 

DataCollectionandPreprocessingPhase

|  |  |
| --- | --- |
| Date | 12 July 2024 |
| TeamID | 740686 |
| ProjectTitle | Lymphography Classification Using ML |
| Maximum Marks | 2 Marks |

# Data Quality Report:

The Data Quality Report will summarize data quality issues from the selected source, includingseverity levels and resolution plans. It will aid in systematically identifying and rectifying datadiscrepancies.

# Data Quality Report:

|  |  |  |  |
| --- | --- | --- | --- |
| **DataSource** | **Data Quality Issue** | **Severity** | **Resolution Plan** |
| kaggleDataset | 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) |
| KaggleDataset | Categorical data in the dataset | Moderate | Encoding has to be done inthe data. |