**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 28 November 2024 |
| Team ID | 739996 |
| Project Title | Deep Fruit Veg: Automated Fruit And Veg Identification |
| 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** |
| Dataset | Duplicate images or image-label pairs. | Moderate | Implement a duplicate detection algorithm (e.g., using hash functions or image comparison techniques). Remove duplicates from the dataset to prevent bias. |
| Dataset | Incorrect image dimensions or resolution variations. | Low | Resize all images to a consistent size (e.g., 200x260 pixels) during preprocessing to maintain uniformity. |
| Dataset | Inconsistent image formats (JPEG, PNG, TIFF). | Low | Convert all images to a consistent format (e.g., JPEG) using image processing tools like OpenCV or PIL. |