Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	SWTID1720027196
Project Title	Greenclassify: Deep Learning-Based Approach For Vegetable Image Classification
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	Know fundamental concepts and techniques of Convolutional Neural Networks. Gain a broad understanding of image data. Know how to pre-process/clean the data using different data preprocessing techniques. know how to build a web application using the Flask framework.

Data Collection Plan	
Raw Data Sources Identified	This dataset contains three folders: ? train (15000 images) ? test (3000 images) ? validation (3000 images) each of the above folders contains subfolders for diffe rent vegetables wherein the images for respective vegetables are present.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	The initial experiment is done with 15 types of common vegetables that are found throughout the world. The vegetables that are chosen for the experimentation are-	https://www.kaggle.com/datasets/misrakahmed/vegetable-image-dataset	jpj	571.74 MB	Public

	bean, bitter gourd, bottle gourd, brinjal, broccoli, cabbage, capsicum, carrot, cauliflower, cucumber, papaya, potato, pumpkin, radish and tomato. A total of 21000 images from 15 classes are used where each class contains 1400 images of size 224×224 and in *.jpg format. The dataset split 70% for training, 15% for validation, and 15% for testing purpose.				