Data Collection and Preprocessing Phase

Date	21 June 2024
Team ID	740005
Project Title	Estimating presence or absence of smoking Through Bio Signals
Maximum Marks	2 Marks

Data Quality Report:

This data quality report serves as a foundational document, outlining the steps taken to ensure the reliability and usability of biosignal data in the project. It underscores the commitment to maintaining high data standards and reliability in achieving the project's objectives effectively.

Data Quality Report:

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	Missing values in 'Gender', 'tartar', 'oral' columns.	Moderate	Use mean/median imputation to fill missing values in numerical columns (tartar, oral). For categorical columns (Gender), use mode imputation to fill missing values.

Kaggle Dataset	Categorical data in the dataset	Moderate	Perform encoding on categorical variables to convert them into numerical format suitable for modeling. For instance, use one-hot encoding or label encoding depending on the nature of the categorical data and the machine learning algorithm used.