

## Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	xxxxxx
Project Title	Chronic Kidney Disease
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 classification algorithm for Chronic Kidney disease analysis within medical organizations , analyzing important factors like hemoglobin levels, red blood cell count, which allows fore an quick and easy diagnosis
Data Collection Plan	Extract medical records from hospital databases which contains Blood cell records , hemoglobin levels , whether the patient suffers from pedal edema or not
Raw Data Sources Identified	The raw data source for this project is obtained from Kaggle, a popular platform for data science competitions and repositories, The dataset is accessible at <a href="#">Kaggle CKD analysis</a> , it has varialbes such as age , hemoglobin levels, whilte blood cell count which are critical for prediction of Chronic Kidney Disease in patients

### Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	Description of the data in this source.	<a href="#">Chronic Kidney Disease dataset (kaggle.com)</a>	CSV	10 kb	Public