**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 15 April 2024 |
| Team ID | Team-738205 |
| Project Title | [Dog Breed Identification Using Transfer Learning](https://dinkarvidya27.atlassian.net/browse/DBIUTL) |
| 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 identify dog breeds using transfer learning. Objectives include leveraging pre-trained models, addressing data challenges, and achieving accurate breed classification for diverse images. |
| Data Collection Plan | We have collected data from Kaggle's Dog Breed Identification it have a large collection of dog images categorized into different breeds. |
| Raw Data Sources Identified | Raw data sources include Kaggle's Dog Breed Identification Challenge, providing benchmark images; online repositories like Unsplash, Flickr, and Pixabay for diverse breed representations; breed-specific websites for quality images; and crowdsourced platforms like Dog CEO's Dog API for collaborative data collection. |

**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/data?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/data?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 | 482.06 KB | 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 | 25.2 MB | Public |