 

Data Collection and Preprocessing Phase

|  |  |
| --- | --- |
| Date | 12 July 2024 |
| Team ID | 740022 |
| Project Title | Lymphography Classification Using ML |
| Maximum Marks | 2 Marks |

# Data Quality Report:

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 Quality Report:

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Source** | **Data Quality Issue** | **Severity** | **Resolution Plan** |
| kaggle Dataset | Outliers in the 'bl.of lymph.c', ‘bl.of lymph.s’,’by pass’, ‘regenaration of’, ‘lymph.nodes dimin’, ‘exclusion of no’ and ‘no. of nodes in ’ columns. | Moderate | IQR (Interquartile Range) |
| Kaggle Dataset | Categorical data in the dataset | Moderate | Encoding has to be done in the data. |