

Data Collection and Preprocessing Phase

Date	10 July 2024
Team ID	SWTID1720092248
Project Title	Revolutionizing Liver Care: Predicting Liver Cirrhosis Using Advanced Machine Learning Techniques
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 machine learning project aims to develop a predictive model for liver cirrhosis. The objective is to utilize patient data, including demographics, medical history, and lab results, to predict the likelihood of liver cirrhosis. The model will help in early diagnosis and improve treatment outcomes.
Data Collection Plan	The data will be collected from various sources, including public healthcare datasets, private hospital records, and online medical repositories. Specific details about each source, including access permissions and data formats, are outlined in the Raw Data Sources Template.
Raw Data Sources Identified	A comprehensive list of raw data sources has been identified, each described with relevant details such as location, format, size, and access permissions

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	Public healthcare dataset containing demographic and medical data of patients.	Healthcare Dataset	CSV	2 GB	Public