**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 12 July 2024 |
| Team ID | SWTID1720157891 |
| 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 | The "Rice Type Classification Using CNN" project aims to develop a machine learning model to accurately classify different types of rice from images using convolutional neural networks (CNNs) and transfer learning. This project aims to automate rice type identification for quality control and distribution in the agricultural industry. |
| Data Collection Plan | The data will be collected from various public datasets, agricultural research institutions, and online image repositories. Steps include identifying data sources, downloading or scraping images, reaching out to institutions for specialized datasets, ensuring accurate labeling, and preprocessing images for consistency. |
| Raw Data Sources Identified | **Kaggle: Rice Image Dataset**   * A comprehensive collection of images of various rice types. * Source: Kaggle |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| Dataset 1 | Kaggle Dataset | https://www.kaggle.com/datasets/muratkokludataset/rice-image-dataset | CSV | 230MB | Public |