

Step 1

Dataset kelulusan_mahasiswa.csv

Isi 10 baris data dengan kolom IPK, Jumlah_Absensi, Waktu_Belajar_Jam, dan Lulus.

Step 2

Baca data menggunakan pandas

Perintah(`pd.read_csv("kelulusan_mahasiswa.csv")`)

Data tampil di terminal (10 baris, 4 kolom)

Step 3

Cek nilai kosong, tidak ada

Hapus duplikat, tidak ada yang terhapus

Tampilkan boxplot IPK, hasilnya normal

Step 4

Lihat statistik data dengan `df.describe()`

Buat histogram IPK, scatterplot IPK vs Waktu_Belajar_Jam, dan heatmap korelasi

Semua grafik tampil tanpa error

Step 5

Tambahkan dua kolom baru

$\text{Rasio_Absensi} = \text{Jumlah_Absensi} / 14$

$\text{IPK_x_Study} = \text{IPK} * \text{Waktu_Belajar_Jam}$

Simpan hasil ke `processed_kelulusan.csv`

Step 6

Pisahkan dataset jadi train, validation, dan test

Proporsi: 70%, 15%, 15%

Output shape: (7,6), (1,6), (2,6)