

Data Collection and Preprocessing Phase

Date	12 July 2024
Team ID	SWTID1720085076
Project Title	Rice Type Classification using CNN
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification

Elevating 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	The rice image classification project aims to accurately classify rice quality using Convolutional Neural Networks (CNNs). Utilizing a dataset of rice images with various quality attributes, the objective is to develop a model that reliably distinguishes between different Categories of rice. This will streamline the Categorical assessment process, ensuring consistency and efficiency in rice quality evaluation for distributors and farmers. The model will be integrated into a user-friendly application, enabling real-time classification and enhancing operational decision-making in the agricultural sector.
Data Collection Plan	<ul style="list-style-type: none"> Look for publicly available datasets related to rice quality and classification. Explore agricultural research databases and academic publications.

	<ul style="list-style-type: none"> Consider platforms like Kaggle and UCI Machine Learning Repository
Raw Data Sources Identified	<p>The raw data sources for this project include datasets obtained from Kaggle and UCI, the popular platforms for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing images of Arborio, Basmati, Ipsala, Jasmine, and Karacadag rice varieties. A total of 75,000 grain images, with 15,000 from each variety, are included.</p>

Raw Data Sources

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	<p>the dataset contains 4 categories of rice -</p> <p>In this study, Arborio, Basmati, Ipsala, Jasmine and Karacadag, which are five different varieties of rice often grown in Turkey, were used. A total of 75,000</p>	https://www.kaggle.com/datasets/muratkokluda/taset/rice-image-dataset	CSV	230 MB	Public

	grain images, 15,000 from each of these varieties, are included in the dataset.				
--	---	--	--	--	--