

NAMA : APRIL H MARBUN

NIM : 1103202039

Dalam kasus dataset penyakit jantung, EDA (Exploratory Data Analysis) akan memiliki fokus yang lebih spesifik terkait dengan kondisi medis. Berikut adalah beberapa poin yang perlu diperhatikan dalam EDA dataset penyakit jantung:

1. **Pemahaman Variabel:** Penting untuk memahami makna setiap variabel dalam dataset. Misalnya, apa yang diukur oleh masing-masing variabel, apakah itu faktor risiko, gejala, atau hasil tes medis.
2. **Statistik Deskriptif:** Anda harus melihat statistik deskriptif untuk variabel numerik seperti usia, tekanan darah, kadar kolesterol, denyut jantung, dan lainnya. Ini akan membantu Anda memahami distribusi dan karakteristik utama dari variabel-variabel ini.
3. **Distribusi Kategori:** Untuk variabel kategorikal seperti jenis kelamin, riwayat merokok, dan jenis nyeri dada, Anda perlu memeriksa distribusi kategori dan jumlah penderita penyakit jantung dalam masing-masing kategori.
4. **Korelasi:** Anda mungkin ingin menganalisis korelasi antara variabel-variabel tertentu dengan penyakit jantung. Apakah ada korelasi yang signifikan antara usia dan risiko penyakit jantung, misalnya?
5. **Visualisasi:** Menggunakan grafik seperti histogram, box plot, dan scatter plot, Anda dapat memvisualisasikan distribusi variabel dan hubungan antar variabel. Ini dapat membantu Anda melihat tren dan pola dalam data.
6. **Pengolahan Nilai yang Hilang:** Anda perlu memeriksa apakah ada data yang hilang (missing values) dalam dataset. Jika ada, Anda perlu memutuskan cara mengatasinya, seperti mengisi nilai-nilai yang hilang atau menghapus baris atau kolom yang mengandung nilai yang hilang.
7. **Mendeteksi Outlier:** Outlier (data ekstrem) dapat memengaruhi analisis. Anda perlu mencari tahu apakah ada outlier dalam data, dan memutuskan apakah Anda ingin membuangnya atau memperlakukannya secara khusus.
8. **Analisis Khusus Gender:** Jika dataset berisi informasi gender, Anda mungkin ingin memeriksa perbedaan dalam risiko penyakit jantung antara pria dan wanita.
9. **Menganalisis Hubungan antara Faktor Risiko dan Hasil:** Anda dapat mengeksplorasi hubungan antara faktor risiko seperti merokok, tekanan darah tinggi, kadar kolesterol, dan lainnya dengan hasil diagnosa penyakit jantung. Apakah faktor-faktor ini berpengaruh signifikan terhadap risiko penyakit jantung?
10. **Membuat Hipotesis Medis:** Berdasarkan hasil EDA, Anda dapat mengembangkan hipotesis medis yang dapat diuji dengan metode analisis lebih lanjut.

EDA dalam dataset penyakit jantung akan membantu memahami karakteristik populasi yang berisiko terhadap penyakit jantung, faktor-faktor risiko yang signifikan, dan dapat membantu mempersiapkan data untuk tahap selanjutnya dalam pengembangan model prediksi atau analisis statistik.