Model Development Phase Template

Date	17 July 2024
Team ID	SWTID1720074204
Project Title	Prediction and Analysis of Liver Patient Data Using Machine Learning
Maximum Marks	5 Marks

Feature Selection Report Template

Feature	Description	Selected (Yes/No)	Reasoning
Age	Age is a crucial feature as liver disease prevalence can vary significantly with age.	Yes	Age is a crucial feature as liver disease prevalence can vary significantly with age.
Total Bilirubin	Total Bilirubin is a key indicator of liver function.	Yes	Elevated levels can indicate liver dysfunction or bile duct obstruction. This feature is typically selected because it directly reflects the liver's ability to process and clear bilirubin, making it highly relevant for predicting liver disease.
Alkaline Phosphatase (ALP)	ALP is an enzyme related to the bile ducts.	Yes	High levels can indicate liver damage or bile duct obstruction. This feature is often selected due to its strong correlation with liver disease. Tree-based models typically assign high importance to ALP, reflecting its diagnostic value.

Albumin	Albumin is a protein produced by the liver, and its levels can indicate liver function.	Yes	Low albumin levels can be a sign of chronic liver disease. This feature is selected because it provides insight into the liver's synthetic capacity, which is crucial for diagnosing liver conditions.
Direct Bilirubin	Direct Bilirubin measures the bilirubin that is processed by the liver.	Yes	Elevated levels can indicate liver dysfunction or bile duct obstruction. This feature is selected due to its strong correlation with liver disease and high importance in tree-based models.
Total Cholesterol	While cholesterol levels can be influenced by liver function, they are not as directly indicative of liver disease as other features.	No	Correlation analysis might show a weaker relationship, and tree-based models might rank it lower in importance.
Aspartate Aminotransferase (AST)	AST is an enzyme found in the liver and other tissues.	Yes	High levels can indicate liver damage. This feature is selected because it is a direct marker of liver cell injury and is highly relevant for predicting liver disease.
Alanine Aminotransferase (ALT)	ALT is another enzyme that is primarily found in the liver.	Yes	Elevated ALT levels are a clear indicator of liver damage. This feature is selected due to its strong association with liver disease and high importance in feature selection techniques.
Total Proteins	Total Proteins measure the total amount of proteins in the blood, including albumin and globulin.	Yes	Low levels can indicate liver disease. This feature is selected because it provides insight into the liver's synthetic function.

Globulin	Globulin is a group of proteins in the blood, including antibodies.	Yes	Abnormal levels can indicate liver disease. This feature is selected due to its relevance in assessing liver function and its importance in tree-based models.
Albumin/Globulin Ratio (A/G Ratio)	The A/G Ratio compares the levels of albumin and globulin.	Yes	An abnormal ratio can indicate liver disease. This feature is selected because it provides additional information about liver function and is often highlighted in feature importance analysis.
Triglycerides	Similar to cholesterol, triglyceride levels can be affected by liver function but are not as directly indicative of liver disease.	No	This feature might show a weaker correlation and lower importance in feature selection techniques.
Body Mass Index (BMI)	BMI can be a risk factor for liver disease, especially non-alcoholic fatty liver disease (NAFLD).	No	However, it may not be as strong a predictor as direct liver function tests. Correlation analysis and feature importance from tree-based models might rank it lower.