Data Collection and Preprocessing Phase

Date	2nd July 2024
Team ID	SWTID1719992739
Project Title	Visual Diagnostics: Detecting Tomato Plant Diseases through Leaf Image Analysis
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 objective of this machine learning project is to develop a model capable of detecting diseases in tomato leaves using image classification techniques. This will involve using convolutional neural networks (CNNs) to identify various diseases in tomato plants from leaf images, helping farmers to take timely action and improve crop yield.
Data Collection Plan	Data will be collected from the Kaggle dataset "Tomato Leaf" which contains images of tomato leaves categorized by disease type. This dataset provides a diverse range of leaf conditions necessary for training a robust model.
Raw Data Sources Identified	The primary raw data source is a publicly available dataset on Kaggle.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Tomato Leaf Dataset	This dataset contains images of tomato leaves categorized into various disease types and a healthy category. The images are labeled and organized into directories for each category.	https://www.kaggle.com/datasets/kaustubhb999/tomatoleaf	Image (JPG)	188 MB	Public