

TRAINING MELEK DATA

INTRODUCTION

Ikang Fadhli

Nutrifood Indonesia

26 November 2021

- 1 *CONCEPTUAL FRAMEWORK*
- 2 *INTRODUCTION TO DATA*
- 3 *KNOWLEDGE ABOUT DIFFERENT DATA TYPES AND FORMATS*
- 4 *BASIC KNOWLEDGE AND UNDERSTANDING DATA*
- 5 *APPLICATIONS OF DATA*

Section 1

CONCEPTUAL FRAMEWORK

CONCEPTUAL FRAMEWORK

Materi pada *training* ini mencakup:

- *Introduction to data*
- *Knowledge about different data types and formats*
- *Basic knowledge and understanding data*
- *Understanding of relevance and applications of data*

Section 2

INTRODUCTION TO DATA

Introduction to Data

Definisi

Data adalah representasi faktual dari observasi (pengamatan).

Data bisa dikelompokkan sesuai dengan tipe dan karakteristiknya. Dengan memahami tipe data yang kita hadapi, kita akan lebih mudah dalam menentukan metode analisa yang tepat.

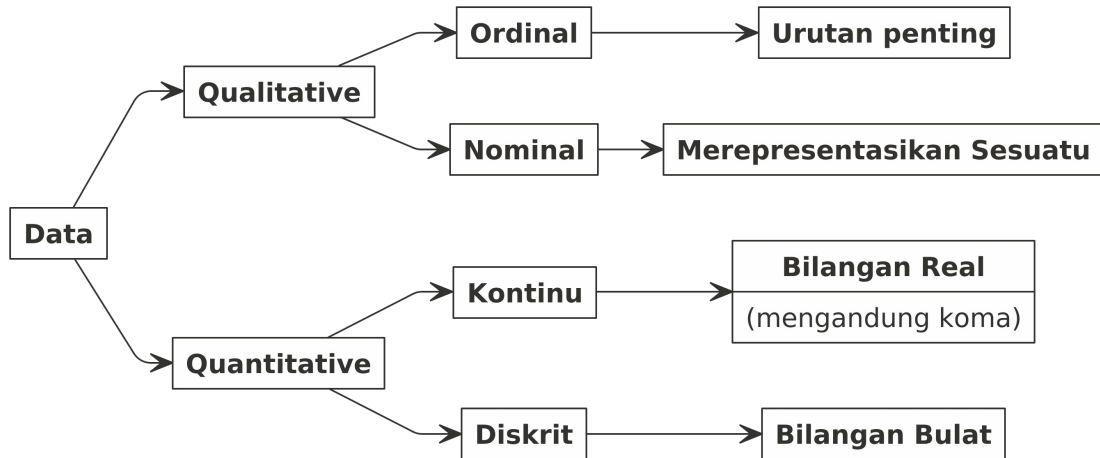
Section 3

KNOWLEDGE ABOUT DIFFERENT DATA TYPES AND FORMATS

Subsection 1

Pengelompokkan Data Berdasarkan Tipe (Statistik)

Pengelompokkan Data Berdasarkan Tipe (Statistik)



Data Kualitatif

Definisi

Data Kualitatif adalah data yang tidak bisa dilakukan operasi aritmatika (penjumlahan, pengurangan, pembagian, dan perkalian). Data seperti ini bisa juga disebut sebagai data kategorik.

- Nominal; Representasi dari sesuatu.
 - Contoh: data seperti *gender*. Misalkan angka 1 saya tulis sebagai representasi dari pria dan 2 sebagai wanita.
- Ordinal; Urutan dari data menjadi penting.
 - Contoh: skala *likert*, misalkan angka 1 - 6 sebagai representasi dari tingkat kesukaan atau kesetujuan (sangat suka sampai sangat tidak suka).

Data Kuantitatif

Definisi

Data Kuantitatif adalah data yang bisa dilakukan operasi aritmatika (penjumlahan, pengurangan, pembagian, dan perkalian). Data seperti ini, kita akan sebut sebagai data numerik.

- Diskrit; bilangan bulat (*integer*).
 - Contoh: banyaknya anak, banyaknya karyawan, dll.
- Kontinu; bilangan *real* (mengandung koma).
 - Contoh: tinggi badan, berat badan, dll.

Contoh Penerapan

Misalkan dalam satu kelas *webinar* berisi 10 orang peserta dengan *gender*:

- 1 7 orang wanita
- 2 3 orang pria

Apakah pernyataan berikut ini benar / salah?

Rata-rata peserta webinar tersebut adalah wanita.

Contoh Penerapan

Pernyataan tersebut **SALAH!**

Contoh Penerapan

Kenapa?

Bagaimana cara menghitung rata-rata?

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Sekarang kita hitung data kita:

Misalkan saya tulis pria sebagai 1 dan wanita sebagai 2. Maka didapatkan:

$$\sum_{i=1}^n x_i = 1 + 1 + 1 + 2 + 2 + 2 + 2 + 2 + 2 + 2 = 17$$

$$n = 10$$

$$\bar{x} = \frac{17}{10} = 1.7$$

Contoh Penerapan

Apa arti 1.7?

Pria yang terlalu kewanitaan?

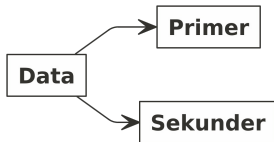
Wanita yang agak kelaki-lakian?

Pernyataan yang **Tepat** adalah **Mayoritas** peserta *webinar* adalah wanita.

Subsection 2

Pengelompokkan Data Berdasarkan Sumber

Pengelompokkan Data Berdasarkan Sumber



Definisi

Data primer adalah data yang diambil dari sumber pertama.

Data sekunder adalah data yang berasal dari sumber yang pernah diolah oleh orang lain.

Contoh Data Primer dan Sekunder

Data Primer

- ➊ Data yang kita himpun sendiri dari hasil interview menggunakan quesioner (*survey*),
- ➋ Data yang kita ambil dari mesin,
- ➌ Data hasil *web scrape*, dll.

Data Sekunder

- ➊ Data kependudukan hasil sensus BPS,
- ➋ Laporan absensi karyawan,
- ➌ Data Covid-19 yang tersedia di *website* resmi pemerintah, dll.

Masalah Seputar Data Primer dan Sekunder

Pertanyaannya:

*Kapan kita memakai **data primer**? Kapan kita memakai **data sekunder**?*

Guidance untuk menjawab pertanyaan:



Penjelasan *Guidance*

Prinsipnya adalah:

Kita akan menggunakan data primer saat data sekunder tidak ada!

Sejujurnya mencari data primer itu relatif sulit. Setidaknya kita membutuhkan waktu, tenaga, dan biaya untuk mencari data langsung dari sumbernya. Contoh: jika saya ingin mencari tahu berapa banyak orang yang teridentifikasi **COVID-19** di suatu kota. Alih-alih saya datang ke semua RS yang ada di kota tersebut, saya cukup cek saja *website* resmi penanganan **Covid-19** pemerintah.

Jadi, jika data sekundernya sudah tersedia kita bisa mempertimbangkan untuk memakai data tersebut daripada mengambil data primer. **TAPI** jika ternyata karakteristik data yang kita mau cari tersebut **sangat dinamis dan cepat berubah** ATAU ada **perbedaan kondisi, situasi atau limitasi** maka kita harus **mempertimbangkan untuk mencari data primer dan tidak menggunakan data sekunder**.

Subsection 3

Pengelompokkan Data Berdasarkan Tipe (Data Sains)

Pengelompokkan Data Berdasarkan Tipe (Data Sains)

Di dalam dunia data sains, ada beberapa tipe data yang sering digunakan. Secara hierarki, bisa diurutkan sebagai berikut:

`character > numeric > integer > logical`

Pengelompokkan Data Berdasarkan Tipe (Data Sains)

Berikut penjelasannya:

- ➊ **character**: merupakan tipe data berupa karakter atau *string*. Semua data bisa dilihat sebagai *character*. Oleh karena itu, secara hierarki tipe data ini ditempatkan di urutan paling atas. Namun, data tipe ini tidak bisa dilakukan operasi aritmatika.
- ➋ **numeric**: merupakan tipe data angka berupa bilangan *real*. Tipe data ini mirip dengan data numerik di poin sebelumnya.
- ➌ **integer**: merupakan tipe data angka berupa bilangan bulat. Sekilas mirip dengan tipe data diskrit di poin sebelumnya. Namun di beberapa kondisi, tipe data ini bisa dijadikan data **kategorik** sehingga kita bisa sebut tipenya menjadi *factor*.
- ➍ **logical**: merupakan tipe data *boolean*. Hanya berisi TRUE atau FALSE. Tipe data ini sangat berguna saat kita melakukan *if conditional*, *looping*, atau membuat *regex* (*regular expression*).

Section 4

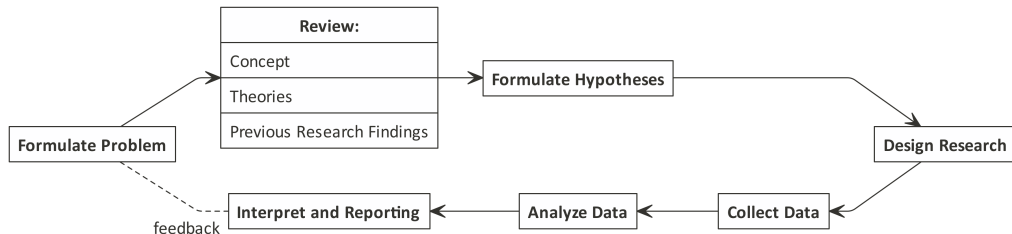
BASIC KNOWLEDGE AND UNDERSTANDING DATA

Subsection 1

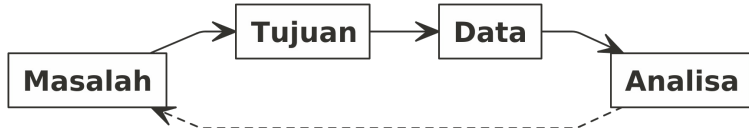
Research Workflow

Proper Steps

Dalam setiap riset atau penelitian, alur pengerjaannya relatif sama seperti ini:



Hubungan Masalah dengan Data



Relevansi Data

Grafik di atas memberikan suatu arahan bahwa data yang dikumpulkan dan dianalisa harus sesuai dengan masalah yang dihadapi.

Subsection 2

Kualitas dari Suatu Data

Kualitas dari Suatu Data

Pertanyaannya:

Data saya sudah bagus belum ya?

Ada beberapa parameter yang bisa digunakan untuk menjawab pertanyaan tersebut:

- ➊ Kualitas dari suatu data diukur dari **seberapa bisa data tersebut memenuhi kebutuhan analisis**. Setelah itu baru dicek secara statistik: *blank* data atau *extreme values*.
- ➋ Data yang bagus biasanya berasal dari *random sampling*.
- ➌ Untuk data yang diambil secara berulang-ulang (harian, mingguan, bulanan, tahunan, atau lainnya), pastikan bahwa cara pengambilan data harus sama.

Subsection 3

Importance of Disclaimer

Importance of Disclaimer

Pertanyaannya:

Apa yang menyebabkan hasil survey / study / penelitian / analisa data tidak sesuai dengan kondisi real?

Ingat bahwa setiap penelitian memiliki **limitasi** dan **asumsi**.

- 1 Perhatikan dengan seksama kondisi yang dilalui saat melakukan penelitian. Apakah ada perbedaan dengan kondisi *real*?
- 2 Perhatikan cara pengambilan data. Apakah ada *bias* yang kita lakukan saat mengambil data?

Contoh

Misalkan:

Survey mengenai makanan favorit orang Indonesia tapi survey dilakukan di RM Padang.

Pertanyaannya:

Kira-kira makanan apa yang menjadi favorit orang Indonesia tersebut?

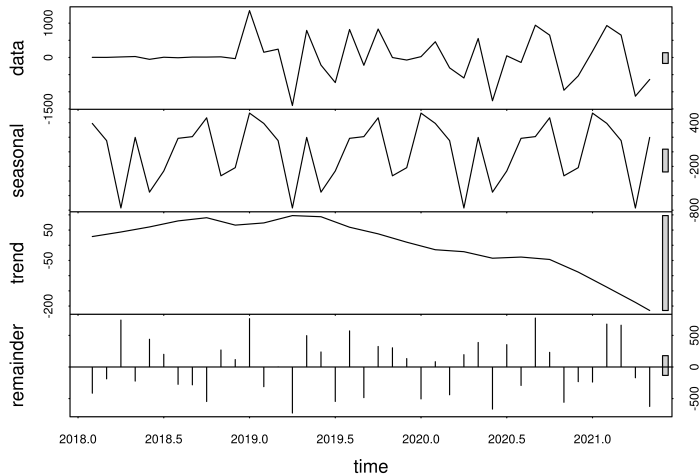
Section 5

APPLICATIONS OF DATA

Subsection 1

Aplikasi Data di *Sales*

Aplikasi Data di Sales



Tim sales memiliki data omset bulanan suatu produk tertentu. Dari data tersebut, bisa kita analisa tren penjualan, efek seasonal, efek promosi, pengelompokkan customer, dll.

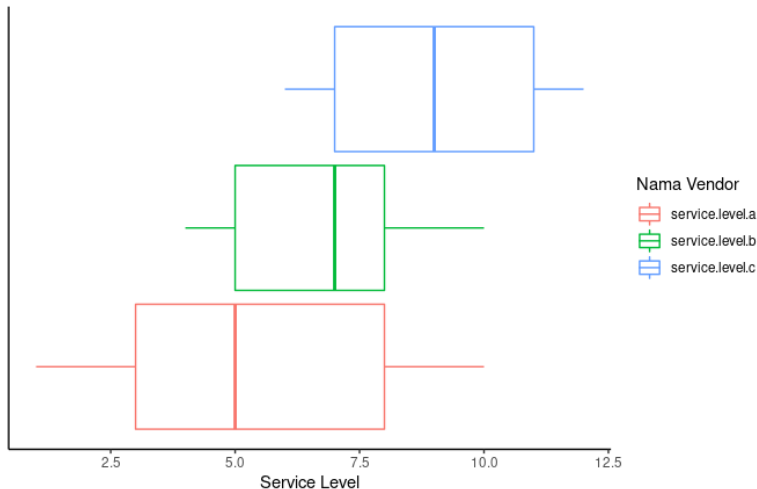
Subsection 2

Aplikasi Data di *Procurement*

Aplikasi Data di *Procurement*

Box Plot dari Masing-Masing Vendor

Data Tiga Vendor



Tim procurement hendak mengambil keputusan strategic terkait 3 vendor yang selama ini digunakan. Mereka mencoba service level antara ketiganya. Siapa yang terbaik? Apakah ketiganya masih sesuai dengan spek minimal yang dipersyaratkan?