**KLASIFIKASI PENYAKIT JANTUNG PADA INDIVIDU MENGGUNAKAN ALGORITMA DECISION TREE**



Disusun Oleh :

|  |  |
| --- | --- |
| Dhea Sabilla | A11.2022.14102 |

**TEKNIK INFORMATIKA**

**UNIVERSITAS DIAN NUSWANTORO**

**2025**

**DAFTAR ISI**

**ISI**

1. Judul / Topik Project & Identitas
2. Judul / Topik Project : Klasifikasi Risiko Penyakit Jantung
3. Identitas :

Nama : Dhea Sabilla

Nim : A11.2022.14102

Kelompok :

1. Ringkasan dan Permasalahan Project + Tujuan yang akan dicapai + Model / Alur Penyelesaian

C:\Users\HP\Downloads\DM.drawio (1).png

1. Data Collection (Pengumpulan Data) :

Proses mengumpulkan data mentah dari berbagai sumber.

1. Data Preparation (Persiapan Data) :

Pengolahan data mentah agar siap digunakan dalam analisis atau pemodelan. Menghapus nilai duplikat dalam dataset, mengganti nilai kosong dengan rata-rata, atau mengonversi data kategori menjadi numerik.

1. Feature Engineering :

Klasifikasi dan akan mengukur keakuratan klasifikasinya, memisahkan data tanpa label/data atribut (X) dan data label (Y).

1. Model Training :

Setelah data diproses (tahap pengumpulan, persiapan dan rekayasa fitur) untuk melatih model machine learning.

Proses pelatihan melibatkan mengoptimalkan parameter model agar sesuai dengan data pelatihan. Tugas ini penulis menggunakan algoritma Decision Tree. Algoritma ini bekerja dengan membangun model berbentuk pohon keputusan (decision tree), di mana setiap simpul (node) merepresentasikan fitur tertentu, setiap cabang (branch) merepresentasikan hasil uji atau keputusan, dan setiap daun (leaf) merepresentasikan hasil atau prediksi.

1. Model Evaluation :

Memastikan model yang dibangun dapat membuat prediksi yang akurat dan berguna pada data yang belum pernah dilihat sebelumnya (data baru atau data tes). Salah satu caranya dengan Akurasi, Akurasi dapat mempresentasikan prediksi yang benar secara keseluruhan.

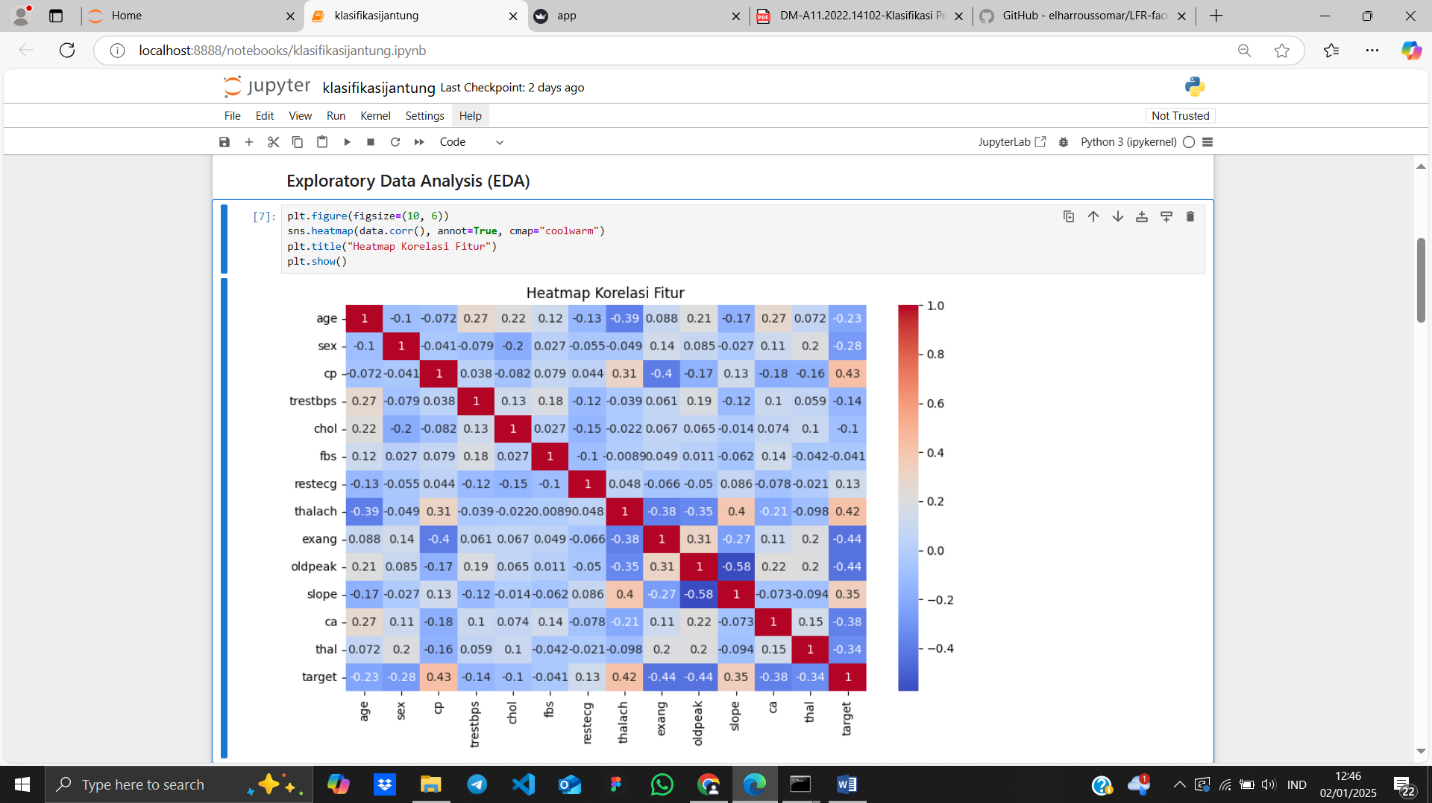
1. Model Deployment :

Membangun model adalah agar dapat digunakan secara praktis dalam aplikasi nyata.

1. Penjelasan Dataset, EDA dan Proses Features Dataset
2. Penjelasan Dataset

* age = Usia
* sex = Jenis kelamin
* cp = Jenis nyeri dada (4 nilai) = Rentang 0 - 3 seberapa nyeri dada pasien
* trestbps = Tekanan darah saat istirahat (Satuan dalam mmHg )
* chol = Kadar Kolestrol serum dalam mg/dl
* fbs = Gula darah puasa > 120 mg/dl ? (1 = Ya, 0 = Tidak)
* restecg = Hasil elektrokardiografi setelah istirahat (nilai 0,1,2)
* thalach = Detak jantung maksimum yang dicapai
* exang = Mengalami angin setelah berolahraga? (1= Ya, 0 = Tidak)
* oldpeak = Depresi Segmen ST yang disbebakan oleh olahraga relatif terhadap istirahat
* slope = Gradien/Kemiringan Segmen ST
* ca = Jumlah pembuluh darah utama (0-3) yang diwarnai dengan flourosopi
* thal = Penyakit Menurun/Genetik Thalasemia : 1 = normal; 2 = cacat ; 3 = cacat yang dapat diperbaiki(carrier)
* target = 1(Ada Penyakit Jantung) / 0(Tidak Ada Penyakit Jantung)

1. EDA



1. Proses Learning / Modeling
2. Performa Model
3. Diskusi Hasil dan Kesimpulan
4. Link Github
5. Link Streamlit