High-Level Design (HLD) Report

Project Overview

The goal of this project is to create a predictive model that determines the probability of a patient having heart disease or a heart attack. The model will also identify the most important drivers for heart disease or heart attacks. This will aid in early detection and preventive measures for individuals at risk.

Dataset

The dataset includes the following features:

- Age
- HighBP (High Blood Pressure)
- PhysHlth (Physical Health Score)
- HighChol (High Cholesterol)
- Diabetes
- Smoker
- Sex
- MentHlth (Mental Health Score)
- BMI (Body Mass Index)
- HeartDiseaseorAttack (Target Variable)

Data Preprocessing

1. **Data Cleaning**: Handle missing values, outliers, and incorrect data entries.
2. **Feature Engineering**: Create new features and transform existing ones to better represent the underlying patterns in the data.
3. **Normalization/Standardization**: Apply scaling techniques to ensure all features contribute equally to the model.

Model Selection

Several models were evaluated to find the best-performing one for predicting heart disease or heart attack:

1. **Logistic Regression**
2. **Decision Tree**
3. **Random Forest**
4. **Gradient Boosting**

Based on the evaluation metrics, the RandomForest model was chosen due to its superior performance with an accuracy of 89%.

Model Evaluation

The RandomForest model was evaluated using the following metrics:

- Accuracy
- Precision
- Recall
- F1 Score

The model achieved an accuracy of 89%, making it the best choice for this prediction task.

Feature Importance

The importance of each feature in predicting heart disease or heart attack was determined. The most important features included Age, High Blood Pressure, Physical Health Score, High Cholesterol, and Diabetes.

UI Creation

An additional UI was created to allow users to input their health data and get a prediction on their risk of heart disease or heart attack. The UI was developed using HTML, CSS, Flask, and Python. The web application leverages the trained RandomForest model to provide real-time predictions.

Technologies Used:

- **HTML and CSS**: For creating the frontend user interface.
- **Flask**: For handling web requests and rendering templates.
- **Python**: For backend processing and integration with the RandomForest model.

End-to-End Workflow

1. **User Input**: Users enter their health data through the web form.
2. **Data Processing**: The input data is processed and formatted for prediction.
3. **Model Prediction**: The processed data is fed into the RandomForest model to generate a prediction.
4. **Result Display**: The prediction result is displayed back to the user on the web page.

Conclusion

The project successfully developed a predictive model with an accuracy of 89% for identifying individuals at risk of heart disease or heart attack. The additional user interface enhances the project's usability by providing an easy-to-use platform for health risk assessment.