FATA UNIVERSITY

Department of Computer Science

Project:

CARDIOVASCULAR DISEASE PREDICTION USING MACHINE LEARNING

Name	Rizwan Ullah & Namreen
Roll No	09 & 32
Department	CS
Semester	6th
Submittetd to	Dr.Bilal Jan
Dated	4 July, 2025

🩺 Cardiovascular Disease Prediction using Machine Learning

🔨 Problem Statement

Cardiovascular diseases (CVDs) are the leading cause of death globally. Early prediction and diagnosis can significantly reduce mortality rates by enabling timely medical intervention. Traditional diagnosis often relies on clinical tests that can be costly, invasive, or time-consuming.

🔍 Objective:

To build a machine learning model that predicts the risk of cardiovascular disease in individuals using basic health indicators like age, gender, blood pressure, cholesterol, glucose level, and lifestyle habits.

📂 Dataset Description

- **Source**: Kaggle - Cardiovascular Disease Dataset
- **File**: `cardio_train.csv`
- **Size**: ~70,000 rows and 13 columns
- **Target Column**: `cardio` (1 = has disease, 0 = healthy)

📄 Features:

Feature	Description
Age	Age (in days)

Gender	1 = Female, 2 = Male
height	Height in cm
weight	Weight in kg
ap_hi	Systolic blood pressure
ap_lo	Diastolic blood pressure
cholesterol	1 = normal, 2 = high, 3 = very high
gluc	Glucose level (same as cholesterol)
smoke	1 = smokes, 0 = no
alco	1 = drinks alcohol, 0 = no
active	1 = physically active, 0 = not
cardio	🔴 Target: 1 = disease, 0 = healthy

⚙️ Methodology

1. Data Preprocessing

- Converted age from days to years.
- Removed unnecessary id column.
- Scaled all features using StandardScaler.

2. Model Selection

- Chose **Random Forest Classifier** due to its robustness and performance on tabular data.

3. Model Training

- Split data: 80% training, 20% testing.
- Trained with 100 decision trees using RandomForestClassifier.

4. Evaluation Metrics

- Accuracy Score
- Classification Report (Precision, Recall, F1-Score)
- Confusion Matrix Visualization

☑ Model Performance

Metric	Value
Accuracy	~70.37%
Precision	70% – 71%
Recall	70% – 71%
F1-Score	70%

🔍 Confusion Matrix:

	Predicted Healthy	Predicted Disease
Actual Healthy	✅	❌
Actual Disease	❌	✅

A heatmap is also saved as `confusion_matrix.png` for reporting.

💾 Saved Files

File Name	Description
`heart_disease_model.pkl`	Trained Random Forest model
`heart_disease_scaler.pkl`	Scaler to transform input features
`confusion_matrix.png`	Visual evaluation of predictions

app.py	Streamlit Web App for predictions

🔬 How to Use the Prediction App

1. Run the model with:

```
streamlit run app.py
```

2. Enter patient information:
   a. Age, height, weight
   b. Blood pressure, cholesterol, glucose
   c. Smoking/alcohol/activity habits
3. Click "**Predict**"
   a. ✅ Low Risk
   b. ⚠️ High Risk

💡 Conclusion

This machine learning model provides a simple and accessible way to assess heart disease risk using basic health data. It can serve as a **supportive screening tool** for early detection and preventive healthcare.