

**PREDIKSI PENYAKIT JANTUNG MENGGUNAKAN MODEL LOGISTIC REGRESSION DAN KNN**

**DISUSUN OLEH**

**NAMA :NARAKARTI NUGRAHA**

**NIM :A11.2024.15787 KELOMPOK :A11.4519**

**Topik / Judul**

“ PREDIKSI PENYAKIT JANTUNG DENGAN LOGISTIC REGRESSION DAN KNN “

# Deskripsi Singkat

Dataset ini berisi 11 fitur yaitu Age, Sex, Chest Pain Type, Resting Blood Pressure, Cholesterol, Fasting Blood Sugar, Resting Electrocardiographic Results, Maximum Heart Rate Achieved, Exercise Induced Angina, Oldpeak, Slope, Number of Major Vessels, dan Thalassemia, serta label HeartDisease yang menunjukkan apakah mempunyai gagal jantung (1) atau tidak (0). Proses pengolahan dan analisis data ini merujuk pada pendekatan yang dijelaskan dalam literatur https://github.com/RizkyAdhiNugroho/heart-disease-prediction-with-Logistic-Regression-KNN untuk deteksi penyakit jantung.

# Masalah

Penyakit Jantung adalah penyebab kematian nomor 1 di dunia. Sepertiga dari kematian ini terjadi pada orang yang berusia di bawah 70 tahun. Gagal jantung adalah kejadian umum yang disebabkan oleh CVD, dan dataset ini memuat 11 fitur yang dapat digunakan untuk memprediksi kemungkinan penyakit jantung.

# Tujuan:

Melatih model machine learning Logistic Regression dan Model K-Nearest Neighboor (KNN) yang dapat memprediksi penyebab penyakit jantung.

# Penjelasan Dataset

* + **Penjelasan Atribut**:
    - Age = umur

# Tipe Data: Numerik .

* + - Sex: jenis kelamin, dengan kategori (M = laki, F = perempuan ).

# Tipe Data: Kategorikal

* + - ChestPain: tipe nyeri dada (**NAP**: **Non-Anginal Pain**, **ATA**: **Angina Tipe Akut**, **ASY**: **Angina Stabil** atau **Angina yang Asimtomatik**, **TA**: **Tachycardia Angina**).

# Tipe Data: Kategorikal

* + - RestingBP: Trestbps: tekanan darah istirahat (dalam mm Hg ).

# Tipe Data: Numerik

* + - Cholesterol: zat lemak yang terdapat dalam tubuh (mg/dL).

# Tipe Data: Numerik

* + - FastingBS: Kadar gula darah puasa **1**: Jika FastingBS ≥ 120 mg/dL (menunjukkan gula darah tinggi). **0**: Jika FastingBS < 120 mg/dL (normal).

# Tipe Data: Biner

* + - RestingECG: pemeriksaan non-invasif untuk merekam aktivitas listrik jantung saat tubuh berada dalam kondisi istirahat.

# Tipe Data: Kategorikal

* + - MaxHR: (Maximum Heart Rate) atau Denyut Jantung Maksimum.

# Tipe Data: Numerik

* + - ExerciseAngina: rasa nyeri atau ketidaknyamanan di dada yang terjadi selama aktivitas fisik atau olahraga.

# Tipe Data: Kategorikal

* + - Oldpeak: parameter yang menggambarkan depresi segmen ST pada elektrokardiogram (ECG) yang diukur setelah latihan atau tes stres jantung.

# Tipe Data: Numerik

* + - ST\_Slope: parameter dari elektrokardiogram (ECG) yang menggambarkan kemiringan segmen ST pada grafik ECG.

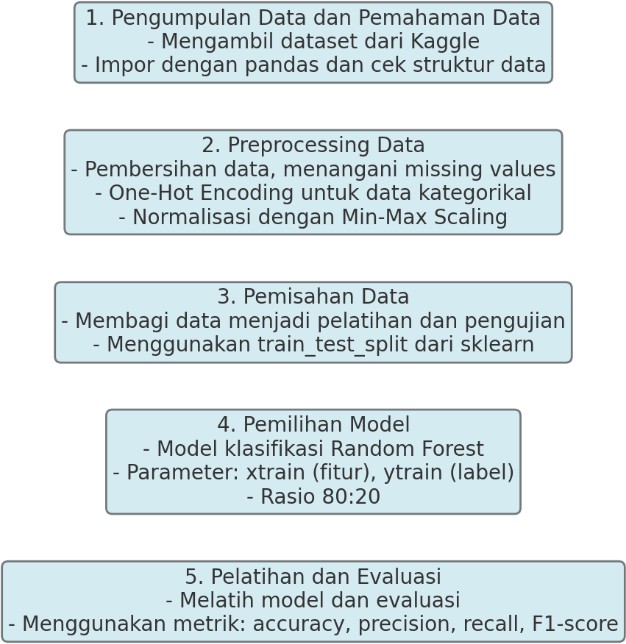
# Tipe Data: Biner

* + **Penjelasan Label**:
    - HeartDisease: Label yang menunjukkan apakah terindikasi penyakit jantung (1) atau tidak (0).

# Tipe Data: Biner

1. **Alur / Tahapan / Kerangka Eksperimen**
2. **Pengumpulan Data dan Pemahaman Data**:Mengambil dataset dari kaggle **https://**[**www.kaggle.com/datasets/bhadramohit/credit-card-fraud-**](http://www.kaggle.com/datasets/bhadramohit/credit-card-fraud-detection)[**detection**](http://www.kaggle.com/datasets/bhadramohit/credit-card-fraud-detection). Impor dataset dengan pandas dan cek struktur data dengan metode seperti .head(), .info(), dan .describe()
3. **Preprocessing Data**: Melakukan pembersihan data, menangani missing values (jika ada), encoding data kategorikal dengan mengubah data kategorikal seperti TransactionType dan Location menjadi angka menggunakan **One-Hot Encoding**, dan normalisasi data menggunakan Teknik Min-Max Scaling untuk memastikan semua fitur memiliki skala yang seragam.
4. **Pemisahan Data**: Membagi dataset menjadi data pelatihan dan pengujian menggunakan train\_test\_split dari sklearn.
5. **Pemilihan Model**: Menggunakan model klasifikasi random forest dengan parameter xtrain yang mewakili fitur (atau variabel independen) dan ytrain parameter yang mewakili label (atau variabel dependen) dari dataset. Menggunakan rasio 80:20..
6. **Pelatihan dan Evaluasi**: Melatih model dan mengevaluasi kinerjanya menggunakan metrik seperti seperti **accuracy** , **precision** , **recall**, dan **F1-score** dengan Parameter: yTest (label asli), yPred (label prediksi dari model).

Bagan



1. **Timeline Eksperimen**

# Pertemuan 1: Selasa, 12 November 2024

* **Kegiatan**: Pengumpulan Data dan Pemahaman Data
  + Mengunduh dataset dari Kaggle.
  + Mengimpor data dengan pandas dan mengecek struktur data menggunakan

.head(), .info(), dan .describe().

# Pertemuan 2: Selasa, 19 November 2024

* **Kegiatan**: Preprocessing Data (Bagian 1)
  + Membersihkan data: menangani missing values dan menghapus data duplikat (jika ada).

# Pertemuan 3: Selasa, 26 November 2024

* **Kegiatan**: Preprocessing Data (Bagian 2)
  + Normalisasi data dengan Min-Max Scaling.

# Pertemuan 4: Selasa, 3 Desember 2024

* **Kegiatan**: Pemisahan Data
  + Membagi dataset menjadi data pelatihan dan pengujian dengan train\_test\_split (rasio 80:20).

# Pertemuan 5: Selasa, 10 Desember 2024

* **Kegiatan**: Pemilihan Model dan Konfigurasi
  + Memilih model Logistic regression dan K-Nearest Neighbors dengan parameter xtrain yang mewakili fitur (atau variabel independen) dan ytrain parameter yang mewakili label (atau variabel dependen) dari dataset

# Pertemuan 6: Selasa, 17 Desember 2024

* **Kegiatan**: Pelatihan dan Evaluasi Model
  + Melatih model dan mengevaluasi kinerjanya.
  + Menghitung metrik evaluasi seperti accuracy, precision, recall, dan F1- score dengan Parameter: yTest (label asli), yPred (label prediksi dari model).

# Pertemuan 7 (Opsional): Selasa, 24 Desember 2024

* **Kegiatan**: Review dan Perbaikan
  + Pengecekan Kembali jika ada kesalahan serta perbaikan Kembali.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Kegiatan | Minggu Ke | | | | | | |
| I | II | III | IV | V | VI | VII |
| Pertemuan 1: Pengumpulan Data dan Pemahaman Data |  |  |  | | | | |
| Pertemuan 2: Preprocessing  Data (Bagian 1) |  |  |  | | | | |
| Pertemuan 3: Preprocessing  Data (Bagian 2) |  |  |  |  | | | |
| Pertemuan 4: Pemisahan Data |  |  |  |  |  | | |
| **Kegiatan**: Pemilihan Model dan Konfigurasi |  |  |  | |
| Pertemuan 6: Pelatihan dan Evaluasi Model |  |  |  |  | |  |  |
| Pertemuan 7: Review dan Perbaikan |  |  |  |  |  |  |  |

# Sumber Dataset https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction

**Literasi https://github.com/RizkyAdhiNugroho/heart-disease-prediction-with-Logistic-Regression-KNN**