Doctors Annual Salary Prediction

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
Problem Statement: Healthcare institutions lack an accurate and reliable method to predict the annual salaries of doctors, which leads to challenges in budgeting, resource allocation, and maintaining competitive and fair compensation packages.

Ref. template: Click Here
Doctors Annual Salary Prediction Statement Report: Click Here

Activity 2: Project Proposal (Proposed Solution)
The healthcare sector faces challenges in managing workforce compensation due to the complexity of factors influencing doctors' salaries. Accurate salary predictions can enhance financial planning and ensure competitive and equitable compensation packages. This project proposes the development of a predictive model to address these needs.

Ref. template: Click Here
Doctors Annual Salary Prediction Project Proposal Report: Click Here

Activity 3: Initial Project Planning

The initial project planning phase sets the groundwork for developing a machine learning model to predict doctors' annual salaries. This phase involves defining objectives, identifying stakeholders, outlining the scope, establishing timelines, and assessing potential risks. Establish the primary objective of creating an accurate salary prediction model and secondary objectives like identifying key salary influencers and generating actionable insights. Create a structured approach to guide the project from initiation through to completion.

Ref. template: Click Here
Doctors Annual Salary Prediction 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 loan

application 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 "Doctors Annual Salary Prediction" is sourced from Kaggle. It includes speciality, satisfied income, annual salary . 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](#)
Doctors Annual Salary Prediction Data Collection Report: [Click Here](#)

Activity 2: Data Quality Report

The dataset for " Doctors Annual Salary Prediction " 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.

Ref. template: [Click Here](#)
Doctors Annual Salary Prediction Data Quality Report: [Click Here](#)

Activity 3: Data Exploration and Preprocessing

Data Exploration involves analyzing doctors annual salary prediction 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 salary prediction.

Ref. template: [Click Here](#)
Doctors Annual Salary Prediction Data Exploration and Preprocessing Report: [Click Here](#)

Milestone 3: Model Development Phase

The Model Development Phase entails crafting a predictive model for salary prediction. It encompasses strategic feature selection, evaluating and selecting models (Linear Regression,

Random Forest, Decision Tree, 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. speciality, satisfaction, income) for the salary prediction model. It evaluates relevance, importance, and impact on predictive accuracy, ensuring the inclusion of key factors influencing the model's ability to discernsalary prediction.

Ref. template: Click Here
Doctors Annual Salary Prediction Feature Selection Report: Click Here
Activity 2: Model Selection Report
The Model Selection Report details the rationale behind choosing Linear Regression, Random Forest, Decision Tree, and XGB models for salary 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
Doctors Annual Salary Prediction 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 salary prediction 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 salary outcomes.

Ref. template: Click Here
Doctors Annual Salary Prediction 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 Decision Tree 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 Decision Tree 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 Decision Tree the ultimate model. Its exceptional accuracy, ability to handle complexity, and successful hyperparameter tuning align with project objectives, ensuring optimal salary predictions.

Ref. template: Click Here
Doctors Annual Salary Prediction 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.