Data Collection and Preprocessing Phase

Date	24 June 2025
Team ID	SWTID1749708868
Project Title	Revolutionizing Liver Care : Predicting Liver Cirrhosis Using Advanced Machine Learning Techniques
Maximum Marks	2 Marks

Data Quality Report Template

The Data Quality Report will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Data Source	Data Quality Issue	Severity	Resolution Plan
Dataset	['S.NO', 'Age', 'Gender', 'Place(location where the patient lives)', 'Duration of alcohol consumption(years)', 'Quantity of alcohol consumption (quarters/day)', 'Type of alcohol consumed', 'Hepatitis B infection', 'Hepatitis C infection', 'Diabetes Result', 'Blood pressure (mmhg)', 'Obesity', 'Family history of cirrhosis/ hereditary',	High	Resolved null values by substituting with median or mode.

	'TCH', 'TG', 'LDL', 'HDL', 'Hemoglobin (g/dl)', 'PCV (%)', 'MCV (femtoliters/cell)', 'Total Count', 'Polymorphs (%)', 'Lymphocytes (%)', 'Monocytes (%)', 'Eosinophils (%)', 'Basophils (%)', 'Platelet Count (lakhs/mm)', 'Total Bilirubin (mg/dl)', 'Direct (mg/dl)', 'Indirect (mg/dl)', 'Total Protein (g/dl)', 'Albumin (g/dl)', 'Globulin (g/dl)', 'A/G Ratio', 'AL.Phosphatase (U/L)', 'SGOT/AST (U/L)', 'SGPT/ALT (U/L)', 'USG Abdomen (diffuse liver or not)', 'Predicted Value(Out Come-Patient suffering from liver cirrosis or not)']		
…	Categorical Data	Easy	Converted Categorical data to numeric using appropriate en-coding practices