Data Collection and Preprocessing Phase

Date	10 July 2024
Team ID	SWTID1721205662
Project Title	Early Prediction of Chronic Kidney Disease Using Machine Learning
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

Data Collection Plan Template

Section	Description
Project Overview	The machine learning project aims to predict the presence of chronic kidney disease based on patient information. Using a dataset with features such as age, blood pressure, specific gravity, albumin, sugar, and various blood test results, the objective is to build a model that accurately classifies the disease status, facilitating early diagnosis and treatment.
Data Collection Plan	• The data is sourced from medical records containing information relevant to chronic kidney disease. • The dataset includes various patient attributes and test results necessary for the prediction model. • Prioritize datasets with diverse demographic and clinical information to ensure model robustness.

Raw Data Sources Identified	The raw data sources for this project include datasets obtained from Kaggle, the popular platform for data science competitions. The provided sample data represents a subset of the collected information, encompassing variables such as Age, BP, Urea and other Kidney Health Related factors.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset comprises patient details (Age, BP, Urea, Sugar Level etc.) and Chronic Kidney Disease Outcomes.	https://drive.google.com/file/d/1mPl4yaTKuKZ3017YfYC19Ni7Y964eCNI/view?usp=sharing	CSV	42.5 KB	Public