

Nama : Muhammad Fauzi Ardiansyah  
Nim : A11..2022.14711

## Langkah CRISP-DM Prediksi Penyakit Jantung

Dataset:

<https://www.kaggle.com/code/desalegngeb/heart-disease-predictions/input>

### 1. Business Understanding (Pemahaman Bisnis)

Tujuan dari proyek ini adalah untuk memprediksi apakah seseorang berisiko terkena penyakit jantung berdasarkan berbagai fitur kesehatan dan demografi. Prediksi ini dapat membantu tenaga medis dan individu dalam mengambil tindakan pencegahan dan membuat keputusan yang tepat. Sasaran bisnis utama adalah:

- Meningkatkan deteksi dini: Mengidentifikasi pasien yang berisiko terkena penyakit jantung secara akurat.
- Mengoptimalkan alokasi sumber daya: Mengarahkan sumber daya kesehatan ke pasien berisiko tinggi berdasarkan prediksi.
- Mengurangi biaya kesehatan: Mencegah penyakit jantung dengan diagnosis dini dan mengurangi pengobatan tahap lanjut yang mahal.

### 2. Data Understanding (Pemahaman Data)

Dataset yang digunakan adalah dataset penyakit jantung, yang biasanya mencakup fitur-fitur seperti umur, jenis kelamin, jenis nyeri dada, tekanan darah istirahat, kolesterol, gula darah puasa, hasil elektrokardiogram saat istirahat, detak jantung maksimal yang dicapai, angina akibat olahraga, ST depresiasi, kemiringan segmen ST puncak, jumlah pembuluh darah utama, thalassemia, dan target (label). Langkah-langkah utama mencakup memeriksa nilai yang hilang, distribusi variabel, dan korelasi antara fitur dengan target.

### 3. Data Preparation (Persiapan Data)

Langkah-langkah penting dalam persiapan data mencakup penanganan nilai yang hilang, feature encoding, normalisasi atau standarisasi, deteksi outlier, pemilihan fitur, dan pembagian dataset menjadi set pelatihan dan pengujian. Langkah ini sangat penting untuk memastikan akurasi model yang dihasilkan.

#### 4. Modeling (Pemodelan)

Beberapa model machine learning yang dapat digunakan untuk prediksi adalah Regresi Logistik, Pohon Keputusan, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), dan Neural Networks. Model dilatih menggunakan set pelatihan, dan hyperparameter dioptimalkan menggunakan cross-validation.

#### 5. Evaluation (Evaluasi)

Metrik evaluasi yang digunakan meliputi akurasi, presisi, recall, F1-Score, dan ROC-AUC. Evaluasi dilakukan menggunakan set pengujian untuk memastikan model dapat melakukan generalisasi dengan baik dan tidak overfitting.

#### 6. Deployment (Penerapan)

Langkah-langkah penerapan meliputi integrasi API, pemantauan model, pengembangan antarmuka pengguna yang sederhana, dan menyediakan dokumentasi yang jelas mengenai penggunaan model, batasannya, serta tindakan yang direkomendasikan berdasarkan prediksi

