

## Data Collection and Preprocessing Phase

Date	16 july 2024
Team ID	SWTID1720074204
Project Title	prediction and analysis of liver patient data using ml
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 leverage machine learning to analyze liver patient data. The goal is to predict the presence or absence of liver disease and potentially identify key factors influencing it.
Data Collection Plan	Data is collected from kaggle
Raw Data Sources Identified	This data set contains 416 liver patient records and 167 non liver patient records collected from North East of Andhra Pradesh, India. The "Dataset" column is a class label used to divide groups into liver patient (liver disease) or not (no disease). This data set contains 441 male patient records and 142 female patient records.

	Any patient whose age exceeded 89 is listed as being of age "90".
--	---

### Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	Description of the data in this source.	<a href="https://www.kaggle.com/datasets/uciml/indian-liver-patient-records">https://www.kaggle.com/datasets/uciml/indian-liver-patient-records</a>	CSV	8KB	Public
Dataset 2	Description of the data in this source.	Link of Dataset 2	Excel	YY GB	Private (with access)
...	...	...	...	...	...