Name: Mayuresh Marade
Roll No: 281052
Batch: A2

Assignment 4

Statement

Q. Apply an appropriate Machine Learning 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. To apply a machine learning model for predicting heart disease risk.
2. To understand how to evaluate classification models using **accuracy**, **precision**, **recall**, and **F1-score**.
3. To gain hands-on experience in **data preprocessing**, **model training**, and **performance evaluation**.

Resources Used

- **Software**: Google Colab
- **Libraries**: Pandas, Scikit-learn, Matplotlib, Seaborn

Introduction to Machine Learning for Heart Disease Prediction

Machine learning plays a crucial role in modern healthcare by leveraging patient data to predict medical risks, including **heart disease**. Using classification algorithms, it's possible to predict high-risk individuals and facilitate timely treatment.

Key dataset attributes often include:

- Age, Sex
- Blood Pressure, Cholesterol
- Heart Rate, Diabetes, Smoking History, etc.

ML models like **Logistic Regression**, **Random Forest**, and **SVM** are commonly used for such predictions.

Methodology

1. **Data Collection and Preprocessing**
    - Load the dataset containing heart health details.
    - Handle missing values, normalize data, and encode categorical features.

- Split the data into training and testing sets (e.g., 80% train, 20% test).

2. **Model Selection and Training**
    - Chose suitable classification algorithms like **Logistic Regression** or **Random Forest**.
    - Trained the model on the training dataset.
3. **Model Evaluation**
    - Made predictions on the test data.
    - Computed the **Confusion Matrix**.
    - Calculated **Accuracy**, **Precision**, **Recall**, and **F1-score** using Scikit-learn metrics.

Confusion Matrix and Performance Metrics

1. **Accuracy**: Overall correctness of the model.
    - **Formula**: (TP + TN) / (TP + TN + FP + FN)
2. **Precision**: Correctness of positive predictions.
    - **Formula**: TP / (TP + FP)
3. **Recall (Sensitivity)**: Ability to find all positive instances.
    - **Formula**: TP / (TP + FN)
4. **F1-score**: Balance between precision and recall.
    - **Formula**: 2 * (Precision × Recall) / (Precision + Recall)

Advantages of ML in Heart Disease Prediction

1. Enables **early diagnosis** and treatment planning.
2. Supports **data-driven** healthcare decisions.
3. Improves **patient care** and reduces complications with timely intervention.

Disadvantages

1. **Large, high-quality datasets** are needed for reliable predictions.
2. Model effectiveness may drop with **imbalanced** or noisy data.

Results

(Add your actual results here from Colab – include the confusion matrix and the following values)

- **Accuracy**: e.g., 0.87
- **Precision**: e.g., 0.84

- **Recall**: e.g., 0.88
- **F1-score**: e.g., 0.86

(Optional: Include screenshot of confusion matrix plot and classification report.)

Conclusion

In this assignment, a machine learning classification model was successfully implemented to predict **heart disease risk**. We evaluated the model using key performance metrics: **accuracy**, **precision**, **recall**, and **F1-score**. These metrics helped us assess the model's performance and effectiveness. Through this practical application, we understood how machine learning can assist healthcare professionals in monitoring heart conditions, enabling faster and more informed medical interventions.