Data Collection and Preprocessing Phase

Date	20 July 2024
Team ID	SWTID1720163161
Project Title	Hydration Essentials: Classifying Water Bottle Images
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 machine learning project aims to develop a robust image classification model to identify different types of water bottles according to the level of water in it (Half, Full or Overflow). The main goal is to develop a model that can accurately classify water bottles into categories based on their water levels (half, full, overflow). This involves utilizing image processing and machine learning techniques to analyze images of water bottles.
Data Collection Plan	Data for this project will be collected from various publicly available datasets, primarily from the Kaggle platform. The datasets contain labeled images of different water bottle levels. These images are classified as half, full and overflow images of water in various bottles.
Raw Data Sources Identified	Kaggle Water bottle classification dataset

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	This dataset contains labeled images of three different types of water levels, namely half, full and overflowing bottles. A total of 486 images are present in the dataset.	https://www.kaggle.com/datasets/chethuhn/water-bottle-dataset	Image	67 MB	Public