**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 08 july 2024 |
| Team ID | SWTID1720193784 |
| Project Title | Early Prediction Of Chronic Kidney Disease 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** |
| Dataset | |  | | --- | | Missing values in ‘Red Blood  Cells’ feature. |  |  | | --- | |  | | High | |  | | --- | | Impute missing values using  mode (most frequent value)  since it's a categorical feature. |  |  | | --- | |  | |
| Dataset | Outliers in ‘Blood Pressure’ feature. | Moderate | |  | | --- | | Cap outliers at the 1st and 99th  percentiles to reduce their  impact. |  |  | | --- | |  | |
| Dataset | Inconsistent data in ‘Age’ feature. | Low | Validate age values to ensure they fall within a realistic range (e.g., 0-120 years) and correct any discrepancies. |
| Dataset | Incorrect data type for ‘Serum Creatinine’ feature. | Moderate | Convert ‘Serum Creatinine’ to a numerical data type to facilitate proper analysis. |