Date	20 June 2024
Team ID	739903
Project Title	Mental health prediction
Maximum Marks	2 Marks

Data Collection and Preprocessing Phase

Data Collection Plan & Raw Data Sources Identification Report:

Identify and gather clinical records, surveys, behavioral data, public datasets, and research studies. Ensure ethical compliance and data quality throughout the process.

Section	Description
Project Overview	The machine learning project aims to predict mental health outcomes based on personal and demographic information. Using a dataset with features such as age, gender, employment status, mental health history, and lifestyle factors, the objective is to build a model that accurately predicts mental health status (e.g., presence of mental health condition, severity). This will facilitate early intervention and personalized treatment planning. |
Data Collection Plan	- Search for datasets related to mental health assessments, psychological surveys, and demographic details. - Prioritize datasets with diverse demographic information, including age, gender, employment status, and lifestyle factors. .
Raw Data Sources Identified	The raw data sources for this project include datasets obtained from mental health surveys and repositories like Kaggle, UCI Machine Learning Repository, and public health databases. The provided sample data represents a subset of the collected information, encompassing variables such as age, gender, employment status, mental health history, and lifestyle factors for machine learning analysis.

Data Collection Plan:

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset comprises personal details (age, gender, employment status), mental health metrics (survey responses, diagnoses), and lifestyle factors (exercise frequency, sleep patterns).	https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey	CSV	TBD	Public