



KMMI 2021

Eksplorasi dan Visualisasi Data

Pertemuan 3:
Tipe Data

Pembahasan

1. Definisi atribut/variabel
2. Atribut nominal beserta contohnya
3. Atribut biner beserta contohnya
4. Atribut ordinal beserta contohnya
5. Atribut numerik beserta contohnya
6. Atribut diskrit vs kontinu
7. Identifikasi tipe atribut

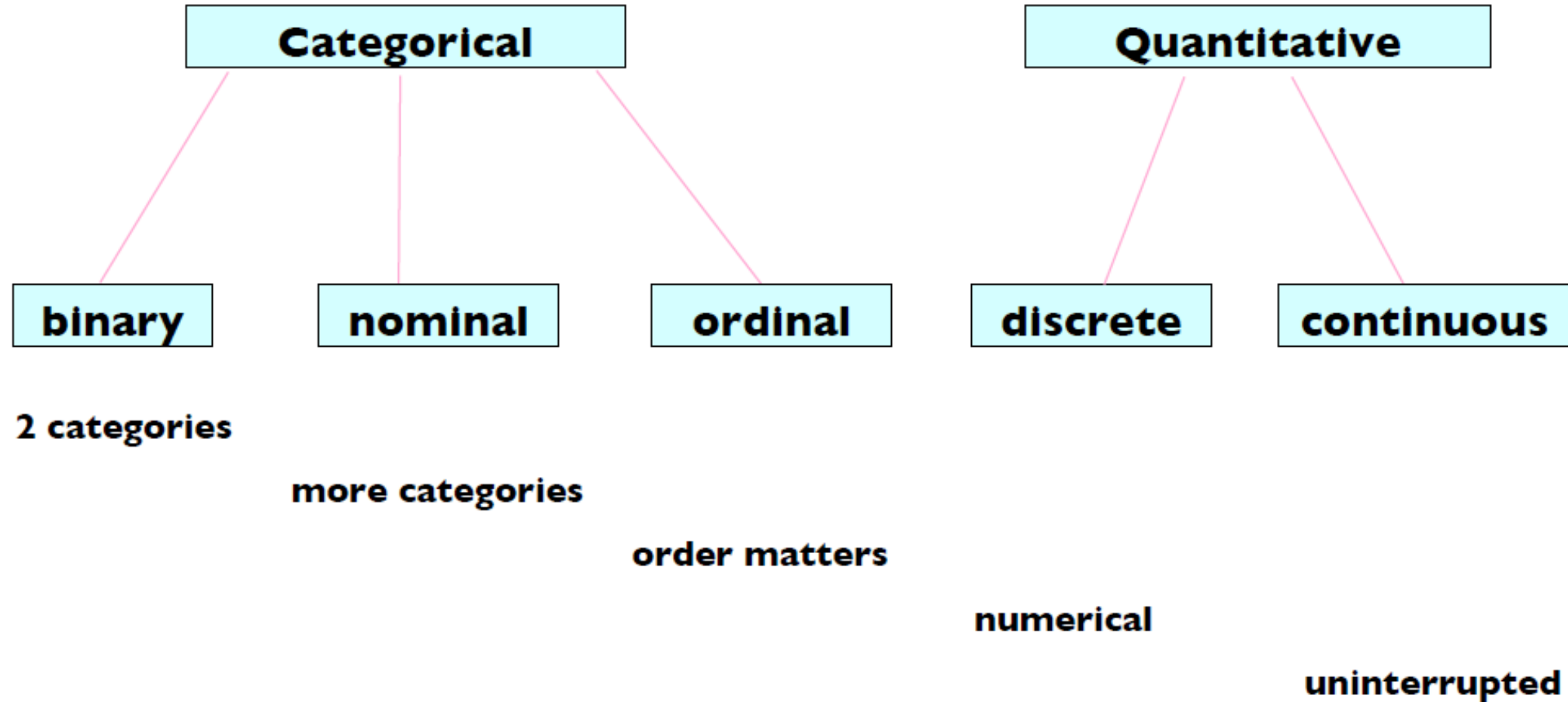
Latar Belakang

- Hal pertama yang harus dilakukan pada dataset apa pun yang anda miliki adalah mengenalinya terlebih dahulu.
- Selanjutnya, yang harus anda lakukan adalah mengidentifikasi tipe data yang anda miliki untuk setiap variabel, misalnya, apakah termasuk dalam tipe data numerik atau kategorik.
- Proses ini juga membantu anda untuk menentukan jenis visualisasi apa yang dapat dibuat dari data tersebut.
- Proses ini juga dilakukan tidak hanya untuk membiasakan diri anda dalam memahami semua data yang telah dikumpulkan, tetapi juga untuk mengurangi beban kerja selama proses analisis nanti.

Atribut

- Atribut mewakili **karakteristik atau fitur** dari objek data.
- Kata lain dari atribut adalah ***dimension, feature, and variable***.
- Istilah dimensi umumnya digunakan dalam data warehouse, fitur untuk pembelajaran mesin (machine learning), dan istilah variabel lebih digunakan dalam bidang statistika.
- Contoh atribut yang menjelaskan objek/entitas pelanggan:
 - ID pelanggan, nama dan alamat.
- Distribusi data yang melibatkan satu atribut (atau variabel) disebut sebagai **univariat**.
- Distribusi data yang melibatkan dua atribut disebut sebagai **bivariat**.

Tipe Data



Data Diskrit

- Data diskrit adalah jenis data kuantitatif yang mengandalkan aspek jumlah.
- Data diskrit hanya berisi nilai-nilai terbatas, mencakup nilai-nilai yang hanya dapat dihitung dalam bilangan bulat atau tidak dapat dipecah menjadi pecahan atau desimal.
- Contoh:
 - jumlah siswa di UNAIR
 - jumlah mobil di tempat parkir
 - jumlah komputer di laboratorium computer
 - jumlah hewan di kebun binatang
 - dll.

Data Kontinum

- Data kontinu adalah data yang diperoleh dari hasil pengukuran, yaitu data yang besarannya dapat menempati semua nilai yang ada di antara dua titik (nilai dari suatu rentang tertentu).
- Contoh:
 - jarak tempuh dari rumah A ke kampus C UNAIR (km)
 - hasil panen petani A (ton)
 - prestasi belajar mahasiswa (IPK)
 - dll.
- Contoh lainnya adalah data terkait usia, tinggi atau berat badan seseorang, suhu, waktu, uang, pendapatan dll.

Data Biner

- Data biner hanya memiliki dua kategori seperti ya/tidak, kepala/ekor, dan sebagainya.
- Saat menggunakan data biner dalam analisis, umumnya dikodekan ke dalam format 0 dan 1.
- Data biner sebenarnya adalah tipe khusus dari data nominal, dengan hanya dua kategori saja.

Data Nominal

- Data nominal juga sering disebut sebagai data kategorik, ini adalah data yang berisi beberapa kelompok (kategori) yang dapat diberikan nilai numerik tetapi tidak memiliki urutan alami.
- Misalnya, berbagai jenis kendaraan seperti sepeda, mobil, truk, kapal, pesawat, dapat diberi nomor 1 sampai 5 untuk memudahkan analisis.
- Namun nilai-nilai itu sendiri tidak ada artinya dan tidak memiliki urutan, mobil memiliki nilai 1 lebih besar dari sepeda hanya berdasarkan penugasan dan bukan karena "lebih baik."

Data Ordinal

- Di sisi lain, data ordinal juga dapat diberikan nilai numerik tetapi memiliki urutan alami.
- Misalnya, reaksi terhadap bahan kimia bisa bernilai “tidak ada”, “ruam”, atau “melepuh” dan mereka dapat diberi nilai 1 hingga 3 dengan 3 jelas lebih parah dari 1, tetapi tidak harus 3 kali lebih parah.
- Contoh lain dari data ordinal adalah skala Likert pada kuesioner yang bernilai tidak setuju sampai sangat setuju.

Tipe Data

- Data juga dapat dikategorikan berdasarkan **sumber datanya** menjadi:
 - Data primer
 - Data sekunder
- Data berdasarkan ketergantungannya terhadap **waktu**:
 - Time Series Data
- Data yang tergantung terhadap **lokasi**:
 - Data Geospasial

Data Primer

- Data primer merupakan jenis data yang dikumpulkan oleh peneliti langsung dari sumber utama.
- Data didapat secara langsung misalnya melalui eksperimen, sensor, observasi langsung, atau survey.
- Data primer bersifat real-time dan berupa data mentah.
- Interpretasi data primer biasanya lebih baik dan kuat dibandingkan dengan data sekunder.
- Hal ini dikarenakan pengambilan data primer dilakukan secara spesifik untuk menjawab suatu hipotesis tertentu.

Data Sekunder

- Data sekunder adalah jenis data yang tidak diambil atau dikumpulkan secara langsung oleh peneliti.
- Data sekunder didapat misalnya dari database/datawarehouse atau berasal dari penelitian-penelitian sebelumnya.
- Data sekunder biasanya bersifat lampau dan terkadang sudah tidak relevan.
- Proses pengumpulannya lebih cepat, murah, dan mudah, namun tidak selalu ada (tersedia).
- Dari segi akurasi data sekunder juga biasanya kurang akurat dan kurang bisa diandalkan dibandingkan dengan data primer.
- Hal ini dikarenakan proses pengumpulan datanya tidak dilakukan secara langsung melainkan kita hanya tinggal menggunakan data yang sudah ada atau sudah diolah.
- Hal ini tentunya dapat menimbulkan bias yang cukup tinggi.

Tipe Data

- Tipe data berdasarkan skala pengukuran data:

Jenis		Deskripsi	Contoh
Kategorik atau kualitatif	Nominal	Skala pengukuran paling rendah dan angka-angka dalam skala nominal tidak mengukur besaran (tidak bermakna kuantitas) tetapi hanya sebagai penggolongan saja (agar dapat dibedakan).	Jenis kelamin, agama, negara, kode pos, warna.
	Ordinal	Mirip dengan skala nominal, akan tetapi angka 1 memiliki nilai lebih tinggi dibandingkan 0. Namun, jarak antara 0 dan 1 tidak terdefinisi dengan jelas. Artinya, tidak memiliki sifat selisih/interval dan perbandingan.	Tingkat pendidikan, ranking, nilai huruf, pangkat militer, tingkat kepuasan: 1 (sangat tidak puas), 2 (tidak puas), 3 (cukup puas), 4 (puas), dan 5 (sangat puas)

Tipe Data

- Tipe data berdasarkan skala pengukuran data:

Jenis		Deskripsi	Contoh
Numerik atau kuantitatif	Interval	Tidak memiliki nol mutlak, memiliki sifat selisih namun tidak memiliki perbandingan.	IQ/EQ/SQ Nilai TOEFL Nilai GRE Nilai TPA Suhu (Celcius dan Kelvin)
	Rasio	Berbeda dengan interval, rasio memiliki nol mutlak dan sifat perbandingan.	Berat badan Tinggi badan Pendapatan/gaji Umur



Terima Kasih