

**Name:** Akash Patel  
**Roll no:** 281029  
**Batch:** A2

## **Assignment 4**

### **Statement:**

Q. Apply an appropriate ML algorithm on a dataset consisting of heart health details to predict the risk of heart disease. Create a confusion matrix based on the above data and find:

- a) Accuracy
- b) Precision
- c) Recall
- d) F1-score

### **Objective:**

1. This assignment aims to implement a machine learning model for predicting heart disease risk.
2. Learn how to assess classification models using metrics such as accuracy, precision, recall, and F1-score.
3. Gain hands-on experience in data preprocessing, model training, and performance evaluation.

### **Resources Used:**

1. Software: Jupyter notebook
2. Libraries: Pandas, Scikit-learn, Matplotlib, Seaborn

### **Introduction to Machine Learning for Heart Disease Prediction:**

1. Machine learning plays a vital role in healthcare by analyzing patient data to predict health risks such as heart disease.
2. Classification models help identify individuals at high risk, enabling early medical intervention.
3. The dataset comprises heart health parameters, including age, sex, blood pressure, cholesterol levels, heart rate, and other relevant indicators.

### **Methodology:**

1. Data Collection and Preprocessing:
  - Load the dataset containing heart disease-related attributes.
  - Handle missing values and encode categorical variables.
  - Split the dataset into training and testing sets.
2. Model Selection and Training:

- Choose an appropriate classification algorithm, such as Logistic Regression, Decision Tree, Random Forest, or SVM.
  - Train the selected model using the training dataset.
3. Model Evaluation:
- Generate predictions on the test dataset.
  - Use a confusion matrix to analyze model performance.
  - Compute accuracy, precision, recall, and F1-score.

### Confusion Matrix and Performance Metrics:

1. Accuracy: Measures overall model correctness.
  - Formula:  $(TP + TN) / (TP + TN + FP + FN)$
2. Precision: Evaluates the proportion of correctly predicted positive cases.
  - Formula:  $TP / (TP + FP)$
3. Recall (Sensitivity): Assesses the model's ability to identify actual positive cases.
  - Formula:  $TP / (TP + FN)$
4. F1-score: Balances precision and recall.
  - Formula:  $2 * (Precision * Recall) / (Precision + Recall)$

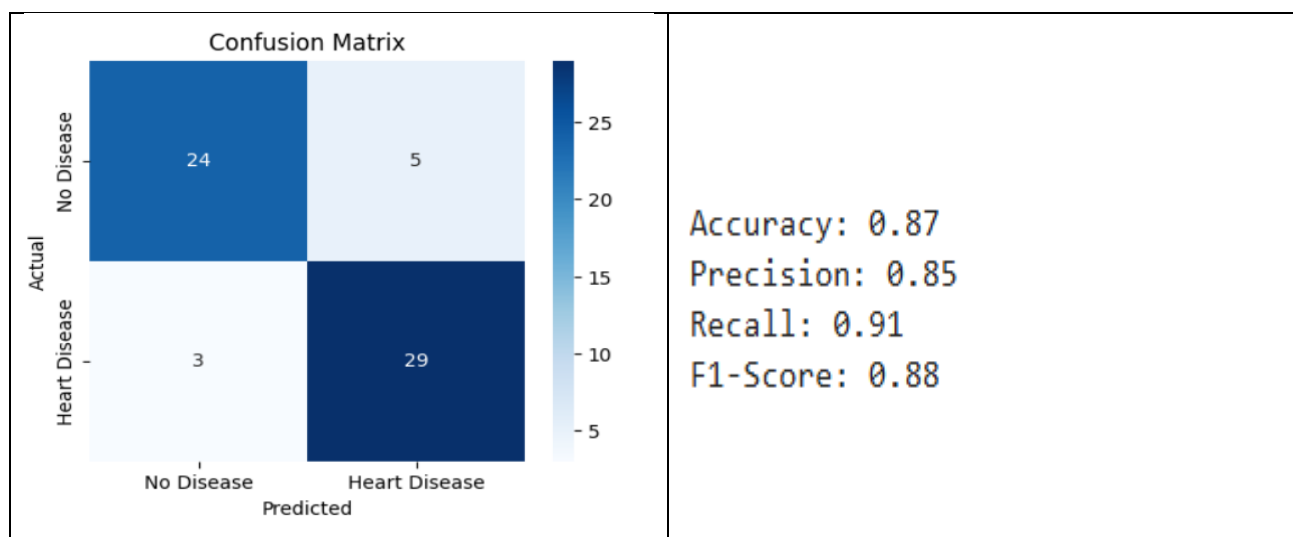
### Advantages of Machine Learning in Heart Disease Prediction:

1. Facilitates early detection, potentially saving lives.
2. Supports healthcare professionals in making data-driven decisions.
3. Enhances patient care through timely medical interventions.

### Disadvantages:

1. Requires a large, high-quality dataset for optimal accuracy.
2. Imbalanced datasets can affect model performance.

### Results:



**Conclusion:**

In this assignment, we developed a machine learning classification model to predict heart disease risk. We evaluated its performance using accuracy, precision, recall, and F1-score. By leveraging machine learning, healthcare professionals can improve risk monitoring and provide timely interventions. These techniques are crucial in enhancing healthcare outcomes and reducing heart disease-related risks.