Data Collection and Preprocessing Phase

Date	6th July 2024
Team ID	SWTID1720090815
Project Title	Early Prediction Of Chronic Kidney Disease Using Machine Learning
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

A robust and comprehensive dataset is the cornerstone of any successful machine learning project. For Chronic Kidney Disease, data should ideally encompass a wide range of patient demographics, clinical parameters, and outcomes.

Data Collection Plan Template

Section	Description
Project Overview	**Project Overview:** Develop a Python-based machine learning model to predict Chronic Kidney Disease (CKD) progression. **Objective:** Utilize patient data including clinical parameters, demographics, and medical history to build a predictive model that can assist in early detection and management of CKD, potentially improving patient outcomes.
Data Collection Plan	The data is collected from the verified source - UC Irivine machine learning repository
Raw Data Sources Identified	UC Irvine ML repository – Most suitable dataset needed for the project

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
UC Irivine machine learning repository	This dataset can be used to predict the chronic kidney disease and it can be collected from the hospital nearly 2 months of period.	https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease	CSV	49 KB	Public