**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 08 JULY 2024 |
| Team ID | SWTID1720193784 |
| 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 | It is possible with such approaches to train an ML model and recognize the signs of chronic kidney disease, however, it is necessary to include the doctors into the ML model development to consider the real healthcare setting at the moment of the model building. |
| Data Collection Plan | Identify the necessary features (variables) related to CKD diagnosis (e.g., serum creatinine, blood pressure, urine protein).  Determine the sample size required for reliable model training. |
| Raw Data Sources Identified | The raw data sources for this project include datasets obtained from  Kaggle & UCI. The provided sample data represents a subset of  the collected information, encompassing variables such as age,bp,rbc,pcc,hemo etc.. details for machine learning  analysis. |

**Raw Data Sources Report:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| Kaggle Dataset | **This dataset contains detailed health information for 1,659 patients diagnosed with Chronic Kidney Disease (CKD). The dataset includes demographic details, lifestyle factors, symptoms, quality of life scores, environmental exposures, and health behaviors. Each patient is uniquely identified by a Patient ID, and the data includes a confidential column indicating the doctor in charge.** | https://www.kaggle.com/datasets/rabieelkharoua/chronic-kidney-disease-dataset-analysis | CSV | 1.0 MB | Public |
| Kaggle Dataset | This data contains **medical** **history, clinical measurements, medication usage.** | https://www.kaggle.com/datasets/mansoordaku/ckdisease?resource=download | CSV | 49 KB | Public |