 

Data Collection and Preprocessing Phase

|  |  |
| --- | --- |
| Date | 21/06/25 |
| Team ID | SWTID1750180744 |
| Project Title | Smart Sorting: Transfer Learning for  Identifying Rotten Fruits and Vegetables |
| 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 | The model may perform poorly in real-world conditions | Low | * Enhance training data variability using more aggressive and realistic augmentations * Supplement the training dataset with actual field images * Introduce synthetic backgrounds or noise to simulate different environments |
| Dataset | The model may overfit to background cues and struggle when applied to environments where fruits are in baskets, stores, or natural settings. | High | * Use Strong Data Augmentation on Background * Use Class Activation Mapping (CAM) to Inspect Bias * Train With Background Clutter Dataset |
| Dataset | The model could fail to detect defects it hasn’t seen before, making it unreliable when generalized to other food items or regions. | High | * Expand the Dataset with Diverse Defects and Fruits * Use Transfer Learning + Fine- Tuning |
| Dataset | Labels are derived from folder names, and there’s no annotation for severity of rot or affected region. | Low | * Introduce Multi-Class Severity Labels * Add Bounding Box or Segmentation Masks * Crowdsourcing or Expert Labeling for Severity |