

Data Collection and Preprocessing Phase

Date	08 July 2024
Team ID	SWTID1720190579
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 ' Pus Cell ' feature.	Moderate	Impute missing values using mode (most frequent value) since it's a categorical feature.
Dataset	Outliers in ' serum ' feature.	Moderate	Use log transformation to reduce the effect of outliers.
Dataset	Duplicate entries in the dataset.	Low	Remove duplicate entries to Ensure data uniqueness and accuracy.
Dataset	Incorrect entries in ' Pus Cell ' feature.	High	Impute missing values using the most frequent value or a predictive model.
Dataset	Incorrect entries in ' Blood Glucose Random ' feature.	Moderate	Correct data entries that fall Outside the physiological rage

			by cross verifying with other related features.
Dataset	Missing values in ' Blood Urea ' feature.	High	Impute missing values using mean/median or a predictive model.