Project Initialization and Planning Phase

Date	23 September 2024
Team ID	LTVIP2024TMID25021
Project Title	Prediction and Analysis of liver patient data using machine learning
Maximum Marks	3 Marks

Project Proposal (Proposed Solution) report

The proposal report aims to transform prediction and analysis of liver disease in patients using machine learning, boosting efficiency and accuracy. It tackles system inefficiencies, promising better operations, reduced risks, and happier customers. Key features include a machine learning-based credit model and real-time decision-making.

Project Overview	
Objective	It evaluate the overall quality of the dataset by addressing key aspects such as missing data, outliers, and anomalies, and ensure the data meets the requirements for building a machine learning model for liver disease prediction..
Scope	The scope of the liver disease prediction project includes collecting a diverse and comprehensive dataset, conducting thorough exploratory analysis, and developing machine learning models to predict liver disease outcomes. The project will provide valuable insights into the key factors influencing liver health, support early diagnosis, and help personalize treatment plans, ultimately improving patient care. Challenges related to data quality, privacy, and model accuracy will be managed throughout the process to ensure reliable outcomes.
Problem Statement	
Description	The prediction and analysis of liver patient data using machine learning involves utilizing clinical features and biochemical markers to develop predictive models that identify the presence of liver disease. By applying various algorithms, the approach aims to enhance early diagnosis and improve treatment strategies based on data-driven insights. This process includes data preprocessing, feature selection, model evaluation, and interpretation to inform healthcare professionals effectively..

Impact	The ability to make accurate predictions, identify risk factors, and improve early detection can lead to better patient outcomes, lower healthcare costs, and more effective public health strategies. It also empowers healthcare providers with data-driven tools, enhancing decision-making and patient care.
Proposed Solution	
Approach	This approach involves predicting liver disease using machine learning through data collection, preprocessing, and exploratory analysis to handle missing data, outliers, and select key features. Various models like Logistic Regression, Random Forest, and SVM are trained and evaluated using accuracy, precision, recall, F1-score, and ROC-AUC. The best model is used for predictions, with post-prediction analysis to identify important features using SHAP or LIME. Visualizations and dashboards assist in interpretation, while the model is deployed and monitored for real-time predictions in healthcare. Continuous refinement ensures improved diagnostic accuracy and risk factor identification.
Key Features	Key features for predicting liver disease include **biochemical markers** such as bilirubin levels, alanine transaminase (ALT), aspartate transaminase (AST), alkaline phosphatase (ALP), albumin, and total proteins, along with **demographic information** like age and gender. Additional features may include the **ALT/AST ratio** and other relevant clinical indicators that contribute to assessing liver health. Analyzing these features helps in identifying patterns and risk factors associated with liver disease.

Resource Requirements

Resource Type	**Description**	**Specification/Allocation**
Hardware		
Computing Resources	CPU/GPU specifications, number of cores	T4 GPU
Memory	RAM specifications	8 GB
Storage	Disk space for data, models, and logs	1 TB SSD
Software		

Frameworks	Python frameworks	Flask
Libraries	Additional libraries	scikit-learn, pandas, numpy, matplotlib, seaborn
Development Environment	IDE	Jupyter Notebook, pycharm
Data		
Data	Source, size, format	Kaggle dataset, 614, csv UCI dataset, 690, csv