# Data Collection and Preprocessing Phase

| 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 for improved healthcare outcomes. |

| Data Collection Plan | • Data will be collected from patient records in collaboration with healthcare institutions. |
| --- | --- |
| | • Prioritize datasets with diverse demographic information. |
| | • Data from clinical trials and liver disease research studies. |

| | |
| --- | --- |
| Date | 5 july 2024 |
| Team ID | 739651 |
| 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**

| Raw Data Sources Identified | The raw data sources for this project include datasets obtained from Kaggle & UCI, the popular platforms for data science competitions and repositories. The provided indian liver patient data represents a subset of the collected information, encompassing the variables such |
|---|---|



| | as Age, Gender, Total_Bilirubin, Direct_Bilirubin, Alkaline_Phosphotase, Alamine_Aminotransferase, Aspartate_Aminotransferase, Total_Protiens, Albumin, Albumin_and_Globulin_Ratio and Dataset |
|---|---|

**Raw Data Sources Template**

| Source Name | Description | Location /URL | Format | Size | Access Permissions |
|---|---|---|---|---|---|
| | | | | | |

| Kaggle Dataset | The dataset contains demographic and clinical data related to liver patients(Age, Gender, Total_Bilirubin, Direct_Bilirubin, Alkaline_Phosphotase, Alamine_Aminotransferase, Aspartate_Aminotransferase, Total_Protiens, Albumin, Albumin_and_Globulin_Rat io and Dataset). | https://w ww.kagg le.com/u ciml/indi an-liverpatientrecords | CSV | 24 KB | Public |
|---|---|---|---|---|---|