

**Prediksi Risiko Penyakit Jantung Menggunakan Algoritma  
Klasifikasi Pada Dataset Medis**



Disusun Oleh :

Faathir El Tasleem

A11.2022.14624

A11.4504

PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS DIAN NUSWANTORO  
SEMARANG 2024

## DESKRIPSI

Proyek ini bertujuan untuk menggunakan metode klasifikasi untuk menilai risiko penyakit jantung pada pasien. Kami akan menganalisis data medis seperti usia, jenis kelamin, tekanan darah, kadar kolesterol, dan indikator kesehatan lainnya. Data yang digunakan berasal dari Kaggle dan bertujuan untuk membantu tenaga medis dalam mendiagnosis risiko penyakit jantung.

## MASALAH

Penyakit jantung adalah penyebab utama kematian di seluruh dunia. Oleh karena itu, penting untuk memiliki model yang dapat mendeteksi dan memprediksi risiko penyakit ini dengan akurat. Deteksi yang tepat dapat membantu dalam pencegahan dan penanganan penyakit jantung.

## TUJUAN

1. Mengembangkan metode yang dapat mengidentifikasi kemungkinan penyakit jantung pada pasien.
2. Menganalisis faktor-faktor yang paling berpengaruh dalam menentukan risiko penyakit jantung.
3. Memberikan rekomendasi untuk langkah-langkah pencegahan atau penanganan berdasarkan hasil analisis.

## ALUR/TAHAPAN/KERANGKA EKSPERIMEN

- **Pengumpulan Data:**
  - Mengunduh dan memahami dataset dari Kaggle.
- **Preprocessing Data:**
  - Memperbaiki data yang tidak normal atau berkualitas rendah dan menstandarisasi fitur sesuai kebutuhan model.
- **Pemilihan Algoritma Klasifikasi:**
  - Memilih metode klasifikasi yang tepat, seperti K-Nearest Neighbors.
- **Pelatihan Model:**
  - Melatih model dengan data yang sudah diproses dan memilih parameter terbaik.
- **Evaluasi Model:**
  - Mengukur performa model menggunakan metrik seperti akurasi, presisi, recall, dan F1-score.
- **Interpretasi dan Analisis Hasil:**
  - Menginterpretasikan hasil prediksi dan menganalisis faktor-faktor yang paling signifikan.
- **Visualisasi dan Kesimpulan:**
  - Menyajikan hasil dalam bentuk visual dan menarik kesimpulan terkait prediksi dan variabel penting.

## PENJELASAN DATASET

Dataset **Heart Disease Classification Dataset** dari Kaggle mengandung beberapa fitur penting yang digunakan untuk prediksi, antara lain:

- **Age:** Usia pasien.
- **Sex:** Jenis kelamin pasien.
- **Chest Pain Type:** Jenis nyeri dada, yang merupakan salah satu indikator risiko.
- **Resting Blood Pressure:** Tekanan darah pasien saat beristirahat.
- **Cholesterol:** Kadar kolesterol dalam tubuh.
- **Fasting Blood Sugar:** Kadar gula darah puasa.
- **Electrocardiographic Results:** Hasil elektrokardiografi.
- **Maximum Heart Rate Achieved:** Denyut jantung maksimum yang dicapai.
- **Exercise Induced Angina:** Indikator apakah pasien mengalami angina saat berolahraga.
- **Oldpeak:** Depresi ST yang diinduksi oleh olahraga.
- **Slope of the Peak Exercise ST Segment:** Kemiringan segmen ST pada puncak latihan.
- **Number of Major Vessels:** Jumlah pembuluh darah utama.
- **Thal:** Indikator thalassemia.

## Timeline Eksperimen

| No | Kegiatan                       | Minggu ke-1 | Minggu ke-2 | Minggu ke-3 | Minggu ke-4 | Minggu ke-5 | Minggu ke-6 | Minggu ke-7 |
|----|--------------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 1. | Perencanaan dan Pemahaman Data | ✓           | ✓           | ✓           |             |             |             |             |
| 2. | Persiapan Data                 |             | ✓           | ✓           |             |             |             |             |
| 3. | Permodelan Data                |             |             | ✓           | ✓           | ✓           |             |             |
| 4. | Evaluasi Model                 |             |             |             | ✓           | ✓           | ✓           |             |
| 5. | Visualisasi dan Pelaporan      |             |             |             |             | ✓           | ✓           | ✓           |