Data Collection and Preprocessing Phase

Date	9th June 2024
Team ID	SWTID1720109498
Project Title	Blueberry Yield Predictor
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	Outliers present in the dataset	Moderate	z-scores are used to standardize the data and effectively detect and remove extreme values that may distort model performance.
Kaggle Dataset	Redundant columns in the dataset	Moderate	Through data visualizations like bar graphs, histograms and correlation heatmaps, columns with high correlation are recognized and removed.