Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	SWTID1720113374
Project Title	Predicting Compressive Strength Of Concrete Using Machine Learning
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
Concrete Compressive Strength Dataset	Inconsistent mix proportions (e.g., varying units)	High	Standardize units of measurement for all mix components. Implement data validation tools to ensure consistency in mix proportions.Review and correct discrepancies.
Concrete Compressive Strength Dataset	Missing or incomplete data on curing conditions	Medium	Identify the cause of missing data (e.g., recording errors, data entry Issues.Impute missing values using appropriate methods (e.g., mean/median imputation).

Concrete Compressive Strength Dataset	Inconsistent age reporting	Medium	Standardize age reporting format. Implement data validation checks during data entry. Train data entry personnel on the correct format for age recording.
Concrete Compressive Strength Dataset	Outliers in compressive strength measurements	Medium	Implement outlier detection algorithms to identify extreme values. Review outliers to determine if they are genuine or errors. Consider handling or excluding extreme outliers.
Concrete Compressive Strength Dataset	Class imbalance in mix proportions	Low	Implement data augmentation techniques to balance mix proportion categories.Explore using algorithms that can handle imbalanced data. Consider incorporating prior knowledge into the model.