

Data Collection and Preprocessing Phase

| | |
|---------------|----------------------------------------------------------------------------------------------------|
| Date | 16th June 2024 |
| Team ID | LTVIP2025TMID33140 |
| Project Title | Revolutionizing Liver Care: Predicting Liver Cirrhosis Using Advanced Machine Learning Techniques. |

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 | The machine learning project aims to predict liver cirrhosis based on patient information. Using a dataset with features such as age, gender, medical history, lab results, and lifestyle factors, the objective is to build a model that accurately predicts the likelihood of liver cirrhosis, facilitating early detection and informed decision-making in healthcare. |
| Data Collection Plan | <ul style="list-style-type: none"> ● Search for datasets related to liver cirrhosis, patient health records, and medical history. ● Prioritize datasets with diverse demographic and medical information. |
| Raw Data Sources Identified | The raw data sources for this project include datasets obtained from Kaggle, the popular platforms for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing variables such as age, gender, medical history, lab results, and lifestyle factors for machine learning analysis. |

Raw Data Sources Template

| Source Name | Description | Location/URL | Format | Size | Access Permissions |
|----------------|-----------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|-------|--------------------|
| Kaggle Dataset | The dataset comprises patient details (age, gender, etc.), medical metrics, and liver cirrhosis outcomes. | https://www.kaggle.com/datasets/bhavani priya222/liver-cirrhosis-prediction | XLSX | 47 kB | Public |