DataCollectionandPreprocessingPhase

Date	June
TeamID	LTVIP2025TMID35140
Project Title	Revolutionizing Liver Care: Predicting Liver CirrhosisUsingAdvancedMachineLearning Techniques.
Maximum Marks	2 Marks

Data Quality Report Template

Thisreportsummarizesthedataqualityissuesidentifiedinthelivercirrhosisdataset,alongwith their severity levels and proposed resolution plans. The goal is to systematically identify and rectify discrepancies to ensure high-quality data for accurate predictions.

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	Missingvaluesinallthecolumns of the dataset. (42 columns) ```df.isnull().sum()``` 0.0s S.NO 0 Age 0 Gender 0 Place(location where the patient lives) 134 Duration of alcohol consumption(years) 0 Quantity of alcohol consumption (quarters/day) 0 Type of alcohol consumed 0 Hepatitis B infection 0 Hepatitis C infection 0 Diabetes Result 0 Blood pressure (mmhg) 0 Obesity 0 Family history of cirrhosis/ hereditary 0 TCH 359 TG 359 LDL 359 HDL 368 Hemoglobin (g/dl) 0 PCV (%) 30 RBC (million cells/microliter) 552 MCV (femtoliters/cell) 9 MCH (picograms/cell) 658 MCHC (grams/deciliter) 672 Total Count 10 Polymorphs (%) 0 ... SGOT/AST (U/L) 0 SGPT/ALT (U/L) 0 USG Abdomen (diffuse liver or not) 0 Outcome 54 dtype: int64	High	Usemean/median imputation.
Kaggle Dataset	Categorical data in the dataset	Moderate	Perform encoding (e.g., Label EncodingorOne-HotEncoding).