

Data Collection and Preprocessing Phase

Date	9 July 2024
Team ID	team-739994
Project Title	Precise Coffee Quality Prediction
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, This plan ensures meticulous curation and integrity of data, facilitating informed decision-making in every analysis. Key components include identifying primary data sources, addressing data quality issues such as missing values and duplicates, the plan implements robust resolution strategies to maintain data accuracy and reliability throughout the project.

Data Collection Plan Template

Section	Description
Project Overview	<p>The Coffee Quality Prediction project aims to leverage machine learning techniques to accurately predict the quality of coffee beans based on various attributes. The project involves collecting and preparing data from multiple sources, including coffee farms, cooperatives, and public datasets. By analyzing this data, the project seeks to identify patterns and factors that influence coffee quality.</p> <p>This project aims to provide coffee producers and stakeholders with a powerful tool to assess and enhance the quality of their products, ultimately leading to better market competitiveness and consumer satisfaction.</p>
Data Collection Plan	Dataset is collected through the kaggle which is related to Coffee Quality Prediction and also collected the information regarding Coffee Quality which will be helpful for model development.

Raw Data Sources Identified	Raw data i.e., Dataset for Coffee Quality Prediction is collected from Kaggle which consists of attributes like Aroma, Body, Aftertaste etc.
-----------------------------	--

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The coffee quality prediction dataset is a comprehensive collection of data points representing various attributes of coffee beans like Aroma, Quakers, Color, Aftertaste, Body etc..	https://www.kaggle.com/datasets/thotamanohar/masterful-machines-precise-coffee-quality	CSV	31MB	Public