Data Collection and Preprocessing Phase

Date	11 November 2024
Team ID	team-739757
Project Title	Tomato Plant Disease Detection From Leaf Images Using Deep Learning
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	Developing a deep learning model to predict tomato plant diseases from leaf images.
Data Collection Plan	Gather a diverse dataset of tomato leaf images, labeled with different disease classes (e.g., Early blight, Late blight, Leaf mold, healthy) and additional metadata from Various Data Sources Like kaggle.
Raw Data Sources Identified	Public Datasets: PlantVillage, Tomato Leaf Disease Dataset, etc. Image Scraping: Collect images from online sources using web scraping techniques, focusing on relevant websites like research papers, agricultural forums, and image sharing platforms.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset-1 PlantVillage	A large-scale dataset containing images of tomato leaves affected by various diseases and healthy leaves.	https://www.kaggle.com/datasets/emmarex/plantdisease	JPEG	343.23 MB	Public
Dataset 2: Tomato Leaf Disease Dataset	A dataset specifically curated for tomato leaf disease classification	https://www.kaggle.com/datasets/kaustubhb999/tomatoleaf	JPEG	186.22 MB	Public
Dataset 3: Image Scraping - [Source Website]	Images collected from a specific website using web scraping	https://github.com/laggui/image-search-scraper	JPEG	200 MB	Specific Permissions