Data Collection and Preprocessing Phase

Date	16 JULY 2024
Team ID	SWTID1720075199
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	Chronic Kidney Disease (CKD) is a global health concern affecting millions worldwide. Early detection and intervention are critical to prevent complications and improve patient outcomes. This project aims to develop a predictive model using machine learning techniques to assist in the early diagnosis of CKD based on patient clinical and laboratory data. **Objective:** The primary objective of this project is to build and validate a machine learning model that can accurately predict the presence or absence of CKD in patients based on their demographic information, medical history, and clinical test results. By leveraging predictive analytics, the goal is to enhance healthcare decision-

	making, facilitate timely intervention, and improve patient management strategies.
Data Collection Plan	Search for datasets related to kidney disease and relevant medical details.
Raw Data Sources Identified	The raw data sources identified for the project includes dataset obtained for Kaggle , a popular data science competition and repository. Provided data set represents a subset of collected information like age, blood sugar etc/

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle dataset	The dataset contains relevant details like age,blood sugar,red blood cells count etc	https://www.kaggle.com/datasets/mansoordaku/ckdisease	CSV	42.5 KB	Public