Data Collection and Preprocessing Phase

Date	13 JULY 2024
Team ID	xxxxxx
Project Title	Detection of Autistic Spectrum Disorder: Classification
Maximum Marks	6 Marks

Data Exploration and Preprocessing Template

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	**Description**
Data Overview	Number of samples (rows) and features (columns) arget variable (if it's a classification task, this would be whether a person has ASD or not
Univariate Analysis	Summary statistics (mean, median, mode, range, standard deviation, etc.)
Bivariate Analysis	Grouped summaries or pivot tables for relationships involving categorical variables
Multivariate Analysis	Dimensionality reduction techniques (PCA, t-SNE) to visualize high-dimensional data
Outliers and Anomalies	Box plots, scatter plots, or statistical methods (like Z-score) to detect outliers Strategies for handling outliers (removal, transformation, or treating separately
Data Preprocessing Code Screenshots	

Loading Data	Loading the dataset (e.g., using pandas in Python, read.csv in R)
Handling Missing Data	Handling missing data (imputation strategies such as mean, median, mode, or advanced methods)
Data Transformation	Data transformation (scaling numeric data, normalization techniques)
Feature Engineering	Data transformation (scaling numeric data, normalization techniques)
Save Processed Data	Saving processed data for future use (to CSV, Excel, or database)