

LAPORAN PRAKTIKUM PENAMBANGAN DATA

Eksplorasi Data: Global Coffee Health Dataset

Nama : Wayan Raditya Putra

NRP : 5054241029 (RKA)

Asisten Dosen : Willy

1. Deskripsi Dataset

Dataset yang digunakan adalah **Global Coffee Health Dataset** dengan jumlah data 10.000 baris dan 16 atribut. Atribut yang tersedia meliputi data demografi (usia, gender, negara), kebiasaan (coffee intake, sleep hours, physical activity, smoking, alcohol consumption), serta kesehatan (BMI, heart rate, stress level, sleep quality, health issues). Tipe data terdiri dari numerik (misalnya Age, Coffee_Intake, Sleep_Hours, Heart_Rate) dan kategorikal (misalnya Gender, Sleep_Quality, Stress_Level).

2. Statistik Deskriptif

	ID	Age	Coffee_Intake	Caffeine_mg	Sleep_Hours	BMI	Heart_Rate	Physical_Activity_Hours	Smoking	Alcohol_Consumption
count	10000.00000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.00000	10000.00000	10000.000000
mean	5000.50000	34.949100	2.509230	238.411010	6.636220	23.986860	70.617800	7.48704	0.20040	0.300700
std	2886.89568	11.160939	1.450248	137.748815	1.222055	3.906411	9.822951	4.31518	0.40032	0.458585
min	1.00000	18.000000	0.000000	0.000000	3.000000	15.000000	50.000000	0.00000	0.00000	0.000000
25%	2500.75000	26.000000	1.500000	138.750000	5.800000	21.300000	64.000000	3.70000	0.00000	0.000000
50%	5000.50000	34.000000	2.500000	235.400000	6.600000	24.000000	71.000000	7.50000	0.00000	0.000000
75%	7500.25000	43.000000	3.500000	332.025000	7.500000	26.600000	77.000000	11.20000	0.00000	1.000000
max	10000.00000	80.000000	8.200000	780.300000	10.000000	38.200000	109.000000	15.00000	1.00000	1.000000

Beberapa atribut numerik utama:

- **Age**: rata-rata 34.9 tahun, dengan rentang 18–80 tahun.
- **Coffee_Intake**: rata-rata 2.5 cangkir/hari, maksimum 8.2 cangkir.
- **Sleep_Hours**: rata-rata 6.6 jam, rentang 3–10 jam.
- **BMI**: rata-rata 24, menunjukkan mayoritas responden berada di kategori normal.
- **Heart_Rate**: rata-rata 70.6 bpm, dengan rentang 50–109 bpm.

Nilai rata-rata dan sebaran data menunjukkan populasi cukup beragam, dari non-coffee drinker hingga heavy drinker, serta pola tidur yang bervariasi.

3. Penanganan Data

- **Missing Value**: ditemukan pada kolom Health_Issues dengan 5.941 nilai kosong (>50%). Karena proporsi terlalu besar dan tidak digunakan dalam analisis utama, kolom ini dihapus.

- **Outlier:**

- **Heart_Rate:** ditemukan outlier >95 bpm hingga 109 bpm. Namun masih dalam batas fisiologis (60–100 bpm normal, >100 bpm dapat terjadi karena stress/kafein), sehingga tetap dipertahankan.
- **Coffee_Intake:** terdapat outlier >6 cangkir (maks 8.2 cangkir). Angka ini masih mungkin terjadi pada heavy coffee drinker.
- **Sleep_Hours:** relatif stabil, dengan nilai 3–10 jam yang masih realistis.

Kesimpulan: outlier tidak dihapus karena masih merepresentasikan kondisi nyata.

4. Visualisasi & Interpretasi

a. Histogram Heart Rate

Distribusi heart rate berbentuk normal dengan puncak di sekitar 70 bpm. Mayoritas responden berada pada rentang 60–80 bpm, sesuai dengan kondisi jantung sehat.

b. Scatter Plot Coffee Intake vs Sleep Hours

Sebaran titik menunjukkan tidak ada korelasi kuat antara jumlah kopi yang dikonsumsi dan lama tidur. Namun, terlihat pola bahwa beberapa individu dengan konsumsi tinggi tetap memiliki jam tidur normal.

c. Bar Chart Sleep Quality

Mayoritas responden melaporkan kualitas tidur **Good** (>50%). Sementara itu, kategori **Fair** dan **Excellent** berada di tengah, dan kategori **Poor** relatif sedikit. Hal ini menunjukkan persebaran kualitas tidur cenderung positif.

5. Kesimpulan

Eksplorasi data menunjukkan bahwa:

- Rata-rata responden memiliki pola konsumsi kopi sedang (2–3 cangkir) dan jam tidur normal (6–7 jam).
- Outlier yang ditemukan masih wajar dan tetap dipertahankan.
- Distribusi heart rate cenderung normal.
- Kualitas tidur sebagian besar responden berada pada kategori baik.

Dengan demikian, dataset ini siap digunakan untuk analisis lebih lanjut atau pemodelan terkait hubungan gaya hidup dan kesehatan.