

Project Initialization and Planning Phase

Date	15 March 2024
Team ID	SWTID1720013031
Project Title	Prediction and Analysis of Liver Patient Data Using Machine Learning
Maximum Marks	3 Marks

Project Proposal (Proposed Solution)

This project proposal outlines a solution to address a specific problem. With a clear objective, defined scope, and a concise problem statement, the proposed solution details the approach, key features, and resource requirements, including hardware, software, and personnel.

Project Overview	
Objective	The primary objective of this project is to develop a machine learning model that can accurately predict the presence of liver disease in patients based on various health indicators.
Scope	The scope of this project includes data collection, preprocessing, model training, and evaluation. The project focuses on the "Indian Liver Patient Records" dataset from Kaggle, encompassing demographic information and biochemical markers. The project aims to build and compare multiple classification models, select the best-performing one, and deploy it for practical use.
Problem Statement	
Description	Liver disease is a significant health issue that can lead to severe complications if not diagnosed and treated early. Traditional diagnostic methods can be invasive and costly, making early detection challenging.
Impact	Solving this problem can lead to early detection of liver disease, potentially saving lives and reducing healthcare costs. It will also enhance the ability of healthcare professionals to make informed decisions based on predictive analytics.
Proposed Solution	

Approach	<p>The proposed solution involves several key steps:</p> <ol style="list-style-type: none"> Data Collection: Gather the "Indian Liver Patient Records" dataset from Kaggle. Data Preprocessing: Clean the data, handle missing values, and encode categorical variables. Model Training: Train multiple machine learning models (SVM, Random Forest, K-Nearest Neighbors, and Logistic Regression) on the preprocessed data. Model Evaluation: Compare the models' performance using metrics such as accuracy, precision, recall, and F1-score. Deployment: Select the best-performing model and deploy it as a web application for real-time liver disease prediction.
Key Features	<ul style="list-style-type: none"> Multiple Algorithms: Utilize various machine learning algorithms for robust predictions. Comprehensive Analysis: Integrate demographic and biochemical markers. User-Friendly Application: Deploy a web application for easy access by healthcare professionals and patients. Reliable Predictions: Implement clear input validation and error handling.

Resource Requirements

Resource Type	Description	Specification/Allocation
Hardware		
Computing Resources	CPU/GPU specifications, number of cores	2 x NVIDIA V100 GPUs
Memory	RAM specifications	32 GB RAM
Storage	Disk space for data, models, and logs	2 TB SSD for data, models, and logs
Software		
Frameworks	Python frameworks	Flask
Libraries	Additional libraries	scikit-learn, pandas, numpy, matplotlib, seaborn, pickle,

		warnings
Development Environment	IDE, version control	Jupyter Notebook, pycharm, Git
Data		
Data	Source, size, format	Kaggle dataset, 23.3 KB, .csv