

Data Collection and Preprocessing Phase

Date	12 July 2024
Team ID	SWTID1720083491
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	The machine learning project aims to predict whether the patient has chronic kidney disease or not earlier. Using the dataset with features such as blood pressure, hemoglobin, red blood cells count, white blood cells count, sugar, and other related details, the objective is to build the model that accurately predicts whether the patient has chronic kidney disease or not.
Data Collection Plan	<ul style="list-style-type: none"> ● Search for datasets related to chronic kidney disease with patient details. ● Prioritize datasets with diverse demographic information.
Raw Data Sources Identified	The raw data sources for this project include datasets obtained from Skill Wallet platform - Google Drive the popular cloud storage and

	file sharing platform of Google. The provided sample data represents a subset of the collected information, encompassing variables such as blood pressure, sugar, hemoglobin, red blood cells, white blood cells and other chronic kidney disease related details for machine learning analysis.
--	--

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Skill Wallet Platform(Google Drive) Dataset	The dataset comprises patient details(blood pressure, sugar, albumin, hemoglobin, red blood cells count, white blood cells count, other related details) and classification outcomes.	https://drive.google.com/file/d/1mP14yaTKuKZ3017YfYC19Ni7Y964eCNI/view?usp=sharing	CSV	48 kB	Public