Data Collection and Preprocessing Phase

Date	11 June 2024
Team ID	SWTID1749620488
Project Title	Chronic Kidney Disease Detection 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	This project aims to build a machine learning model that detects Chronic Kidney Disease (CKD) using patient health data such as blood pressure, sugar levels, hemoglobin, and other biomarkers. The objective is early and accurate prediction to support timely medical intervention.
Data Collection Plan	The dataset was collected from publicly available medical repositories (such as UCI Machine Learning Repository or Kaggle). It contains clinical and laboratory records of CKD patients.
Raw Data Sources Identified	One main dataset is used in this project, containing 400+ patient records with both categorical and continuous features.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	Clinical records of patients with/without Chronic Kidney Disease. Contains 25 features including demographic and medical data.	https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease	CSV	~15 KB	Public