**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 17th June 2025 |
| Team ID | LTVIP2025TMID44055 |
| Project Title | Revolutionizing Liver Care : Predicting Liver  Cirrhosis Using Advanced Machine Learning  Techniques |
| 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 Source** |  | **Data Quality Issue** | **Severity** | **Resolution Plan** |
| Kaggle | • | Too many NULL values | Moderate | Filling the Numerical Columns with Mean and removing abnormalities from categorical column. |
|  |  |  |  | Changing the data type by type casting.  Ex- |
| Kaggle | • | Wrong Data Type | Moderate |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Kaggle | • | Ambiguous string entries in multiple column | Low | Dropping the below rows |
| Kaggle | • | Highly imbalanced outcome | Moderate | * Synthetically generating 300 rows with minority class            * Checking the value counts.        * Filling the null values with   MODE. |
| Kaggle | • | Many unimportant Features | Moderate | • Removing the features by analyzing the importance scores. |
|  |  | |  | • Dropping all the unnecessary columns.    • |