HEART DISEASE PREDICTION MODEL

NAME: Samruddhi Sunil Shinde

REGISTRATION NUMBER: 25BAI10332

COURSE CODE: CSE1021

COURSE NAME: Introduction to Problem Solving

SLOT: C21+F11+F12

Identify a real-world problem or need:

A Heart Disease Prediction Model significantly helps to address a real-world problem by assisting early detection, personalized prevention strategies and efficient allocation of healthcare resources.

There are 4.77 million cardiovascular disease (CVD) deaths in India in 2020. These death could have been prevented if there was an early detection of heart defect. The prevalence of coronary heart disease varies, ranging from 1.6% to 7.4% in rural areas and 1% to 13.2% in urban areas. Heart failure estimates range from 1.3 to 4.6 million cases.

WAYS IN WHICH THIS MODEL HELPS TO SOLVE A REAL-WORLD PROBLEM:

Early Risk Detection: Machine learning models can sift through enormous amounts of patient data (e.g. age, blood pressure, cholesterol values, EKG results, lifestyle factors, etc.) to decrease the risk of developing heart disease, which can occur long before any clinical symptoms become apparent, enabling an earlier intervention and lifestyle changes that could be important in preventing a future major cardiac event.

Tailored Treatment Plans: When risk factors are narrowed down to the patient level, the physician can develop a more tailored prevention and treatment plan. Rather than delivering general advice based on population studies, there may be some specific duration, diet, exercises or medications for the patient to heed to ensure optimum health.

Resource Utilizaion: Predictive modelling can improve triaging and utilization resources within a hospital and healthcare system. With an accurate prediction of individuals needing urgent and immediate treatment, hospitals and clinics can evaluate patient flow, maximize efficiency utilization of physician time and specialty equipment and mitigate the impact of local emergency departments.

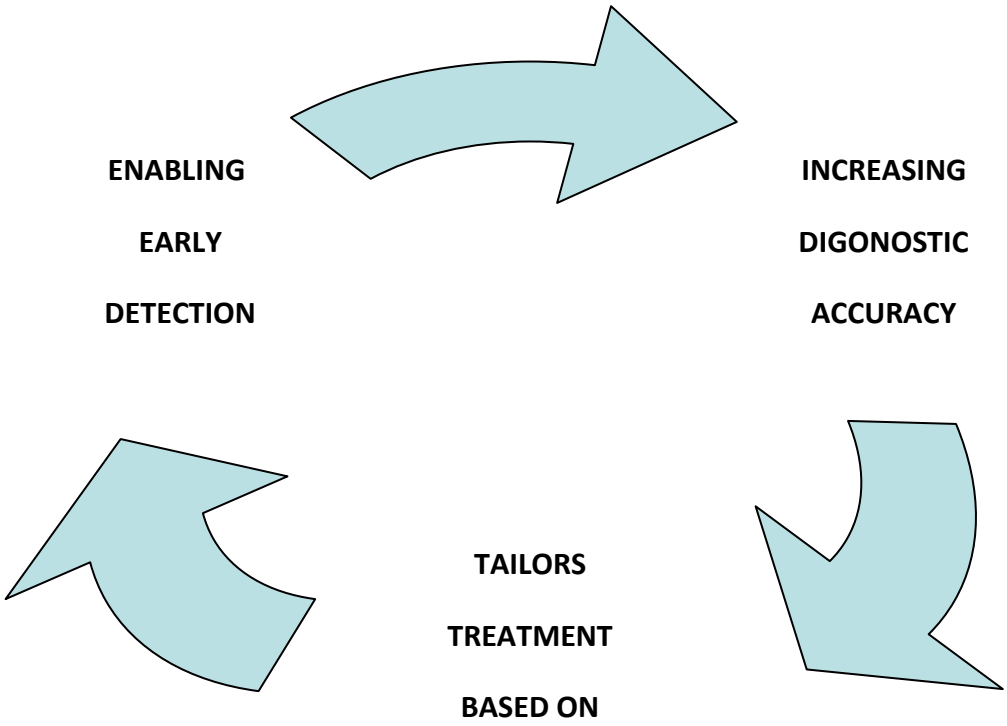
Lower Cost: Proactive prevention of heart disease and even heart attacks, based on early risk prediction, is far less expensive than reactive, acute care, and pricing for emergency and urgent procedures. Effective early detection gets the heart disease treatment treatment diverted, prior to a larger expense on the system.

Empowering Patients: Offering people a specific, data-driven assessment of their risk may be a strong motivator for adopting healthier behaviors. A risk score provided in numbers can transform health warnings from an abstract to a specific, legitimate reason to change.

Supporting Clinical Decision-Making: These models may be valuable support tools for physicians acting as an objective, data-driven second opinion that can improve the accuracy of a diagnosis and treatment decisions.

Define clear objectives and expected outcomes:

OBJECTIVES:



EXPECTED OUTCOMES:

This model takes input from the user such as age, blood pressure, chest pain type, sex, etc. and returns output that predicts if the patient has heart defect or not. Machine learning model helps us to prevent heart diseases.

Apply concepts learned in the course to design the solution:

CONCEPTS USED:

- **Data Structures**: Using built-in structures like lists, dictionaries and tuples for temporary data storage, but primary relying on Pandas for organized and tabular data manipulation.
- **Object-Oriented Programming (OOP):** Concepts like classes and objects are utilized, when working with models and functions from libraries like Scikit-learn.
- **Function and Modules:** Defining modules and functions for cleaning data or model evaluation. Creating Logistic Regression object from the Logistic Regression class.
- **File I/O:** Reading the dataset using file input/output operations, typically handled by the pandas library's read_csv() function. The data set used in this model was downloaded from kaggle website.

Tools, libraries, or programming techniques used:

TOOLS USED:

1. Programming Language: Python is the industry standard for ML and data science due to its simplicity, extensive library system and strong community support.
2. Integrated Development Notebooks:

- Jupyter Notebook are popular for exploratory data analysis, model building, and documentation because they allow code, output and explanatory text to be combined.
- Anaconda is often used as a distribution platform platform to manage environments and packages(like Python and its libraries).

LIBRARIES:

1. Data Manipulation and Analysis:

- Pandas: Used for loading, manipulating, cleaning and analysing structured data .
- NumPy: Provides support for large, multi-dimensional arrays and matrices, along with high-level mathematical functions to operate on these arrays

2. Machine Learning:

- Scikit-learn(sklearn): The most crucial library for classical ML. It provides a consistent interface for classification, regression, clustering, dimensionality reduction and pre-processing.

PROGRAMMING TECHNIQUES:

1. Data Acquisition and cleaning
2. Data preprocessing
3. Data analysis
4. Model Training
5. Model Evaluation

Problem definition:

The problem statement of predicting heart disease by using machine learning is to come up with a model that is not only accurate but also quick to figure out the chances of heart disease occuring in a person using their health data and this way eventually enable early intervention and thus saving human lives. Traditional methods combined with the facts that heart disease is the number one cause of death worldwide make this problem even more urgent. The idea is to defeat the challenges of complex medical data and differences in personal risk factors to be able to get a dependable prediction tool which can be used by doctors to make decisions.

Key components of the problem definition:

- Accurate prediction
- Early detection
- Efficiency and accessibility
- Data-driven approach
- Handling complex data

Top-Down design:

1. USER INTERFACE
2. APPLICATION LAYER
3. THE PREDICTION ENGINE
4. DATA STORAGE AND ACCESS

ALGORITHM:

1. Start
2. Importing the dependency - Import different libraries like pandas , numpy and sklearn
3. Data collection and processing – load the csv data to pandas dataframe using ‘read_csv()’
4. Splitting the features and target – create a variable ‘x’ and remove the column target from the csv file and then store the column target in variable ‘y’.
5. Splitting the data into training data and test data .
6. Training the logistic regression model using training dataset .
7. Final prediction : Heart Disease present or not.
8. End

Implementation:

1. Data Collection and Preprocessing

- **Gather Data:** Collect patient data, including factors like age, sex, blood pressure, cholesterol, and other medical history. **Clean Data:** Handle missing values and outliers in the dataset.
- **Feature Engineering:** Select the most relevant features that have a strong relationship with the target variable (heart disease). Techniques like chi-square evaluation can help with this.
- **Split Data:** Divide the dataset into a training set and a testing set (e.g., 70% or 80% for training and 20% or 30% for testing).

2. Model Training

- **Choose an Algorithm:** Logistic Regression
- **Train the Model:** Use the training data to train the chosen algorithm to learn the patterns between the input features and the presence of heart disease.

3. Model Evaluation

- **Test the Model:** Use the testing set (which the model has not seen before) to evaluate its performance.
- **Calculate Metrics:** Analyze the model's accuracy, sensitivity, and specificity to understand how well it predicts heart disease.
- **Refine:** If performance is not satisfactory, you may need to adjust the features, try a different algorithm, or fine-tune the model's parameters and retrain

Testing:

```
Enter Patient's Details
ENTER AGE 63
ENTER SEX - 0: female , 1: male 1
ENTER Chest Pain Type - 0: Typical Angina , 1: Atypical Angina , 2:Non-Anginal Pain , 3:Asymptomatic 3
ENTER Resting Blood Pressure 145
ENTER Serum Cholesterol 233
ENTER Fasting Blood Sugar - 0: <= 120 mg/dl , 1: > 120 mg/dl 1
ENTER Resting Electrocardiographic Results - 0: Normal , 1: Abnormal 0
Enter Maximum Heart Rate Achieved =150
Enter Exercise Induced Angina - 0: No , 1: Yes =0
Enter ST Depression Induced by Exercise Relative to Rest =2.3
Enter Slope of the Peak Exercise ST Segment - 0:upsloping , 1:flat , 2:downsloping =0
Enter Number of Major Vessels (0-3) Colored by Fluoroscopy =0
Enter Thal - 1: Normal , 2: Fixed defect , 3: Reversible defect =1
[1]
The person has HEART DEFECT
/usr/local/lib/python3.12/dist-packages/sklearn/utils/validation.py:2739: UserWarning: X does not have valid feature names, but LogisticRegression was f
  warnings.warn(
```

THANK YOU…