters and detect the source language using machine learning models like Google’s Compact Language Detector. The prepared text is passed to DeepTranslator, which interfaces with APIs like Google Translate or Microsoft Translator. These APIs use neural machine translation (NMT) with encoder-decoder architectures and attention mechanisms to translate the text into English. Transformer-based models enhance accuracy by capturing context and long-range dependencies. Post Processing ensures quality by validating completeness, correcting errors, and preserving formatting. The final translated

| **Fig. 7.8 DFD Level 2 for translation to English** |
| --- |

text is displayed or saved, ensuring accuracy and readability. This pipeline integrates NLP, deep learning, and error handling for reliable translations.

**Audio Mood Analysis**

| **Fig. 7.9 DFD Level 2 for audio mood analysis** |
| --- |

The analyze audio mood function begins with the user providing a video file path. The audio is extracted using the extract audio from video function and loaded into memory with the Librosa library. Mel-frequency cepstral coefficients (MFCCs) are com puted using librosa.feature.mfcc to capture frequency patterns for mood analysis. The MFCC array is segmented into four frequency bands—low, mid-low, mid-high, and high—and the scalar mean of each band is calculated to simplify data for classification. The mood is classified (e.g., normal, calm, anxious) based on the dominant frequency characteristics. Results can be further enhanced using the Gemini API to provide tone, mood, and summary details for the audio. This process combines audio feature extraction and classification for comprehensive mood analysis.

**Prediction to Wellbeing Mapping**

| **Fig. 7.10 DFD Level 2 for prediction wellbeing mapping** |
| --- |

The system first processes user input (text, behavioral data, or health metrics) through preprocessing and feature extraction. Machine learning models then predict the most likely mental health condition along with associated probabilities. The top predicted issue and its probability are then fed into GEMINI 2.0 FLASH with a structured prompt to generate wellbeing insights based on Ryff’s six parameters. To refine these insights, an association matrix maps the probabilities of all predicted issues to specific wellbeing parameters, selecting 1 to 3 key parameters (e.g., autonomy, personal growth, or self-acceptance). This ensures that the user receives targeted recommendations for improving their psychological wellbeing. There is also an additional feature where Retrieval Augmented Generation (RAG) is used to provide personalized recommendations based on the user’s mental health condition. The system retrieves relevant information from a knowledge base and generates tailored suggestions. The above 6 main functionalities are reused in the application for the options available to the user. These include Text analysis, Image analysis, Video analysis, PDF analysis, analysis of User response to Image, analysis of Reddit and Twitter user profiles. Wellbeing Survey Option and RAG for wellbeing Insights have been added under the *Implementation* section.

Below are the flow diagrams for the various Analysis options that the web application provides.

|  |
| --- |
| **Fig. 7.11 Text Classification Flow Diagram** |
|  |
| **Fig. 7.12 PDF Upload Flow Diagram** |

|  |
| --- |
| **Fig. 7.13 Image Classification Flow Diagram** |
|  |
| **Fig. 7.14 User Response to Image Flow Diagram** |

| **Fig. 7.15 Video Classification Flow Diagram** |
| --- |

| **Fig. 7.16 Reddit and Twitter username Classification Flow Diagram** |
| --- |

# **8. Implementation**

| **Fig. 8.1 Workflow for getting the model for the web application** |
| --- |

## **8.1 Data Collection and Dataset Preparation**

|  |
| --- |

|  |
| --- |

| **Fig. 8.2 Obtained Dataset** |
| --- |

The dataset for mental health classification was compiled from subreddit communities, initially containing 385,800 records across five categories: anxiety (54 subreddits), PTSD (38), depression (80), bipolar disorder (60), and normal mental states (53). After data cleaning to ensure quality and relevance, the dataset was reduced to 167,279 records. A subset of 18,596 cleaned records was used for analysis to address computational constraints.

| **Fig. 8.3 Collected Data Statistics** |
| --- |

such as memory and processing power, which would make training on the full dataset infeasible on standard hardware. This approach balances efficiency and performance, enabling faster experimentation and iterative model improvement while maintaining a representative sample.

## **8.2 Data Cleaning and Feature Extraction**

|  |
| --- |

|  |
| --- |

| **Fig. 8.4 Output of the TF-IDF Vectorizer** |
| --- |

**Datasets:**

**Before Cleaning**: mental health.csv

**After Cleaning:** preprocessed mental helth.csv

|  |
| --- |

The matrix dimensions for Bag of Words are determined by the number of records, which is 18,597 in this case, and the size of the vocabulary, which represents the number of unique words in the cleaned text column. Similarly, for TF-IDF, the dimensions of the matrix are the same as BOW, calculated as the number of records multiplied by the vocabulary size. For N-gram, the matrix size depends on the range of n-grams used. For example, a unigram produces dimensions equivalent to BOW, while bigram or trigram models increase the vocabulary size due to the inclusion of word combinations. Word2Vec, on the other hand, creates a dense vector representation for each word, with dimensions based on the predefined vector size, such as 100 or 200. Aggregating these vectors at the sentence level, typically by averaging, results in a matrix of dimensions equal to the number of records multiplied by the vector dimensions. For LIWC, the dimensions are determined by the number of predefined LIWC categories, which is typically around 70. The resulting matrix dimensions are the number of records multiplied by the number of LIWC categories.

## **8.3 Machine Learning Models**

Algorithms like Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Random Forest, XGBoost, and Naive Bayes are implemented for the classification of mental health issues. The code algorithm below demonstrates the implementation. The models are evaluated on the test set using metrics like accuracy and classification reports. Hyperparameter tuning is performed using RandomizedSearchCV to optimize the model performance. The Naive Bayes model, after hyperparameter tuning, is selected along with the basic Logistic Regression, Support Vector Machine, and XGBoost for the final ensemble model for the web application.

|  |
| --- |

## **8.4 Deep Learning Models**

|  |
| --- |

|  |
| --- |

| **Fig. 8.5 Output for the LSTM Epochs** |
| --- |

|  |
| --- |
| **Fig. 8.6 LSTM Validation Loss and Accuracy** |
|  |
|  |
| **Fig. 8.7 LSTM Random and Learned Embeddings** |

|  |
| --- |
| **Fig. 8.8 LSTM model architecture** |
|  |
|  |

|  |
| --- |
|  |
| **Fig. 8.9 Transformer Epoch, Loss and Accuracy** |

| **Fig. 8.10 Transformer Model Random and Learned Embeddings** |
| --- |

## **8.5 Ensemble Model**

|  |
| --- |

|  |
| --- |

## **8.6 Wellbeing Survey and Association Matrix**

| **Fig. 8.11 Sample Response Collection Sheet** |
| --- |

|  |
| --- |

|  |
| --- |

| **Fig. 8.12 Sample Association Matrix** |
| --- |

|  |
| --- |

|  |
| --- |

## **8.7 RAG for Wellbeing Insights**

|  |
| --- |
| **Fig. 8.13 Overview of RAG for generating Insights** |
|  |

|  |
| --- |

|  |
| --- |
|  |
|  |
| **Fig. 8.14 Nearest Neighbours among the Embeddings** |

|  |
| --- |
| **Fig. 8.15 Count of Matches and Top Score for an input** |
|  |
| **Fig. 8.16 Instruction Embeddings in 2D space** |

| **Fig. 8.17 Visualization of Query and Top-k Matches** |
| --- |

# **9. Test Plans, Results and Analysis**

|  |
| --- |

|  |
| --- |

|  |
| --- |
|  |
|  |

## **9.1 Results from Base Models**

|  |
| --- |

|  |
| --- |

|  |
| --- |

| 1. Logistic Regression | 1. Naive Bayes |
| --- | --- |
| (c) Support Vector Machine | (d) Random Forest |
| (e) XGboost | (f) K-Nearest Neighbours |
| **Fig. 9.1 Confusion Matrices for ML Models** | |

| 1. Long Short Term Memory | 1. Transformer based model |
| --- | --- |
| **Fig. 9.2 Confusion Matrices for DL Models** | |
| 1. Logistic Regression | 1. K-Nearest Neighbours |
| (c) Support Vector Machine | (d) Naive Bayes |
| **Fig. 9.3 Confusion Matrices for Models after Hyperparameter Tuning** | |

| 1. Logistic Regression | 1. Naive Bayes |
| --- | --- |
| (c) Support Vector Machine | (d) Random Forest |
| (e) XGboost | (f) K-Nearest Neighbours |
| **Fig. 9.4 ROC AUC for ML Models** | |

| 1. Long Short Term Memory | 1. Transformer Based Model |
| --- | --- |
| **Fig. 9.5 ROC AUC for DL Models** | |
| 1. Logistic Regression | 1. K-Nearest Neighbours |
| (c) Support Vector Machine | (d) Naive Bayes |
| **Fig. 9.6 ROC AUCfor Models after Hyperparameter Tuning** | |

|  |
| --- |

|  |
| --- |

K-Nearest Neighbors (KNN) performs poorly on this mental health dataset compared to other algorithms, even after hyperparameter tuning. KNN relies on distance-based metrics, which struggle with the sparse, high-dimensional nature of text data, making it difficult to capture meaningful relationships between Reddit posts and mental health labels. Despite optimizing parameters like weights='distance', n\_neighbors=10, and metric='euclidean', KNN’s performance remains suboptimal, with lower accuracy, precision, recall, and F1-score than Logistic Regression, Naive Bayes, or SVM. Its inability to handle non-linear, context-dependent patterns in text—especially for nuanced categories like anxiety, PTSD, or bipolar disorder—highlights its limitations in text classification tasks. KNN’s reliance on proximity without considering textual context likely explains its poor performance.

|  |
| --- |
| **Fig. 9.7 Result Comparison of the Algorithms** |
|  |
|  |
| **Fig. 9.8 Result Comparison of the Algorithms after Hyperparameter Tuning** |

## **9.2 Comparison of different tokenizations**

|  |
| --- |

In ensemble learning, combining models like Logistic Regression (LR), Naive Bayes (NB), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), XGBoost (XGB), Random Forest (RF), and Long Short-Term Memory (LSTM) poses challenges in computational efficiency, model size, and accuracy. Each model has unique strengths and limitations, making their integration complex. LR, NB, and SVM perform well with simpler vectorization methods like Bag of Words (BoW), which efficiently represents text data as sparse vectors. BoW’s simplicity ensures these models remain computationally lightweight while maintaining good classification accuracy. In contrast, XGBoost benefits from more complex feature extraction methods like TF-IDF (Term Frequency-Inverse Document Frequency), which captures nuanced relationships by weighting words based on their importance across documents. However, combining TF-IDF-based models like XGBoost with BoW-based models can degrade ensemble performance, as the feature representations may not align well. KNN is excluded due to its computational intensity, as it requires storing the entire training dataset and performs poorly with large, high-dimensional datasets. Similarly, Random Forest is omitted because its multiple decision trees increase model size and inference time, outweighing its individual performance benefits. XGBoost, while highly accurate with N-Gram features, is computationally expensive and impractical for real-time systems or large-scale deployment. The choice of feature extraction methods—BoW for LR, NB, and SVM, and TF-IDF for XGBoost—balances simplicity and complexity. BoW ensures efficient, interpretable representations for simpler models, while TF-IDF provides richer, context-aware features for XGBoost. This combination leverages the strengths of both methods, ensuring each model operates effectively without excessive computational strain. Ultimately, this approach strikes a balance between accuracy and efficiency, enabling the ensemble to perform well while remaining scalable and practical for real-world applications.

## **9.3 Results from Ensemble Model Training and Testing**

|  |
| --- |

|  |
| --- |

| 1. Ensemble Model 1 | 1. Ensemble Model 2 |
| --- | --- |
| (c) Ensemble Model 3 | (d) Ensemble Model 4 |
| (e) Ensemble Model 5 | (f) Ensemble Model 6 |
| **Fig. 9.9 Confusion Matrices for Ensemble Models 1 to 6** | |

| 1. Ensemble Model 1 | 1. Ensemble Model 2 |
| --- | --- |
| (c) Ensemble Model 3 | (d) Ensemble Model 4 |
| (e) Ensemble Model 5 | (f) Ensemble Model 6 |
| **Fig. 9.10 ROC AUCfor Ensemble Models 1 to 6** | |

|  |  |
| --- | --- |
| **Fig. 9.11 Confusion Matrix, ROC AUC for Ensemble Model 7 (used in web app)** | |

| **Fig. 9.12 Comparison of Base Models and Ensemble Model 7** |
| --- |

|  |
| --- |

|  |
| --- |

| Below is a comparison of all the ensemble models for reference    **Fig. 9.13 Comparison of all Ensemble Models** |
| --- |

## **9.4 Results from Hierarchical Ensemble Models**

Ensemble Model 7 is applied on different subsets of a very large dataset to create hierarchical ensemble models. The performance of each model is evaluated, and the results are compared to the global ensemble model.

|  |
| --- |

| **Fig. 9.14 Scalable Distributed Architecture 1** |
| --- |
|  |

| **Fig. 9.15 Scalable Distributed Architecture 2** |
| --- |
|  |

|  |
| --- |

## **9.5 User Interface of the Application**

Below are some snapshots from the web application.

| **Fig. 9.16 List of Options for the User** |
| --- |

| 1. Text Input Option | 1. Image Upload Option |
| --- | --- |
| (c) Upload Video Option | (d) PDF Upload Option |
| (e) User Response to Image Option | (f) Questions Related to Image |
| (g) Reddit User Analysis | (h) Twitter User Analysis |
| **Fig. 9.17 Visuals from user inputs in web application** | |

| 1. Generate Image Caption | 1. Social Media User Analysis |
| --- | --- |
| (c) Model Retraining Result | (d) Emotion Analysis |
| (e) Knowledge Graph | (f) Euclidean Distance Analysis |
| (g) Weighted Sum Analysis | (h) Cosine Similarity Analysis |
| **Fig. 9.18 Visuals from analysis of the user inputs** | |

| 1. Prediction and Insights | 1. Specific Parameter Based Insights |
| --- | --- |
| (c) Wellbeing Survey Option | (d) Wellbeing Survey Questions |
| (e) Wellbeing Survey Results | (f) Updated Association Matrix after survey |
| (g) RAG based wellbeing insights | (h) Match visualization and dataset update |
| **Fig. 9.19 Visuals of wellbeing survey and insights** | |

# **10. Conclusion**

|  |
| --- |

# **11. References**

|  |
| --- |