Data Collection and Preprocessing Phase

Date	4th June 2024
Team ID	SWTID1720164961
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	To minimize the impact of Chronic Kidney Disease (CKD), it is essential to employ a machine learning model for early detection, which can identify the disease at its earliest stages. Early detection facilitates timely medical intervention, significantly slowing disease progression and improving patient outcomes.
Data Collection Plan	1. Gather patient demographics, medical history, and clinical biomarkers from electronic health records and public health databases. 2. Ensure data quality through cleaning, normalization, and feature engineering while maintaining patient privacy and ethical compliance
Raw Data Sources Identified	1. *Electronic Health Records (EHRs): Comprehensive patient information including demographics, medical history, and clinical test results from Kaggle 2. *Laboratory Test Results: Detailed biomarker data such as serum creatinine, eGFR, and urine albumin-to-creatinine ratio

Raw Data Sources Template

Dataset 1	Description of the data in this source.	Link of Dataset 1	format	data	Access permissions
Kaggle dataset	This data comprises age bp RBC count WBC count and other major factors to detect Chronic Kidney Disease	https://www.kaggle.com/code/niteshyadav3103/chronic-kidney-disease-prediction-98-accuracy	csv	10 kb	public