**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 20 August 2025 |
| Team ID | Sneha S |
| Project Title | mushroom |
| 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 | |  |  |  | | --- | --- | --- | | **Class Imbalance** – Slightly fewer poisonous images than edible ones. | Moderate | Applied real-time **data augmentation** (rotation, flip, zoom) on the minority class to balance training batches. |  |  | | --- | | folders. | | Moderate | Applied real-time **data augmentation** (rotation, flip, zoom) on the minority class to balance training batches. |
|  | **Duplicate Images** – Some images appear in both train and validation folders. | Moderate | Used a hash-based script to detect and remove duplicates before training. |