

Data Collection and Preprocessing Phase

Date	4 July 2024
Team ID	SWTID1720000747
Project Title	Detection Of Autistic Spectrum Disorder: Classification
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
Dataset 1	NULL values in age, ethnicity, and relation	High	Impute missing values using appropriate methods like mean/median for age and mode for categorical columns.
Dataset 1	Age outliers (e.g., 383)	High	Remove or correct outliers using statistical methods or domain knowledge.
Dataset 1	Incorrect data types for age, relation, jaundice, and family autism history	Medium	Convert object types to appropriate data types (e.g., integer for age, binary for jaundice and autism history).

Dataset1	Unnecessary columns (e.g., ethnicity, country_of_res, relation)	Medium	Drop unnecessary columns to streamline the dataset and focus on relevant features.
Dataset 1	Non-integer categorical data	Medium	Apply encoding techniques like OneHotEncoder or LabelEncoder to convert categorical data to numeric form.