**Project Initialization and Planning Phase**

|  |  |
| --- | --- |
| Date | July 15, 2024 |
| Team ID | 739761 |
| Project Title | Doctor’s Annual Salary Prediction |
| Maximum Marks | 3 Marks |

**Project Proposal (Proposed Solution) template**

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 project aims to develop a machine learning model to accurately predict doctors' annual salaries based on factors like experience, specialization, and location. This will help healthcare administrators with reliable salary forecasts for better budgeting, salary planning, and competitive compensation strategies. |
| Scope | The scope of this project includes data collection, preprocessing, model development, validation, and deployment. It encompasses the creation of a user-friendly interface for healthcare administrators to input data and obtain salary predictions. The project will also ensure data security and compliance with relevant regulations. |
| **Problem Statement** | |
| Description | Healthcare administrators face challenges in predicting doctors' annual salaries due to the complexity and variety of factors that influence salary levels. This difficulty in prediction leads to challenges in budgeting, salary planning, and maintaining competitive compensation packages. |
| Impact | Accurate salary predictions will enable healthcare administrators to make informed decisions regarding budgeting and salary planning. It will help maintain competitive compensation packages, thus attracting and retaining skilled medical professionals. |
| **Proposed Solution** | |
| Approach |  **Data Collection:**   * Gather historical salary data from multiple healthcare facilities. * Collect additional relevant data such as doctors' experience, specialization, location, education, and performance metrics.    **Data Preprocessing:**   * Clean and normalize the data to ensure consistency. * Handle missing values and outliers appropriately. * Encode categorical variables and scale numerical features.    **Model Training:**   * Split the dataset into training and testing sets. * Explore various machine learning algorithms (e.g., linear regression, decision trees, random forests, and gradient boosting) to identify the best-performing model. * Fine-tune the model using cross-validation and hyperparameter optimization.    **Model Validation:**   * Evaluate the model's performance using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R²). * Conduct validation with unseen data to ensure generalizability.    **Deployment:**   * Deploy the final model in a cloud-based or on-premise environment. * Develop a user-friendly interface for healthcare administrators to input data and obtain salary predictions. * Implement security measures to protect sensitive data. |
| Key Features | * **Accurate Salary Predictions:** Utilize advanced machine learning algorithms to provide reliable salary predictions. * **User-Friendly Interface:** Create an intuitive interface for easy data input and result interpretation. * **Scalability:** Ensure the solution can handle large datasets and scale as needed. * **Security:** Protect sensitive salary data with robust security measures. |

**Resource Requirements**

|  |  |  |
| --- | --- | --- |
| **Resource Type** | **Description** | **Specification/Allocation** |
| **Hardware** | | |
| Computing Resources | CPU/GPU specifications, number of cores | e.g., 2 x NVIDIA V100 GPUs |
| Memory | RAM specifications | e.g., 8 GB |
| Storage | Disk space for data, models, and logs | e.g., 1 TB SSD |
| **Software** | | |
| Frameworks | Python frameworks | e.g., Flask |
| Libraries | Additional libraries | e.g., tensorflow |
| Development Environment | IDE, version control | e.g., Jupyter Notebook, Git |
| **Data** | | |
| Data | Source, size, format | e.g., historical salary data, doctors' profiles, CSV format |