

# Doctors Annual Salary Prediction

## Milestone 1: Project Initialization and Planning Phase

The project initialization and planning phase for predicting doctors' annual salaries using machine learning involves setting clear objectives, defining the scope, and outlining the problem statement. The project will collect and preprocess data, select relevant features, and choose suitable machine learning algorithms. The model will be trained, evaluated, and deployed as a web application or API, with continuous monitoring and updates. Key features include a user-friendly interface, real-time predictions, data visualization, and scalability. The project requires high-performance servers, software like Python and TensorFlow, and a skilled team.

### Activity 1: Define Problem Statement

Problem Statement: " The problem statement for predicting doctors' annual salaries highlights the lack of a reliable method in the healthcare industry to accurately determine salaries. This results in discrepancies and inefficiencies in salary distribution. By developing a machine learning model that considers factors like experience, specialization, and location, the project aims to provide a data-driven solution to predict salaries more accurately, addressing these issues.

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

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

"The proposed solution for predicting doctors' annual salaries involves developing a machine learning model. This model will use data on factors like experience, specialization, and location to make accurate salary predictions. The process includes collecting and preprocessing data, selecting relevant features, and choosing suitable machine learning algorithms. The model will be trained, evaluated, and deployed as a web application or API, with continuous monitoring and updates to ensure accuracy. Key features include a user-friendly interface, real-time predictions, data visualization, and scalability

"

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

### Activity 3: Initial Project Planning

The initial project planning for predicting doctors' annual salaries using machine learning involves setting clear objectives, defining the scope, and outlining the problem statement. The project will collect and preprocess data, select relevant features, and choose suitable machine learning algorithms. The model will be trained, evaluated, and deployed as a web application or API, with continuous monitoring and updates. Key features include a user-friendly interface, real-time predictions, data visualization, and scalability. The project requires high-performance servers, software like Python and TensorFlow, and a skilled team.

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

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

The data collection and preprocessing phase for predicting doctors' annual salaries using machine learning involves gathering data from reliable sources such as salary surveys, healthcare databases, and public records. This data includes factors like years of experience, specialization, geographic location, and education. Preprocessing involves cleaning the data to handle missing values and outliers, normalizing it for consistency, and encoding categorical variables. This phase ensures that the data is accurate and ready for model training, ultimately improving the model's performance and reliability.

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

The data collection plan for predicting doctors' annual salaries using machine learning involves identifying the necessary data, determining how to collect it, and ensuring its quality. Raw data sources include salary surveys, healthcare databases, and public records, which provide information on factors like experience, specialization, and location. The data quality report assesses the completeness, accuracy, and consistency of the collected data, addressing any issues such as missing values or outliers to

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

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

The Data Quality Report for predicting doctors' annual salaries using machine learning assesses the completeness, accuracy, and consistency of the collected data. It identifies and addresses issues such as missing values, outliers, and inconsistencies. This ensures that the data is reliable and suitable for model training, ultimately improving the model's performance and the accuracy of salary predictions.

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

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

The data exploration and preprocessing phase for predicting doctors' annual salaries using machine learning involves examining the collected data to understand its structure and key characteristics. This includes visualizing data distributions, identifying patterns, and detecting anomalies. Preprocessing steps include handling missing values, removing outliers, normalizing numerical features, and encoding categorical variables. These steps ensure the data is clean, consistent, and ready for model training, ultimately enhancing the model's performance and accuracy.

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

## **Milestone 3: Model Development Phase**

The initial model development phase for predicting doctors' annual salaries using machine learning involves setting up the machine learning environment and importing necessary libraries. This phase includes defining the model architecture, selecting appropriate algorithms, and preparing the data for training. The data is split into training and testing sets, and the model is trained on the training data. Initial evaluations are conducted to assess the model's performance, and adjustments are made to improve accuracy. This phase lays the foundation for further optimization and tuning to enhance the model's predictive capabilities.

### **Activity 1: Feature Selection Report**

The Feature Selection Report for predicting doctors' annual salaries using machine learning identifies the most relevant factors influencing salary predictions. This involves analyzing various attributes such as experience, specialization, location, and education to determine their impact on salary. Statistical methods and domain knowledge are used to select features that significantly contribute to the model's accuracy. By focusing on these key features, the model can make more precise and reliable salary predictions.

**Feature Selection Report:** [Click Here](#)

## Activity 2: Model Selection Report

The Model Selection Report for predicting doctors' annual salaries using machine learning involves evaluating various algorithms to determine the best fit for the data. Commonly considered models include Linear Regression, Decision Trees, Random Forests, and Gradient Boosting. Each model's performance is assessed using metrics like Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared. The report highlights the model that provides the highest accuracy and reliability for salary predictions, ensuring the chosen model effectively captures the relationships between the input features and the target variable.

**Model Selection Report:** [Click Here](#)

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

The initial model training code for predicting doctors' annual salaries involves setting up the machine learning environment, importing necessary libraries (like scikit-learn, pandas, and numpy), and defining the model architecture. The training process includes splitting the data into training and testing sets, fitting the model to the training data, and tuning hyperparameters for optimal performance.

The model validation and evaluation report assesses the model's performance using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared. This report highlights the model's accuracy, identifies any potential overfitting or underfitting issues, and suggests improvements. Continuous evaluation ensures the model remains reliable and accurate over time.

**Model Development Phase Template:** [Click Here](#)

## Milestone 4: Model Optimization and Tuning Phase

The initial model optimization and tuning phase for predicting doctors' annual salaries using machine learning involves fine-tuning the chosen model to improve its performance. This includes adjusting hyperparameters, such as learning rate and regularization parameters, to minimize errors and enhance accuracy. Techniques like grid search or random search are used to systematically explore different hyperparameter combinations. Cross-validation is employed to ensure the model generalizes well to unseen data. This phase aims to achieve the best possible predictive performance before deploying the model.

## **Activity 1: Hyperparameter Tuning Documentation**

The initial hyperparameter tuning documentation for predicting doctors' annual salaries using machine learning involves detailing the process of optimizing the model's hyperparameters to improve performance. This includes selecting techniques like grid search or random search to systematically explore different hyperparameter combinations. The documentation outlines the chosen hyperparameters, such as learning rate, regularization parameters, and tree depth, and explains how they were adjusted to minimize errors and enhance model accuracy. It also includes the results of cross-validation to ensure the model generalizes well to unseen data, providing a comprehensive guide to achieving the best predictive performance.

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

The Performance Metrics Comparison Report for predicting doctors' annual salaries using machine learning evaluates the effectiveness of different models. It compares metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared across various algorithms like Linear Regression, Decision Trees, and Random Forests. This report highlights which model performs best in terms of accuracy and reliability, providing insights into the strengths and weaknesses of each approach. The goal is to select the model that offers the most precise and consistent salary predictions.

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

The final model selection justification for predicting doctors' annual salaries using machine learning involves choosing the model that offers the best balance of accuracy, reliability, and interpretability. After evaluating various models such as Linear Regression, Decision Trees, and Random Forests, the chosen model is selected based on performance metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared. The selected model demonstrates superior predictive accuracy and generalizes well to unseen data, ensuring it effectively captures the relationships between input features and salary outcomes. This decision is supported by thorough testing and validation to ensure robustness and reliability.

**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

The project demonstration for predicting doctors' annual salaries using machine learning involves showcasing the entire workflow from data collection to model deployment. This includes presenting how data is gathered and preprocessed, explaining the feature selection process, and demonstrating the training and evaluation of the machine learning model. The demonstration highlights the model's accuracy and reliability through performance metrics and visualizations. Finally, it shows the deployment of the model as a web application or API, allowing users to input relevant data and receive real-time salary predictions. This comprehensive demonstration ensures stakeholders understand the project's capabilities and benefits.