

Data Collection and Preprocessing Phase

Date	8 July 2024
Team ID	SWTID1720078167
Project Title	Rice Type Classification using CNN
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	The machine learning project aims to develop a robust image classification model to identify different types of rice grains (Arborio, Basmati, Ipsala, Jasmine, and Karacadag). The objective is to create an efficient and accurate model that can classify rice grain images into their respective categories to assist in quality control and inventory management in the agricultural sector.
Data Collection Plan	Data for this project will be collected from various publicly available datasets, primarily from the Kaggle platform. The datasets contain labeled images of different rice grain types. Additionally,

	supplementary data sources such as agricultural research databases and open-source image repositories may be used to enhance the dataset.
Raw Data Sources Identified	Kaggle Rice Image Dataset

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	This dataset contains labeled images of five different types of rice grains: Arborio, Basmati, Ipsala, Jasmine, and Karacadag. Each category has around 15000 images, which will be used for training and validation purposes.	https://www.kaggle.com/datasets/muratkokluturk/rice-image-dataset/data	Image	230 MB	Public