**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 15 june 2024 |
| Team ID | 740003 |
| Project Title | Disease prediction using Machine learning |
| Maximum Marks | 2 Marks |

**Data Quality Report Template**

The Data Quality Report 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** |
| Kaggle Dataset | Missing values in the  ‘fluid\_overload’,  ‘Unnamed: 133’ columns | High | Use drop columns |
| Kaggle dataset | Categorical data in the dataset | moderate | Encoding has to be done in the data |