

One Year Life Expectancy post on Thoracic Surgery 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 wellorganized and efficiently executed machine learning project, ensuring clarity, alignment, and proactive measures for potential challenges.

Activity 1: Define Problem Statement

Given patient data Our goal is to predict whether the patient will survive for one year after the surgery or not.

Ref. template: [Click here](#)

Problem Statement Report: [Click here](#)

Activity 2: Project Proposal (Proposed Solution)

The proposal report aims to predict one-year life expectancy after thoracic surgery, which is crucial for patient well-being and clinical decision-making. Machine learning (ML) offers advanced capabilities to analyze complex datasets and provide accurate predictions, aiding in better management and outcomes for patients undergoing thoracic surgery. .

Ref. template: [Click here](#)

Project Proposal Report: [Click here](#)

Activity 3: Initial Project Planning

Initial Project Planning involves outlining key objectives, defining scope, and identifying stakeholders for a loan approval system. It encompasses setting timelines, allocating resources, and determining the overall project strategy. During this phase, the team establishes a clear understanding of the dataset, formulates goals for analysis, and plans the workflow for data processing. Effective initial planning lays the foundation for a systematic and well-executed project, ensuring successful outcomes.

Ref. template: [Click here](#)

Project Planning Report: [Click here](#)

Milestone 2: Data Collection and Preprocessing Phase

The Data Collection and Preprocessing Phase involves executing a plan to gather the patients datasets

1

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 "thoracic surgery patients- Prediction for Life survival" is sourced from datasets. It includes patients details and report. Data quality is ensured through thorough verification, addressing missing values, and maintaining adherence to ethical guidelines, establishing a reliable foundation for predictive modeling.

Ref. template: [Click here](#)

Data Collection Report: [Click here](#)

Activity 2: 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.

Ref. template: [Click here](#)

Data Quality Report: [Click here](#)

Activity 3: Data Exploration and Preprocessing

Data Exploration involves analyzing the patients diagnosis 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 predicting model project.

Ref. template: [Click here](#)

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,

Decision Tree, KNN, XGB), 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., pain, PAD, diagonosis, etc for the training Model. It evaluates relevance, importance, and impact on predictive accuracy, ensuring the inclusion of key factors influencing the model's ability to predict the given dataset .

Ref. template: [Click here](#)

Feature Selection Report: [Click here](#)

2

Activity 2: Model Selection Report

The Model Selection Report details the rationale behind choosing Random Forest, Decision Tree, KNN, and XGB models for loan approval 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.

Ref. template: [Click here](#)

Model Selection Report: [Click here](#)

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

The Initial Model Training Code algorithms on the training dataset, setting the foundation for predictive modeling. The subsequent Model Validation and Evaluation Report rigorously assesses model performance, doorelation metrics like accuracy and precision to ensure reliability and effectiveness in predicting life for one year.

Ref. template: [Click here](#)

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 Gradient Boosting 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 Gradient Boosting 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 Gradient Boosting as the ultimate model. Its exceptional accuracy, ability to handle complexity, and successful hyperparameter tuning align with project objectives, ensuring optimal loan approval predictions.

Ref. template: [Click here](#)

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. [Click here](#)

For the documentation, Kindly refer to the link. [Click here](#)

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.

