Data Collection and Preprocessing Phase

Date	08 July 2024
Team ID	SWTID1720174640
Project Title	Early Prediction of Chronic Kidney Disease
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	Here in this project we have used logistic regression for a classification task in which there are 2 clases one is ckd and other is not-ckd. here we have first performed data preprocessing in which we handle missing data and perform EDA . later we have do the spliting of data and then tranied our model and then evaluated model based on metrics like MSE,MAE,R Squared ect in this results we have achieved accuracy of 0.98 to 1 when we run this file mulitple times.
Data Collection Plan	We have used kaggle inorder to find out the dataset.

Raw Data Sources Identified	SmartInternz, Kaggle, UCI Machine Learning Repository.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle	The chronic kidney disease dataset comprises multivariate data collected over a two-month period from a hospital. This dataset is designed for classification tasks, enabling the prediction of chronic kidney disease in patients. It includes 400 instances and 24 real-valued features, reflecting a wide range of medical	https://www.kaggle.com/datasets/mansoordaku/ckdisease	CSV	48.55 kB	Public

	and demographic information. The diverse feature set provides a robust basis for building predictive models, aiding in early detection and diagnosis. Researchers and practitioners can leverage this dataset to develop and validate machine learning algorithms, ultimately contributing to improved patient outcomes in the field of nephrology.				