

Data Collection and Preprocessing Phase

Date	19 May 2025
Team ID	SWTID1750233055
Project Title	Mental Health Prediction
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Report:

Elevate your data strategy with the Data Collection Plan and the Raw Data Sources Report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and prediction of mental health conditions.

Data Collection Plan:

Section	Description
Project Overview	The machine learning project aims to predict mental health risk based on individual responses to survey questions. Using a dataset containing features such as age, gender, family history, work interference, and access to mental health resources, the objective is to build a model that can accurately identify individuals at risk. This supports early detection, intervention, and mental wellness planning.
Data Collection Plan	<ul style="list-style-type: none">🔍 Search for publicly available mental health datasets from sources like Kaggle or UCI Machine Learning Repository.🔍 Prioritize datasets that include diverse demographic and workplace-related features relevant to mental health.🔍 Ensure data is anonymized, clean, and ethically sourced for responsible use.

Raw Data Sources Identified	<p>The raw data sources for this project include datasets obtained from Kaggle, specifically the “<i>Mental Health in Tech</i>” survey dataset. Kaggle is a widely recognized platform for open datasets, competitions, and collaborative data science.</p> <p>The sample data includes information such as age, gender, family history of mental illness, workplace support, and availability of mental health resources. These variables are used to train the machine learning model (AdaBoost) to assess mental health risks effectively.</p>
-----------------------------	---

⛶ Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permission
Kaggle Dataset	The dataset contains mental health survey responses from individuals working in the tech industry. It includes details like age, gender, family history of mental illness, treatment sought, and workplace support. These features are used to train	https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey	CSV	15 kB	Public
	the model to identify individuals at risk of mental health issues.				