Data Collection and Preprocessing Phase

Date	8 July 2024
Team ID	SWTID1720116242
Project Title	Predicting Compressive Strength Of Concrete Using Machine Learning
Maximum Marks	2 Marks

Data Quality Report Template

The Data Quality Report Template will compile an assessment of data quality issues sourced from selected datasets, outlining severity levels and proposed resolution strategies. This structured approach will facilitate the systematic identification and remediation of data inconsistencies, ensuring enhanced accuracy and reliability for analytical purposes.

Data Source	Data Quality Issue	Severity	Resolution Plan
Concrete Compressive Strength Dataset	Inconsistent mix proportions (e.g., varying units)	High	Standardize units of measurement across all mix components. Implement rigorous data validation to ensure consistency. Review and correct discrepancies systematically.
Concrete Compressive Strength Dataset	Missing or incomplete data on curing conditions	Medium	Investigate the root cause of missing data (e.g., recording errors, data entry issues). Utilize appropriate methods (e.g., imputation) to fill missing values. Implement robust data validation procedures to prevent future discrepancies.

Concrete Compressive Strength Dataset	Inconsistent age reporting	Medium	Establish standardized reporting formats for age data. Enforce validation checks during data entry and provide training to ensure compliance with format standards.
Concrete Compressive Strength Dataset	Outliers in compressive strength measurements	Medium	Deploy outlier detection algorithms to identify extreme values. Evaluate and handle outliers based on their impact and validity. Document procedures for outlier management.
Concrete Compressive Strength Dataset	Class imbalance in mix proportions	Low	Apply data augmentation techniques to balance mix proportion categories. Employ algorithms designed to handle imbalanced data effectively. Incorporate domain knowledge to refine modeling approaches