Data Collection and Preprocessing Phase

Date	5 July 2024
Team ID	SWTID1720082525
Project Title	Early Prediction of Chronic Kidney Disease Using Machine Learning
Maximum Marks	2 Marks

Data Quality Report Template

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
Skill Wallet Dataset	Categorical data in the dataset	Moderate	Encoding must be done in the data.
Skill Wallet Dataset	Missing values in the 'red blood', 'pus cell', 'pus cell clumps bacteria', 'hypertension', 'diabetes mellitus', 'coronary artery disease', 'appetite', 'pedal edema' and 'anemia' columns.	High	Using Random Value Imputation and Mode Imputation.

Skill Wallet Dataset	Incorrect data types	Low	Convert columns to appropriate data types using pd.to_numeric() function.