Data Collection and Preprocessing Phase

Date	13 June 2025
Team ID	SWTID1749713922
Project Title	Early prediction for chronic kidney disease detection: A progressive approach to health management
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 focuses on building a machine learning model to detect chronic kidney disease based on structured data. The primary objective is to apply preprocessing, feature selection, and modeling techniques to derive meaningful insights and accurate predictions
Data Collection Plan	The dataset used in this project was obtained from Kaggle, a widely-used platform for data science projects. The dataset offers a realistic context for applying data preprocessing, feature engineering, and modeling techniques.
Raw Data Sources Identified	The dataset is publicly available and includes both categorical and numerical medical features relevant to diagnosing chronic kidney disease.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle	The dataset contains medical diagnostic information, including categorical and numerical variables. It serves as the foundational input for training and evaluating the predictive model.	https://www.kaggle.com/datasets/mansoordaku/ckdisease	CSV	48 KB	Public