

## Data Collection and Preprocessing Phase

Date	19 June 2025
Team ID	SWTID1750170729
Project Title	<b>Deepfruitveg: Automated Fruit And Veg Identification</b>
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 classify images of fruits and vegetables using a deep learning model (EfficientNetB3) trained on a curated dataset.
Data Collection Plan	The dataset was manually downloaded from Kaggle Website ( <a href="https://www.kaggle.com/datasets/yudhaislamisulistya/plants-type-datasets/code">https://www.kaggle.com/datasets/yudhaislamisulistya/plants-type-datasets/code</a> ) and contains images of various fruits and vegetables across 30 categories.

Raw Data Sources Identified	The raw data includes separate folders for training, validation, and testing, each containing subfolders for individual classes (e.g., Apple, Banana, etc.).
-----------------------------	--

### Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Train Dataset, Test Dataset, Validation Dataset (For Fruits And Vegetables)	Fruit and vegetable image dataset containing 30 classes, with separate folders for train, test, and validation sets.	<a href="#">Kaggle - Plants Type Dataset</a>	Image directory	937 MB	Public (Manual Download)