IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# Mental Health Tracker Using Data Science And Artificial Intelligence And Machine Learning

Dr.Bindiya M K
Dept of CSE,
SJB Institute of Technology,
Bengaluru, India

Sathwika T
Dept of CSE,
SJB Institute of Technology,
Bengaluru, India

Raksha A H
Dept of CSE,
SJB Institute of Technology,
Bengaluru, India

Sanjana M S
Dept of CSE,
SJB Institute of Technology,
Bengaluru, India

Sharanya A
Dept of CSE,
SJB Institute of Technology,
Bengaluru, India

*Abstract* — The unique strategy to tracking and monitoring mental health that this paper presents makes use of data science and artificial intelligence/machine learning (AI/ML) approaches. The project's goal is to produce a complete mental health tracker system that can analyze various data sources and offer chances for early intervention and individualized insights. The system integrates information from several sources, including wearable technology, social media activity, calendar events, and self-reported mood assessments, by utilizing AI/ML algorithms. The tracker uses sophisticated data analysis and predictive modelling to find patterns, trends, and possible risk factors related to variations in mental health. In addition, the system makes use of feedback mechanisms and real-time monitoring to enable people to take proactive measures to maintain their mental health. The project offers a scalable and easily available solution for ongoing mental health monitoring and support, helping to close the gap between conventional mental health evaluation techniques and cutting-edge technologies.

**Keywords**— logistic regression, K-nearest neighbors (KNN), behavioral analysis.

## I. Introduction

Due to its substantial influence on people's lives, communities, and societies at large, mental health has attracted more attention globally in recent years. There is increasing interest in using data science and artificial intelligence/machine learning (AI/ML) approaches to develop novel solutions for mental health monitoring and intervention as a result of the development of digital health technology and the availability of data.This introduction presents the rationale and framework for a Mental Health Tracker project, which aims to harness the power of data science and AI/ML algorithms, including logistic regression, K-nearest neighbors (KNN) to create a comprehensive tool for tracking and managing mental well-being.

The requirement for easily accessible, adaptable, and customized methods of mental health monitoring and assistance is what spurred the creation of the Mental Health Tracker project. Self-reporting or recurring clinician exams are two common components of traditional mental health assessment techniques, which can be biased, have inadequate data collecting, or have therapeutic delays. In contrast, the capacity to analyze massive volumes of data, such as physiological signs, behavioural patterns, and contextual elements, to provide real-time insights into people's mental health status is presented by the merging of data science and AI/ML approaches.

The core of the Mental Health Tracker project is the deployment of many AI/ML models, each with unique advantages in tracking and predicting mental health outcomes.Logistic regression is a frequently used statistical technique for binary classification, which simulates the chance of experiencing specific mental health symptoms or diseases based on various criteria.

The K-nearest neighbors (KNN) algorithm, on the other hand, compares an individual's attributes with those of comparable persons in the dataset, using similarity measurements to identify an individual's mental health condition. Finally, by combining a number of trained decision trees on various portions of information, the random forest technique, which is renowned for its ensemble learning methodology, can provide strong predictions.

People can get individualized insights and recommendations for taking care of their mental health by combining these models into a unified Mental Health Tracker platform. Using a variety of data sources, including wearable technology, social media activity, and self-reported surveys, the platform may include functions like behavioral analysis, stress level monitoring, and mood tracking. The program also aims to observe ethical standards, prioritize user privacy and data security, and provide informed consent for data gathering and analysis.

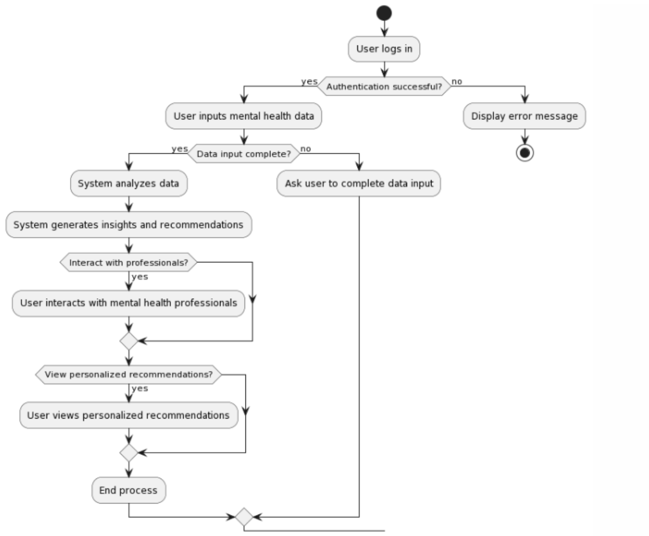
The potential of utilizing contextual data for forecasting mood swings is highlighted by the study on mood prediction based on calendar events using multitask learning. This methodology emphasizes the significance of taking into account a variety of information sources in order to improve prediction precision and tailored treatments.

The examination of mental health in engineering education sheds emphasis on the intersectional issues that affect mental health in addition to demography disparities. These findings demonstrate the necessity for tailored interventions that address the unique requirements of diverse groups.

## II. Related Works

Research by A Danowitz, K Beddoes [1] studies mental health in engineering education, finding engineering students may experience similar rates of mental health issues as other students, but with less likelihood to seek help.It emphasizes the significance of addressing mental health due to the demanding engineering culture and the impact on graduation rates. Tateyama, Naoki, Rui Fukui, and Shin'ichi Warisawa [2] proposes a method for predicting mood swings using information from daily schedules. When compared to conventional methods, the study shows how effective this methodology is at improving the precision of mood prediction. The research focuses on using multitask learning, which allows the model to learn from multiple related tasks simultaneously. In this instance, the tasks involve not only predicting mood but also extracting features from calendar events that are indicative of emotional state. This combined learning improves the overall accuracy of mood prediction. Guo, Teng, et al *[3]* proposed a method to identify students who may be struggling mentally by combining information from various educational data sources. The authors test their method on real-world data and show that it can effectively detect students with mental health problems. Their multimodal educational data fusion method has the potential to effectively identify students at risk and make connections with suitable support systems for them. Luoluo Liu et al *[4]* proposed a novel algorithm called Minimum Similarity Association Rules (MSAR) to determine which combinations of comorbidities are most closely linked to recurrent ED visits and hospital admissions. MSAR is designed to be interpretable by medical professionals. It goes beyond just identifying frequently occurring conditions together and instead highlights combinations that are most specific to frequent patient readmissions. Research by Nash, Christian, Rajesh Nair, and Syed Mohsen Naqvi [5] focused on surveying the existing research on applying machine learning (ML) techniques to diagnose ADHD and depression. The survey summarized the applications of Machine Learning in diagnosing ADHD and Depression. Banna, Md Hasan Al, et al [6] explores the potential of a hybrid deep learning model to predict the impact of COVID-19 regarding mental health by analyzing social media big data. The proposed model combines different deep learning techniques to handle the complexities of social media data. This data include text, emojis, and even posting frequencies. By processing all this information, the model can hopefully extract meaningful patterns that correlate with mental health.

## III. DESIGN




Start:The process begins with the user logging into the system.

Authentication successful?:The system checks if the user's authentication is successful.If authentication is successful, the process continues; otherwise, an error message is displayed, and the process stops.

User inputs mental health data:Upon successful authentication, the user inputs their mental health data into the system.

Data input complete:The system checks if the data input process is complete.If data input is complete, the system proceeds to analyze the data; otherwise, the user is prompted to finish the input.

System analyzes data:Once the data input is complete, the system analyzes the user's mental health data.

System generates insights and recommendations:Based on the analysis, the system generates insights and personalized recommendations for the user.

Interact with professionals:The system checks if the user wants to communicate with experts on mental health for further support or guidance.If the user chooses to interact, they can interact with professionals through the system.

View personalized recommendations?:The system checks if the user wants to view personalized recommendations generated based on their mental health data.If the user chooses to view recommendations, they are reachable by the system.

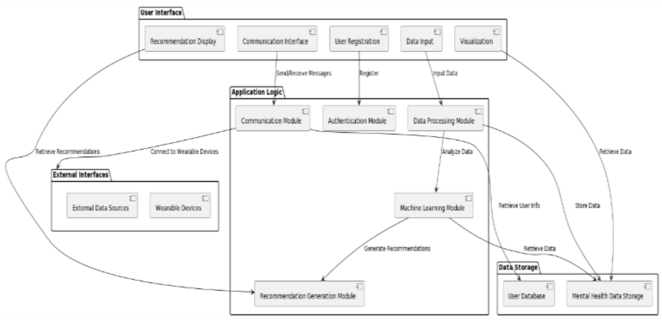
End process:The process ends after the user has received insights, recommendations, and any desired conversations with mental health specialists.

## IV.SYSTEM ARCHITECTURE

*User Interface*: This section represents the components responsible for handling user interaction.

*User Registration*: Allows users to register an account within the system.

*Data Input*: Enables users to input their mental health data, such as mood, sleep patterns, stress levels, etc.




*Visualization*: Provides visual representations of the user's mental health data, allowing them to track their metrics over time.

*Recommendation Display* : personalized recommendations generated by the system according to the analysis of user data are displayed.

*Communication Interface*: Facilitates communication between users and mental health professionals, if needed.

*Application Logic*: This package contains the core logic of the system responsible for processing user data, performing analysis, generating recommendations, and handling communication.

*Authentication Module*: Manages user authentication and ensures secure access to the system's functionalities.

*Data Processing Module*: Handles the processing and storage of user input data, preparing it for analysis.

*Machine Learning Module*: Utilizes machine learning algorithms to analyze the user's mental health data and extract insights.

*Recommendation Generation Module*: Generates personalized recommendations based on the analysis results to assist users in improving their mental well-being.

*Communication Module*: Manages communication channels between users and mental health professionals, additionally external interfaces.

*Data Storage*: Represents the database components in charge of keeping user and mental health data stored.

*User Database*: Stores user account information such as usernames, passwords, and profile details.

*Mental Health Data Storage*: Stores the user's input data, analysis results, and generated recommendations for future reference and analysis.

*External Interfaces*: Represents connections to external devices or data sources that interact with the system.

*Wearable Devices*: Interfaces with wearable devices such as fitness trackers or smartwatches to collect additional data about the user's behavior and physiology.

*External Data Sources*: Interfaces with other resources, like social media sites or medical databases, to obtain background data that could have an impact on the user's mental health.

## V. METHODLOGY

This section contains the presentation of the proposed overflow of the project, which includes steps like dataset collection, data pre-processing, segmentation, feature extraction and classification.

### A. *Data Collection*

The first step in methodology involves the acquisition of a diverse and representative dataset for training and testing our deep learning model. The data is taken from the Kaggle. The information in this dataset about users' experiences, behaviors, and moods should be included. Eighty percent of the dataset is used for training, while twenty percent is used for testing.

### B. *Pre-processing*

Prior to model training, the collected dataset undergoes preprocessing to ensure consistency and quality. Preprocessing is an essential phase in preparing your mental health tracker data for analysis. Improving the quality of the input data and speeding up the model's learning process are the objectives. Simple explanation of preprocessing steps in this context:

1. Data Cleaning: Ensure that there are no outliers in the data. Determine which entries have missing data points, then use statistical procedures to solve them.

2. Data Transformation: For consistency, all text submissions will be changed to lowercase and have all punctuation deleted. All dates will have the same format. To ensure that every feature contributes equally, numerical data will either be scaled (e.g., mood rating 0-1) or normalized (e.g., sleep time in minutes). This guarantees that the dataset is homogeneously formatted and ready for analysis.

3. Feature Engineering: By extracting new features from preexisting data points, data enrichment can be accomplished. This procedure makes it possible to investigate user trends and behavior in greater detail. For example, figuring out the average mood score over a certain time frame (like a week) offers important information on general mood swings. Underlying tendencies in sleep patterns, including differences in sleep duration or quality over time, can be uncovered by time-series analysis of sleep patterns. Through the integration of these derived features, scholars acquire a more all-encompassing comprehension of the dynamics present in the data, hence facilitating a more intricate and refined analysis.

4. Data Validation: Data validation is the final step in preprocessing the mental health tracker data. It ensures the transformations and cleaning procedures haven't introduced errors or inconsistencies, guaranteeing reliable data for analysis.

### C. *Train and Test split:*

It is imperative to partition the dataset into a training set and a test set throughout the training and testing phase of a mental health project. This division is typically done randomly, with a common ratio being 80% for training and 20% for testing.

Equally important is ensuring that the distribution of mental health-related features is similar in both the training and test sets. This helps prevent bias and ensures that the model is evaluated accurately across all relevant variables.

By maintaining a balanced distribution of features between the two sets, researchers can better gauge the model's performance and its ability to generalize to new, unseen data. This approach fosters robustness in model evaluation and enhances the reliability of the project's outcomes.

### C. *Model Selection:*

Choose logistic regression as your baseline model. Logistic regression is a wise decision when dealing with binary classification problems, which is often the case in mental health tracking.

### D. Model Training:

Train the logistic regression model on the training dataset. The model will learn the relationship between the input features (questions and answers) and the target variable (level of mental stress).

### E. Model Evaluation :

To make sure the model generalizes well beyond the training data it was developed on, this stage is essential. We use metrics such as F1-score, recall, accuracy, and precision to evaluate the model's performance on the testing set of data.

Accuracy: - Its implementation as a performance metric for classification techniques is very prevalent. It can be characterized as the ratio of all predictions made to the number of accurate predictions achieved.

Precision: - The number of accurate documents produced by our ML model can indeed be considered as precision, which is employed in document retrieval of information. The ratio of true positives to the total of true positives and false positives may be utilized to calculate the precision.

F1-measure: - We can determine the harmonic mean of recall and precision using this score. The precision and recall-adjusted averages make up the F1 score mathematically. F1 would have a greatest value of 1 and a worst value of 0. Precision plus recall added together by a factor of two may be utilized to calculate the F1 measure.

### F. Interpretability:

The ability to be interpreted guarantees that the knowledge derived from prediction models is clear and useful. Through an analysis of the logistic regression model's coefficients, researchers can determine which factors have the greatest predictive power for mental stress. Interpretability promotes educated decision-making and raises user and healthcare professional confidence in the model's predictions, enhancing the efficacy of mental health tracking and intervention techniques.

## III. Conclusions

The mental stress tracker project was successful in creating a sophisticated application that included live chat help, tailored advice, and the use of logistic regression as the method for predicting stress levels. The purpose of the application is to help people monitor and manage their mental health in an effective manner. The program offers useful insights and resources for stress reduction through precise estimates of stress levels and tailored advice. With the addition of live chat support, clients now have a convenient option to seek advice from medical experts when they're feeling stressed.Positive experiences with the application's functionality, usability, and intuitive interface are indicated by user feedback. Users have expressed satisfaction with the individualized recommendations, which provide them with hobbies, music, and resources specifically designed to alleviate stress.

Despite the fact that there are still some drawbacks, like the requirement for a bigger dataset to train the logistic regression model, upcoming improvements can concentrate on growing the dataset and incorporating real-time physiological data, like heart rate variability or sleep patterns, to offer more thorough insights into a user's stress levels. The app may be able to gather data in real-time and offer individualized recommendations based on the user's physiological and activity patterns through integration with wearable technology or sensors. In summary, the logistic regression-driven mental stress tracker application provides a useful tool for tracking and controlling stress levels. A thorough approach to stress management is facilitated by the live chat help, tailored recommendations, and precise stress level projections.

## References


[1] Andrew Danowitz and Kacey Beddoes, "Mental Health in Engineering Education: Identifying Population and Intersectional Variation," 3 August 2022, volume 65,
doi: 10.1109/TE.2022.3182626

[2] Naoki tateyama , Rui Fukui and Shin'ichi Warisawa
"Mood Prediction Based on Calendar Events Using Multitask Learning," 25 July 2022, volume 10,
doi: 10.1109/ACCESS.2022.3193778

[3] Luoluo Liul , Dennis Swearingen , Eran Simhonl , Chaitanya Kulkarni , David Noren , Ronny Mans,
"Interpretable Identification of Comorbidities Associated with Recurrent ED and Impatient Visit,"
11-15 July 2022,

[4] Md. Hasan Al Banna , Tapotosh Ghosh , Md. Jaber Al Nahian , M. Shamim Kaiser , Mufti Mahmud , Kazi Abu Taher , Mohammad

Shahadat Hossain , Karl Andersson, " A Hybrid Deep Learning Model to Predict the Impact of COVID-19 on Mental Health From Social Media Big Data," 16 February 2023, volume 11,

doi : 10.1109/ACCESS.2023.3243104

[5] Teng Guo , Wenhong Zhao , Mubarak Alrashoud , Amr Tolba , Selena Firmin , Feng Xia, " Multimodal Educational Data Fusion for Students' Mental Health Detection," 8 July 2022, volume 10,

Doi: 10.1109/ACCESS.2022.3187502

[6] Christian Nash , Rajesh Nair , Syed Mohsen Naqvi, "Machine Learning in ADHD and Depression Mental Health Diagnosis: A Survey," 8 July 2022, volume 11

doi: 10.1109/ACCESS.2023.3304236