**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 15 March 2024 |
| Team ID | SWTID1720110358 |
| Project Title | Rice Classification using CNN |
| 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 | A machine learning project utilizing Convolutional Neural Networks (CNN) to classify different types of rice grains accurately, aiming to improve efficiency and accuracy in rice quality assessment and sorting processes |
| Data Collection Plan | Kaggle  https://www.muratkoklu.com/datasets/ |
| Raw Data Sources Identified | Description: This dataset contains labeled images of five different types of rice grains.  Types of Rice: Arborio, Basmati, Jasmine, Ipsala, and Karacadag.  Content: High-resolution images, each labeled with the corresponding type of rice.  Usage: Ideal for training and testing Convolutional Neural Networks (CNN) for rice classification tasks. |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| Dataset 1 | This dataset has 75K images including 15K pieces from each rice variety. | https://www.muratkoklu.com/datasets/ | Image | 0.24 GB | Public |
| Dataset 2 | This dataset has 12 morphological, 4 shape and 90 color features. | https://www.muratkoklu.com/datasets/ | Image | 0.24 GB | Public |