

## Data Collection and Preprocessing Phase

Date	26 June 2024
Team ID	SWTID1720073336
Project Title	Dog Breed Identification Using Transfer Learning
Maximum Marks	2 Marks

### Data Collection Plan & Raw Data Sources Identification Report:

For an effective data collection plan in a dog breed identification project, leveraging diverse and reputable sources is key. Utilizing online image databases such as Kaggle, ImageNet, and Google Images provides a vast array of labeled dog breed images. Identifying raw data sources is crucial for maintaining data integrity and diversity. Publicly available datasets, including the Stanford Dogs Dataset and the Oxford-IIIT Pet Dataset, offer well-curated images and annotations for multiple dog breeds. Combining these sources ensures a comprehensive, high-quality dataset essential for reliable dog breed identification.

### Data Collection Plan:

Section	Description
Project Overview	The primary objective of this project is to create a machine learning-based system for accurate and efficient dog breed identification. By automating the breed identification process with transfer learning, we aim to enhance user satisfaction and trust through quick and reliable results. Key features include advanced image recognition and real-time processing.
Data Collection Plan	<ul style="list-style-type: none"> <li>● Search for datasets related dog breeds which are present globally .</li> <li>● Prioritize datasets with diverse demographic information.</li> </ul>
Raw Data Sources Identified	The raw data sources for this project include datasets obtained from Kaggle & UCI, the popular platforms for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing variables such as breed

	height, color, and region details for machine learning analysis.
--	--

### Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset comprises applicant details of the dog like (breed name, height, region, color, etc.)	<a href="https://www.kaggle.com/datasets/jessicali9530/stanford-dogs-dataset">https://www.kaggle.com/datasets/jessicali9530/stanford-dogs-dataset</a>	CSV	787 MB	Public