

Data Collection and Preprocessing Phase

| | |
|---------------|---------------------------------------------------|
| Date | 14 December 2024 |
| Team ID | 740062 |
| Project Title | Plant seedling classification using Deep learning |
| 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 | Resizing | High | Use the target_size parameter in the flow_from_directory method to resize all images to a consistent size (e.g., (128, 128)). |
| Dataset | Training the dataset | Moderate | Use techniques like flipping, rotation, scaling, and cropping to artificially expand the dataset and address class imbalance. |
| Dataset | Data augmentation overfitting | Low | Dropout, Early Stopping |

| | | | |
|---------|--------------------------|-----|------------------------------------------------------------------------------------|
| Dataset | Inadequate Preprocessing | Low | This is handled with the rescale=1./255 parameter in the ImageDataGenerator. |
|---------|--------------------------|-----|------------------------------------------------------------------------------------|