****

**Assessment Report**

on

**“Heart Disease Prediction Using Machine Learning”**

submitted as partial fulfillment for the award of

**BACHELOR OF TECHNOLOGY**

**DEGREE**

SESSION 2024-25

in

**Introduction of AI**

By

Shani Kumar (202401100300224)

**Under the supervision of**

“Mr. Abhishek Shukla”

**KIET Group of Institutions, Ghaziabad**

Affiliated to

**Dr. A.P.J. Abdul Kalam Technical University, Lucknow**

(Formerly UPTU)

**May, 2025**

**Problem Statement:**  
To develop a machine learning model that can accurately predict whether a patient is likely to develop heart disease based on their medical attributes.

**Introduction:**

Cardiovascular diseases are a major cause of death worldwide. Early detection and timely medical intervention are essential for improving survival rates. Machine Learning (ML) offers powerful techniques to analyze large amounts of clinical data and predict outcomes like heart disease presence.

The objective of this project is to use machine learning techniques to predict heart disease using patient data. The dataset used in this study contains medical features such as age, sex, chest pain type, resting blood pressure, cholesterol levels, and more. The goal is to classify patients into two categories: those with heart disease and those without.

This report details the steps taken to prepare, process, and analyze the data using Random Forest Classifier, a robust ensemble learning method. It also includes evaluation metrics to validate model performance.

**Methodology:**

The following methodology was adopted to solve the problem:

1. **Dataset Collection**: The dataset was obtained from a standard source and contains 303 entries and 14 attributes, including a binary target variable (target) indicating the presence or absence of heart disease.
2. **Data Preprocessing**:
   * Verified that there were no missing values in the dataset.
   * Features (independent variables) and the target (dependent variable) were separated.
   * The dataset was split into training and testing subsets in an 80:20 ratio.
3. **Model Selection**: Random Forest Classifier was chosen because:
   * It handles both numerical and categorical data.
   * It provides high accuracy.
   * It reduces overfitting due to ensemble learning.
4. **Training and Prediction**:
   * The model was trained using the training subset.
   * Predictions were made on the unseen testing subset.
5. **Evaluation**:
   * Accuracy Score
   * Classification Report (precision, recall, f1-score)
   * Confusion Matrix

**Code:**

# Import necessary libraries

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import accuracy\_score, classification\_report, confusion\_matrix, ConfusionMatrixDisplay

import matplotlib.pyplot as plt

# Load the dataset

df = pd.read\_csv("4. Predict Heart Disease.csv")

# Display the first few rows of the dataset (optional)

print("Dataset Preview:")

print(df.head())

# Separate features (X) and target variable (y)

X = df.drop("target", axis=1)  # All columns except 'target'

y = df["target"]               # Target column

# Split the data into training and testing sets (80% train, 20% test)

X\_train, X\_test, y\_train, y\_test = train\_test\_split(

    X, y, test\_size=0.2, random\_state=42

)

# Initialize the Random Forest Classifier

model = RandomForestClassifier(random\_state=42)

# Train the model on the training data

model.fit(X\_train, y\_train)

# Make predictions on the test data

y\_pred = model.predict(X\_test)

# Calculate the accuracy of the model

accuracy = accuracy\_score(y\_test, y\_pred)

# Generate a detailed classification report

report = classification\_report(y\_test, y\_pred)

# Create a confusion matrix and plot it

conf\_matrix = confusion\_matrix(y\_test, y\_pred)

disp = ConfusionMatrixDisplay(confusion\_matrix=conf\_matrix, display\_labels=model.classes\_)

disp.plot(cmap='Blues')

plt.title(f"Confusion Matrix (Accuracy: {accuracy:.2f})")

plt.tight\_layout()

plt.show()

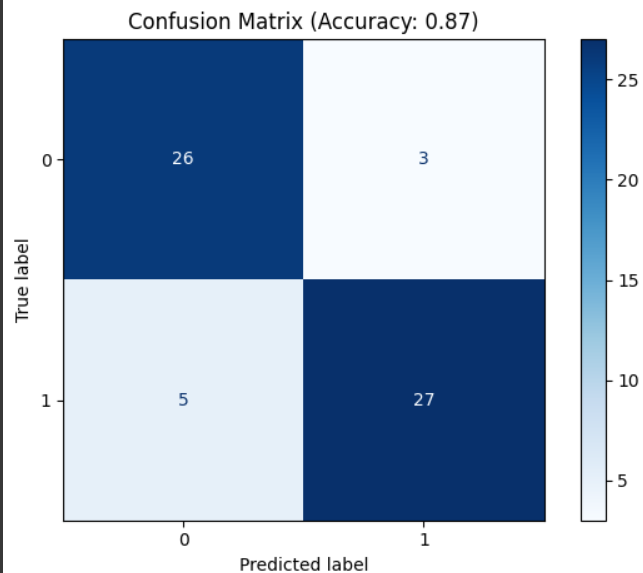
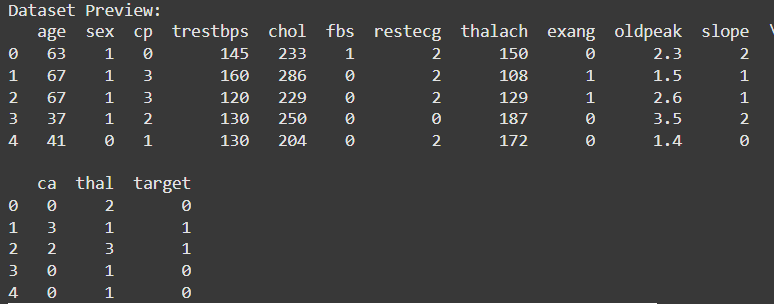
# Print the evaluation results

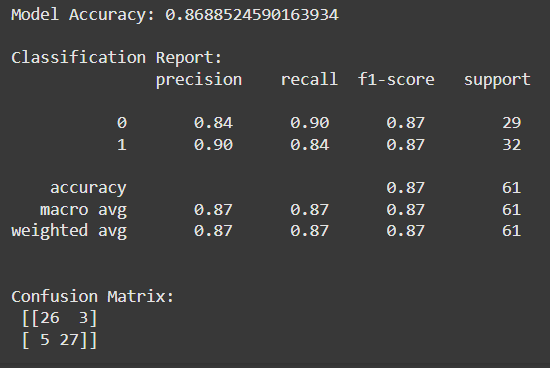
print("\nModel Accuracy:", accuracy)

print("\nClassification Report:\n", report)

print("\nConfusion Matrix:\n", conf\_matrix)

**Output / Result:**

****

****

**References / Credits:**

* **Dataset Source: UCI Machine Learning Repository - Heart Disease Dataset**
* **Libraries Used: pandas, scikit-learn**
* **Model: Random Forest Classifier**
* **IDE: Google Colab**
* **Guidance: Mr. Abhishek Shukla**