

Data Collection and Preprocessing Phase

Date	9th July 2024
Team ID	SWTID1720243396
Project Title	Panic Disorder Detection
Maximum Marks	2 Marks

Data Quality Report Template

The Data Quality Report Template will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	Inconsistent Categorical Values in : Gender, Family History, Personal History, Current Stressors, Symptoms, Severity, Impact on Life, Demographics, Medical History, Psychiatric History, Substance Use, Coping Mechanisms, Social Support, and Lifestyle Factors	Moderate	Standardize Categorical Values
Kaggle Dataset	Redundant Information: Some columns may contain redundant information	Low	Feature Selection and Engineering

Kaggle Dataset	Imbalance in Target Variable: 4.2% diagnosed vs. 95.8% not diagnosed This indicates potential imbalance in the dataset	High	Either oversampling the minority class or undersampling the majority class. Using methods like SMOTE Use algorithms that are robust to class imbalance or adjust the class weights in algorithms that support this parameter.
----------------	---	------	---