**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 15 March 2024 |
| Team ID | LTVIP2024TMID24955 |
| Project Title | SMS Spam Detection - AIML |
| 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 | Mention the issues faced in the selected dataset. | Low/ Moderate/ High | Give the solution for that issue technically. |
| Dataset | Duplicate Records | Moderate | Implement deduplication processes to identify and remove duplicate entries from the dataset to ensure unique records. |
| Dataset | Inconsistent Formatting | High | Standardize formats for text fields (e.g., phone numbers, dates) to ensure uniformity across the dataset. |
| Dataset | Irrelevant Data | Low | Review the dataset to identify and remove unnecessary columns or records that do not contribute to the analysis. |
| Dataset | Missing Values | High | Identify and fill in missing values using imputation methods (mean, median, mode) or remove affected records if the missing percentage is significant. |