Data Collection and Preprocessing Phase

Date	24 June 2025
Team ID	SWTID1749634408
Project Title	Early Prediction for Chronic Kidney Disease Detection: A Progressive Approach to Health Management
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Report:

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:

Section	Description
Project Overview	This machine learning project aims to predict the early onset of Chronic Kidney Disease (CKD) based on patient information such as age, blood pressure, glucose levels, and medical history. The goal is to develop a model that identifies high-risk individuals at an early stage, facilitating timely diagnosis and preventive healthcare.
Data Collection Plan	● Search for datasets related to CKD, patient health records, and medical diagnostics. ● Prioritize datasets containing diverse demographic and clinical data, including lab results and medical history. ● Ensure inclusion of labeled data for supervised learning (e.g., CKD vs. non-CKD).
Raw Data Sources Identified	Dataset includes 400 patient records with features such as age, blood pressure, specific gravity, albumin, sugar levels, red blood cell count, and classification of CKD stages.

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset comprises applicant details (age, blood pressure, specific gravity, albumin, sugar levels, red blood cell count), and classification of CKD stages.	https://www.kaggle.com/datasets/mansoordaku/ckdisease	CSV	49 kB	Public