**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 12 JULY 2024 |
| Team ID | SWTID1720438678 |
| Project Title | Early prediction of chronic kidney disease. |
| Maximum Marks | 2 Marks |

**Data Collection Plan & Raw Data Sources Identification**

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 model to predict the likelihood of chronic disease onset. The model will target a specific disease or a group of related diseases and use various data points to assess an individual's risk. | |  | | |  | |
| Data Collection Plan | |  | | --- | | The data for this project will be collected from multiple sources. The primary source will likely be Electronic Health Records (EHR) systems, containing valuable details like demographics, medical history, and lab results. Additionally, depending on the chosen disease, anonymized data from public health databases or wearable devices (with user consent) might also be incorporated. | |
| Raw Data Sources Identified | Csv file (47 kb) from skillwallet platform. |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| Dataset 1 | A csv file that consist of data collected from about 400 people that may or may not have ckd. | Link of Dataset 1 | CSV | 47kb | Public |