Data Collection and Preprocessing Phase

Date	14 July 2024
Team ID	SWTID1720171884
Project Title	Predicting Compressive Strength Of Concrete Using Machine Learning
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

Data Collection Plan Template

Section	Description
Project Overview	This project aims to predict the compressive strength of concrete based on its ingredients, using a dataset with features like cement, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, fine aggregate, and the age of the concrete.. The analysis involves univariate, bivariate, and multivariate approaches to identify patterns and relationships. Outliers are managed to enhance model accuracy, and a Gradient Boosting Regressor is used for predictions. The model is deployed via a Flask web application, allowing users to input data and obtain predictions interactively.
Data Collection Plan	• Search Datasets related to concrete strength. • Accessing dataset provided in project workspace.
Raw Data Sources Identified	The raw data sources for this project include datasets obtained from Kaggle, the popular platforms for data science competitions and repositories. The dataset contains components of concrete mixture like cement, blast furnace slag, fly ash, water,

	superplasticizer, coarse aggregate, fine aggregate, and the age of the concrete.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset contains details of components for concrete mixture like ('cement', 'blast_furnace_slag', 'fly_ash', 'water', 'superplasticizer', 'coarse_aggregate', 'fine_aggregate ', 'age','concrete_comp ressive_strength')	https://www.kag gle.com/datasets /elikplim/concret e-compressive-strength-data-set	CSV	58KB	Public