**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 15 March 2024 |
| Team ID | SWTID1720110358 |
| Project Title | Rice type classification using CNN |
| 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** |
| Rice Dataset Kaggle 1 | Data not split into train, validation and test directories. | High | Using os and shutil libraries. The os library will be used to create new directories for train, validation and test. The shutil library will be used to copy the files from the source folder to their destination train, validation and test folders. |
| Rice Dataset Kaggle 2 | Image dimensions are too large. | Moderate | Images will have to be resized to a dimension of 224x224. |