**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 12 JULY 2024 |
| Team ID | SWTID1720438678 |
| Project Title | Early prediction of chronic kidney disease. |
| Maximum Marks | 2 Marks |

**Data Quality Report**

The dataset exhibits significant quality issues impacting its reliability. Missing data is prevalent across multiple variables, hindering comprehensive analysis. Inconsistent formatting complicates data cleaning and preprocessing efforts. Data inconsistencies, such as variations in date formats and numerical representations, further reduce data integrity. These challenges necessitate robust data cleaning and imputation techniques to ensure data quality and validity for subsequent modeling. Addressing these issues is crucial to derive meaningful insights from the dataset.

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Source** | **Data Quality Issue** | **Severity** | **Resolution Plan** |
| Dataset | Missing values | Moderate | Imputation , Deletion |
| Dataset | Inconsistant formating | Low | Standardization , Data transformation |