

Data Collection and Preprocessing Phase

Date	28-07-2025
Team ID	HK
Project Title	Predicting Plant Growth Stages with Environmental and Management Data Using Power BI
Maximum Marks	3 Marks

Data Quality Report

The Data Quality Report 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 Collection and Preprocessing Phase

Data Quality Report Template

Data Source	Data Quality Issue	Severity	Resolution Plan
Smart Farming Data	Missing values in humidity and sunlight columns	Moderate	Use mean imputation for missing values or apply KNN imputation if patterns exist in nearby records.
Fertilizer Usage & Growth	Inconsistent labelling of fertilizer types (e.g., "Org", "Organic", "org.")	Low	Apply data standardization using string normalization techniques to unify all entries (e.g., convert all to lowercase and map synonyms).
Temperature & Humidity Records	Some extreme temperature outliers (e.g., >70°C) that are unrealistic	High	Use Z-score method to detect and remove outliers or cap them using IQRbased clipping.
Soil Performance Data	Duplicate rows with identical soil and irrigation values	Moderate	Use Power Query or pandas. drop_duplicates () to remove duplicate entries and retain unique records.