



HEART DISEASE PREDICTION

Using Machine Learning

Presented by
mazin ahmed mohammed





PROJECT IDEA

Predict heart disease based on patient data using machine learning techniques

Help doctors or users identify risk early through a simple prediction system

Use medical features such as age, cholesterol, blood pressure, and ECG results





DATASET DESCRIPTION



Name of the dataset:

Heart Disease Dataset

Any other relevant details

Source or link to the dataset:

<https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>

Total rows: 1025 samples No missing values •

Number of features:

input features + 1 target variable = 14 columns 13

Target variable:

target → (1 = has heart disease, 0 = no heart disease)



Features Used

- age: عمر المريض
- sex: جنس المريض (1 = ذكر، 0 = أنثى)
- cp: نوع ألم الصدر (0 = نموذجي، 1 = غير نموذجي، 2 = غير قلبي، 3 = لا أعراض)
- trestbps: ضغط الدم أثناء الراحة (بوحدة mm Hg)
- chol: نسبة الكوليسترول في الدم (mg/dl)
- fbs: سكر صائم < 120 (1 = نعم، 0 = لا)
- restecg: نتائج تخطيط القلب أثناء الراحة (0 = طبيعي، 1 = خلل 2، ST-T = تضخم البطين الأيسر)
- thalach: أقصى معدل نبض وصل إليه المريض
- exang: وجود ذبحة صدرية بعد التمرين (1 = نعم، 0 = لا)
- oldpeak: انخفاض مستوى ST الناتج عن التمرين مقارنة بالراحة
- slope: ميل مقطع ST أثناء التمرين (0 = صاعد، 1 = مسطح، 2 = نازل)
- ca: عدد الأوعية الدموية الرئيسية التي تم تلويينها (من 0 إلى 3)
- thal: نوع الثلاسيميا (1 = طبيعي، 2 = عيب ثابت، 3 = عيب قابل للانعكاس)



PROJECT DETAILS

Methods and techniques used

Supervised machine learning, classification models, data scaling

Data cleaning steps

Checked for and removed duplicate rows

Verified that there are no missing values

Data preprocessing steps

Separated features and target variable

Scaled numerical features using StandardScaler

Evaluation metrics

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

Algorithms applied

Logistic Regression

Support Vector Machine
(SVM)

K-Nearest Neighbors
(KNN)

Decision Tree

Random Forest

Naive Bayes





ADDITIONAL EFFORTS



Any modifications made to the dataset

- Verified data quality and consistency
- Removed duplicates if any
- No null values were present
- Scaled the dataset using StandardScaler

Changes in algorithm selection

- Selected Random Forest as final model due to highest accuracy (84.21%)
- Tuned random_state for stability in Decision Tree & Random Forest
- Searching for the best random_state in Decision Tree & Random Forest



**THANK
YOU**