Data Collection and Preprocessing Phase

Date	09 July 2024
Team ID	SWTID1720013031
Project Title	Prediction and Analysis of Liver Patient Data Using Machine Learning
Maximum Marks	2 Marks

Data Quality Report

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
Kaggle https://www.kaggle.com/datasets/uciml/indian-liver-patient-records	Missing Values in 'Albumin_and_Globulin_Ratio' column of 'indian_liver_patients' data	Low	Used mode to fill the values
Kaggle https://www.kaggle.com/datasets/uciml/indian-liver-patient-records	Categorical data in the dataset 'Gender' column has 'Male' and 'Female'	Moderate	Used LabelEncoder to convert categorical data to numerical data in 'Gender' column