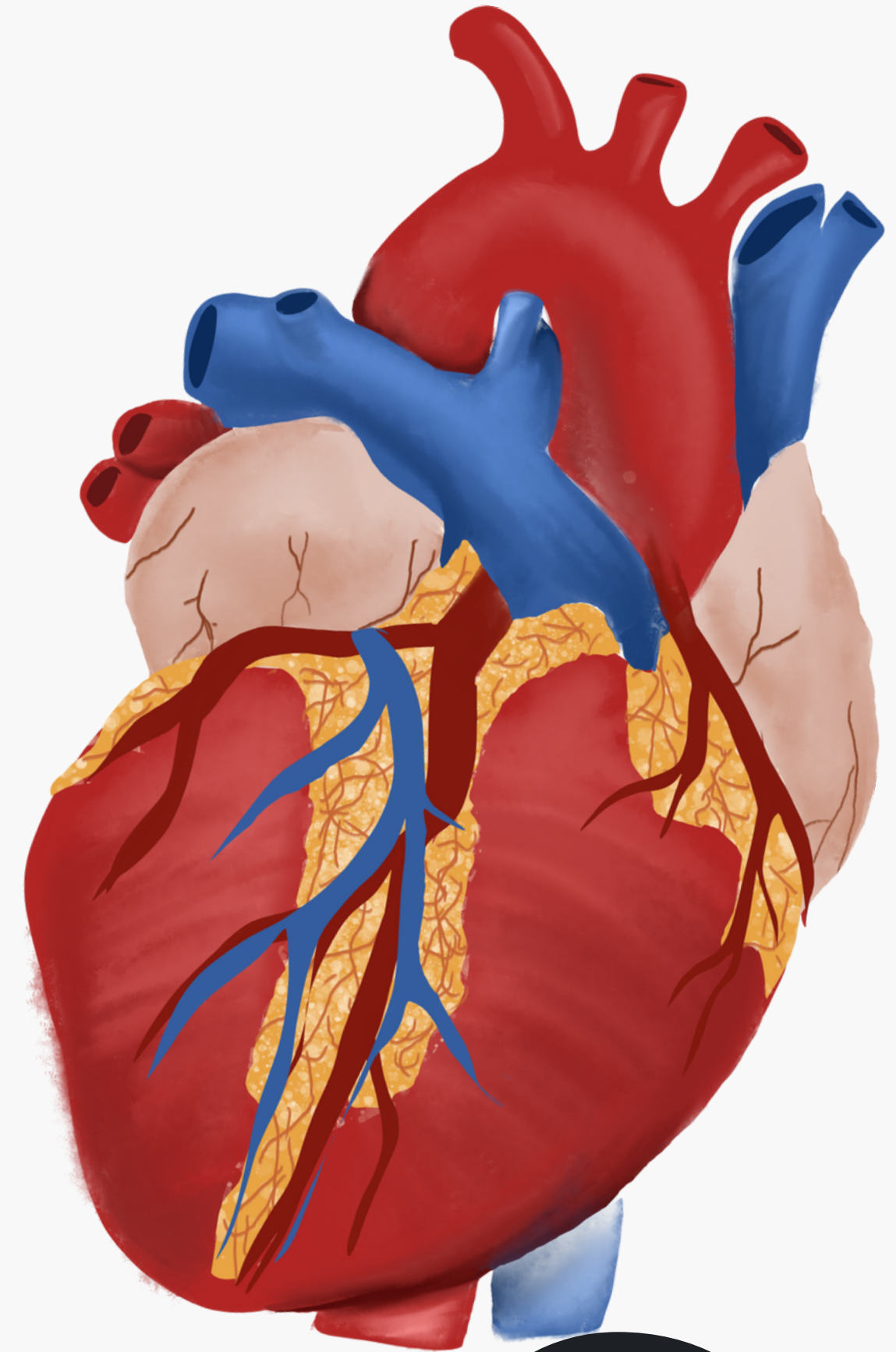


*H*ear*D*isease *P*rediction

Leveraging data to predict and prevent heart disease early

Presented by
Yara Lihawa
14 May 2025

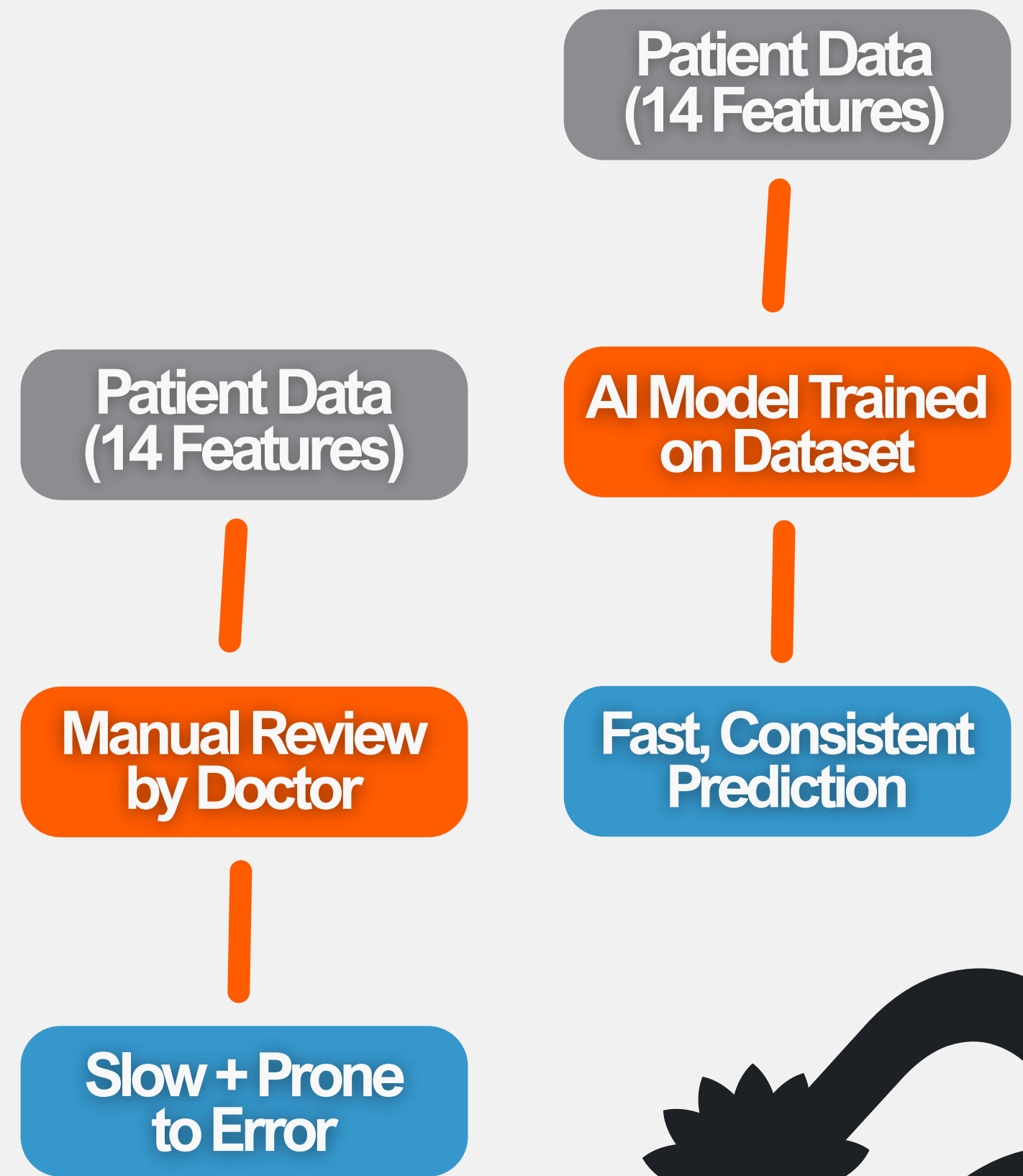


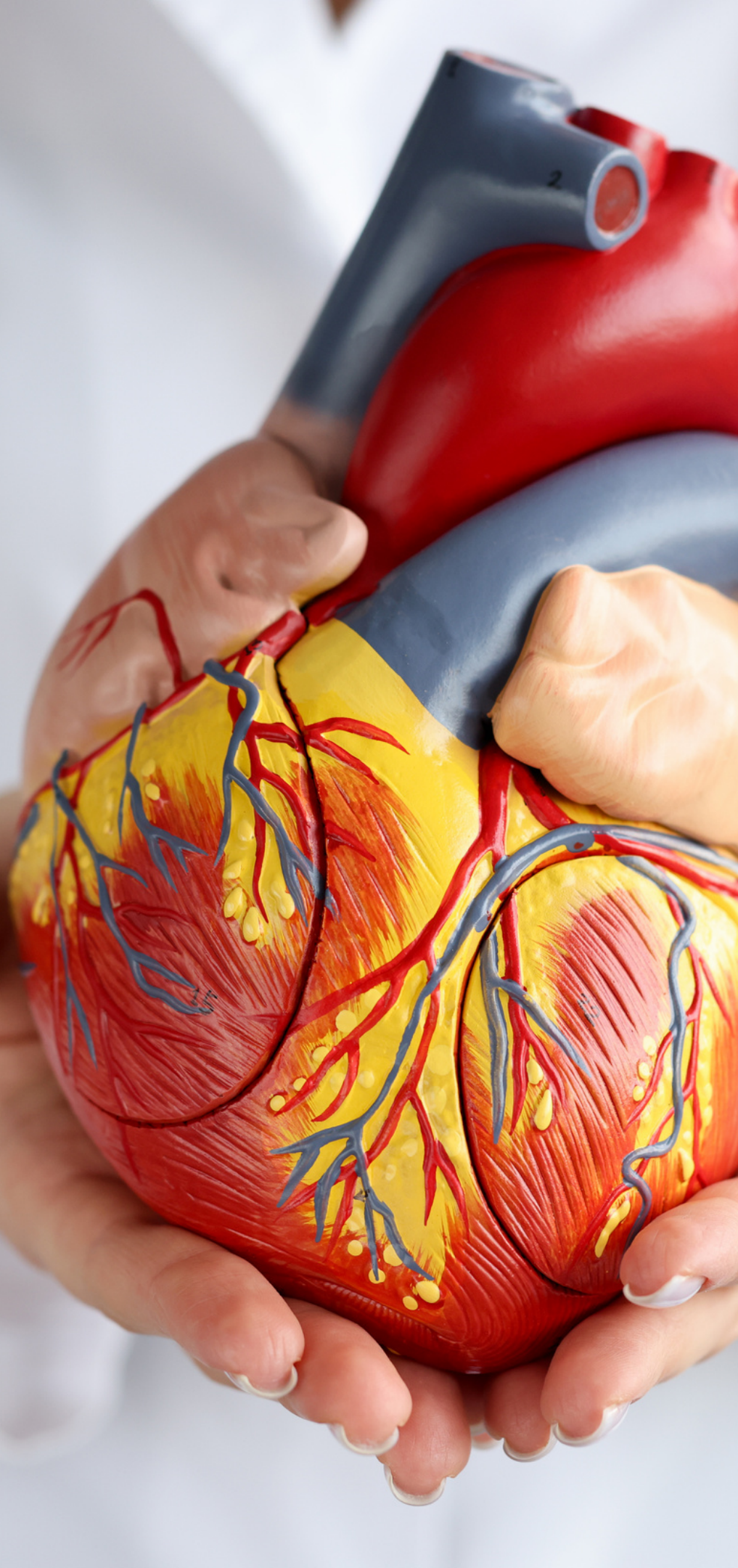
Background Problem

Industri kesehatan saat ini **menghadapi tantangan besar dalam diagnosis penyakit jantung.**

Masalah utama adalah bagaimana memprediksi dengan akurat apakah seorang pasien memiliki penyakit jantung atau tidak, berdasarkan data medis yang kompleks dan beragam.

Diagnosis manual memakan waktu dan rentan terhadap kesalahan manusia. Dengan solusi berbasis AI, kami ingin membangun model prediktif yang dapat menganalisis 14 atribut klinis utama (seperti tekanan darah, kolesterol, dan detak jantung maksimum) untuk memberikan prediksi yang cepat dan akurat.





Objectives



Memprediksi apakah seorang pasien menderita penyakit jantung berdasarkan 14 atribut klinis.



Meningkatkan efisiensi dan kecepatan dalam proses diagnosis.



Mengurangi risiko kesalahan diagnosis akibat faktor manusia.



Memberikan dukungan data untuk membantu pengambilan keputusan oleh tenaga medis.



Menggunakan pendekatan machine learning klasifikasi binari (0 = tidak sakit, 1 = sakit).

Dataset Overview

1. age: Usia pasien dalam tahun.
2. sex: Jenis kelamin (1 = laki-laki, 0 = perempuan).
3. cp: Jenis nyeri dada (0 = typical angina, 1 = atypical angina, 2 = non-anginal pain, 3 = asymptomatic).
4. trestbps: Tekanan darah saat istirahat (mm Hg).
5. chol: Kadar kolesterol dalam darah (mg/dl).
6. fbs: Gula darah puasa > 120 mg/dl (1 = ya, 0 = tidak).
7. restecg: Hasil elektrokardiogram istirahat (0 = normal, 1 = ST-T abnormal, 2 = left ventricular hypertrophy).
8. thalach: Detak jantung maksimum saat tes stres.
9. exang: Apakah pasien mengalami angina saat olahraga (1 = ya, 0 = tidak).
10. oldpeak: Depresi ST akibat aktivitas dibanding saat istirahat.
11. slope: Kemiringan segmen ST saat tes stres (0 = menurun, 1 = datar, 2 = naik).
12. ca: Jumlah pembuluh darah utama (0–3) terlihat oleh fluoroskopi.
13. thal: Hasil tes thalassemia (1 = fixed defect, 2 = normal, 3 = reversible defect).
14. target: Apakah pasien memiliki penyakit jantung (1 = ya, 0 = tidak).

Jumlah data: 303 entri, 14 fitur

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
4	57	0	1	130	236	0	0	174	0	0.0	1	1	2	0

Dataset dibagi menjadi:

- Fitur (X): Semua kolom kecuali target
- Target (Y): Kolom target, yaitu label apakah pasien memiliki penyakit jantung

Kemudian dibagi menjadi:

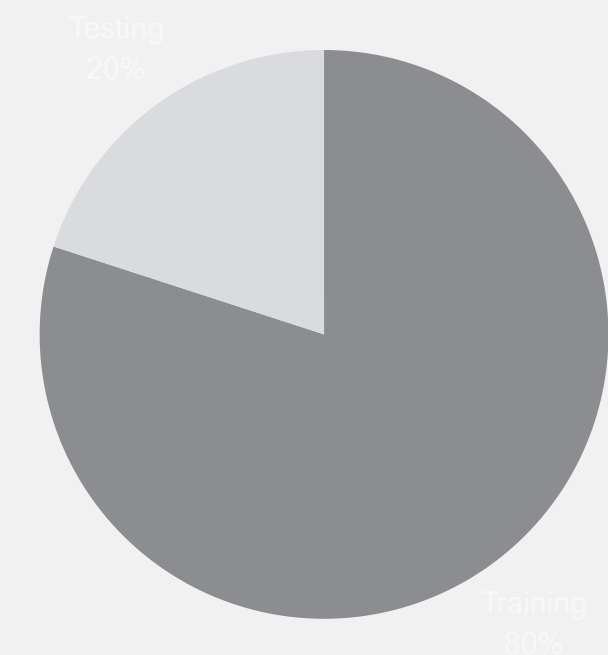
- 80% untuk pelatihan (training)
- 20% untuk pengujian (testing)

Pembagian menggunakan stratify=Y agar proporsi target tetap seimbang

```
from sklearn.model_selection import train_test_split

# Pisahkan fitur dan label
X = hdata.drop(columns='target', axis=1)
Y = hdata['target']

# Split ke train dan test
X_train, X_test, Y_train, Y_test = train_test_split(
    X, Y, test_size=0.2, stratify=Y, random_state=2
)
```



Splitting the Features(X) and Target(Y)

Training Logistic Regression Model

Model yang Digunakan

Model yang digunakan adalah Logistic Regression karena cocok untuk masalah klasifikasi biner seperti prediksi penyakit jantung.

Model dilatih menggunakan data training (X_train dan Y_train) yang sudah dipisah sebelumnya.

```
from sklearn.linear_model import LogisticRegression  
  
model = LogisticRegression()  
model.fit(X_train, Y_train)
```

Data
Training

Model

Trained Model





Model diuji menggunakan data testing (X_test dan Y_test).

Akurasi dihitung untuk melihat seberapa baik model memprediksi penyakit jantung.

Akurasi model dalam proyek ini mencapai sekitar 85.33%.

```
from sklearn.metrics import accuracy_score  
  
y_predict = model.predict(X_test)  
acc = accuracy_score(Y_test, y_predict)  
print(f"Akurasi model: {acc * 100:.2f}%")
```

Model Evaluation



Predicting Heart Disease for New Patients

- ✱ Sistem ini menerima data input dari pasien baru, seperti usia, tekanan darah, kolesterol, dan lainnya.
- ✱ Model memprediksi apakah pasien tersebut menderita penyakit jantung atau tidak.
- ✱ Contoh data input:
(75, 0, 2, 145, 233, 1, 0, 150, 0, 2.3, 0, 0, 1)
- ✱ Prediksi dilakukan dengan merubah bentuk data input dan memberikannya ke model yang sudah dilatih.

```
inputData = (75, 0, 2, 145, 233, 1, 0, 150, 0, 2.3, 0, 0, 1)
input_data_np = np.asarray(inputData)
input_data_reshaped = input_data_np.reshape(1, -1)
prediction = model.predict(input_data_reshaped)
print(prediction)
```

```
if (prediction[0]==1):
    print('The Person has a Heart Disease')
else:
    print('The Person does not have Heart Disease')
```

The Person has a Heart Disease

Conclusion



Model Logistic Regression berhasil mencapai akurasi sekitar 85.33% dalam memprediksi penyakit jantung.



Dataset digunakan langsung tanpa preprocessing tambahan seperti encoding atau scaling.



Model ini dapat menjadi dasar untuk pengembangan metode lebih lanjut dengan teknik yang lebih kompleks.

Model ini cocok digunakan sebagai alat bantu diagnosis awal.





Contact Me

Terima kasih atas perhatian Anda!
Jangan ragu untuk menghubungi saya jika ada pertanyaan, masukan, atau jika Anda ingin berkolaborasi.

Let's connect.

Yara Lihawa

✉ yaralihawa@gmail.com

 Yara Lihawa



Thank You!

