

Data Collection and Preprocessing Phase

| | |
|---------------|--|
| Date | 15 july 2024 |
| Team ID | 739952 |
| Project Title | Prediction and Analysis of Liver Patient Data Using Machine Learning |
| 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 |
|--------------------|---|-----------------|--|
| Kaggle Dataset | Missing Values in Critical Column 'Albumin_and_Globulin_Ratio' | High | Impute missing values using appropriate methods (mean, median, or KNN imputation). |

| | | | |
|----------------|--|--------|--|
| Kaggle Dataset | Duplicate Records | Medium | Identify and remove duplicate records to ensure data uniqueness. |
| Kaggle Dataset | Potential Outliers in Medical Test Results | Medium | Apply statistical methods to detect and handle outliers. |