Data Collection and Preprocessing Phase

Date	10 July 2024
Team ID	Team-740058
Project Title	Masterful Machines: Precise Coffee Quality Predictons Through ML
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Report:

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:

Section	Description
Project Overview	This project aims to develop a machine learning model to predict coffee quality based on physical, chemical, and sensory data of coffee beans. By utilizing various ML algorithms and feature engineering, we seek to create a robust and accurate prediction system. The end goal is to deploy a user-friendly application for real-time quality assessments, benefiting producers and consumers alike.
Data Collection Plan	Collect data on coffee beans from sources like the Coffee Quality Institute (sensory evaluations), International Coffee Organization (production details), and farm-level records (growing conditions). Preprocess the data by handling missing values, normalizing features, and encoding categorical variables for model readiness.
Raw Data Sources	Raw data sources identified include the Coffee Quality Institute (CQI) for sensory evaluations,

Identified	the International Coffee Organization (ICO) for production details, and farm-level records for growing conditions and practices. Additional sources include chemical analysis reports of coffee bean composition.

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset comprises Features of coffee i.e, Aroma, Flavor,aftertaste, Acidity,body, Balance,Uniformity, Quakers, Color_Encoded	https://drive.google.com/file/d/1H_U9WGqZqluXxkg1GoY0fqJGAVLaQSoP/view?usp=drive_link	CSV	15 kB	Public