**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 09 July 2024 |
| Team ID | SWTID1720023141 |
| Project Title | Prediction and Analysis of Liver Patient  Data 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 | This project aims to enhance liver disease diagnosis and treatment by leveraging machine learning techniques to analyze patient data, providing accurate predictions and personalized treatment plans to improve healthcare outcomes. |
| Data Collection Plan | - Collect data from patient records in collaboration with healthcare institutions.  - Prioritize datasets with diverse demographic information.  - Include data from clinical trials and liver disease research studies. |
| Raw Data Sources Identified | The raw data sources for this project include datasets obtained from Kaggle and UCI, popular platforms for data science competitions and repositories. The provided Indian liver patient data represents a subset of the collected information, encompassing variables such as Age, Gender, Total\_Bilirubin, Direct\_Bilirubin, Alkaline\_Phosphatase, Alamine\_Aminotransferase, Aspartate\_Aminotransferase, Total\_Proteins, Albumin, Albumin\_and\_Globulin\_Ratio, and additional datasets. |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| Kaggle Dataset | The raw data sources for this project include datasets obtained from Kaggle and UCI, popular platforms for data science competitions and repositories. The provided Indian liver patient data represents a subset of the collected information, encompassing variables such as Age, Gender, Total\_Bilirubin, Direct\_Bilirubin, Alkaline\_Phosphatase, Alamine\_Aminotransferase, Aspartate\_Aminotransferase, Total\_Proteins, Albumin, and Albumin\_and\_Globulin\_Ratio. | https://w  ww.kagg  le.com/u  ciml/indi  an-liverpatientrecor | CSV | 24 KB | Public |