

Rizqi Alif Permana Putra

Kelompok 2

10681152



Eksplorasi Data

Eksplorasi data adalah langkah awal yang sangat penting dalam proses analisis data. Pada tahap ini, tujuannya adalah untuk memahami karakteristik dataset secara menyeluruh, termasuk struktur, distribusi, dan pola yang ada dalam data. Dengan melakukan eksplorasi data, kita dapat mengidentifikasi pola yang berharga, menemukan hubungan antar variabel, mendeteksi anomali, serta melihat kemungkinan adanya data yang hilang atau outliers. Proses ini juga membantu dalam mempersiapkan data agar lebih siap untuk dianalisis lebih lanjut, seperti dalam pembuatan model atau pengambilan keputusan berbasis data.



Kenapa Eksplorasi Data Penting?

1 Memahami Dataset

Eksplorasi data membantu kita memahami struktur, pola, dan keunikan dataset yang akan dianalisis.

2 Identifikasi Masalah

Proses eksplorasi dapat mengungkap isu-isu dalam data seperti data hilang, outlier, atau inkonsistensi.

3 Perencanaan Analisis

Wawasan yang diperoleh dari eksplorasi data dapat membantu merencanakan langkah-langkah analisis yang tepat.



Mempersiapkan Dataset

Unduh Data

Pastikan Anda memiliki akses ke dataset yang relevan dengan proyek Anda.

Periksa Struktur

Pahami format data, jumlah baris/kolom, dan tipe data dari setiap fitur.

Pembersihan Awal

Lakukan pembersihan awal untuk menangani kesalahan atau inkonsistensi dalam data.

Mengeksplorasi Statistik Dasar

1

Statistik Deskriptif

Pelajari ukuran pemusatan dan penyebaran data seperti mean, median, standar deviasi, dll.

2

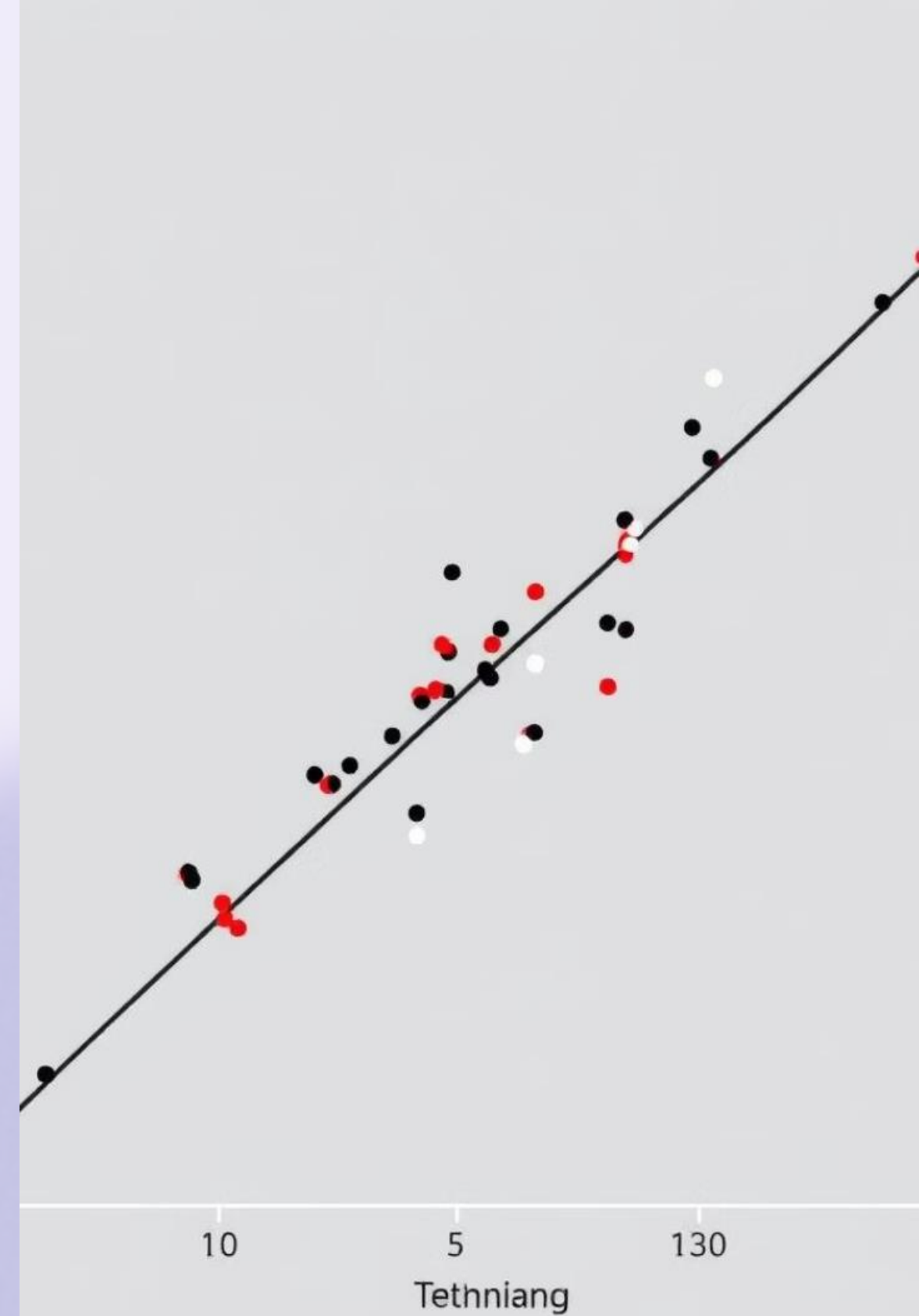
Distribusi Data

Gunakan histogram, box plot, atau plot lainnya untuk melihat distribusi variabel numerik.

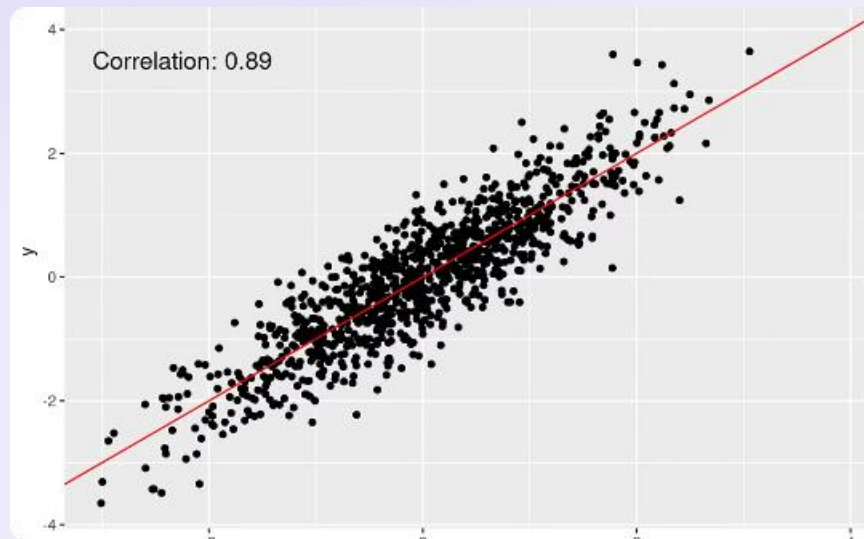
3

Korelasi Antar Fitur

Cari tahu hubungan antar fitur dengan menghitung koefisien korelasi.

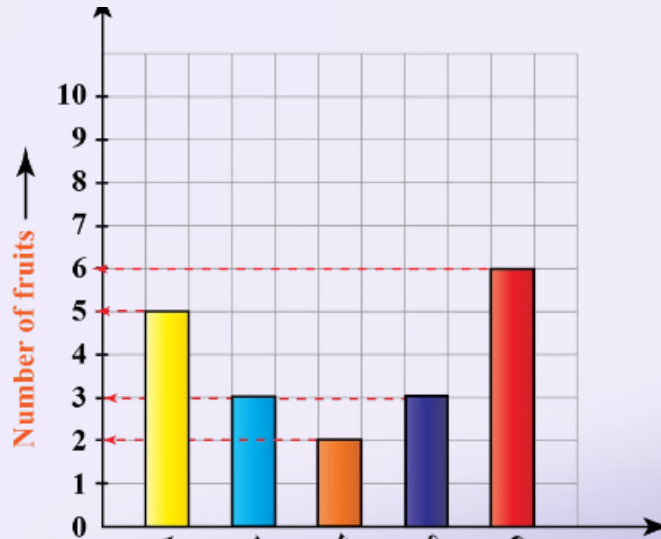


Visualisasi Data Awal



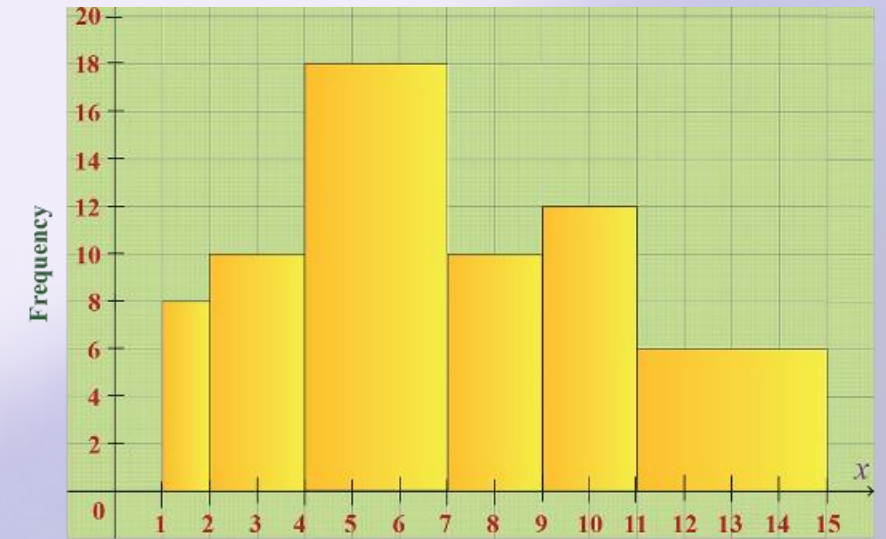
Scatterplot

Memetakan hubungan antara dua variabel numerik.



Bar Chart

Membandingkan nilai-nilai kategorikal atau frekuensi.



Histogram

Melihat distribusi dan frekuensi variabel numerik.

Menemukan Hubungan Antar Variabel

1

Korelasi

Analisis korelasi untuk mengidentifikasi hubungan linier antar variabel.

2

Regresi

Model regresi dapat memprediksi nilai satu variabel berdasarkan variabel lain.

3

Anova

Uji ANOVA untuk melihat perbedaan signifikan antar kelompok dalam data.





Messing

Values 1 1 Lpag
abus 1 x 1 Datg

Menangani Data Hilang dan Outlier

Identifikasi Data Hilang

Kenali pola dan proporsi data hilang untuk memutuskan strategi penanganannya.

Strategi Pengisian

Isi data hilang dengan nilai tengah, regresi, atau teknik lain yang sesuai.

Deteksi Outlier

Gunakan metode statistik untuk menemukan dan menangani nilai-nilai outlier.

Kesimpulan



Rangkum Wawasan

eksplorasi data dan pemrosesan dataset merupakan langkah-langkah krusial dalam analisis data. Eksplorasi data membantu kita memahami karakteristik dataset, mengidentifikasi masalah seperti data hilang atau outliers, dan merencanakan langkah-langkah analisis yang tepat. Mempersiapkan dataset melibatkan pengunduhan data, pemeriksaan struktur, dan pembersihan awal untuk memastikan data bersih dan siap untuk dianalisis. Statistik dasar memberikan gambaran tentang ukuran pemusatan, penyebaran, dan hubungan antar fitur, sementara visualisasi data awal, seperti scatterplot, bar chart, dan histogram, memungkinkan kita untuk memetakan hubungan antar variabel, membandingkan nilai-nilai kategorikal, dan melihat distribusi variabel numerik. Semua langkah ini sangat penting untuk menghasilkan analisis yang akurat dan informatif.

