

Data Collection and Preprocessing Phase

Date	16th June 2025
Team ID	SWTID1749740962
Project Title	Dog Breed Identification Using Transfer 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	Brief overview of the machine learning project and its objectives.
Data Collection Plan	Mention from which sources the data are going to be collected.
Raw Data Sources Identified	List the raw data sources with relevant details (as a short description).

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Train Dataset	The training data for the Kaggle Dog Breed Identification competition consists of labeled images of dogs belonging to various breeds, serving as the primary dataset for model training	https://www.kaggle.com/competitions/dog-breed-identification/dataset?select=train	ZIP(images)	10.2k	Public
Test Dataset	The test data for the Kaggle Dog Breed Identification competition comprises unlabeled images of dogs, used for evaluating the trained models' performance and generating predictions for breed identification.	https://www.kaggle.com/competitions/dog-breed-identification/dataset?select=train	ZIP(images)	10.4k	Public

Labels	The labels data for the Kaggle Dog Breed Identification competition provides breed annotations for the training images	https://www.kaggle.com/competitions/dog-breed-identification/data?select=train	.CSV	482KB	Public
Sample Submission	The sample submission data for the Kaggle Dog Breed Identification competition outlines the required format for result submissions	https://www.kaggle.com/competitions/dog-breed-identification/data?select=train	.CSV	25.2MB	Public