AI Project

Heart Disease Prediction

Name – ID:

**1 – بيشوي امجد فوزي عبدالملاك - 2023170147**

**2 – بيشوي كرم اسعد ذكي - 2023170149**

**3 – جهاد خالد عيد محمد - 2023170170**

**4 – جورج اسامة فؤاد رياض - 2023170173**

**5 – شيرين وائل مصطفي حمزة - 2023170305**

Heart Disease Prediction Using UCI Dataset: Summary

# Project Overview

This project develops an AI model to predict heart disease using the UCI Heart Disease Dataset (303 records, 13 features). The goal is to classify patients as having heart disease

(1) or not (0) using Custom Linear Regression, SVM, Decision Tree (ID3), and KNN, comparing their accuracy to identify the best model.

# Dataset

The UCI Heart Disease Dataset includes:

* + **Target**: target (0 = no heart disease, 1 = heart disease)
  + **Features**: age, sex, chest pain type (cp), resting blood pressure (trestbps), choles- terol (chol), fasting blood sugar (fbs), resting ECG (restecg), max heart rate (tha- lach), exercise-induced angina (exang), ST depression (oldpeak), slope, major ves- sels (ca), thalassemia (thal)

# Preprocessing

* + Load heart.csv using pandas
  + Convert chol to numeric, fill missing values with column means
  + Remove outliers using IQR method
  + Scale features to [0,1] with MinMaxScaler
  + Select top 8 features using SelectKBest (chi2)
  + Visualize correlation matrix and feature importance

# Model Implementation

* + **Custom Linear Regression**: Gradient descent, threshold=0.5 (learning\_rate=0.01,

n\_iterations=1000)

* + **SVM**: RBF kernel, tuned with GridSearchCV (C: [0.1, 1, 10, 100], gamma: [scale, auto, 0.01, 0.001])
  + **Decision Tree**: Entropy criterion (ID3 style), random\_state=42
  + **KNN**: Tuned with GridSearchCV (n\_neighbors: [3, 5, 7, 9])

# Evaluation

* + **Metrics**: Accuracy, confusion matrix, classification report
  + **Data Split**: 80% train, 20% test (random\_state=42)
  + **Visualizations**: Confusion matrices and accuracy comparison bar plot