Data Collection and Preprocessing Phase

Date	06 July 2024
Team ID	SWTID1720082372
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	This project aims to develop a machine learning system for early detection of chronic kidney disease (CKD) risk in a large urban hospital. By analyzing patient data from electronic health records, lab results, and demographics, the system will enable faster and more accurate identification of at-risk individuals. This approach allows for timely preventive measures, potentially improving patient outcomes and quality of life. The project involves data collection, model development, and integration with existing systems. Challenges include ensuring data privacy, obtaining regulatory approvals, and clinical validation. The goal is a reliable system for earlier interventions and reduced healthcare costs.
Data Collection Plan	We will compile comprehensive data on current diagnoses (such as diabetes and hypertension), prescribed medications, and hospital admission records to pinpoint potential risk factors. Serum analyses (creatinine concentrations) and urinalyses (protein-to-creatinine ratios) will supply vital biomarkers for the early identification of chronic kidney disease.

Raw Data Sources Identified	Information will be extracted from Skillwallet's Chronic Kidney Disease Repository. This dataset encompasses demographic details, clinical histories, laboratory findings, and lifestyle indicators, all of which will be utilized for machine learning-based prognostication of chronic kidney disease.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Skill wallet	The dataset comprises patient blood related details (blood glucose, albumin, wbc, haemoglobin), urine details (urea, albumin), and age.	https://drive.google.com/file/d/1mPl4yaTKuKZ3017YfYC19Ni7Y964eCNI/view?usp=sharing	CSV	48 KB	Public