

# Prediction and Analysis of Liver Patient Data Using Machine Learning

## Milestone 1: Project Initialization and Planning Phase

The "Project Initialization and Planning Phase" marks the project's outset, defining goals, scope, and stakeholders. This crucial phase establishes project parameters, identifies key team members, allocates resources, and outlines a realistic timeline. It also involves risk assessment and mitigation planning. Successful initiation sets the foundation for a well-organized and efficiently executed machine learning project, ensuring clarity, alignment, and proactive measures for potential challenges.

### Activity 1: Define Problem Statement

The problem statement is formally defined as:

‘Given a dataset containing various attributes of 584 Indian patients, use the features available in the dataset and define a supervised classification algorithm which can identify whether a person is suffering from liver disease or not. This data set contains 416 liver patient records and 167 non- liver patient records. The data set was collected from north east of Andhra Pradesh, India. This data set contains 441 male patient records and 142 female patient records. Any patient whose age exceeded 89 is listed as being of age "90".

**Problem Statement Report:** [Click here](#)

### Activity 2: Project Proposal (Proposed Solution)

This seems to be a classic example of supervised learning. We have been provided with a fixed number of features for each data point, and our aim will be to train a variety of Supervised Learning algorithms on this data, so that , when a new data point arises, our best performing classifier can be used to categorize the data point as a positive example or negative. Exact details of the number and types of algorithms used for training is included in the 'Algorithms and Techniques' sub-section of the 'Analysis' part.

**Project Proposal Report:** [Click here](#)

### Activity 3: Initial Project Planning

In problems of disease classification like this one, simply comparing the accuracy, that is, the ratio of correct predictions to total predictions is not enough. This is because depending on the context like severity of disease, sometimes it is more important that an

algorithm does not wrongly predict a disease as a non-disease, while predicting a healthy person as diseased will attract a comparatively less severe penalty.

**Project Planning Report:** [Click here](#)

## **Milestone 2: Data Collection and Preprocessing Phase**

The Data Collection and Preprocessing Phase involves executing a plan to gather relevant Liver Patient data from Kaggle, ensuring data quality through verification and addressing missing values. Preprocessing tasks include cleaning, encoding, and organizing the dataset for subsequent exploratory analysis and machine learning model development.

### **Activity 1: Data Collection Plan, Raw Data Sources Identified, Data Quality Report**

The dataset for "Liver Patient Analysis" is sourced from Kaggle. It includes applicant details and financial metrics. Data quality is ensured through thorough verification, addressing missing values, and maintaining adherence to ethical guidelines, establishing a reliable foundation for predictive modeling.

**Data Collection Report:** [Click here](#)

### **Activity 2: Data Quality Report**

The dataset for "Liver Patient Analysis" is sourced from Kaggle. It includes applicant details and financial metrics. Data quality is ensured through thorough verification, addressing missing values, and maintaining adherence to ethical guidelines, establishing a reliable foundation for predictive modeling.

**Data Quality Report:** [Click here](#)

### **Activity 3: Data Exploration and Preprocessing**

Data Exploration involves analyzing the loan applicant dataset to understand patterns, distributions, and outliers. Preprocessing includes handling missing values, scaling, and encoding categorical variables. These crucial steps enhance data quality, ensuring the reliability and effectiveness of subsequent analyses in the Liver Patient Analysis project.

**Data Exploration and Preprocessing Report:** [Click here](#)

## Milestone 3: Model Development Phase

The Model Development Phase entails crafting a predictive model for loan approval. It encompasses strategic feature selection, evaluating and selecting models (Random Forest, Logistic Regression, KNN, SVC), initiating training with code, and rigorously validating and assessing model performance for informed decision-making in the lending process.

### Activity 1: Feature Selection Report

The Feature Selection Report outlines the rationale behind choosing specific features (e.g., Gender, Age, Total Bilirubin, Direct Bilirubin, etc) for the Liver Patient Analysis model. It evaluates relevance, importance, and impact on predictive accuracy, ensuring the inclusion of key factors influencing the model's ability to discern credible Liver Patient Analysis.

Feature Selection Report: [Click here](#)

### Activity 2: Model Selection Report

The Model Selection Report details the rationale behind choosing Random Forest, Logistic Regression, KNN, and SVM models for Liver Patient prediction. It considers each model's strengths in handling complex relationships, interpretability, adaptability, and overall predictive performance, ensuring an informed choice aligned with project objectives

Model Selection Report: [Click here](#)

### Activity 3: Initial Model Training Code, Model Validation and Evaluation Report

The Initial Model Training Code employs selected algorithms on the Liver Patient dataset, setting the foundation for predictive modeling. The subsequent Model Validation and Evaluation Report rigorously assesses model performance, employing metrics like accuracy and precision to ensure reliability and effectiveness in predicting Liver Patient Disease outcomes.

Model Development Phase Template: [Click here](#)

## Milestone 4: Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase involves refining machine learning models for peak performance. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency

### Activity 1: Hyperparameter Tuning Documentation

The SVC model was selected for its superior performance, exhibiting high accuracy during hyperparameter tuning. Its ability to handle complex relationships, minimize overfitting, and optimize predictive accuracy aligns with project objectives, justifying its selection as the final model

### **Activity 2: Performance Metrics Comparison Report**

The Performance Metrics Comparison Report contrasts the baseline and optimized metrics for various models, specifically highlighting the enhanced performance of the SVC model. This assessment provides a clear understanding of the refined predictive capabilities achieved through hyperparameter tuning.

### **Activity 3: Final Model Selection Justification**

The Final Model Selection Justification articulates the rationale for choosing SVC as the ultimate model. Its exceptional accuracy, ability to handle complexity, and successful hyperparameter tuning align with project objectives, ensuring optimal loan approval predictions.

**Model Optimization and Tuning Phase Report:** [Click here](#)

## **Milestone 5: Project Files Submission and Documentation**

For project file submission in Github , Kindly click the [LINK](#) and refer to the flow.

For the documentation, Kindly refer to the [LINK](#).

## **Milestone 6: Project Demonstration**

In the upcoming module called Project Demonstration, individuals will be required to record a video by sharing their screens. They will need to explain their project and demonstrate its execution during the presentation.

**Source Code :** [Click here](#)

**Project Demo Link :** [Click here](#)