**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 08 July 2024 |
| Team ID | SWTID1720174640 |
| Project Title | Early Prediction of Chronic Kidney Disease |
| Maximum Marks | 6 Marks |

**Data Exploration and Preprocessing Template**

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

|  |  |
| --- | --- |
| **Section** | **Description** |
| Data Overview | The dataset `chronickidneydisease.csv` has 400 rows and 26 columns, including both numerical and categorical data. It contains various patient attributes like age, blood pressure, specific gravity, albumin, sugar, red blood cells, hemoglobin, and more, along with a classification label indicating chronic kidney disease status. There are missing values in several columns. Basic statistics show that attributes like age, blood pressure, and others have varying ranges and distributions. |
| Univariate Analysis |  |
| Bivariate Analysis |  |
| Multivariate Analysis |  |
| **Data Preprocessing Code Screenshots** | |
| Loading Data |  |
| Handling Missing Data |  |
| Data Transformation |  |
| Feature Engineering |  |
| Save Processed Data |  |