

Data Collection and Preprocessing Phase

Date	11 July 2024
Team ID	SWTID1720067113
Project Title	Dog Breed Identification using Transfer Learning
Maximum Marks	2 Marks

Data Quality Report Template

The Data Quality Report Template will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	Missing Labels	Moderate	Verify the integrity of the dataset by checking for any missing or null values in the labels file. Use pandas to identify and handle missing labels by either removing the corresponding images or imputing the missing values if applicable.
Kaggle Dataset	Variability in Image Sizes	Low	Resize all images to a consistent size (224x224) using the 'ImageDataGenerator' to ensure uniform input dimensions for the neural network.
Kaggle Dataset	Variability in Image Quality	Moderate	Apply image preprocessing techniques such as normalization

			to scale pixel values and possibly denoising filters to improve image quality.
Kaggle Dataset	High Dimensionality	High	Use feature extraction from pre-trained models (e.g., VGG19, EfficientNetB7, MobileNetV2) to reduce the dimensionality of the images and focus on the most informative features.
Kaggle Dataset	Imbalanced Dataset	Least	Use data augmentation techniques such as rotation, shifting, zooming, and flipping to artificially increase the number of images for underrepresented classes.