

Data Collection and Preprocessing Phase

Date	23 september 2024
Team ID	LTVIP2024TMID25021
Project Title	Prediction and Analysis of liver patient data using machine learning
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Report:

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:

Section	Description
Project Overview	The project focuses on leveraging machine learning techniques to predict liver disease in patients by analyzing clinical and biochemical data, aiming for early diagnosis and improved treatment strategies. Through data preprocessing, feature selection, and model evaluation, it seeks to identify significant predictors of liver health and enhance decision-making in healthcare..
Data Collection Plan	The data collection plan involves sourcing liver patient data from medical institutions, public repositories, or health databases, focusing on key features like age, gender, and biochemical markers (bilirubin, ALT, AST). Data will be ethically gathered, anonymized, and stored securely for analysis using machine learning model.
Raw Data Sources Identified	Raw data sources for predicting and analyzing liver patients include public datasets like the UCI Liver Disorder Dataset and medical records from hospitals and clinics, which provide comprehensive clinical and biochemical information. Additionally, health databases and research studies offer valuable data for training machine learning models.

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	Based on chemical compounds(bilirubin,albumin,protiens,alkaline phosphatase) present in human body and tests like SGOT , SGPT the outcome mentioned whether person is patient ie needs to be diagnosed or not..	https://www.kaggle.com/datasets/uciml/indian-liver-patient-records/data https://www.kaggle.com/datasets/rishikeshkonapure/home-loan-approval?select=loan_sanction_train.csv	CSV	15 kB	Public