

EKMA6214  
Edisi 1

MODUL 01

# Eksplorasi Data

Dr. Fatia Fatimah, S.Si., M.Pd.

# Daftar Isi

<b>Modul 01</b>	<b>1.1</b>
Eksplorasi Data	
<b>Kegiatan Belajar 1</b>	<b>1.4</b>
Pengenalan Data	
<b>Latihan</b>	<b>1.10</b>
<b>Rangkuman</b>	<b>1.10</b>
<b>Tes Formatif 1</b>	<b>1.11</b>
<b>Kegiatan Belajar 2</b>	<b>1.12</b>
Presentasi Data	
<b>Latihan</b>	<b>1.23</b>
<b>Rangkuman</b>	<b>1.24</b>
<b>Tes Formatif 2</b>	<b>1.24</b>
<b>Kunci Jawaban Tes Formatif</b>	<b>1.26</b>
<b>Daftar Pustaka</b>	<b>1.27</b>



## Pendahuluan

**M**ahasiswa Program Doktor Universitas Terbuka, selamat datang pada mata kuliah Analisis dan Interpretasi Data. Saat ini kita sedang mempelajari Modul 1. Modul 1 memuat dua kegiatan belajar. Kegiatan belajar pertama berkaitan dengan sekilas pandang tentang pengenalan data. Adapun kegiatan belajar kedua membahas tentang presentasi data.

Meski informasi tentang eksplorasi data sudah pernah Anda pelajari pada jenjang Strata Satu dan Dua namun masih perlu untuk dikaji ulang kembali agar perspektif tentang data dapat diimplementasikan sesuai kebutuhan disertasi Anda. Eksplorasi data menjadi penting karena hal ini merupakan tahapan awal dalam memahami data sebelum Anda melakukan analisis.

Keingintahuan tentang data perlu dimiliki oleh seorang peneliti agar informasi tidak bersifat tunggal melainkan data didapat dari berbagai pihak. Pada kondisi tertentu, kita temui bahwa pemilik data bahkan tidak tahu masalah yang sedang mereka hadapi. Data yang banyak menjadi sekedar rutinitas sehingga tidak menganggap ada keganjilan atau merasa tidak perlu melakukan inovasi perbaikan. Misal seorang CEO (*Chief Executive Officer*) memiliki ratusan bahkan ribuan data transaksi penjualan setiap bulannya tetapi tidak tahu cara menggunakan data tersebut untuk mendapatkan informasi berguna terkait peningkatan penjualan agar dapat lebih baik lagi.

Setelah mempelajari modul ini diharapkan mahasiswa mampu menguraikan penggunaan eksplorasi data. Secara khusus, mahasiswa diharapkan mampu untuk:

1. mengidentifikasi berbagai jenis data,
2. menguraikan tahapan agar diperoleh data bermakna,
3. mempresentasikan data untuk satu variabel,
4. mempresentasikan data untuk variabel lebih dari satu.

**Selamat membaca!**

## Pengenalan Data

**D**ata dapat berasal dari berbagai sumber. Metode pengumpulan data juga beragam. Sebagai contoh, data yang diambil dari instrumen berupa survei atau wawancara. Instrumen tersebut dibutuhkan untuk mendapatkan informasi tentang pendapat, preferensi dan perilaku. Proses untuk mendapatkan data dapat dilakukan secara langsung dimana peneliti berhadapan tatap muka dengan responden, atau peneliti cukup menghubungi responden melalui telepon, atau data hanya diperoleh melalui isian formulir secara daring atau peneliti bertemu responden secara virtual menggunakan aplikasi webinar.

Untuk menghindari bias, responden terpilih dapat diambil secara acak. Selanjutnya pertanyaan atau pernyataan kuesioner tidak boleh mengarahkan ke jawaban yang diharapkan. Hal yang perlu menjadi perhatian juga adalah waktu pengambilan survei apakah pagi, siang, sore, atau malam sesuai karakteristik jenis pertanyaan dan profil responden. Serta perlu kejelasan perkiraan biaya yang ditimbulkan selama proses pengumpulan data. Pada Kegiatan Belajar (KB) 1 ini beberapa istilah sudah umum Anda kenal dan bersifat pengulangan informasi tentang data yang disarikan dari beberapa referensi (Albright & Winston, 2017; Anderson, 2011; Keller, 2018).

### A. KONSEP DASAR TENTANG DATA

Pada bagian ini, kita akan mengulas singkat tentang konsep dasar tentang data. Sebelum membahas tentang data, kita awali dengan konsep populasi dan sampel.

Populasi terdiri dari semua item atau individu yang ingin didapat kesimpulannya. Parameter adalah ukuran yang menggambarkan karakteristik populasi. Beberapa contoh populasi sebagai berikut: semua pembeli produk X, semua pemilih potensial dalam pemilihan presiden atau kepala daerah, semua mahasiswa baru, semua pelanggan jasa konsultasi, semua pasien positif Covid-19 (*corona virus disease* 2019). Pada kondisi tertentu, sangat sulit mendapatkan informasi dari semua anggota populasi. Seperti halnya akan memakai biaya dan waktu yang lama untuk mewawancarai semua warga tentang hak mereka dalam pemilihan presiden atau kepala daerah. Begitu juga halnya akan sangat berisiko untuk mendapatkan informasi dan riwayat semua pasien positif Covid-19 di Indonesia untuk dianalisa.

Oleh karena itu, kita dapat mendekati karakteristik populasi melalui sampel. Ukuran yang menggambarkan karakteristik sampel dinamakan statistik. Sampel adalah bagian dari populasi yang akan dianalisis sehingga sampel haruslah representatif dari populasi agar dengan cukup mendalami karakteristik sampel, kita diperbolehkan mengambil kesimpulan untuk populasi.

Variabel adalah karakteristik dari sebuah item atau individu yang merupakan anggota dari populasi. Variabel dapat berbeda-beda sesuai kebutuhan atau fakta di lapangan. Contoh variabel nilai mata kuliah Analisis dan Interpretasi Data (AID) dimana tidak semua mahasiswa mendapatkan skor nilai yang sama untuk mata kuliah ini. Contoh lain harga saham masing-masing UMKM (Usaha Mikro, Kecil, dan Menengah) setiap harinya. Pengamatan merupakan daftar semua nilai variabel anggota populasi. Contoh nilai mata kuliah AID merupakan angka bulat antara 0 sampai 100. Nilai harga saham merupakan bilangan real sesuai mata uang misalnya Rupiah. Himpunan data biasanya disajikan dalam bentuk tabel dimana kolom menyatakan variabel dan baris menampilkan pengamatan.

Klasifikasi variabel berdasarkan jenis data. Pengetahuan tentang jenis data yang digunakan diperlukan untuk presentasi data dan tahapan analisis selanjutnya. Data nominal menggambarkan variabel yang nilainya terbatas dan tidak dapat diurutkan. Data nominal biasanya berbentuk kualitatif atau kategori. Contoh nilai tentang status pekerjaan menghasilkan pilihan jawaban “wiraswasta”, “pegawai swasta”, “PNS (Pegawai Negeri Sipil)”, dan “tidak bekerja”. Data interval menggambarkan interval antar nilai dapat diperbandingkan. Data interval berbentuk kuantitatif atau numerik. Contoh data interval seperti nilai mahasiswa, pendapatan, jarak tempuh, dan berat badan. Data ordinal merupakan variabel yang nilainya dapat diurutkan atau diberi peringkat. Contoh data ordinal adalah nilai pelayanan jasa yaitu “luar biasa”, “bagus”, “cukup baik”, dan “jelek”. Meskipun data interval juga dapat diurutkan seperti halnya data ordinal namun nilai pada data interval memiliki perbedaan yang konsisten. Misalnya untuk data interval nilai mahasiswa antara skor 80 dan 90 sama dengan skor 60 dan 70 yaitu memiliki perbedaan sebanyak 10 angka. Sementara nilai antara “luar biasa” dengan “bagus” dan “cukup baik” dengan “jelek” dapat bervariasi jarak perbedaannya sesuai pendapat masing-masing individu.

Variabel kategori mempunyai nilai dalam bentuk kategori seperti di bawah rata-rata, rata-rata atau di atas rata-rata. Variabel numerik memiliki nilai yang menggambarkan kuantitas. Variabel numerik diidentifikasi menjadi variabel diskret dan variabel kontinu. Variabel diskret memiliki nilai numerik yang dapat dicacah perhitungannya. Sebagai contoh jumlah karyawan sebuah perusahaan. Variabel kontinu menghasilkan tanggapan numerik yang diambil berdasarkan proses ukuran. Contoh variabel kontinu adalah waktu yang dibutuhkan oleh *front desk* Universitas Terbuka dalam melayani mahasiswa atau tamu yang datang.

Namun demikian variabel kategori dapat diberi kode secara numerik untuk memudahkan proses analisis data. Misalkan variabel kategori jenis kelamin dibuatkan

kode yaitu perempuan diberi angka 1 dan laki-laki diberi angka 0. Perubahan tampilan data setelah dilakukan numerisasi tersebut dikenal dengan istilah variabel dummy. Sebaliknya, variabel numerik juga dapat dikategorikan. Sebagai contoh variabel numerik usia dikategorikan sebagai berikut: usia 0-1 tahun disebut bayi, usia 2-10 tahun disebut anak-anak, usia 11-19 tahun disebut remaja, usia 20-60 tahun dikatakan dewasa, dan lebih dari 60 tahun disebut lanjut usia. Proses perubahan tampilan variabel numerik menjadi kategori dikenal dengan istilah diskritisasi.

## B. KEBERMAKNAAN DATA

Agar data yang dikumpulkan berarti atau bermakna maka perhatikan tahapan-tahapan berikut:

### 1. Identifikasi Masalah

Masalah penelitian harus terdefinisi dengan baik dan tidak rancu. Hal ini penting agar tujuan penelitian terarah secara jelas dan dapat diukur. Sehingga rencana tahapan-tahapan untuk pencapaian tujuan dapat dipahami secara bersama oleh seluruh tim dan dieksekusi sesuai pembagian tugas masing-masing termasuk biaya yang harus dikeluarkan.

Mengenal masalah dengan baik menjadi syarat utama. Pengenalan masalah dapat diperoleh dari observasi dan wawancara awal dengan sumber informasi. Hal ini dilakukan untuk menggali asumsi mereka dan asumsi peneliti. Agar peta arah jalan penelitian benar. Identifikasi dapat diawali dengan pertanyaan-pertanyaan untuk menggali informasi.

Contoh ilustrasi, CEO menyampaikan masalahnya ke Anda yaitu bagaimana meningkatkan penjualan *online* produk X. Tujuan penelitian untuk memberikan rekomendasi yang tepat agar penjualan produk X secara *online* meningkat. Identifikasi informasi yang dibutuhkan untuk menyusun rencana diantaranya: 1) media online yang paling sering dikunjungi oleh pembeli produk X apakah website atau media sosial; 2) kelompokkan pembeli berdasarkan pola foto yang pernah dilihat baik pada website maupun media sosial; 3) tentukan keterkaitan antara nomor 1) dan 2) berdasarkan keputusan orang tersebut membeli produk X; 4) pelajari risiko masing-masingnya serta biaya yang akan ditimbulkan; 5) buat tugas pokok dan fungsi (tupoksi) masing-masing tim meliputi ketua tim, anggota penelitian, sumber data, pakar pengolahan data, dan pembeli; 6) buat tabel jadwal pelaporan masing-masing kegiatan oleh setiap individu dan dijalankan secara disiplin. Meskipun *progress* belum sesuai target jika sudah sampai pada jadwal untuk *cross check* kegiatan maka laporan tetap diberikan dan rapat dilaksanakan sesuai waktu yang disepakati di awal.

## 2. Persiapan Data

Persiapan data meliputi beberapa tahapan. Pertama adalah pengumpulan data. Cara pengumpulan data sesuai dengan tujuan yang ingin dicapai. Kedua, mengelompokkan data sesuai karakteristiknya berdasarkan instrumen penelitian dan data mentah hasil pengumpulan data. Sebagai contoh data tentang identitas dan perilaku pembeli meliputi jenis kelamin, usia, pendidikan, pekerjaan, warna yang dipilih, spesifikasi barang, manfaat barang yang diharapkan, harga, dan lainnya dikelompokkan berdasarkan jenis data. Ketiga, jika diperlukan lakukan pembersihan data atau transformasi sebelum dilakukan proses pengolahan data. Namun dengan syarat bahwa transformasi data dilakukan tidak boleh mengeliminir informasi penting.

Jika pada suatu kejadian, peneliti menemukan data yang tidak mendukung asumsi awal, maka gali kembali informasi dari sumber awal untuk menyamakan asumsi karena tentu peneliti tidak ingin cepat berada pada posisi akhir dengan menggunakan peta jalan yang salah. Terkadang apa yang responden isi pada kuesioner tidak selalu mencerminkan perilaku mereka sebenarnya. Sehingga selain menggunakan survei dalam bentuk isian perlu juga didampingi dengan wawancara untuk menggali motivasi responden serta didukung data sekunder. Ketiga hal tersebut dikombinasikan untuk mendapatkan informasi utuh.

## 3. Analisis

Pemilihan data analisis sangat penting karena hasilnya akan sangat berbeda jika model yang digunakan tidak sesuai dengan tujuan dan karakteristik data. Setelah model yang akan digunakan sudah dipelajari dengan baik, lakukan optimalisasi teknik-tekniknya (khusus bagi mahasiswa program Doktor disarankan menggunakan teknik yang direnovasi sendiri sesuai penelitian yang digeluti sehingga dapat memberikan kebaruan lebih baik pada ilmu pengetahuan) agar diperoleh hasil terbaik. Ada banyak model dan metode analisis data seperti pohon keputusan, aturan asosiasi, regresi linear, regresi berganda, *k-nearest neighbor*, perluasan teori *fuzzy sets*, *rough sets*, *soft sets*, *N-soft sets* dan seterusnya sesuai kebutuhan yang didasari pada tujuan penelitian.

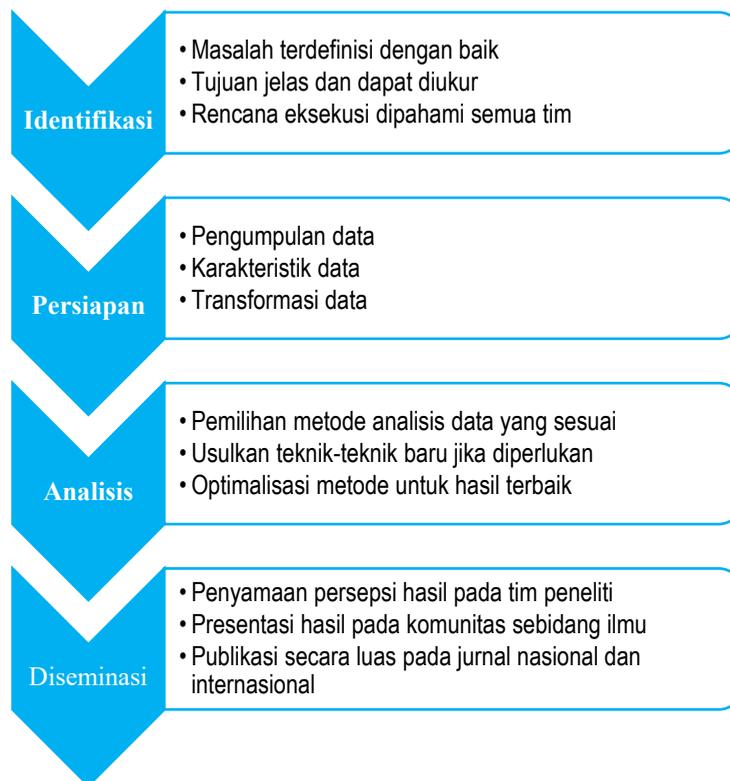
## 4. Diseminasi

Setelah didapat hasil analisis data maka sebaiknya presentasikan ke semua tim peneliti agar mempunyai persepsi yang sama. Selanjutnya lakukan sosialisasi pada komunitas bidang ilmu dan instansi yang berkepentingan agar mendapat masukan sesuai target penelitian. Terakhir, publikasikan hasil secara luas baik di jurnal nasional terakreditasi maupun jurnal internasional bereputasi. Kenapa dua tipe jurnal itu dibutuhkan sekaligus? Hal ini agar proyek penelitian dapat membawa manfaat lebih luas baik bagi masyarakat Indonesia maupun warga dari negara lain.

Diseminasi juga sebaiknya menyebutkan dengan jujur apa yang sudah dilakukan dan apa yang tidak dapat dikerjakan. Untuk hal yang belum atau tidak dapat dikerjakan karena sesuatu dan lain hal maka dibuat sebagai saran untuk penelitian lanjutan. Saran

seperti itu termasuk bagian dari kontribusi kita pada ilmu pengetahuan karena memberikan ide-ide penelitian kepada orang lain yang juga dikenal dengan istilah *open problems*.

Secara ringkas dapat dikatakan bahwa kita sebagai peneliti harus berupaya maksimal agar data-data yang diberikan atau diperoleh menjadi bermakna dan berfungsi dengan baik bagi orang lain. Data tersebut bukan hanya dimengerti oleh peneliti saja. Ringkasan tahapan-tahapan agar data menjadi bermakna seperti telah diuraikan di atas dapat dilihat pada Gambar 1.1.



Gambar 1.1  
Fase Kebermaknaan Data

### C. BUKTI EMPIRIS

Kita sering menjumpai jenis data pada artikel-artikel jurnal. Berikut diambil beberapa contoh penerapan tentang konsep yang kita pelajari pada KB 1 yang ditemui pada artikel jurnal.

Swain (2016) membahas tentang model yang menghasilkan dukungan terhadap strategi intervensi awal dan kebijakan pengambilan keputusan untuk administrasi dan

tenaga profesional petugas medis. Variabel yang digunakan Swain (2016) merupakan variabel kategori. Deskripsi variabelnya dapat dilihat pada Tabel 1.1.

Tabel 1.1  
Daftar Variabel pada Artikel Swain (2016)

Variabel	Keterangan	Nilai
AGE1	Usia (dalam tahun)	1 = 18-29 2 = 30-44 3 = 45-64 4 = lebih besar dari 65
AVEDRNK2	Rata-rata minum alkohol per hari selama 1 bulan	1= lebih dari atau sama dengan 1 0 = selainnya
EDUCA	Tingkat pendidikan: PAUD < Educa1 < Sekolah Menengah Atas	1 = Iya 0 = Tidak
	Educa4 > kuliah minimal empat tahun lebih	1 = Iya 0 = Tidak
EMPLOY	Status Pekerjaan Employ1 = Sesuai usia produktif	1 = Iya 0 = Tidak
	Employ8 = tidak mampu bekerja	1 = Iya 0 = Tidak
EMTSUPRT	Berapa sering membutuhkan bantuan	1 = Selalu 2 = biasanya 3 = kadang-kadang 4 = jarang 5 = tidak pernah
EXERANY2	Olahraga rutin setiap hari selama satu bulan	1 = Iya 0 = Tidak
GENDER	Jenis kelamin	1= perempuan 0 = laki-laki
GENHLT	Kesehatan umum	1= luar biasa 2= sangat bagus 3 = bagus 4 = cukup 5 = jelek
HISPANIC	Perlu penanganan khusus	1 = Iya 0 = Tidak
HLTHPLAN	Punya riwayat penyakit	1 = Iya 0 = Tidak
MARITAL	Status perkawinan	1 = Iya, kawin 0 = Tidak kawin
SMOKER31	Apakah perokok	1 = Iya 0 = Tidak

Worthman, dkk. (2016) membahas tentang pengembangan pemasok dengan mempertimbangkan rantai pasok yang terdiri dari satu produsen dan satu pemasok. Adapun daftar variabel yang menjadi acuan pada artikel ini disajikan pada Tabel 1.2.

Dapat dilihat bahwa variabel yang digunakan berbeda dengan artikel Swain (2016), pada artikel ini menggunakan variabel numerik.

Tabel 1.2  
Daftar Variabel pada Artikel Worthman, dkk. (2016)

Variabel	Keterangan	Nilai
$T$	Periode kontrak	60
$a$	Harga yang dilarang	200
$b$	Elastisitas harga	0.01
$c_M$	Biaya variabel per unit (M)	70
$c_0$	Biaya variabel per unit (S)	100
$r$	Margin Keuntungan Tetap	15
$c_{SD}$	Biaya pengembangan pemasok per unit	100000
$w$	Ketersediaan sumberdaya	1
$m$	Kecepatan pembelajaran	-0.1



### Latihan

Untuk memperdalam pemahaman Anda mengenai materi Kegiatan Belajar 1, kerjakanlah latihan berupa aktivitas berbasis artikel jurnal sebagai berikut!

Silahkan cari dan unduh minimal dua artikel dari jurnal internasional bereputasi sesuai dengan ketertarikan ide penelitian (usulan proposal) Anda dimana masing-masing artikel menggunakan jenis data berbeda!

#### *Petunjuk Jawaban Latihan*

- 1) Buatlah rangkuman tentang perbandingan jenis data yang digunakan pada masing-masing artikel tersebut!
- 2) Tentukan jenis data yang akan digunakan pada disertasi Anda dan kemukakan alasannya!
- 3) Latihan ini dibuat dalam bentuk file Word menggunakan bahasa sendiri dan didukung literatur terkini (lima tahun terakhir). Hasil latihan dikirimkan kepada dosen Anda sebelum jadwal mempelajari Kegiatan Belajar 2.



### Rangkuman

Pengenalan data sangat diperlukan sebelum melakukan proses tahapan analisis data. Pemahaman tentang populasi, sampel, variabel, data nominal, data ordinal, dan data interval perlu dipahami dengan baik agar interpretasi dan analisis data tepat sasaran.

Data yang dikumpulkan agar dapat bermakna dapat menggunakan tahapan berikut.

1. Identifikasi.
2. Persiapan.
3. Analisis.
4. Diseminasi.



### Tes Formatif 1

---

Kerjakan tes formatif berupa aktivitas berikut.

Tindak lanjut dari Latihan KB1, silahkan Anda buat sebuah artikel literatur review tentang pengenalan data yang dikaitkan dengan topik disertasi yang Anda pilih! Aturannya sebagai berikut.

- 1) Artikel ilmiah yang memuat judul, abstrak, kata kunci, pendahuluan, isi, kesimpulan, dan referensi.
- 2) Artikel dalam bentuk file Microsoft Word.
- 3) Font Times New Roman dengan ukuran 12, spasi 1.5, margin kiri, kanan, atas, dan bawah yaitu 3 cm.
- 4) Jumlah halaman minimal 5 halaman di luar referensi
- 5) Artikel ini dikirim ke dosen Anda sebelum jadwal mempelajari Modul 2.

Jika Anda telah mengerjakan Latihan dan Tes Formatif pada Kegiatan Belajar 1 dalam Modul 1, serta Dosen Anda telah menyetujuinya, **Bagus!** Anda dapat melanjutkan mempelajari Kegiatan Belajar 2 Modul 1 ini. Jika Dosen Anda belum menyetujui pekerjaan yang Anda kirimkan maka sebaiknya Anda mengulangi kembali materi Kegiatan Belajar 1, terutama bagian yang belum dikuasai.

## Presentasi Data

Pentingnya presentasi data apapun jenis penelitiannya. Apakah metode statistika deskriptif atau statistika inferensia, kita tetap memerlukan menyajikan rangkuman data agar dapat dibaca secara cepat dan bermakna. Dalam kehidupan sehari-hari, kita sangat sering melihat presentasi data di televisi, internet, media sosial, majalah, koran, artikel dan buku.

Pada Kegiatan Belajar 2, kita akan menyajikan data agar bermakna bagi siapapun yang melihat. Materi disajikan secara sederhana dan mudah dimengerti namun data awal (mentah) tidak selalu sederhana. Karena di era digital seperti sekarang, data real sangat komplikasi dengan jumlahnya yang besar dan karakteristik data yang belum terstruktur. Bahkan untuk data yang kecil sekalipun tetap diperlukan keterampilan dan kecermatan agar data berguna bagi yang membutuhkan. Misalnya pada perusahaan, baik CEO, Manager ataupun karyawan ingin menggunakan penyajian data yang menarik dan tepat guna untuk diketahui oleh agen penjualan, *supplier*, dan konsumen.

Pada bagian A, membahas tentang beberapa cara presentasi data untuk satu variabel. Pada bagian B, presentasi data untuk variabel tidak tunggal. Bagian C, bukti empiris.

Presentasi data diperlukan untuk meringkas data secara visual agar mendapatkan informasi berguna untuk pengambilan keputusan. Sebagai contoh, manajer yang memiliki data tentang konsumen produk terdiri dari jenis kelamin, usia, domisili, pekerjaan, kesukaan tipe produk, waktu berbelanja dan lain sebagainya. Manajer akan lebih mudah memahami masalah yang sedang dihadapi jika data disajikan secara visual dan bentuk ringkasan lainnya yang mewakili kesimpulan tentang data. Sehingga secara sederhana dapat dikatakan bahwa visualisasi data adalah daya tarik pertama dalam penyampaian gagasan atau hasil.

### A. PRESENTASI DATA SATU VARIABEL

#### 1. Tendensi Pusat

Sebagai bagian dari fungsi manajemen, tendensi pusat berguna untuk mengetahui secara umum dan sekilas mengenai posisi nilai pengamatan. Beberapa tendensi pusat yang sering digunakan adalah modus, median dan *mean*.

Modus dihitung berdasarkan nilai yang sering muncul. Jika ada lebih dari satu angka yang sering muncul maka yang diambil adalah titik tengah dari data yang sering muncul tersebut. Median adalah nilai tengah setelah data diurutkan. Jika jumlah bilangan yang diurutkan genap maka median adalah rata-rata dari dua nilai tengah yang berdekatan.

*Mean* adalah rata-rata dari semua nilai. Rata-rata data yang merupakan sampel dari populasi besar disebut *mean* sampel dan dinotasikan dengan  $\bar{X}$ . Jika data merupakan populasi maka disebut *mean* populasi dengan notasi  $\mu$ . *Mean* diperoleh dengan menjumlah semua nilai dan dibagi dengan banyaknya data. Rumus *mean* ditulis sebagai berikut dimana  $n$  adalah jumlah pengamatan dan  $X_i$  adalah nilai pengamatan ke- $i$  (Rumus 1.1).

$$\text{Mean} = \frac{\sum_{i=1}^n X_i}{n} \quad 1.1$$

Penggunaan *mean* perlu kepekaan dan kehati-hatian. Karena tidak semua data sets nilai mean hampir mendekati nilai modus atau mediannya. Terkadang kita temui ada data pencilan yang nilainya saat ekstrim atau jauh lebih rendah/tinggi dibandingkan nilai-nilai lain pada data yang sedang diamati. Akibatnya jika ukuran *mean* menjadi dasar pengambilan keputusan menjadi tidak tepat.

## 2. Keragaman

Kuartil membagi variabel kontinu menjadi empat bagian sama besar. Kuartil satu, dua dan tiga. Kuartil kedua sama dengan nilai median.

Ragam menggambarkan penyebaran data dan mengukur seberapa jauh nilai variabel dari rata-rata. Jika semua data pengamatan mendekati rata-rata maka ragam data akan relatif kecil. Sementara itu jika terdapat beberapa pengamatan yang nilainya jauh dari rata-rata maka data dapat dikatakan beragam. Berikut rumus untuk menghitung ragam (Rumus 1.2 dan Rumus 1.3).

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad 1.2$$

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{N} \quad 1.3$$

Dimana  $s$  adalah simpangan baku sampel,  $\sigma$  simpangan baku populasi,  $x_i$  adalah nilai aktual data,  $\bar{x}$  adalah nilai rata-rata sampel,  $\mu$  nilai rata-rata populasi dan  $n$  merupakan jumlah pengamatan sampel dan  $N$  jumlah populasi. Semakin tinggi nilai simpangan baku maka semakin menyebar data dari rata-rata.

Nilai simpangan baku perlu menjadi kewaspadaan saat menjadi dasar pengambilan keputusan khususnya ketika data memiliki pola yang sangat menyimpang seperti kurva tidak berdistribusi normal atau terdapat banyak pencilan pada data. Akibatnya, kita perlu dasar atau acuan penggunaan simpangan baku. Diantaranya dapat menggunakan aturan empiris.

Aturan empiris (berdasarkan pengamatan) untuk interpretasi simpangan baku jika nilai dari variabel diasumsikan terdistribusi normal sebagai berikut:

Sekitar 68% pengamatan berada satu simpangan baku dari nilai rata-rata yaitu  $\bar{X} \pm s$ . Sekitar 95% pengamatan berada dua simpangan baku dari nilai rata-rata yaitu  $\bar{X} \pm 2s$ . Sekitar 99% pengamatan berada tiga simpangan baku dari nilai rata-rata yaitu  $\bar{X} \pm 3s$ .

Nilai kemiringan (*skewness*) dari nol mengindikasikan distribusi simetris. Jika luas kurva di kiri lebih luas daripada luas di bawah kurva sebelah kanan maka nilai kemiringan bertanda positif. Dengan kata lain, mayoritas pengamatan berada di kiri nilai rata-rata. Sementara sebaliknya jika luas di bawah kurva sebelah kanan lebih luas daripada sebelah kiri maka skor kemiringan bernilai negatif. Hal ini berarti mayoritas pengamatan berada di kanan nilai rata-rata. Rumus untuk mencari kemiringan secara manual menggunakan Rumus 1.4 berikut.

$$\text{Kemiringan} = \left( \frac{\sqrt{n \times (n-1)}}{n-2} \right) \times \frac{\frac{1}{n} \times \sum_{i=1}^n (x_i - \bar{x})^3}{\left( \frac{1}{n} \times \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{3/2}} \quad 1.4$$

Selanjutnya istilah kurtosis. Distribusi dengan nilai kurtosis tinggi bermakna bahwa data memiliki lebih banyak pengamatan ekstrim. Kurtosis diperlukan untuk mengidentifikasi apakah ada data ekstrim atau tidak wajar sehingga pengambilan keputusan perlu mengambil tindakan atau penanganan khusus. Contoh paling populer tentang manfaat kurtosis adalah saat Wall Street “jatuh” pada akhir tahun 2008 dimana analisis keuangan mengacu pada distribusi normal. Padahal faktanya kurva distribusi memiliki “ekor” yang lebih tebal artinya terdapat kejadian ekstrim. Rumus 1.5 adalah formula untuk menghitung kurtosis.

$$\text{kurtosis} = \frac{n-1}{(n-2) \times (n-3)} \times \left( (n+1) \times \frac{\frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n}}{\left( \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \right)^2} - 3 \right) + 6 \quad 1.5$$

Nilai kemiringan dan kurtosis mendekati nol mengindikasikan bahwa distribusi frekuensi variable mendekati distribusi normal. Akibatnya perlu meninjau kembali asumsi metode analisis data yang digunakan.

Tendensi pusat dan simpangan baku menggambarkan hanya satu aspek dari variabel numerik. Untuk melihat ringkasan data secara visual maka data disajikan dalam bentuk grafik atau diagram.

### 3. Grafik

Untuk skala nominal, diagram batang biasa digunakan untuk menyajikan frekuensi relatif untuk nilai berbeda. Sumbu X dari diagram batang menyajikan nilai sementara sumbu Y menyajikan frekuensi, proporsi dan persentase. Urutan sumbu X ditampilkan secara acak, namun demikian agar gambar nyaman dilihat dan memudahkan pencarian bagi yang sedang mengamati digaram batang maka sumbu X dapat disusun berdasarkan alfabet atau nilai frekuensi dari rendah ke tinggi atau sebaliknya.

#### Contoh 1.1

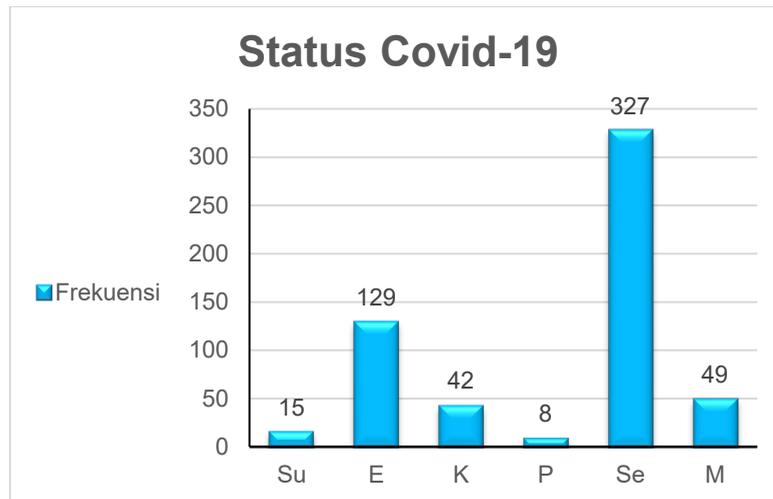
Data status Covid-19 di Kecamatan Demak Kabupaten Demak per tanggal 22 Desember 2020 dapat dilihat pada Tabel 1.3. Data dapat diakses pada link berikut <https://data.go.id/dataset/sebaran-covid-19-kabupaten-demak-per-22-desember-2020>.

Tabel 1.3  
Status Covid-19 di Kecamatan Demak Kabupaten Demak

Status Covid-19	Frekuensi
Kasus Suspek (Su)	15
Kontak Erat (E)	129
Konfirmasi Covid-19 ditemukan di Demak yang belum sembuh (K)	42
Positif Covid-19 ditemukan diluar Demak yang belum sembuh (P)	8
Sembuh (Se)	327
Meninggal (M)	49
<b>Total</b>	<b>570</b>

Sumber: Dinas Kesehatan Kabupaten Demak per 22 Desember 2020

Cara membuat diagram batang Contoh 1.1. menggunakan Excel dengan cara blok kolom “Status Covid-19” dan “Frekuensi” pada Tabel 1.3 kemudian klik Insert dan pilih 2-D Column. Selanjutnya lakukan edit judul, keterangan sumbu horizontal dan vertikal sesuai kebutuhan. Diagram batang Contoh 1.1. seperti terlihat pada Gambar 1.2.



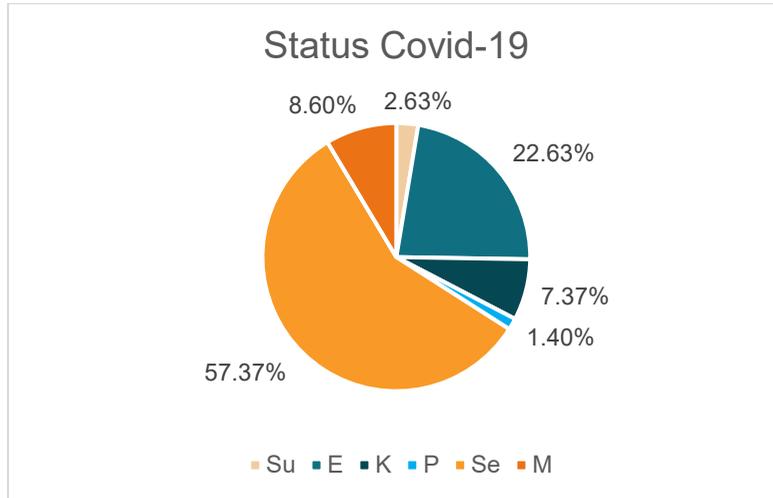
Gambar 1.2  
Status Covid-19 di Kecamatan Demak per 22 Desember 2020

Selanjutnya dicari frekuensi relatif untuk masing-masing status Covid-19. Untuk Kasus suspek, frekuensi relatifnya adalah  $(15 / 570) \times 100 \% = 2,63\%$  dan seterusnya. Hasil perhitungan frekuensi relatif Contoh 1.1. disajikan pada Tabel 1.4.

Tabel 1.4  
Frekuensi Relatif Status Covid-19 di Kecamatan Demak  
per 22 Desember 2020

Status Covid-19	Frekuensi Relatif (%)
Kasus Suspek (Su)	2,63
Kontak Erat (E)	22,63
Konfirmasi Covid-19 ditemukan di Demak yang belum sembuh (K)	7,37
Positif Covid-19 ditemukan diluar Demak yang belum sembuh (P)	1,4
Sembuh (Se)	57,37
Meninggal (M)	8,6
<b>Total</b>	<b>100</b>

Berdasarkan Tabel 1.4 dibuat diagram lingkaran menggunakan Excel dengan cara blok kolom “Status Covid-19” dan “Frekuensi Relatif” kemudian klik Insert dan pilih Pie Chart. Selanjutnya edit keterangan judul, dan letak angka pada diagram lingkaran sesuai posisi yang diinginkan. Misal angka dapat ditempatkan di dalam lingkaran atau di luar lingkaran. Diagram lingkaran Contoh 1.1. seperti terlihat pada Gambar 1.3.



Gambar 1.3  
Status Covid-19 di Kecamatan Demak per 22 Desember 2020

## B. PRESENTASI DATA UNTUK VARIABEL LEBIH DARI SATU

Pada bagian ini akan dibahas cara penyajian data untuk variabel lebih dari satu. Pertama, kumpulkan data yang terdiri dari beberapa variabel sesuai tujuan penelitian. Selanjutnya buat tabel yang memuat semua data dalam satu tabel. Tabel ini disebut sebagai tabel klasifikasi silang yang mencantumkan frekuensi setiap nilai dari dua variabel atau lebih. Perhatikan contoh berikut.

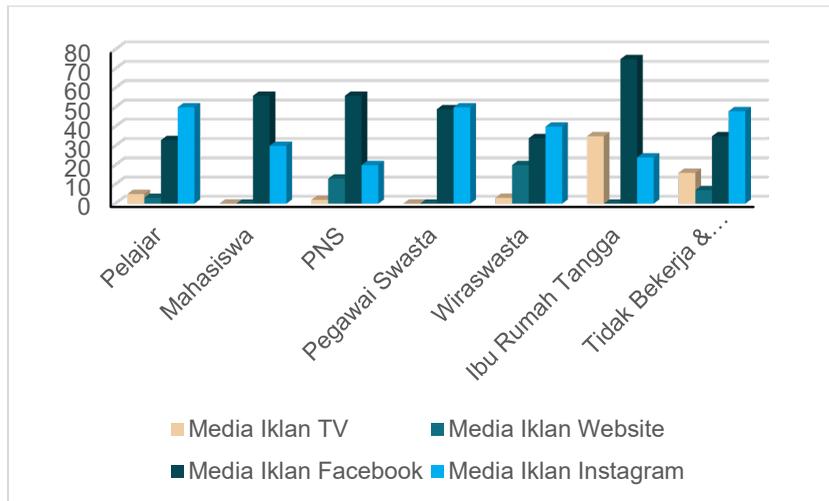
### Contoh 1.2

CEO ingin mengetahui hubungan antara jenis pekerjaan konsumen yang membeli produk X dengan sumber iklan tentang produk X yang dilihat oleh konsumen. Hasil pengumpulan data dapat dilihat pada Tabel 1.5.

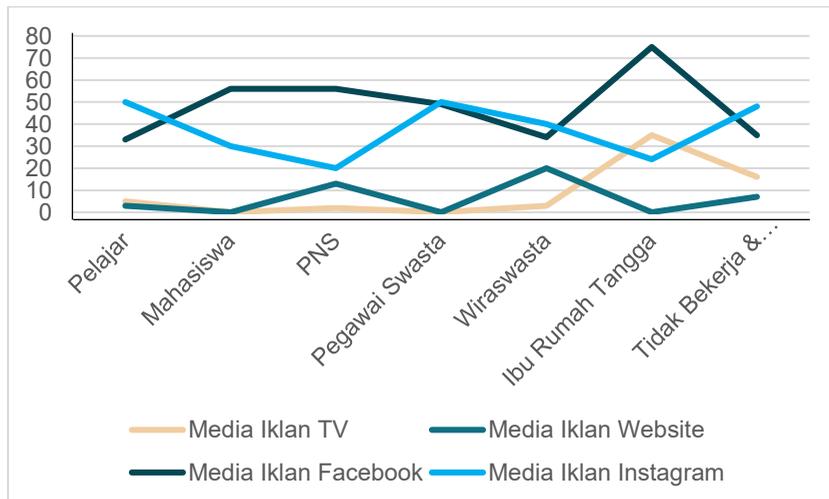
Tabel 1.5  
Klasifikasi Silang Contoh 1.2

Jenis Pekerjaan	Media Iklan				Total
	TV	Website	Facebook	Instagram	
Pelajar	5	3	33	50	91
Mahasiswa	0	0	56	30	86
PNS	2	13	56	20	91
Pegawai Swasta	0	0	49	50	99
Wiraswasta	3	20	34	40	97
Ibu Rumah Tangga	35	0	75	24	134
Tidak Bekerja & Tidak sedang Studi	16	7	35	48	106
<b>Total</b>	<b>61</b>	<b>43</b>	<b>338</b>	<b>262</b>	<b>704</b>

Diagram batang dari Tabel 1.5 disajikan pada Gambar 1.4 dan diagram garis diperlihatkan pada Gambar 1.5. Pada dua diagram ini, dapat dengan mudah dan cepat kita simpulkan bahwa iklan menggunakan media sosial *facebook* paling efektif dalam penjualan produk X dibandingkan media lain untuk semua jenis pekerjaan yang menjadi target produk ini. Berdasarkan saran ini, maka CEO memutuskan untuk memasang iklan lebih banyak di facebook dan mengatur strategi lebih baik lagi dari segi kreativitas, biaya dan lainnya.



Gambar 1.4  
Diagram Batang Contoh 1.2



Gambar 1.5  
Diagram Garis Contoh 1.2

Terkadang kita berminat untuk melihat hubungan antar variabel. Peneliti tertarik untuk mengetahui apakah variabel tertentu dapat berubah ketika variabel lainnya berubah. Misalnya, apakah ada hubungan antara pendidikan dan pendapatan, bekerja dengan belanja secara online. Untuk hal ini diperlukan ukuran koefisien korelasi.

Koefisien korelasi adalah ukuran numerik antara -1 dan 1 yang menyatakan keterkaitan antara dua variabel. Jika nilai koefisien korelasi semakin mendekati -1 dan 1 maka artinya terdapat hubungan yang sangat kuat antara dua variabel. Sebaliknya, jika nilai korelasi mendekati 0 maka hubungan antar variabel rendah. Hubungan searah bernilai positif (+) dan hubungan berlawanan arah diberi tanda negatif (-).

Terdapat dua jenis koefisien korelasi yang sering digunakan yaitu korelasi *Pearson Product Moment* dan korelasi *Rank Spearman*. Korelasi *Pearson Product Moment* digunakan untuk data berjumlah besar dan memerlukan asumsi normal. Sementara korelasi *Rank Spearman* digunakan untuk data kurang dari 30, data diurutkan terlebih dulu dan dibuat peringkat serta data tidak harus berdistribusi normal. Rumus koefisien korelasi *Pearson Product Moment* (Rumus 1.6) dan korelasi *Rank Spearman* (Rumus 1.7).

$$r_{xy} = \frac{n \sum XY - \sum X \sum Y}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}} \quad 1.6$$

Keterangan:

- $r_{xy}$  = koefisien korelasi antara  $X$  dan  $Y$
- $n$  = jumlah sampel
- $X$  = skor setiap individu variabel  $X$
- $Y$  = skor setiap individu variabel  $Y$

$$r_s = 1 - \frac{n \sum d^2}{n(n^2 - 1)} \quad 1.7$$

Keterangan:

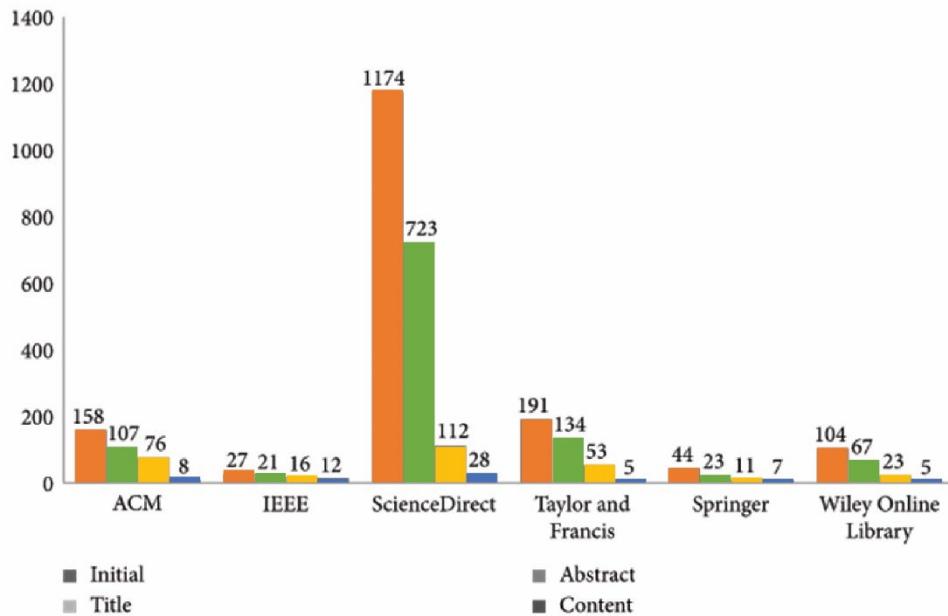
- $r_s$  = koefisien korelasi *Rank Spearman*
- $n$  = jumlah sampel
- $d^2$  = selisih peringkat

### C. BUKTI EMPIRIS

Teknik presentasi data dapat beragam mulai dari hal yang sederhana sampai ke penyajian yang rumit menggunakan *software* tertentu. Namun demikian, bukan berarti semakin rumit presentasi data yang ditampilkan maka artikel tersebut semakin bagus kualitas penyajiannya. Hal ini bergantung kepada topik dan tujuan penelitian. Berikut

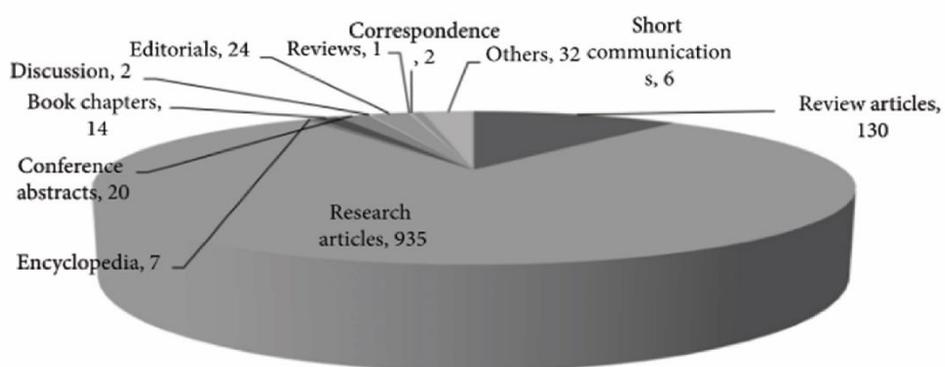
beberapa bukti dari artikel jurnal yang memperlihatkan bahwa peneliti menggunakan presentasi data untuk menyampaikan rangkuman tentang data dan hasil temuan.

Yang, dkk. (2020) menggunakan diagram batang (Gambar 1.6) untuk memperlihatkan rangkuman dari hasil pencarian dan filter berdasarkan judul, abstrak dan isi pada artikel-artikel di ACM, IEEE, ScienceDirect, Taylor and Francis, Springer dan Wiley Online Library. Nama-nama prosiding beserta jumlah publikasinya disajikan menggunakan diagram lingkaran. Diagram lingkaran juga dililih oleh penulis untuk menampilkan jenis artikel beserta jumlahnya (Gambar 1.7). Diagram garis digunakan untuk menjelaskan tentang jumlah artikel berdasarkan tahun (Gambar 1.8).



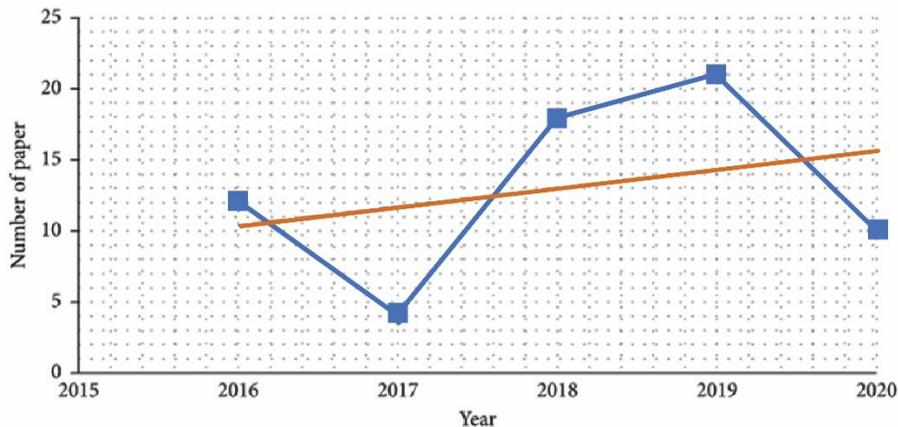
Sumber: Yang dkk (2020).

Gambar 1.6 Hasil Pencarian dan Filter Artikel Jurnal



Sumber: Yang dkk (2020).

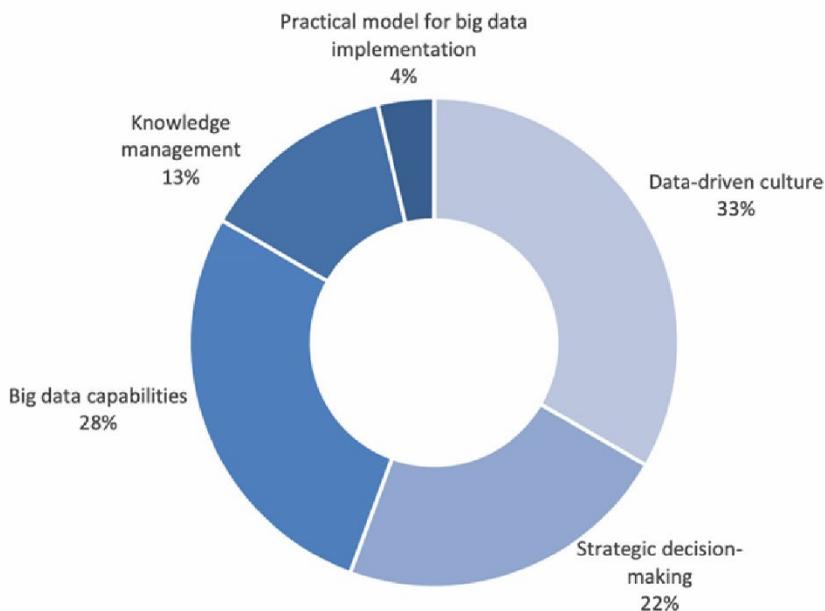
Gambar 1.7 Diagram Lingkaran Jenis Artikel



Sumber: Yang, dkk. (2020)

Gambar 1.8  
Jumlah Artikel per Tahun

Polese, dkk. (2019) menyajikan area penelitian menggunakan diagram lingkaran (Gambar 1.9). Diagram lingkaran tersebut memperlihatkan bahwa 1) pendekatan data-driven 33%; 2) kemampuan big data 28%; 3) pengambilan keputusan strategis yaitu kontribusi integrasi big data pada strategi bisnis dan keputusan sebesar 22%; 4) manajemen pengetahuan hanya 13%; dan sisanya 5) model praktis untuk implementasi big data 4%.

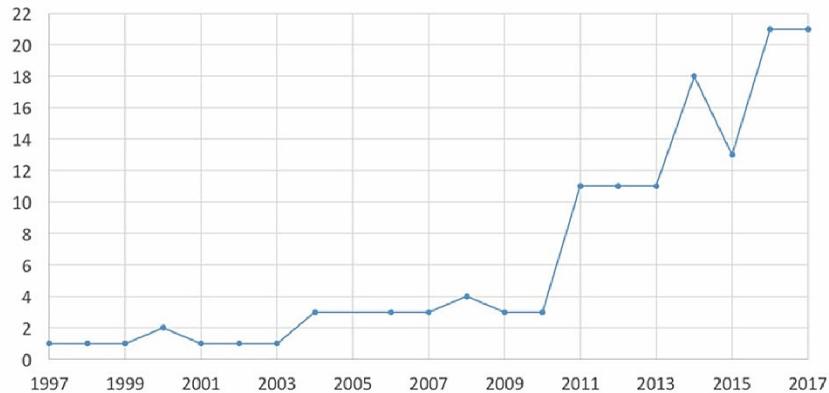


Sumber: Polese, dkk. (2019)

Gambar 1.9  
Diagram Lingkaran Topik Penelitian

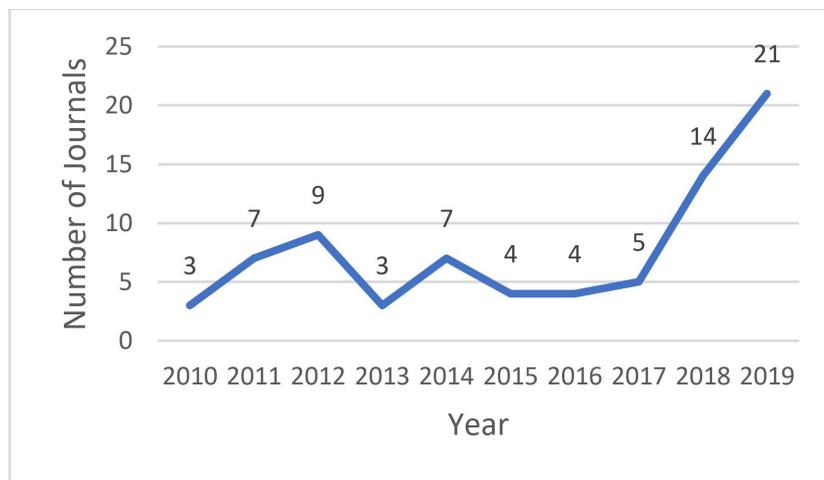
Pelissari, dkk. (2018) menyajikan distribusi frekuensi 134 artikel per tahun mulai 1997 sampai 2017 menggunakan diagram garis (Gambar 1.10). Diagram garis tersebut memperlihatkan bahwa jumlah artikel yang terbit mengalami kenaikan. Presentasi data dengan diagram garis juga dilakukan oleh Aamer, dkk. (2021) untuk menampilkan hasil *trend* naik artikel *machine learning* khususnya topik peramalan yang diterbitkan pada jurnal mulai 2010 sampai 2019 (Gambar 1.11).

R. Pelissari et al. / Intl. Trans. in Op. Res. 00 (2018) 1–37



Sumber: Pelissari dkk (2018).

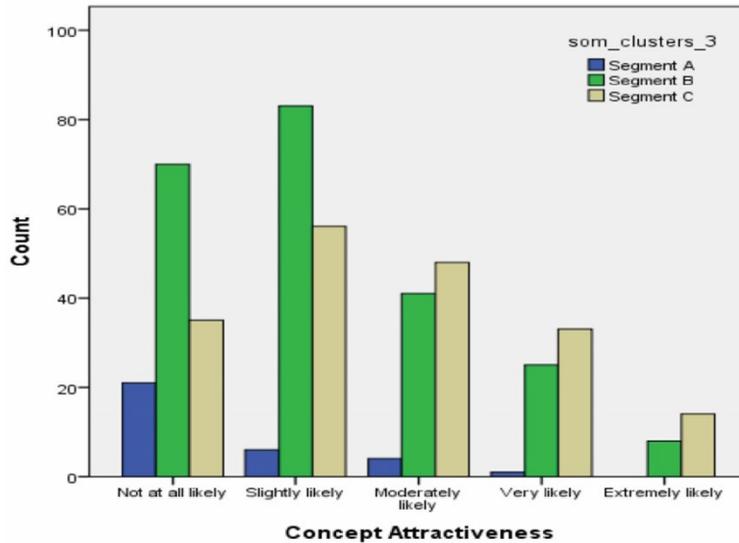
Gambar 1.10  
Diagram Garis Publikasi Artikel



Sumber: Aamer dkk (2021).

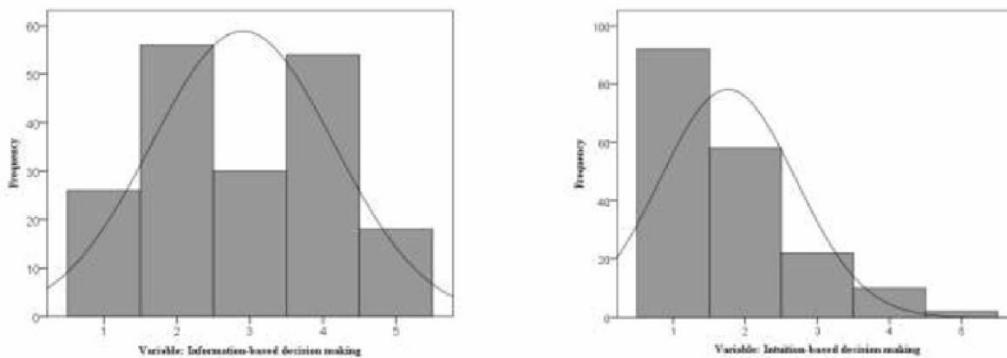
Gambar 1.11  
Artikel Peramalan Menggunakan Machine Learning

Arunachalam & Kumar (2018) menggunakan diagram batang untuk menyajikan data yang mengindikasikan segmen atraktif terkait *pub* non alkohol (Gambar 1.12). Sementara itu, Tarka (2016) menampilkan histogram tentang variabel terikat yaitu intuisi dan informasi yang diterapkan pada model regresi ordinal (Gambar 1.13).



Sumber: Arunachalam & Kumar (2018).

Gambar 1.12  
Daya Tarik *Pub* Non Alkohol



Sumber: Tarka (2016).

Gambar 1.13  
Histogram Variabel Intuisi dan Informasi



## Latihan

Untuk memperdalam pemahaman Anda mengenai materi di atas, kerjakanlah latihan berupa aktivitas berbasis artikel jurnal berikut!

Silahkan cari dan unduh minimal dua artikel dari jurnal internasional bereputasi sesuai dengan ketertarikan ide penelitian (usulan proposal) Anda dimana masing-masing artikel harus terdapat minimal satu jenis presentasi data seperti yang dipelajari

pada Kegiatan Belajar 2 ini! Artikel boleh menggunakan artikel yang sama saat mengerjakan latihan Kegiatan Belajar 1 dengan syarat artikel tersebut menyajikan data secara visual.

#### *Petunjuk Jawaban Latihan*

- 1) Buatlah tabel rangkuman tentang perbandingan jenis presentasi data yang digunakan pada masing-masing artikel tersebut!
- 2) Tentukan jenis presentasi data yang akan Anda gunakan pada disertasi dan kemukakan alasannya!
- 3) Latihan ini dibuat dalam bentuk file Word menggunakan bahasa sendiri dan didukung literatur terkini. Hasil latihan dikirimkan kepada dosen pengampu sebelum jadwal mempelajari Modul 2.



#### Rangkuman

Presentasi data perlu dilakukan untuk memberikan gambaran ringkas data dengan menggunakan visualisasi. Ada banyak variasi visualisasi data tergantung. Namun presentasi data umum dan sering dilakukan minimal wajib diketahui oleh peneliti. Seperti modus, median, mean, diagram batang, garis, lingkaran, dan histogram.



#### Tes Formatif 2

Kerjakan tes formatif berupa aktivitas berbasis video berikut

1. Silahkan pelajari video dari laman berikut tentang standar deviasi. <https://www.youtube.com/watch?v=MRqtXL2WX2M>. Rangkumlah video tersebut dalam bentuk narasi dan temukan apakah Anda sudah memahami isi dari video tersebut.
2. Tindak lanjut dari Latihan KB 2, silahkan Anda buat sebuah artikel literatur review tentang presentasi data yang dikaitkan dengan topik disertasi yang Anda pilih! Aturannya sebagai berikut.
  - Artikel ilmiah yang memuat judul, abstrak, kata kunci, pendahuluan, isi, kesimpulan, dan referensi.
  - Artikel dalam bentuk file Microsoft Word.
  - Font Times New Roman dengan ukuran 12, spasi 1.5, margin kiri, kanan, atas, dan bawah yaitu 3 cm.
  - Jumlah halaman minimal 5 halaman di luar referensi.

Artikel ini dikirim ke dosen Anda sebelum jadwal mempelajari Modul 2.

Jika Anda telah mengerjakan Latihan KB 2 dan Tes Formatif 2 pada Kegiatan Belajar 2 dalam Modul 1, serta Dosen Anda telah menyetujuinya, **Bagus!** Anda dapat melanjutkan mempelajari Modul 2. Jika Dosen Anda belum menyetujui pekerjaan yang Anda kirimkan maka sebaiknya Anda mengulangi kembali materi yang belum dikuasai.

## Kunci Jawaban Tes Formatif

### *Tes Formatif 1*

Setiap mahasiswa menghasilkan sebuah artikel jurnal minimal 5 halaman mengulas tentang jenis-jenis data yang digunakan sesuai topik disertai masing-masing.

Contoh: Literatur Review Jenis Data yang Sesuai untuk Penelitian Sumber Daya Manusia.

### *Tes Formatif 2*

- 1) Tiga aturan empiris simpangan baku yaitu 68%, 95%, dan 99%.
- 2) Setiap mahasiswa menghasilkan sebuah artikel jurnal minimal 5 halaman mengulas tentang presentasi yang ada hubungannya dengan topik disertai masing-masing.

## Daftar Pustaka

- Aamer, A. M., Yani, L. P. E., & Priyatna, I. M. A. (2021). Data analytics in the supply chain management: Review of machine learning applications in demand forecasting. *Operations and Supply Chain Management*, 14(1), 1-13.
- Albright, S. C., & Winston, W. L. (2017). *Business analytics: Data analysis and decision making* (6<sup>th</sup> ed.). USA: Cengage Learning.
- Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2011). *Statistics for business and economics* (11<sup>th</sup> ed.). South-Western: Cengage Learning.
- Arunachalam, D., & Kumar, N. (2018). Benefit-based consumer segmentation and performance evaluation of clustering approaches: An evidence of data-driven decision-making. *Expert Systems with Applications*, 111. pp. 11-34. ISSN 0957-4174 <https://doi.org/10.1016/j.eswa.2018.03.007>
- Keller G. (2018). *Statistics for management and economics* (11<sup>th</sup> ed.). USA: Cengage Learning.
- Myatt, G. J., & Johnson, W. P. (2014). *Making sense of data I: A practical guide to data visualization, advances data mining methods, and applications* (2<sup>nd</sup> ed.). New Jersey: John Wiley & Sons, Inc.
- Nevo, D. (2017). *Making sense of data through statistics (An introduction)* (2<sup>nd</sup> ed.). USA: Legerity Digital Press.
- Ott, L., & Longnecker, M. (2015). *An introduction to statistical methods & data analysis* (7<sup>th</sup> ed.). USA: Cengage Learning.
- Pelissari, R., Oliveira, M. C., Abackerli, A. J., Abackerli, S., & Assumpcao, M. R. P. (2018). Techniques to model uncertain input data of multi-criteria decision-making problems: A literature review. *International Transactions in Operational Research*, 1-37. <https://www.researchgate.net/publication/327863939>
- Polese, F., Troisi, O., Grimaldi, M., Romeo, E. (2019). A big data-oriented approach to decision making: a systematic literature review. *22<sup>nd</sup> International Conference Proceedings*, ISBN 9788890432798, 472-496. <https://sites.les.univr.it/eisic/wp-content/uploads/2019/11/31-Polese-Troisi-Grimaldi-Romeo-1.pdf>
- Proch, M., Worthmann, K., Schlüchtermann, J. A. (2017). Negotiation-based algorithm to coordinate supplier development in decentralized supply chains. *Eur. J. Oper. Res.*, 256, 412–429 <https://doi.org/10.1016/j.ejor.2016.06.029>

- Swain, A. K. (2016) Mining big data to support decision making in healthcare. *Journal of Information Technology Case and Application Research*, 18:3, 141-154. <http://dx.doi.org/10.1080/15228053.2016.1245522>
- Tarka, P. (2016). Intuition or marketing research information usefulness in business organizations in the light of ordinal regression analysis. *Journal of Emerging Trends in Marketing and Management*, 1(1): 41-51. [http://www.etimm.ase.ro/RePEc/aes/jetimm/2016/ETIMM\\_V01\\_2016\\_53.pdf](http://www.etimm.ase.ro/RePEc/aes/jetimm/2016/ETIMM_V01_2016_53.pdf)
- Worthmann, K., Proch, M., Braun, P., Schlüchtermann, J., & Pannek, J. (2016). Towards dynamic contract extension in supplier development. *Logist. Res.*, 9, 1-12.
- Yang, M., Nazir, S., Xu, Q., & Ali, S. (2020). Deep learning algorithms and multicriteria decision-making used in big data: a systematic literature review. *Hindawi Complexity*. <https://doi.org/10.1155/2020/2836064>.