**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 20 June 2025 |
| Team ID | SWTID1749826875 |
| Project Title | |  | | --- | | Dog Breed Identification using Transfer Learning | |  | |
| Maximum Marks | 2 Marks |

**Data Collection Plan & Raw Data Sources Identification**

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 | Dog Breed Identification using Transfer Learning" aims to develop a robust machine learning model for accurately classifying dog breeds from images. The project leverages transfer learning, a technique that utilizes pre-trained deep learning models as feature extractors, to overcome the challenges of limited training data and computational resources. By fine-tuning a pre-trained convolutional neural network (CNN) on a dataset of dog images, the model learns to distinguish between different breeds with high accuracy. The resulting system provides a valuable tool for dog breed recognition in various applications, including pet care, veterinary medicine, and animal welfare. |
| Data Collection Plan | The datasets are acquired from Kaggle |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| Kaggle Dataset 1 | The training dataset consists of folders corresponding to each class containing image data. | https://www.kaggle.com/competitions/dog-breed-identification/ | Image | 344 MB | Public |
| Label | Contains the label for all the images in the dataset | https://www.kaggle.com/competitions/dog-breed-identification/ | CSV | 471 KB | Public |
| Kaggle Dataset 2 | The testing dataset consists of folders corresponding to each class containing image data. | https://www.kaggle.com/datasets/gpiosenka/70-dog-breedsimage-data-set | Image | 344 MB | Public |
| Dogs | Contains the label for all the images in the dataset | https://www.kaggle.com/datasets/gpiosenka/70-dog-breedsimage-data-set | CSV | 471 KB | Public |