**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 03 February 2025 |
| Team ID | xxxxxx |
| Project Title | Implementing AgriPrediction using Machine 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** |
| Agricultural Data | Missing values (soil pH, rainfall, temperature) | High | Mean/mode imputation, forward/backward fill for time-series. |
| Agricultural Data | Duplicate records | Moderate | Remove using df.drop\_duplicates(). |
| Weather Data | Inconsistent units (temperature, rainfall) | High | Standardize to Celsius and mm. |
| Soil Data | Outliers in pH, nutrients | High | Detect via Z-score/IQR, cap extreme values. |