Heart Disease Prediction Application

Project Documentation

1. Project Overview

This document provides a detailed explanation of the Heart Disease Prediction Application. The project is designed to predict a user's risk of heart disease based on their health metrics. It uses a machine learning model trained on a standard health dataset and provides a simple, interactive web interface for users.

2. Project Workflow

The entire process, from data handling to prediction, follows a systematic workflow.

Step-by-Step Workflow:

1. Data Collection:

- The model is trained using the **heart.csv** dataset, which contains anonymous patient data with various health attributes like age, cholesterol, blood pressure, etc.

2. Model Training:

- The **train_model.py** script is executed.
- It loads the **heart.csv** data using the Pandas library.
- The data is cleaned by handling any missing values.
- The dataset is then split into features (input data like age, sex) and a target (output: whether the patient has heart disease).
- A **Random Forest Classifier** algorithm from the Scikit-learn library is used to train the model on this data.

3. Model Saving:

- After training, the ready-to-use model is saved as a single file: **heart_disease_model.pkl**. This process is done using the Joblib library.

4. User Interaction:

- The user opens the web application, which is run by the **app.py** script using Streamlit.
- The application displays a user-friendly form with input fields for health data (e.g., Age, Blood Pressure, Cholesterol).

5. Prediction:

- When the user fills in their details and clicks the "Predict" button:
- The web application loads the pre-trained **heart_disease_model.pkl** file.
- It takes the user's input and converts it into the format required by the model.
- The model processes this input and returns a prediction: either "Low Risk" or "High Risk" of heart disease.
- This final result is displayed to the user on the screen.

3. Project Architecture

The project is built on a simple yet effective multi-layered architecture.

1. Data Layer:

Component: heart.csv

Description: This is the foundation of the project, containing all the raw data used for training the machine learning model.

2. Model Training Layer:

Component: train_model.py

Description: This script acts as the "brain" of the project. It handles data preprocessing, training the model, and saving the final model file.

3. Model Layer:

Component: heart_disease_model.pkl

Description: This is the final, trained, and saved machine learning model. It is a lightweight and portable file ready for making predictions.

4. Application Layer:

Component: app.py

Description: This is the user-facing part of the project. It uses the Streamlit framework to create an interactive web interface and uses the saved .pkl model to provide live predictions to the user.

4. Conclusion

This project successfully combines data science and web development to create a practical tool for health risk assessment. Its modular architecture makes it easy to maintain and update in the future.

Heart Disease Prediction Application - Project Documentation