

Buku Referensi

# METODE PENGUMPULAN DATA

UNTUK SAINS DATA



Prof. Dr. Harya Damar Widiuputra, S.T., M.Kom., IPM.





**BUKU REFERENSI**

**METODE  
PENGUMPULAN  
DATA  
UNTUK SAINS DATA**

Prof. Dr. Harya Damar Widiputra, S.T., M.Kom., IPM.

# METODE PENGUMPULAN DATA UNTUK SAINS DATA

---

---

Ditulis oleh:

Prof. Dr. Harya Damar Widiputra, S.T., M.Kom., IPM.

---

Hak Cipta dilindungi oleh undang-undang. Dilarang keras memperbanyak, menerjemahkan atau mengutip baik sebagian ataupun keseluruhan isi buku tanpa izin tertulis dari penerbit.

---



ISBN: 978-634-7662-37-8  
IV + 267 hlm; 18,2 x 25,7 cm.  
Cetakan I, April 2026

**Desain Cover dan Tata Letak:**  
Melvin Mirsal

Diterbitkan, dicetak, dan didistribusikan oleh  
**PT Media Penerbit Indonesia**  
Royal Suite No. 6C, Jalan Sedap Malam IX, Sempakata  
Kecamatan Medan Selayang, Kota Medan 20131  
Telp: 081362150605  
Email: [ptmediapenerbitindonesia@gmail.com](mailto:ptmediapenerbitindonesia@gmail.com)  
Web: <https://mediapenerbitindonesia.com>  
Anggota IKAPI No.088/SUT/2024



# KATA PENGANTAR

---

Sains data merupakan bidang yang berkembang pesat dan berperan penting dalam mendukung pengambilan keputusan berbasis informasi di berbagai sektor. Keberhasilan penerapan sains data tidak hanya ditentukan oleh kecanggihan algoritma atau teknik analisis, tetapi juga sangat bergantung pada kualitas data yang digunakan sebagai dasar pengolahan. Data yang relevan, lengkap, dan akurat menjadi fondasi utama dalam menghasilkan analisis yang dapat dipercaya serta insight yang bernilai strategis, sebagaimana pengumpulan data merupakan tahap awal yang menentukan keseluruhan proses analitik selanjutnya.

Buku referensi ini membahas konsep dasar data dan sains data, prinsip serta tujuan pengumpulan data, klasifikasi data primer dan sekunder, berbagai metode pengumpulan data seperti survei, wawancara, observasi, eksperimen, dan pemanfaatan sumber digital, serta pengelolaan kualitas, validasi, etika, keamanan, dan privasi data. Selain itu, buku referensi ini juga membahas perkembangan pengumpulan data di era digital, big data, dan kecerdasan buatan, serta contoh penerapan praktis untuk mendukung analisis dan pengambilan keputusan berbasis data.

Semoga buku referensi ini dapat menjadi sumber rujukan yang bermanfaat dalam memahami metode pengumpulan data serta mendukung pembelajaran, penelitian, dan pengambilan keputusan berbasis data.

Salam Hangat,

**Penulis**

# DAFTAR ISI

<b>KATA PENGANTAR .....</b>	<b>i</b>
<b>DAFTAR ISI .....</b>	<b>ii</b>
<b>BAB I    PENDAHULUAN .....</b>	<b>1</b>
A. Peran Data dalam Sains Data.....	1
B. Pengumpulan Data dalam Siklus Sains Data.....	3
C. Evolusi Metode Pengumpulan Data .....	6
D. Tantangan Pengumpulan Data di Era <i>Big data</i> .....	10
E. Etika, Legalitas, dan Privasi Data.....	14
F. Struktur dan Sistematika Buku .....	16
<b>BAB II    KONSEP DASAR PENGUMPULAN DATA .....</b>	<b>21</b>
A. Pengertian dan Tujuan Pengumpulan Data.....	21
B. Klasifikasi Data: Data Primer dan Data Sekunder .....	26
C. Metode Konvensional dan Berbasis Teknologi.....	33
D. Kualitas Data: Validitas, Reliabilitas, dan Bias.....	41
E. Reprodusibilitas dan Dokumentasi Proses Pengumpulan Data.....	47
<b>BAB III   DATA PRIMER.....</b>	<b>53</b>
A. Pengertian dan Karakteristik Data Primer .....	53
B. Kelebihan dan Keterbatasan Data Primer.....	56
C. Peran Data Primer dalam Analisis Sains Data.....	61
<b>BAB IV    METODE PENGUMPULAN DATA PRIMER</b>	
<b>KONVENSIONAL .....</b>	<b>67</b>
A. Wawancara .....	67
B. Kuesioner.....	71
C. Observasi .....	75
D. <i>Focus Group Discussion</i> (FGD).....	78
E. Eksperimen .....	80

F.	Digitalisasi Data Primer Konvensional .....	84
G.	Pengolahan Awal Data Primer Menggunakan Python .....	88
<b>BAB V</b>	<b>METODE PENGUMPULAN DATA PRIMER</b>	
	<b>BERBASIS TEKNOLOGI.....</b>	<b>95</b>
A.	Sensor dan <i>Internet of Things</i> (IoT) .....	95
B.	<i>Mobile Data Collection Applications</i> .....	99
C.	<i>Digital Tracking</i> dan <i>Log Data</i> .....	102
D.	<i>Drones</i> dan <i>Remote Sensing</i> .....	106
E.	<i>Machine Learning</i> dan <i>Computer Vision</i> .....	109
F.	Isu Teknis dan Etika Data Primer Digital.....	111
G.	Pengambilan dan Akuisisi Data Primer Digital Menggunakan Python .....	114
<b>BAB VI</b>	<b>DATA SEKUNDER.....</b>	<b>117</b>
A.	Pengertian dan Karakteristik Data Sekunder .....	117
B.	Kelebihan dan Keterbatasan Data Sekunder.....	119
C.	Validitas dan Relevansi Data Sekunder.....	122
D.	Peran Data Sekunder dalam Sains Data.....	126
<b>BAB VII</b>	<b>METODE PENGUMPULAN DATA SEKUNDER</b>	
	<b>KONVENSIONAL .....</b>	<b>129</b>
A.	Studi Dokumen dan Arsip .....	129
B.	<i>Literature Review</i> (Kajian Pustaka).....	132
C.	<i>Database</i> Statistik Resmi.....	134
D.	Data dari Organisasi dan Institusi.....	135
E.	Internet dan Portal <i>Online</i> .....	137
F.	Media Massa dan Publikasi Populer.....	142
G.	Ekstraksi dan Integrasi Data Sekunder Menggunakan Python.....	145
<b>BAB VIII</b>	<b>METODE PENGUMPULAN DATA SEKUNDER</b>	
	<b>BERBASIS TEKNOLOGI.....</b>	<b>149</b>
A.	Konsep Otomatisasi Pengumpulan Data .....	149
B.	Infrastruktur Digital dan <i>Cloud Data</i> .....	153
C.	Tantangan Teknis dan Etika Data Sekunder Digital .....	157

<b>BAB IX</b>	<b><i>WEB SCRAPING DAN WEB CRAWLING</i></b> .....	<b>163</b>
	A. Pengertian <i>Web Scraping</i> dan <i>Web Crawling</i> .....	163
	B. Perbedaan <i>Web Scraping</i> dan <i>Web Crawling</i> .....	168
	C. Arsitektur dan Alur Kerja .....	173
	D. <i>Tools</i> dan Teknologi <i>Web Scraping</i> .....	178
	E. Penerapan dalam Sains Data.....	182
	F. Isu Legal dan Kepatuhan .....	185
	G. Implementasi <i>Web Scraping</i> Menggunakan Python.....	188
<b>BAB X</b>	<b><i>APPLICATION PROGRAMMING INTERFACE (API)</i></b> <b>SEBAGAI SUMBER DATA</b> .....	<b>193</b>
	A. Konsep Dasar API .....	193
	B. Jenis-Jenis API.....	199
	C. Mekanisme Request dan Response.....	203
	D. API sebagai Sumber Data Sekunder.....	206
	E. Isu Keamanan dan Etika .....	210
	F. Pengambilan Data API Menggunakan Python .....	211
<b>BAB XI</b>	<b>INTEGRASI METODE PENGUMPULAN DATA</b> <b>DALAM PROYEK SAINS DATA</b> .....	<b>217</b>
	A. Pemilihan Metode Berdasarkan Tujuan Analisis .....	217
	B. Integrasi Data Primer dan Sekunder .....	220
	C. Pengumpulan Data Multisumber .....	225
	D. Pipeline Pengumpulan Data Berbasis Python.....	228
	E. Studi Kasus Proyek <i>End-to-End</i> .....	231
<b>BAB XII</b>	<b>PENUTUP</b> .....	<b>233</b>
	A. Ringkasan Metode Pengumpulan Data.....	233
	B. Tren Masa Depan Pengumpulan Data .....	234
	C. Implikasi bagi Mahasiswa dan Praktisi Sains Data .....	239
	<b>DAFTAR PUSTAKA</b> .....	<b>245</b>
	<b>GLOSARIUM</b> .....	<b>257</b>
	<b>INDEKS</b> .....	<b>261</b>
	<b>BIOGRAFI PENULIS</b> .....	<b>265</b>
	<b>SINOPSIS</b> .....	<b>267</b>

# BAB I

## PENDAHULUAN

---

Metode pengumpulan data merupakan fondasi utama dalam sains data karena kualitas, relevansi, dan kelengkapan data sangat menentukan akurasi analisis serta keandalan model yang dihasilkan. Pemilihan metode pengumpulan data yang tepat memungkinkan data *scientist* memperoleh data yang sesuai dengan tujuan analitis, baik dari sumber terstruktur maupun tidak terstruktur, sehingga proses pengolahan, analisis, dan pengambilan keputusan berbasis data dapat berjalan secara efektif dan sistematis.

### A. Peran Data dalam Sains Data

Data merupakan elemen inti dalam sains data karena seluruh proses analisis, pemodelan, dan pengambilan keputusan berangkat dari data yang dikumpulkan. Tanpa data yang memadai, sains data tidak dapat menghasilkan wawasan (*insight*) yang bermakna, akurat, dan dapat ditindaklanjuti. Data berfungsi sebagai representasi fenomena nyata yang dianalisis untuk memahami pola, hubungan, dan tren tertentu. Oleh karena itu, kualitas dan karakteristik data sangat menentukan keberhasilan penerapan sains data dalam berbagai bidang. Secara lebih rinci, peran data dalam sains data dapat dijelaskan sebagai berikut:

#### 1. Sumber Informasi Utama

Sumber informasi utama dalam sains data merujuk pada peran data sebagai representasi faktual dari fenomena dunia nyata yang menjadi dasar dalam proses analisis, pemodelan, dan pengambilan keputusan. Data yang diperoleh dari berbagai sumber seperti sistem transaksi, sensor, media digital, maupun interaksi pengguna memungkinkan sains data menangkap kondisi aktual, perilaku, dan dinamika yang terjadi secara objektif dan terukur.

Sebagai sumber informasi utama, data berfungsi untuk mengurangi ketergantungan pada asumsi subjektif dengan menyediakan bukti empiris yang dapat dianalisis secara sistematis. Melalui proses pengolahan dan eksplorasi data, informasi yang tersembunyi dalam kumpulan data mentah dapat diungkap sehingga membantu organisasi memahami pola, tren, serta hubungan sebab akibat yang relevan dengan tujuan analitis dan strategis (Provost & Fawcett, 2023).

## **2. Bahan Baku Analisis**

Bahan baku analisis dalam sains data menempatkan data sebagai *input* utama yang harus melalui berbagai tahapan pengolahan sebelum menghasilkan informasi yang bermakna. Data mentah yang berasal dari beragam sumber umumnya masih mengandung ketidakkonsistenan, *noise*, dan ketidaklengkapan, sehingga perlu diproses melalui pembersihan, transformasi, dan integrasi agar siap digunakan dalam analisis yang andal.

Pada konteks analitik, data sebagai bahan baku berfungsi untuk memungkinkan penerapan teknik statistik, analisis eksploratif, dan algoritma komputasional guna menemukan pola serta hubungan yang relevan. Kualitas bahan baku data secara langsung memengaruhi validitas hasil analisis, karena data yang tidak representatif atau bias akan menghasilkan kesimpulan yang menyesatkan meskipun metode analisis yang digunakan sudah tepat dan canggih (Han, Pei, & Kamber, 2022).

## **3. Dasar Pengambilan Keputusan**

Dasar pengambilan keputusan dalam sains data menempatkan data sebagai landasan objektif untuk menentukan pilihan dan tindakan yang rasional dalam berbagai konteks organisasi maupun bisnis. Dengan memanfaatkan data yang telah diolah dan dianalisis, pengambil keputusan dapat mengurangi ketergantungan pada intuisi semata serta meningkatkan akurasi keputusan berdasarkan fakta dan bukti empiris.

Pada praktiknya, data memungkinkan evaluasi berbagai alternatif keputusan melalui pengukuran kinerja, analisis risiko, dan proyeksi dampak di masa depan. Proses ini membantu organisasi memahami konsekuensi dari setiap pilihan secara lebih terstruktur, sehingga keputusan yang diambil tidak hanya bersifat reaktif, tetapi juga strategis dan berorientasi jangka panjang (Marr, 2021).

#### 4. Media Pembelajaran Model

Media pembelajaran model dalam sains data merujuk pada peran data sebagai sarana utama yang memungkinkan model analitik dan pembelajaran mesin mempelajari pola, struktur, dan hubungan yang terdapat dalam fenomena nyata. Tanpa data yang memadai, representatif, dan relevan, model tidak memiliki dasar empiris untuk melakukan proses pembelajaran secara efektif dan menghasilkan keluaran yang bermakna.

Pada proses pelatihan model, data berfungsi sebagai media yang mentransfer informasi dari dunia nyata ke dalam bentuk matematis dan komputasional. Melalui data latih, data validasi, dan data uji, model secara bertahap menyesuaikan parameter internalnya agar mampu mengenali pola, melakukan prediksi, serta menggeneralisasi pengetahuan ke data baru yang belum pernah ditemui sebelumnya (Géron, 2022).

#### 5. Alat Evaluasi dan Prediksi

Alat evaluasi dan prediksi dalam sains data menempatkan data sebagai sarana utama untuk menilai kinerja masa lalu sekaligus memproyeksikan kondisi di masa depan secara sistematis dan terukur. Melalui analisis data historis, organisasi dapat mengevaluasi efektivitas kebijakan, strategi, atau proses yang telah diterapkan berdasarkan indikator kinerja yang objektif dan berbasis bukti.

Pada fungsi prediksi, data digunakan untuk membangun model analitik dan pembelajaran mesin yang mampu mengidentifikasi pola serta tren yang berulang dari waktu ke waktu. Model prediktif memanfaatkan data sebagai dasar untuk memperkirakan kemungkinan kejadian di masa depan, sehingga membantu pengambil keputusan dalam mengantisipasi risiko, peluang, dan kebutuhan strategis secara lebih proaktif (Shmueli *et al.*, 2020).

### B. Pengumpulan Data dalam Siklus Sains Data

Pengumpulan data dalam siklus sains data merupakan tahap awal yang sangat krusial karena menjadi dasar bagi seluruh proses analisis, pemodelan, dan pengambilan keputusan berbasis data. Pada tahap ini, data *scientist* berfokus pada upaya memperoleh data yang relevan, akurat, dan representatif sesuai dengan tujuan bisnis atau permasalahan analitis yang telah dirumuskan sebelumnya. Kegagalan dalam

pengumpulan data yang tepat dapat menyebabkan bias, kesalahan interpretasi, serta menghasilkan model yang tidak valid atau tidak dapat diandalkan.

Secara konseptual, pengumpulan data mencakup proses identifikasi sumber data, penentuan metode pengambilan data, serta mekanisme penyimpanan data awal sebelum diproses lebih lanjut. Sumber data dalam sains data dapat berasal dari sistem internal organisasi seperti basis data transaksi, laporan operasional, dan log sistem, maupun dari sumber eksternal seperti survei, media sosial, sensor IoT, data publik, dan aplikasi pihak ketiga. Keberagaman sumber ini menuntut pemahaman mendalam terhadap karakteristik data, termasuk format, volume, kecepatan pembaruan, dan tingkat keandalannya. Dalam praktiknya, pengumpulan data dalam siklus sains data memiliki beberapa fokus utama yang saling berkaitan, antara lain:

## **1. Penentuan Kebutuhan Data**

Penentuan kebutuhan data merupakan tahap strategis dalam siklus sains data yang berfokus pada penerjemahan tujuan analisis atau permasalahan bisnis ke dalam spesifikasi data yang jelas dan terukur. Pada tahap ini, analis harus mengidentifikasi variabel apa saja yang dibutuhkan, tingkat detail data yang sesuai, serta periode waktu data agar analisis yang dilakukan benar-benar relevan dengan tujuan yang ingin dicapai. Kejelasan kebutuhan data sejak awal membantu mencegah pengumpulan data yang berlebihan, tidak relevan, atau justru kurang memadai untuk mendukung proses analitik.

Penentuan kebutuhan data juga melibatkan pemahaman terhadap konteks penggunaan data dan hubungan antarvariabel yang akan dianalisis. Proses ini menuntut kemampuan analitis untuk mengaitkan kebutuhan informasi dengan sumber data yang tersedia, baik data primer maupun data sekunder, sehingga data yang dikumpulkan dapat merepresentasikan kondisi nyata secara akurat. Menurut Han, Kamber, dan Pei (2022), perumusan kebutuhan data yang tepat merupakan prasyarat penting agar proses analisis data dapat menghasilkan pola dan pengetahuan yang bermakna serta dapat ditindaklanjuti.

## **2. Pemilihan Sumber Data**

Pemilihan sumber data merupakan proses krusial dalam siklus sains data yang berfokus pada identifikasi asal data yang paling relevan

untuk menjawab tujuan analisis atau kebutuhan bisnis secara tepat. Sumber data dapat berasal dari sistem internal organisasi seperti basis data operasional, data transaksi, dan log sistem, maupun dari sumber eksternal seperti data publik, media sosial, sensor, atau layanan API pihak ketiga, sehingga diperlukan evaluasi menyeluruh terhadap kesesuaian konteks dan karakteristik data. Menurut Provost dan Fawcett (2018), pemilihan sumber data yang tepat sangat menentukan nilai analitik yang dihasilkan karena kualitas dan relevansi sumber data berpengaruh langsung terhadap akurasi serta kegunaan hasil analisis.

Pemilihan sumber data juga harus mempertimbangkan kualitas data yang mencakup kelengkapan, konsistensi, akurasi, dan tingkat kebaruan data agar hasil analisis tidak bias atau menyesatkan. Aspek teknis seperti kemudahan akses, format data, ketersediaan metadata, serta kompatibilitas dengan sistem pengolahan data juga menjadi pertimbangan penting dalam menentukan sumber data yang optimal. Pemilihan sumber data yang tidak tepat dapat meningkatkan kompleksitas proses pembersihan dan integrasi data, sehingga berdampak pada efisiensi dan efektivitas keseluruhan proses sains data.

### **3. Metode Pengumpulan Data**

Metode pengumpulan data dalam konteks sains data merupakan rangkaian pendekatan sistematis untuk mendapatkan informasi yang diperlukan secara efektif dan efisien agar dapat mendukung tujuan analisis serta pengambilan keputusan yang tepat. Metode ini mencakup teknik yang beragam seperti pengambilan data melalui API, web *scraping*, survei, log sistem, atau sensor, yang dipilih berdasarkan kebutuhan data, sumber yang tersedia, serta karakteristik masalah yang dihadapi. Menurut Kotu dan Deshpande (2019), pemilihan metode pengumpulan data yang sesuai menjadi faktor penting karena menentukan kualitas, representativitas, dan reliabilitas data yang akan digunakan dalam seluruh proses analitik.

Penerapan metode pengumpulan data harus mempertimbangkan aspek teknis seperti frekuensi pengambilan, format penyimpanan, serta kemampuan sistem untuk menangani volume dan kecepatan data yang tinggi. Proses ini juga menuntut perencanaan yang matang untuk mengatasi tantangan seperti data yang tidak terstruktur, data hilang, atau potensi bias yang dapat muncul selama pengambilan data. Metode pengumpulan data yang dipilih akan sangat mempengaruhi tahap

pembersihan, transformasi, dan eksplorasi data selanjutnya, sehingga kesalahan pada tahap awal dapat berimbas pada keseluruhan hasil analisis.

#### 4. Pengelolaan Kualitas Data Awal

Pengelolaan kualitas data awal adalah proses penting dalam siklus sains data yang bertujuan memastikan bahwa data yang diperoleh bebas dari kesalahan, lengkap, konsisten, dan sesuai dengan kebutuhan analitis sebelum melanjutkan ke tahap pemrosesan lebih lanjut. Proses ini mencakup identifikasi nilai yang hilang, duplikasi, inkonsistensi, serta anomali lain yang dapat mengganggu keakuratan model dan pemahaman analitik. Menurut Kimball dan Ross (2019), pengelolaan kualitas data di tahap awal sangat menentukan efektivitas dan keandalan hasil analisis karena data yang buruk cenderung menghasilkan *insight* yang menyesatkan dan keputusan yang kurang tepat.

Pengelolaan kualitas data awal memerlukan penerapan teknik validasi dan pembersihan seperti imputasi nilai hilang, standarisasi format, dan penanganan *outlier* untuk memastikan integritas data yang digunakan dalam analisis. Aktivitas ini juga memerlukan dokumentasi yang jelas dan sistematis agar setiap perubahan atau perbaikan data dapat ditelusuri serta dipahami oleh seluruh pemangku kepentingan. Pengelolaan kualitas data awal yang tepat tidak hanya meningkatkan keakuratan analisis tetapi juga mengurangi biaya dan waktu yang diperlukan pada tahap pembersihan dan transformasi data selanjutnya, sehingga mendukung keseluruhan keberhasilan proyek sains data.

### C. Evolusi Metode Pengumpulan Data

Evolusi metode pengumpulan data terjadi seiring dengan perkembangan teknologi, kebutuhan analisis, dan kompleksitas permasalahan yang dihadapi oleh organisasi maupun peneliti. Pada tahap awal, pengumpulan data dilakukan secara manual dan sederhana, kemudian berkembang menjadi berbasis sistem digital hingga memanfaatkan teknologi canggih yang mampu menangani data dalam skala besar dan kecepatan tinggi.

## 1. Era Tradisional (Manual dan Konvensional)

Era tradisional dalam pengumpulan data ditandai oleh penggunaan metode manual dan konvensional yang sangat bergantung pada interaksi langsung antara peneliti dan sumber data, seperti observasi lapangan, wawancara tatap muka, serta penyebaran kuesioner berbasis kertas yang memerlukan pencatatan dan dokumentasi secara fisik. Pendekatan ini menuntut ketelitian tinggi karena seluruh proses, mulai dari pengumpulan hingga pengolahan awal data, dilakukan oleh manusia tanpa dukungan sistem digital, sehingga rentan terhadap bias subjektif, kesalahan pencatatan, dan inkonsistensi data. Meskipun memiliki keterbatasan dari sisi efisiensi dan skala, metode tradisional tetap memiliki keunggulan dalam menggali data yang bersifat mendalam, kontekstual, dan kualitatif, terutama ketika pemahaman sosial dan perilaku menjadi fokus utama penelitian (Saunders *et al.*, 2019).

Pada praktiknya, era tradisional memerlukan waktu dan sumber daya yang besar karena setiap tahapan pengumpulan data dilakukan secara bertahap dan tidak terotomatisasi, mulai dari persiapan instrumen hingga pengarsipan hasil pengumpulan data. Validitas data sangat bergantung pada kemampuan peneliti dalam merancang instrumen yang jelas, melakukan pengamatan secara objektif, serta menjaga konsistensi prosedur di lapangan agar data yang diperoleh tetap dapat dipercaya. Kondisi ini membuat metode manual kurang fleksibel untuk menghadapi kebutuhan data dalam jumlah besar, namun tetap relevan dalam konteks penelitian skala kecil hingga menengah yang menekankan kedalaman informasi dibandingkan kuantitas data.

## 2. Era Digital Awal (Komputerisasi Data)

Era digital awal dalam pengumpulan data ditandai oleh mulai digunakannya komputer sebagai alat utama untuk mencatat, menyimpan, dan mengelola data yang sebelumnya dikerjakan secara manual, sehingga proses administrasi data menjadi lebih terstruktur dan sistematis. Pada fase ini, organisasi dan peneliti mulai memanfaatkan perangkat lunak dasar seperti *spreadsheet* dan *database* sederhana untuk mengurangi kesalahan pencatatan serta meningkatkan kecepatan pengolahan data. Peralihan dari media fisik ke media digital memungkinkan data disimpan dalam jumlah lebih besar dengan tingkat keteraturan yang lebih baik, meskipun masih terbatas pada data yang bersifat terstruktur.

Perkembangan komputerisasi data juga mendorong munculnya sistem informasi yang dirancang untuk mendukung aktivitas operasional dan manajerial, sehingga data tidak hanya dikumpulkan untuk dokumentasi, tetapi juga untuk kebutuhan pelaporan dan pengambilan keputusan. Penggunaan formulir elektronik dan *input* data berbasis komputer mulai menggantikan kuesioner kertas, yang berdampak pada peningkatan efisiensi waktu serta kemudahan dalam proses rekapitulasi data. Menurut Laudon dan Laudon (2020), sistem informasi berbasis komputer pada tahap awal ini berperan penting dalam mengubah data mentah menjadi informasi yang lebih bermakna bagi organisasi.

### **3. Era Internet dan Sistem Informasi Terintegrasi**

Era internet dan sistem informasi terintegrasi menandai perubahan besar dalam metode pengumpulan data karena konektivitas jaringan memungkinkan data dikumpulkan secara cepat, luas, dan lintas batas geografis melalui berbagai platform digital. Pada fase ini, *website*, aplikasi berbasis web, media sosial, dan sistem transaksi daring menjadi sumber data utama yang terus menghasilkan aliran data dalam jumlah besar dan beragam. Kondisi tersebut mendorong organisasi dan peneliti untuk tidak lagi bergantung pada sumber data internal semata, melainkan mulai memanfaatkan data eksternal sebagai dasar analisis yang lebih komprehensif.

Integrasi sistem informasi seperti *Enterprise Resource Planning* (ERP), *Customer Relationship Management* (CRM), dan *Supply Chain Management* (SCM) memungkinkan data dari berbagai fungsi organisasi dikumpulkan secara terpusat dan konsisten. Sistem terintegrasi ini mengurangi duplikasi data, meningkatkan akurasi, serta memungkinkan akses data secara *real-time* untuk mendukung pengambilan keputusan strategis. Menurut Laudon dan Laudon (2021), integrasi sistem informasi berbasis internet memungkinkan organisasi mengoordinasikan proses bisnis dan arus data secara lebih efisien dalam lingkungan yang saling terhubung.

### **4. Era *Big data* dan Otomatisasi**

Era *big data* dan otomatisasi ditandai oleh kemampuan sistem untuk mengumpulkan, menyimpan, dan memproses data dalam volume yang sangat besar, dengan kecepatan tinggi, serta dalam berbagai bentuk yang kompleks, termasuk data terstruktur, semi-terstruktur, dan tidak

terstruktur. Pada tahap ini, aktivitas digital, sensor IoT, media sosial, transaksi daring, dan sistem log menghasilkan aliran data yang terus-menerus tanpa henti. Kondisi tersebut mendorong perubahan paradigma pengumpulan data dari pendekatan periodik menjadi *real-time* dan berkelanjutan.

Gambar 1. *Big Data*



Sumber: *Corporate Training*

Perkembangan teknologi komputasi terdistribusi, *cloud computing*, serta kerangka kerja pemrosesan data seperti sistem paralel memungkinkan proses pengumpulan dan pengolahan data dilakukan secara otomatis dengan intervensi manusia yang minimal. Otomatisasi dalam pengumpulan data membantu meningkatkan efisiensi, konsistensi, dan skalabilitas, sehingga organisasi mampu menangani kompleksitas data yang tidak mungkin dikelola dengan metode konvensional. Menurut Marr (2018), era *big data* memungkinkan organisasi memanfaatkan otomatisasi dan teknologi analitik canggih untuk mengekstraksi nilai strategis dari data dalam skala besar.

## 5. Era Sains Data dan Kecerdasan Buatan

Era sains data dan kecerdasan buatan ditandai oleh pemanfaatan data secara sistematis untuk menghasilkan pengetahuan, prediksi, dan rekomendasi melalui pendekatan analitik lanjutan yang melampaui sekadar pelaporan deskriptif. Pada fase ini, data tidak hanya dikumpulkan dan disimpan, tetapi diproses sebagai aset strategis yang

digunakan untuk membangun model statistik, pembelajaran mesin, dan algoritma cerdas. Integrasi antara metode sains data dan kecerdasan buatan memungkinkan sistem untuk belajar dari data historis serta menyesuaikan diri secara dinamis terhadap pola baru yang muncul.

Gambar 2. Kecerdasan Buatan



Sumber: *Codepolitan*

Perkembangan kecerdasan buatan mendorong otomatisasi yang lebih cerdas dalam seluruh siklus pengelolaan data, mulai dari akuisisi, pembersihan, hingga analisis dan visualisasi hasil. Algoritma pembelajaran mesin dan pembelajaran mendalam memungkinkan sistem mengenali pola kompleks dalam data berskala besar yang sulit diidentifikasi melalui analisis konvensional. Menurut Russell dan Norvig (2021), kecerdasan buatan modern berfokus pada kemampuan sistem untuk belajar dari data dan mengambil keputusan secara rasional berdasarkan pengalaman yang diperoleh.

#### **D. Tantangan Pengumpulan Data di Era *Big data***

Tantangan pengumpulan data di era *big data* menjadi semakin kompleks karena volume, kecepatan, dan variasi data yang sangat besar. Dalam konteks ini, organisasi dan peneliti harus menghadapi beberapa aspek kritis yang mempengaruhi kualitas, efisiensi, dan keamanan pengumpulan data. Berikut penjelasan rinci mengenai tantangan-tantangan tersebut:

## 1. Volume Data yang Sangat Besar

Volume data yang sangat besar menjadi tantangan utama dalam era *big data* karena organisasi kini harus menangani aliran data yang terus meningkat dari berbagai sumber digital, mulai dari transaksi *online*, media sosial, hingga sensor IoT, yang membutuhkan kapasitas penyimpanan dan pemrosesan yang luar biasa untuk memastikan informasi tetap dapat diakses dan dianalisis secara efektif (Chen, 2019). Pengelolaan data dalam skala ini menuntut penggunaan teknologi canggih seperti sistem penyimpanan terdistribusi dan kerangka komputasi paralel agar mampu menangani jutaan hingga miliaran entri data tanpa kehilangan integritas atau kualitas data. Selain itu, pemantauan kualitas dan konsistensi data menjadi semakin penting karena kesalahan kecil dalam volume besar data dapat berdampak signifikan terhadap hasil analisis dan keputusan bisnis.

Volume data yang sangat besar juga menimbulkan kesulitan dalam identifikasi dan pemilihan data yang relevan untuk analisis, karena tidak semua data yang terkumpul memiliki nilai informatif yang sama atau sesuai dengan tujuan bisnis dan penelitian. Proses ekstraksi, transformasi, dan pemuatan data (ETL) menjadi lebih kompleks dan membutuhkan algoritma serta strategi yang efisien agar data yang relevan dapat disaring dari kumpulan data yang sangat besar. Kegagalan dalam menangani volume data secara tepat dapat menyebabkan *bottleneck*, keterlambatan dalam analisis, dan potensi kehilangan wawasan penting yang dapat memengaruhi strategi organisasi.

## 2. Kecepatan Data (*Velocity*)

Kecepatan data (*velocity*) merupakan salah satu tantangan utama dalam era *big data* karena data dihasilkan dan mengalir secara *real-time* dari berbagai sumber digital, termasuk sensor IoT, transaksi *online*, dan media sosial, sehingga organisasi harus mampu menangkap, memproses, dan menganalisis informasi dengan kecepatan tinggi agar tetap relevan dan bermanfaat (Katal, Wazid, & Goudar, 2018). Pemrosesan data secara *real-time* menuntut infrastruktur yang handal, termasuk teknologi *streaming*, pemrosesan paralel, dan algoritma canggih untuk memastikan data yang masuk tidak menumpuk atau hilang, sekaligus meminimalkan keterlambatan dalam pengambilan keputusan berbasis data. Selain itu, manajemen kecepatan data juga membutuhkan strategi penyaringan dan

pengaturan prioritas agar sistem mampu menangani volume data yang terus bertambah tanpa menurunkan kualitas analisis.

Kecepatan tinggi data dapat menimbulkan tantangan tambahan terkait integrasi data dari berbagai sumber yang memiliki frekuensi dan format berbeda, sehingga membutuhkan mekanisme sinkronisasi dan konsolidasi yang efisien. Proses ETL (*Extract, Transform, Load*) tradisional sering kali tidak cukup cepat untuk memenuhi kebutuhan analisis *real-time*, sehingga organisasi perlu mengadopsi teknologi *big data streaming* dan *event-driven architecture*. Kegagalan dalam mengelola kecepatan data dapat menyebabkan hilangnya wawasan kritis, keterlambatan respons terhadap perubahan pasar, dan keputusan yang kurang tepat.

### **3. Variasi dan Heterogenitas Data (*Variety*)**

Variasi dan heterogenitas data (*variety*) merupakan salah satu tantangan utama dalam pengelolaan *big data* karena data berasal dari berbagai format dan sumber, mulai dari data terstruktur seperti basis data relasional hingga data tidak terstruktur seperti teks, gambar, audio, dan video, sehingga memerlukan teknik integrasi dan transformasi yang kompleks untuk menghasilkan informasi yang dapat dianalisis secara konsisten (Gandomi & Haider, 2019). Tantangan ini memaksa organisasi untuk mengembangkan infrastruktur yang mampu menampung dan memproses berbagai jenis data secara simultan, termasuk penggunaan NoSQL, data lakes, dan sistem manajemen data hibrida agar semua data dapat diakses dan dianalisis dengan efisien. Selain itu, perbedaan struktur, format, dan kualitas data dari berbagai sumber meningkatkan risiko inkonsistensi, redundansi, dan kesalahan interpretasi, sehingga penanganan data yang heterogen menjadi faktor kunci dalam keberhasilan analisis *big data*.

Pengelolaan data yang beragam juga menuntut metode pra-pemrosesan yang canggih, termasuk normalisasi, standarisasi, dan pengkodean data agar berbagai jenis data dapat digabungkan dan dianalisis secara efektif. Tanpa strategi yang tepat, organisasi dapat mengalami kesulitan dalam menghubungkan data dari sumber yang berbeda, sehingga wawasan yang diperoleh menjadi terbatas dan tidak mencerminkan realitas yang lengkap. Oleh karena itu, keberhasilan analisis *big data* sangat bergantung pada kemampuan untuk mengelola variasi dan heterogenitas data secara sistematis dan terstruktur.

#### 4. Kualitas Data (*Veracity*)

Kualitas data (*veracity*) menjadi salah satu tantangan paling kritis dalam pengelolaan *big data* karena data yang dikumpulkan sering kali mengandung ketidakakuratan, ketidaklengkapan, duplikasi, atau bias yang dapat memengaruhi hasil analisis dan pengambilan keputusan (Rajaraman & Ullman, 2020). Untuk memastikan integritas data, organisasi perlu menerapkan mekanisme validasi, pembersihan, dan verifikasi data secara sistematis agar informasi yang digunakan dapat diandalkan dan mendukung keputusan berbasis bukti secara akurat. Selain itu, manajemen *veracity* data juga melibatkan pemantauan berkelanjutan terhadap sumber data, kualitas *input*, serta deteksi anomali yang dapat muncul dari volume dan variasi data yang sangat besar.

Tantangan kualitas data semakin kompleks karena data berasal dari berbagai sumber dengan tingkat kepercayaan yang berbeda, sehingga integrasi data memerlukan evaluasi kritis untuk menentukan validitas setiap elemen data. Proses pra-pemrosesan seperti deduplikasi, normalisasi, dan pengisian nilai yang hilang menjadi penting agar dataset yang digunakan memiliki konsistensi dan relevansi yang tinggi. Kegagalan dalam menangani *veracity* data dapat menyebabkan kesalahan prediksi, rekomendasi yang tidak tepat, serta risiko strategis bagi organisasi yang mengandalkan analisis data.

#### 5. Keamanan dan Privasi Data

Keamanan dan privasi data menjadi salah satu tantangan utama dalam era *big data* karena data yang dikumpulkan sering kali mengandung informasi sensitif dan pribadi, sehingga organisasi harus mampu melindungi data dari akses tidak sah, kebocoran, dan penyalahgunaan (Zhou, Zhang, Xie, & Qian, 2019). Upaya ini menuntut penerapan protokol keamanan yang kuat, termasuk enkripsi, kontrol akses, audit jejak, dan sistem deteksi intrusi, agar informasi tetap aman tanpa mengurangi kelancaran aliran data untuk analisis. Selain itu, kepatuhan terhadap regulasi perlindungan data seperti GDPR atau peraturan lokal menjadi aspek penting dalam menjaga reputasi organisasi serta membangun kepercayaan *stakeholder* terhadap pengelolaan data.

Tantangan keamanan semakin kompleks karena volume dan variasi data yang besar meningkatkan risiko serangan siber dan eksploitasi celah keamanan, sehingga setiap titik pengumpulan,

penyimpanan, dan pemrosesan data harus diperiksa dan diamankan secara menyeluruh. Strategi keamanan harus mencakup mitigasi risiko terhadap ancaman internal maupun eksternal, serta sistem pengawasan *real-time* untuk mencegah akses ilegal atau manipulasi data. Kegagalan dalam menjaga keamanan dan privasi data dapat mengakibatkan kerugian finansial, hilangnya kepercayaan pelanggan, dan implikasi hukum yang signifikan bagi organisasi.

## **E. Etika, Legalitas, dan Privasi Data**

Di era digital yang ditandai dengan melimpahnya data, pemanfaatan informasi melalui sains data tidak hanya menuntut kemampuan teknis, tetapi juga memerlukan perhatian serius terhadap aspek etika, legalitas, dan privasi data agar penggunaan data tetap bertanggung jawab dan tidak merugikan pihak mana pun.

### **1. Etika**

Etika pada dasarnya membahas tentang nilai, norma, dan prinsip yang menjadi pedoman bagi manusia dalam menentukan apakah suatu tindakan dapat dianggap benar, adil, dan bertanggung jawab dalam kehidupan pribadi maupun sosial. Dalam konteks perkembangan ilmu pengetahuan dan teknologi, etika menjadi semakin penting karena keputusan yang diambil berbasis data dan sistem modern dapat berdampak luas terhadap individu dan masyarakat. Floridi dkk. menyatakan bahwa etika berperan sebagai kerangka untuk memastikan bahwa penggunaan teknologi dan informasi tetap berpihak pada kesejahteraan manusia dan menjaga nilai-nilai kemanusiaan (Floridi *et al.*, 2018). Pandangan ini menegaskan bahwa etika tidak hanya bersifat normatif, tetapi juga praktis sebagai pedoman dalam menghadapi perubahan yang cepat dan kompleks.

Pada praktiknya, etika membantu individu dan organisasi membedakan antara apa yang sekadar menguntungkan dan apa yang benar secara moral, sehingga keputusan yang diambil tidak hanya berorientasi pada hasil, tetapi juga pada proses yang dapat dipertanggungjawabkan. Etika juga berfungsi sebagai pengendali agar kekuasaan, pengetahuan, dan teknologi tidak digunakan secara sewenang-wenang yang dapat merugikan pihak lain. Dengan adanya etika, setiap tindakan diharapkan mempertimbangkan dampak jangka

pendek maupun jangka panjang terhadap manusia dan lingkungan sosial. Oleh karena itu, etika tidak bisa dipisahkan dari tanggung jawab sosial dan profesional dalam berbagai bidang kehidupan.

## **2. Legalitas**

Legalitas merujuk pada kepatuhan terhadap hukum dan peraturan yang berlaku dalam setiap aktivitas organisasi maupun individu, termasuk dalam pengelolaan informasi dan data di era digital. Dalam konteks ini, legalitas berfungsi sebagai batas normatif yang memastikan bahwa setiap proses pengumpulan, pengolahan, penyimpanan, dan penggunaan data dilakukan secara sah serta dapat dipertanggungjawabkan. Kuner, Bygrave, dan Docksey menyatakan bahwa perlindungan data dalam kerangka hukum modern bertujuan untuk menyeimbangkan antara pemanfaatan informasi dan perlindungan hak-hak fundamental individu (Kuner *et al.*, 2020). Pandangan tersebut menegaskan bahwa legalitas tidak hanya bersifat administratif, tetapi juga memiliki dimensi perlindungan hak dan keadilan.

Pada praktiknya, legalitas menuntut adanya dasar hukum yang jelas untuk setiap tindakan, sehingga tidak ada proses yang dilakukan secara sewenang-wenang atau tanpa legitimasi. Kepatuhan terhadap hukum juga membantu organisasi meminimalkan risiko sengketa, sanksi, dan kerugian reputasi yang dapat timbul akibat pelanggaran aturan. Selain itu, legalitas mendorong terbentuknya tata kelola yang lebih tertib dan terstruktur, karena setiap keputusan harus mengacu pada norma yang telah ditetapkan. Dengan demikian, legalitas berperan sebagai fondasi penting dalam menciptakan kepastian hukum dan kepercayaan publik.

## **3. Privasi Data**

Privasi data merujuk pada perlindungan terhadap informasi pribadi individu agar tidak diakses, digunakan, atau disebarluaskan tanpa dasar yang sah dan tanpa kendali dari pemilik data. Dalam masyarakat digital, data pribadi tidak hanya mencerminkan identitas seseorang, tetapi juga kebiasaan, preferensi, dan pola perilaku yang dapat digunakan untuk berbagai kepentingan, baik yang bermanfaat maupun yang berisiko. Solove menjelaskan bahwa privasi berkaitan dengan pengendalian atas informasi pribadi dan pencegahan terhadap berbagai bentuk penyalahgunaan data yang dapat merugikan individu (Solove,

2021). Pemahaman ini menunjukkan bahwa privasi data bukan sekadar isu teknis, melainkan juga isu perlindungan martabat dan hak dasar manusia.

Pada praktiknya, privasi data menuntut adanya pembatasan yang jelas mengenai siapa yang boleh mengakses data, untuk tujuan apa data digunakan, dan berapa lama data tersebut disimpan. Tanpa pengaturan yang memadai, data pribadi dapat dengan mudah disalahgunakan untuk kepentingan komersial, manipulasi, atau bahkan tindakan kriminal yang merugikan pemilik data. Perlindungan privasi juga mendorong organisasi untuk lebih berhati-hati dalam merancang sistem informasi agar tidak mengumpulkan data secara berlebihan. Dengan demikian, privasi data berfungsi sebagai mekanisme pengaman agar pemanfaatan teknologi tetap sejalan dengan kepentingan dan hak individu.

## **F. Struktur dan Sistematika Buku**

Struktur dan sistematika buku adalah susunan bab, subbab, dan alur pembahasan yang dirancang secara logis agar pembaca dapat memahami materi secara bertahap, dari konsep dasar hingga penerapan praktis. Struktur dan sistematika buku disusun untuk memberikan alur pembelajaran yang logis dan bertahap bagi pembaca, khususnya mahasiswa atau praktisi sains data. Selain itu, sistematika juga berfungsi untuk mengintegrasikan teori, metode, dan praktik agar pembaca dapat melihat hubungan antara konsep dan penerapannya di dunia nyata. Secara umum, buku Metode Pengumpulan Data untuk Sains Data dapat disusun dengan struktur bertahap dari dasar hingga lanjutan, seperti berikut:

### **1. Bagian Pendahuluan**

Pendahuluan berfungsi sebagai bagian awal yang memberikan gambaran umum mengenai ruang lingkup, tujuan, dan arah pembahasan suatu buku atau karya ilmiah. Pada bagian ini, pembaca diperkenalkan pada konteks permasalahan yang melatarbelakangi penulisan sehingga dapat memahami pentingnya topik yang dibahas. Pendahuluan juga membantu membangun kerangka berpikir pembaca dengan menjelaskan posisi materi dalam keseluruhan struktur pembahasan. Selain itu, bagian ini berperan sebagai penghubung antara latar belakang masalah dan pembahasan inti yang akan diuraikan lebih lanjut. Dengan demikian,

pendahuluan menjadi fondasi penting agar pembaca memiliki pemahaman awal yang jelas dan terarah sebelum memasuki pembahasan utama.

## **2. Bagian Konsep Dasar Data dan Sains Data**

Konsep dasar data dan sains data berkaitan dengan pemahaman tentang data sebagai kumpulan fakta atau informasi yang dapat diolah untuk menghasilkan pengetahuan yang bermakna. Data dapat berasal dari berbagai sumber dan memiliki bentuk yang beragam, sehingga perlu dikelola dan diproses secara sistematis agar dapat digunakan secara efektif. Sains data memanfaatkan metode statistik, komputasi, dan pemikiran analitis untuk mengekstrak pola, hubungan, dan informasi penting dari data. Dalam praktiknya, sains data tidak hanya berfokus pada analisis, tetapi juga mencakup proses pengumpulan, pembersihan, pengolahan, dan interpretasi data. Oleh karena itu, pemahaman konsep dasar data dan sains data menjadi fondasi penting dalam mendukung pengambilan keputusan yang lebih tepat dan berbasis informasi.

## **3. Bagian Metode Pengumpulan Data**

Metode pengumpulan data merupakan serangkaian cara yang digunakan untuk memperoleh data yang relevan, akurat, dan sesuai dengan tujuan analisis. Pemilihan metode yang tepat sangat penting karena akan memengaruhi kualitas data serta keandalan hasil yang dihasilkan. Dalam praktiknya, metode pengumpulan data dapat melibatkan pengambilan data langsung dari sumber utama maupun pemanfaatan data yang sudah tersedia dari berbagai sumber. Proses ini harus dilakukan secara sistematis dan terencana agar data yang diperoleh dapat mewakili kondisi yang sebenarnya. Dengan demikian, metode pengumpulan data menjadi tahap awal yang menentukan keberhasilan keseluruhan proses pengolahan dan analisis data.

## **4. Bagian Kualitas, Etika, dan Keamanan Data**

Kualitas, etika, dan keamanan data merupakan tiga aspek penting yang saling berkaitan dalam pengelolaan data agar dapat digunakan secara bertanggung jawab dan andal. Kualitas data menekankan pada keakuratan, kelengkapan, dan konsistensi informasi sehingga hasil analisis yang dihasilkan dapat dipercaya. Etika data berkaitan dengan

penggunaan data secara adil, transparan, dan tidak merugikan pihak mana pun, terutama ketika data menyangkut individu atau kelompok tertentu. Keamanan data berfokus pada upaya melindungi data dari akses tidak sah, kebocoran, atau penyalahgunaan melalui prosedur dan sistem yang memadai. Dengan memperhatikan ketiga aspek ini secara seimbang, pengelolaan data dapat mendukung pengambilan keputusan yang tepat sekaligus menjaga kepercayaan dan tanggung jawab sosial.

## **5. Bagian Pengolahan Awal dan Validasi Data**

Pengolahan awal dan validasi data merupakan tahap penting yang dilakukan setelah data dikumpulkan untuk memastikan data siap digunakan dalam proses analisis. Pada tahap ini, data diperiksa untuk mengidentifikasi kesalahan, duplikasi, nilai yang hilang, atau ketidakkonsistenan yang dapat memengaruhi hasil analisis. Proses pengolahan awal mencakup kegiatan seperti pembersihan, penyaringan, dan penyesuaian format data agar sesuai dengan kebutuhan pengolahan lebih lanjut. Validasi data dilakukan untuk memastikan bahwa data benar-benar merepresentasikan kondisi yang diukur dan memenuhi kriteria yang telah ditetapkan. Dengan pengolahan awal dan validasi yang baik, kualitas data dapat ditingkatkan sehingga hasil analisis menjadi lebih akurat dan dapat diandalkan.

## **6. Bagian Studi Kasus dan Aplikasi**

Bagian studi kasus dan aplikasi berfungsi untuk menunjukkan bagaimana konsep dan metode yang telah dibahas dapat diterapkan dalam situasi nyata. Melalui bagian ini, pembaca dapat melihat hubungan antara teori dan praktik sehingga pemahaman menjadi lebih konkret dan kontekstual. Studi kasus membantu menggambarkan proses pengambilan keputusan, pemilihan metode, serta cara menangani permasalahan yang muncul dalam pengolahan data. Sementara itu, contoh aplikasi memberikan gambaran langkah-langkah implementasi yang dapat dijadikan acuan dalam berbagai bidang. Dengan adanya bagian ini, pembaca diharapkan mampu mengembangkan kemampuan analitis sekaligus keterampilan praktis dalam menerapkan pengetahuan yang diperoleh.

## **7. Bagian Penutup**

Bagian penutup berfungsi untuk merangkum seluruh pembahasan yang telah diuraikan sebelumnya dan menegaskan kembali pokok-pokok penting yang perlu dipahami pembaca. Pada bagian ini, inti gagasan disajikan secara singkat agar pembaca memperoleh gambaran menyeluruh tentang materi yang telah dipelajari. Penutup juga dapat menekankan manfaat praktis dari pembahasan serta relevansinya dengan kebutuhan atau permasalahan yang dihadapi. Selain itu, bagian ini sering digunakan untuk memberikan arahan umum mengenai pengembangan atau penerapan lebih lanjut dari topik yang dibahas. Dengan demikian, bagian penutup menjadi pengikat akhir yang memperkuat pemahaman dan memberikan kesan utuh terhadap keseluruhan isi pembahasan.



# BAB II

## KONSEP DASAR PENGUMPULAN DATA

---

Pengumpulan data merupakan fondasi utama dalam sains data karena kualitas, relevansi, dan kelengkapan data akan sangat menentukan akurasi analisis serta ketepatan pengambilan keputusan. Oleh karena itu, pemahaman terhadap konsep dasar pengumpulan data, mulai dari penentuan sumber data, metode pengambilan, hingga pengelolaan awal data, menjadi langkah penting untuk memastikan bahwa proses analisis berjalan secara sistematis, efisien, dan dapat dipertanggungjawabkan.

### A. Pengertian dan Tujuan Pengumpulan Data

Pengumpulan data merupakan salah satu tahap paling fundamental dalam proses penelitian dan pengolahan informasi karena melalui kegiatan inilah berbagai fakta, angka, dan keterangan yang relevan dapat dihimpun secara sistematis. Pemahaman yang baik mengenai pengertian dan tujuan pengumpulan data menjadi penting agar proses perolehan informasi tidak hanya terarah dan terukur, tetapi juga mampu menghasilkan dasar yang kuat bagi analisis, penarikan kesimpulan, dan pengambilan keputusan.

#### 1. Pengertian Pengumpulan Data

Pengumpulan data merupakan proses sistematis untuk memperoleh informasi yang relevan, akurat, dan dapat digunakan sebagai dasar analisis dalam berbagai kegiatan ilmiah maupun praktis. Proses ini mencakup penentuan sumber data, pemilihan metode, serta prosedur pencatatan agar informasi yang diperoleh mencerminkan kondisi yang sebenarnya dan dapat diolah secara logis. Dalam konteks penelitian, pengumpulan data dipahami sebagai rangkaian aktivitas

terencana yang bertujuan menghimpun bukti empiris secara terstruktur, sebagaimana dinyatakan bahwa “*data collection is a series of interrelated activities aimed at gathering good information to answer emerging research questions*” (Creswell & Creswell, 2018). Dengan demikian, pengumpulan data tidak dapat dipisahkan dari prinsip ketelitian, konsistensi, dan keterukuran agar informasi yang dihimpun memiliki nilai ilmiah.

Secara konseptual, pengumpulan data juga dapat dipandang sebagai jembatan antara realitas yang diamati dengan proses analisis yang akan dilakukan. Kualitas proses ini sangat dipengaruhi oleh kejelasan rancangan, ketepatan instrumen, serta ketertiban dalam pencatatan dan pengelolaan awal data. Jika proses pengumpulan dilakukan secara sembarangan, maka informasi yang dihasilkan berpotensi menimbulkan bias dan kesalahan interpretasi. Oleh karena itu, pengumpulan data harus dipahami sebagai kegiatan inti yang menuntut perencanaan matang, prosedur yang jelas, dan pelaksanaan yang disiplin agar data yang diperoleh layak digunakan dalam proses pengolahan dan penarikan kesimpulan.

## **2. Tujuan Pengumpulan Data**

Tujuan pengumpulan data tidak hanya sekadar “mengumpulkan informasi”, tetapi memiliki fungsi strategis dalam berbagai kegiatan ilmiah dan praktis. Berikut tujuan-tujuannya secara rinci:

### **a. Menyediakan Dasar untuk Analisis**

Pengumpulan data sebagai tujuan menyediakan dasar untuk analisis menegaskan bahwa seluruh proses pengolahan dan penarikan kesimpulan bergantung pada mutu informasi yang dihimpun sejak awal. Data yang dikumpulkan secara sistematis, relevan, dan terukur akan membentuk fondasi yang kokoh bagi analisis, karena dari sanalah pola, hubungan, dan kecenderungan dapat diidentifikasi secara objektif, sebagaimana ditegaskan bahwa “*the quality of the data determines the quality of the analysis and the conclusions that can be drawn from it*” (Hair *et al.*, 2019). Dengan demikian, pengumpulan data tidak dapat dipahami sekadar sebagai aktivitas administratif, melainkan sebagai tahap konseptual yang menentukan apakah analisis mampu menggambarkan realitas secara akurat atau justru menghasilkan kesimpulan yang menyesatkan.

Pada praktiknya, data yang berfungsi sebagai dasar analisis harus disusun melalui prosedur yang jelas, instrumen yang tepat, serta mekanisme pencatatan yang konsisten agar dapat diolah secara logis dan sistematis. Ketepatan dalam memilih sumber data dan metode pengumpulan akan memengaruhi struktur data yang dihasilkan, sehingga analisis dapat dilakukan secara lebih mendalam, terarah, dan dapat dipertanggungjawabkan. Oleh karena itu, tujuan menyediakan dasar untuk analisis menempatkan pengumpulan data sebagai titik awal yang strategis, karena kualitas proses ini akan menentukan kekuatan bukti, ketajaman interpretasi, dan validitas kesimpulan yang dihasilkan.

b. Menjawab Masalah atau Pertanyaan Penelitian

Pengumpulan data dalam penelitian berperan penting dalam menemukan jawaban atas pertanyaan penelitian karena data yang dikumpulkan secara sistematis akan menjadi bahan empiris utama untuk menguji hipotesis atau menjawab rumusan masalah yang telah dirumuskan sejak awal. Ketika data dihimpun dengan memperhatikan pertanyaan penelitian, peneliti dapat menelusuri pola, hubungan, dan bukti yang diperlukan untuk menjawab pertanyaan tersebut secara objektif serta merinci fenomena yang diteliti. Hal ini sesuai dengan pendapat yang menyatakan bahwa pengumpulan data dilakukan untuk mendapatkan informasi yang diperlukan guna menjawab pertanyaan dan menguji hipotesis dalam penelitian, sehingga memberikan dasar kuat bagi interpretasi hasil studi (Taherdoost, 2021). Dengan demikian, proses pengumpulan data tidak hanya sekadar mengumpulkan fakta secara acak, tetapi harus diarahkan secara cermat untuk menjawab fokus penelitian secara tepat dan efektif.

Pada praktik penelitian, data yang relevan terhadap pertanyaan dan tujuan penelitian akan membantu peneliti melakukan analisis yang valid dan menarik kesimpulan yang sah, karena tanpa menyasar pertanyaan penelitian, data yang terkumpul bisa menjadi tidak relevan dan membingungkan saat dianalisis. Karena itu, penentuan teknik, instrumen, dan sumber data yang tepat juga sangat penting agar data dapat menjawab kebutuhan pertanyaan penelitian dengan akurat dan efisien. Ketelitian pada tahap pengumpulan data akan mempermudah

proses selanjutnya dalam penelitian, termasuk pengolahan, interpretasi, dan penarikan kesimpulan yang kredibel sesuai kerangka penelitian yang telah dirancang.

c. Mendukung Pengambilan Keputusan

Pengumpulan data memiliki peran krusial dalam mendukung pengambilan keputusan karena informasi yang akurat dan relevan menjadi dasar untuk memilih alternatif tindakan yang tepat dalam suatu konteks organisasi, bisnis, maupun penelitian. Dengan data yang komprehensif, proses pemilihan strategi, penilaian risiko, serta evaluasi opsi dapat dilakukan secara objektif, sehingga keputusan yang dihasilkan lebih rasional dan dapat dipertanggungjawabkan, terutama ketika menghadapi ketidakpastian. Hal ini sejalan dengan pernyataan bahwa pengambilan keputusan yang efektif bergantung pada ketersediaan dan kualitas informasi yang diperoleh dari proses pengumpulan data yang sistematis dan valid (Sharma, 2019). Oleh karena itu, pengumpulan data menjadi instrumen penting yang meminimalkan asumsi subjektif dan meningkatkan efektivitas keputusan dalam berbagai bidang.

Pada prakteknya, data yang mendukung pengambilan keputusan dapat berbentuk angka, fakta, tren historis, maupun *feedback* langsung dari lingkungan yang bersangkutan, yang kemudian dianalisis untuk memahami konteks serta implikasi potensial dari setiap alternatif keputusan. Ketika data yang disajikan lengkap dan terstruktur, para pengambil keputusan mampu mengevaluasi konsekuensi dari berbagai pilihan dengan mempertimbangkan bukti empiris, bukan sekadar intuisi atau pengalaman semata. Dengan demikian, tujuan pengumpulan data yang berkaitan dengan pengambilan keputusan memberi penekanan bahwa informasi yang tepat pada waktu yang tepat akan meningkatkan kualitas keputusan dan mengurangi risiko kesalahan dalam berbagai situasi.

d. Menggambarkan Kondisi atau Fenomena Nyata

Pengumpulan data bertujuan untuk menggambarkan kondisi atau fenomena nyata secara objektif sehingga realitas yang diteliti dapat direpresentasikan melalui informasi yang terukur dan dapat dianalisis. Dalam kerangka penelitian, data berfungsi

sebagai cerminan keadaan lapangan karena melalui proses pengumpulan yang sistematis peneliti dapat merekam fakta, pola, dan variasi yang benar-benar terjadi, sebagaimana dinyatakan bahwa *“data collection involves gathering information that will represent the reality of the phenomenon being studied”* (Creswell & Creswell, 2018). Dengan demikian, pengumpulan data tidak hanya berperan sebagai tahap teknis, tetapi juga sebagai proses konseptual untuk memastikan bahwa gambaran yang dihasilkan sesuai dengan kondisi aktual dan bukan sekadar asumsi.

Agar mampu menggambarkan fenomena secara akurat, data harus diperoleh dari sumber yang relevan dengan menggunakan metode dan instrumen yang tepat sehingga informasi yang dihimpun tidak menyimpang dari konteks yang diteliti. Ketelitian dalam menentukan prosedur pengumpulan akan memengaruhi sejauh mana data mampu merepresentasikan keadaan sebenarnya, baik dalam bentuk angka, deskripsi, maupun pola hubungan antarvariabel. Oleh karena itu, tujuan menggambarkan kondisi atau fenomena nyata menempatkan pengumpulan data sebagai jembatan penting antara realitas empiris dan proses analisis, karena kualitas gambaran yang dihasilkan sangat bergantung pada ketepatan dan ketertiban pada tahap ini.

e. Menjamin Keakuratan dan Keandalan Hasil

Pengumpulan data bertujuan untuk menjamin keakuratan dan keandalan hasil karena kualitas kesimpulan sangat ditentukan oleh ketepatan informasi yang dihimpun sejak awal proses penelitian atau analisis. Data yang diperoleh melalui prosedur yang sistematis, instrumen yang teruji, serta pencatatan yang konsisten akan meminimalkan kesalahan pengukuran dan bias, sehingga hasil yang dihasilkan lebih dapat dipercaya, sebagaimana ditegaskan bahwa *“the validity and reliability of research findings depend heavily on the quality of the data collection procedures”* (Creswell & Creswell, 2018). Dengan demikian, pengumpulan data tidak dapat dipisahkan dari upaya menjaga ketelitian metodologis agar hasil analisis benar-benar mencerminkan kondisi yang diteliti.

Pada praktiknya, jaminan keakuratan dan keandalan menuntut peneliti atau analis untuk cermat dalam memilih sumber data, merancang instrumen, serta menerapkan teknik

pengumpulan yang sesuai dengan karakteristik objek yang dikaji. Ketika proses ini dilakukan secara disiplin, data yang dihasilkan akan lebih stabil, konsisten, dan dapat direplikasi, sehingga memperkuat kepercayaan terhadap temuan yang diperoleh. Oleh karena itu, tujuan pengumpulan data dalam menjamin keakuratan dan keandalan hasil menempatkan tahap ini sebagai fondasi utama yang menentukan apakah suatu kesimpulan layak dijadikan dasar pemahaman, kebijakan, atau pengambilan keputusan.

## **B. Klasifikasi Data: Data Primer dan Data Sekunder**

Pada kegiatan penelitian, sains data, maupun analisis kebijakan, data dapat diklasifikasikan berdasarkan sumber perolehannya, yaitu menjadi data primer dan data sekunder. Klasifikasi ini penting karena menentukan cara pengumpulan, tingkat keakuratan, biaya, waktu, serta kesesuaian data dengan tujuan analisis yang ingin dicapai. Pemilihan jenis data yang tepat akan sangat memengaruhi kualitas hasil penelitian atau keputusan yang diambil.

### **1. Data Primer**

Data primer adalah data yang diperoleh secara langsung dari sumber pertama oleh peneliti atau pengumpul data untuk tujuan tertentu. Data ini dikumpulkan secara khusus sesuai dengan kebutuhan penelitian, sehingga informasinya lebih spesifik, aktual, dan relevan terhadap masalah yang sedang dikaji. Karena diperoleh langsung dari lapangan atau objek penelitian, data primer mencerminkan kondisi nyata pada saat pengumpulan dilakukan. Dalam konteks penelitian, sains data, maupun evaluasi kebijakan, data primer memiliki peran penting karena peneliti mengendalikan sendiri proses pengumpulannya, mulai dari perancangan instrumen, pemilihan responden, hingga teknik pengambilan data. Hal ini memungkinkan penyesuaian data dengan variabel yang ingin diteliti dan tujuan analisis yang ingin dicapai. Namun, proses pengumpulan data primer biasanya memerlukan perencanaan yang matang, waktu yang lebih lama, serta biaya dan tenaga yang tidak sedikit.

#### **a. Ciri-Ciri Data Primer**

Ciri-ciri data primer dapat dipahami dari cara perolehannya yang langsung dari sumber asli serta dari tingkat kendali peneliti

terhadap proses pengumpulan informasi tersebut. Data primer dikumpulkan secara khusus untuk menjawab kebutuhan penelitian tertentu sehingga bersifat lebih spesifik, kontekstual, dan aktual, sebagaimana dijelaskan bahwa data primer merupakan data yang dikumpulkan langsung oleh peneliti untuk tujuan penelitian yang sedang dilakukan (Taherdoost, 2021). Oleh karena itu, salah satu ciri utamanya adalah relevansi yang tinggi terhadap masalah yang dikaji karena sejak awal data memang dirancang sesuai dengan fokus penelitian.

Data primer memiliki ciri bahwa proses pengumpulannya berada di bawah kendali penuh peneliti, mulai dari penentuan instrumen, pemilihan responden, hingga teknik pengambilan data yang digunakan. Kondisi ini memungkinkan peneliti mengontrol kualitas data, meminimalkan kesalahan pengukuran, serta menyesuaikan prosedur dengan karakteristik objek yang diteliti. Namun, konsekuensinya adalah proses tersebut biasanya memerlukan waktu, biaya, dan tenaga yang lebih besar dibandingkan penggunaan data yang sudah tersedia.

b. Metode Pengumpulan Data Primer

Data primer dapat dikumpulkan melalui berbagai metode, antara lain:

1) Wawancara

Wawancara dalam metode pengumpulan data primer merupakan teknik yang memungkinkan peneliti memperoleh informasi secara langsung, mendalam, dan terarah melalui proses tanya jawab dengan responden, sehingga data yang dihasilkan tidak hanya bersifat faktual tetapi juga kaya akan konteks dan penjelasan mengenai pengalaman maupun pandangan subjek yang diteliti, sebagaimana dijelaskan bahwa wawancara adalah cara untuk mengumpulkan data di mana peneliti menanyakan pertanyaan kepada partisipan dan merekam jawabannya untuk memperoleh informasi yang relevan dengan tujuan penelitian (Saunders, Lewis, & Thornhill, 2019). Oleh karena itu, penggunaan wawancara menuntut kesiapan peneliti dalam menyusun pedoman pertanyaan, membangun komunikasi yang efektif, serta menjaga ketelitian pencatatan agar data yang diperoleh benar-benar mencerminkan kondisi dan pandangan

responden sebagai sumber data primer yang dapat dipertanggungjawabkan.

## 2) Kuesioner/Angket

Kuesioner atau angket merupakan salah satu metode utama dalam pengumpulan data primer yang digunakan untuk memperoleh informasi langsung dari responden melalui serangkaian pertanyaan yang disusun secara sistematis, sehingga peneliti dapat mengumpulkan data dalam jumlah besar secara relatif efisien sekaligus menjaga keseragaman struktur jawaban untuk keperluan analisis, sebagaimana dijelaskan bahwa kuesioner adalah instrumen yang berisi pertanyaan standar yang diajukan kepada responden untuk mengukur sikap, pendapat, atau karakteristik tertentu secara terstruktur (Krosnick & Presser, 2018). Dengan demikian, penggunaan kuesioner menuntut perancangan item yang jelas, bahasa yang mudah dipahami, serta tata cara penyebaran yang tepat agar data yang dihasilkan benar-benar mencerminkan kondisi responden dan layak digunakan sebagai data primer yang akurat dan dapat diandalkan.

## 3) Observasi

Observasi dalam metode pengumpulan data primer merupakan teknik yang dilakukan dengan cara mengamati secara langsung objek, perilaku, atau peristiwa yang diteliti sehingga peneliti dapat memperoleh data yang bersifat faktual, kontekstual, dan mencerminkan kondisi nyata di lapangan, karena melalui pengamatan yang terencana peneliti dapat merekam fenomena sebagaimana adanya, sebagaimana dijelaskan bahwa observasi adalah proses sistematis untuk memperhatikan, mencatat, dan mendokumentasikan kejadian serta perilaku dalam situasi sosial tertentu (Flick, 2018). Dengan demikian, penggunaan observasi menuntut ketelitian dalam menetapkan fokus pengamatan, konsistensi dalam pencatatan, serta kepekaan terhadap konteks agar data yang dihasilkan benar-benar akurat, relevan, dan layak digunakan sebagai data primer dalam proses analisis.

#### 4) Eksperimen/Pengukuran Langsung

Eksperimen atau pengukuran langsung dalam metode pengumpulan data primer merupakan teknik yang dilakukan dengan cara mengendalikan variabel tertentu dan mengamati dampaknya secara sistematis, sehingga peneliti dapat memperoleh data yang bersifat objektif, terukur, dan mencerminkan hubungan sebab-akibat secara lebih jelas, sebagaimana dijelaskan bahwa penelitian eksperimen melibatkan manipulasi satu atau lebih variabel dan pengukuran hasilnya untuk menentukan pengaruh perlakuan terhadap variabel terikat (Gravetter & Forzano, 2020). Dengan demikian, penggunaan eksperimen atau pengukuran langsung menuntut perancangan prosedur yang ketat, pemilihan alat ukur yang tepat, serta pengendalian kondisi penelitian agar data yang dihasilkan benar-benar akurat, reliabel, dan layak digunakan sebagai data primer dalam proses analisis.

#### 5) Survei Lapangan

Survei lapangan dalam metode pengumpulan data primer merupakan teknik yang dilakukan dengan cara menghimpun informasi langsung dari responden di lokasi atau konteks yang diteliti melalui instrumen terstruktur, sehingga peneliti dapat memperoleh gambaran empiris yang representatif mengenai karakteristik, sikap, atau kondisi suatu populasi, sebagaimana dijelaskan bahwa survei adalah metode sistematis untuk mengumpulkan informasi dari sampel responden guna menggambarkan atau menjelaskan karakteristik populasi yang lebih luas (Fowler, 2019). Dengan demikian, pelaksanaan survei lapangan menuntut perencanaan yang cermat dalam penentuan sampel, penyusunan instrumen, serta prosedur pengumpulan data agar informasi yang diperoleh akurat, relevan, dan layak digunakan sebagai data primer dalam proses analisis.

#### c. Kelebihan Data Primer

Kelebihan data primer terletak pada tingkat relevansinya yang tinggi terhadap tujuan penelitian karena data ini dikumpulkan secara khusus sesuai dengan kebutuhan analisis, sehingga variabel, konteks, dan kedalaman informasi dapat

disesuaikan dengan fokus kajian yang sedang dilakukan. Hal ini sejalan dengan pandangan bahwa data primer dikumpulkan untuk tujuan spesifik penelitian yang sedang dikerjakan sehingga memberikan kontrol yang lebih besar terhadap isi dan kualitas data, sebagaimana dinyatakan bahwa “*primary data are collected for the specific purpose at hand*” (Malhotra, Nunan, & Birks, 2019). Dengan kondisi tersebut, peneliti memiliki keleluasaan untuk memastikan bahwa data yang diperoleh benar-benar mencerminkan fenomena yang diteliti dan tidak terikat pada keterbatasan definisi atau struktur data yang dibuat pihak lain.

Data primer memiliki kelebihan dalam hal ketepatan waktu dan ketepatan konteks karena dikumpulkan langsung dari sumber pada saat penelitian berlangsung, sehingga lebih aktual dan sensitif terhadap perubahan situasi di lapangan. Peneliti juga dapat mengontrol instrumen, prosedur, serta kualitas proses pengumpulan data, yang pada akhirnya meningkatkan validitas dan reliabilitas temuan sekaligus meminimalkan risiko bias yang sering muncul ketika menggunakan data yang tidak dirancang khusus untuk tujuan penelitian tertentu.

d. Keterbatasan Data Primer

Keterbatasan data primer terutama berkaitan dengan kebutuhan sumber daya yang besar, karena proses pengumpulannya menuntut perencanaan, pelaksanaan di lapangan, serta pengolahan awal yang memakan waktu dan biaya, sehingga tidak selalu efisien untuk semua jenis penelitian. Hal ini sejalan dengan pandangan bahwa pengumpulan data primer sering kali memerlukan waktu, biaya, dan upaya yang signifikan dibandingkan penggunaan data yang sudah tersedia, sebagaimana dinyatakan bahwa “*collecting primary data is often time-consuming and expensive*” (Saunders, Lewis, & Thornhill, 2019). Kondisi tersebut menunjukkan bahwa meskipun data primer lebih spesifik dan relevan, keterbatasan sumber daya dapat menjadi hambatan serius dalam pelaksanaannya.

Data primer juga memiliki keterbatasan dari sisi cakupan dan risiko kesalahan pengumpulan, karena jumlah responden, lokasi, atau periode pengamatan sering kali dibatasi oleh kemampuan peneliti. Jika instrumen tidak dirancang dengan baik atau pelaksanaan di lapangan tidak konsisten, maka data yang

diperoleh berpotensi mengandung bias atau ketidakakuratan yang dapat memengaruhi kualitas analisis dan kesimpulan. Oleh karena itu, keterbatasan data primer perlu dipahami sebagai konsekuensi dari proses pengumpulan langsung yang menuntut ketelitian metodologis sekaligus kesiapan sumber daya yang memadai.

## 2. Data Sekunder

Data sekunder adalah data yang tidak dikumpulkan langsung oleh peneliti dari sumber pertama, melainkan diperoleh melalui pihak lain atau dari dokumen dan publikasi yang sudah tersedia sebelumnya. Data ini umumnya telah dihimpun oleh lembaga, instansi, atau peneliti lain untuk tujuan tertentu, kemudian dimanfaatkan kembali untuk kepentingan analisis, penelitian, atau pengambilan keputusan. Karena sifatnya yang sudah ada, data sekunder sering digunakan untuk menghemat waktu dan biaya serta sebagai pelengkap atau pembanding terhadap data primer.

Pada praktik penelitian dan sains data, data sekunder berperan penting untuk memberikan gambaran awal, memperluas konteks analisis, serta membantu mengidentifikasi pola atau tren yang berskala lebih luas. Namun, karena peneliti tidak terlibat langsung dalam proses pengumpulannya, kualitas dan kesesuaian data sekunder perlu dievaluasi secara kritis agar tidak menimbulkan bias atau kesalahan interpretasi. Oleh sebab itu, penggunaan data sekunder menuntut ketelitian dalam menilai sumber, tahun penerbitan, metode pengumpulan, dan relevansinya dengan tujuan studi.

### a. Ciri-Ciri Data Sekunder

Ciri-ciri data sekunder dapat dikenali dari sumber dan cara perolehannya, karena data ini tidak dikumpulkan langsung oleh peneliti melainkan berasal dari dokumen, publikasi, atau basis data yang telah tersedia sebelumnya, sehingga peneliti hanya memanfaatkan informasi yang telah dihimpun untuk keperluan penelitian atau analisis tertentu. Hal ini sejalan dengan pandangan bahwa “*secondary data are those which have been collected by someone else for purposes other than the current research project*” (Bryman & Bell, 2019), sehingga data sekunder cenderung lebih mudah diakses dan efisien dari sisi waktu dan biaya dibanding data primer.

Data sekunder memiliki ciri bahwa kualitas, relevansi, dan cakupan informasinya sangat bergantung pada pihak yang awalnya mengumpulkannya, sehingga peneliti perlu melakukan evaluasi kritis terhadap sumber, metode, dan tahun pengumpulan data. Karena tidak berada di bawah kendali langsung peneliti, data sekunder mungkin memiliki keterbatasan konteks atau kesesuaian dengan variabel yang diteliti, namun tetap bermanfaat untuk analisis perbandingan, tren historis, atau sebagai bahan pendukung studi awal yang memperluas wawasan penelitian.

b. Sumber-Sumber Data Sekunder

Sumber-sumber data sekunder mencakup berbagai dokumen, publikasi, dan basis data yang telah dikumpulkan oleh pihak lain untuk keperluan tertentu, sehingga peneliti dapat memanfaatkannya kembali untuk mendukung penelitian atau analisis tanpa harus melakukan pengumpulan data dari awal. Hal ini sejalan dengan pandangan bahwa data sekunder dapat diperoleh dari literatur, laporan penelitian sebelumnya, arsip institusi, atau basis data daring yang tersedia, dan dapat digunakan untuk analisis tambahan, perbandingan, atau pengayaan informasi (Saunders, Lewis, & Thornhill, 2019).

Sumber data sekunder dapat bersifat resmi maupun nonresmi, termasuk statistik pemerintah, laporan tahunan perusahaan, jurnal akademik, dan publikasi organisasi internasional, sehingga memberikan fleksibilitas bagi peneliti untuk memilih sumber yang relevan dengan tujuan penelitian. Penggunaan berbagai sumber ini memungkinkan data sekunder menyediakan cakupan yang lebih luas dan historis, sekaligus memperkuat validitas analisis dengan membandingkan atau memadukan informasi dari berbagai referensi yang tersedia.

c. Kelebihan Data Sekunder

Kelebihan data sekunder terletak pada efisiensi waktu dan biaya, karena data ini sudah tersedia dan tidak memerlukan proses pengumpulan langsung di lapangan, sehingga peneliti dapat segera menggunakannya untuk analisis atau sebagai bahan pembanding dengan penelitian lain. Hal ini sejalan dengan pandangan bahwa “*secondary data offer advantages in terms of cost and time savings, making it practical for exploratory studies and preliminary research*” (Collis & Hussey, 2018), sehingga

penggunaan data sekunder memungkinkan penelitian menjadi lebih cepat dan ekonomis tanpa mengurangi nilai informasinya.

Data sekunder sering kali menyediakan cakupan yang lebih luas dan historis, termasuk data statistik, laporan tahunan, atau arsip penelitian sebelumnya, sehingga memungkinkan peneliti untuk menganalisis tren, membandingkan kondisi antarperiode, atau memperkuat validitas temuan dengan memadukan berbagai sumber. Dengan ketersediaan data yang beragam ini, peneliti dapat memperoleh wawasan yang lebih lengkap dan mendalam tanpa harus menghadapi keterbatasan sumber daya yang biasanya terkait dengan pengumpulan data primer secara langsung.

d. Keterbatasan Data Sekunder

Keterbatasan data sekunder terutama berkaitan dengan relevansi dan kesesuaian data terhadap tujuan penelitian saat ini, karena data yang dikumpulkan oleh pihak lain mungkin tidak sepenuhnya sesuai dengan variabel atau konteks yang ingin dianalisis, sehingga peneliti harus melakukan evaluasi kritis sebelum menggunakannya. Hal ini sejalan dengan pandangan bahwa “*secondary data may lack relevance or accuracy for the current research objectives, requiring careful assessment of its suitability*” (Saunders, Lewis, & Thornhill, 2019), sehingga meskipun data sekunder lebih mudah diakses, kualitas dan kecocokannya tetap menjadi pertimbangan penting dalam penelitian.

Data sekunder juga memiliki keterbatasan dari sisi keandalan dan ketepatan waktu, karena proses pengumpulan awal, metode, dan konteksnya ditentukan oleh pihak lain, sehingga informasi yang diperoleh mungkin sudah tidak mutakhir atau mengandung bias tertentu. Oleh karena itu, penggunaan data sekunder menuntut kehati-hatian dalam menilai sumber, validitas, dan kecocokan data agar hasil analisis tetap dapat dipertanggungjawabkan secara ilmiah.

### C. Metode Konvensional dan Berbasis Teknologi

Metode pengumpulan data dalam sains data dapat dilakukan melalui dua pendekatan utama, yaitu metode konvensional dan metode berbasis teknologi, yang masing-masing memiliki karakteristik,

kelebihan, dan keterbatasan tersendiri. Metode konvensional menekankan pengumpulan data secara manual melalui wawancara, kuesioner, observasi, dan studi dokumen, sedangkan metode berbasis teknologi memanfaatkan perangkat digital, sistem otomatis, dan aplikasi untuk memperoleh, menyimpan, dan mengelola data secara cepat, efisien, dan dalam skala besar.

## 1. Metode Konvensional

Metode konvensional adalah pendekatan pengumpulan data yang dilakukan secara manual atau tradisional, dengan mengandalkan interaksi langsung antara peneliti dan sumber data. Pendekatan ini menekankan pada observasi, pencatatan, dan pengambilan informasi melalui proses tatap muka, kuesioner cetak, wawancara, atau studi dokumen. Tujuan utamanya adalah memperoleh data yang akurat dan relevan dengan konteks, baik untuk penelitian kuantitatif maupun kualitatif.

### a. Karakteristik Metode Konvensional

Metode konvensional dalam pengumpulan data memiliki karakteristik utama berupa interaksi langsung antara peneliti dan responden, di mana informasi diperoleh melalui wawancara, observasi, kuesioner cetak, atau studi dokumen yang memungkinkan peneliti menangkap konteks dan nuansa perilaku yang sulit diakses melalui metode digital (Saunders *et al.*, 2019). Pendekatan ini menekankan kualitas data yang tinggi dengan kemampuan untuk melakukan probing dan klarifikasi jawaban secara *real-time*, sehingga informasi yang diperoleh lebih mendalam dan relevan untuk analisis kualitatif maupun kuantitatif.

Metode konvensional bersifat fleksibel karena peneliti dapat menyesuaikan pertanyaan atau teknik pengumpulan data sesuai situasi di lapangan, memungkinkan adaptasi terhadap respons dan kondisi unik dari setiap responden. Metode ini juga membutuhkan sumber daya manusia yang cukup, baik enumerator maupun supervisor, serta waktu yang relatif lama untuk pengumpulan dan pencatatan data, sehingga skalabilitasnya terbatas jika dibandingkan dengan metode berbasis teknologi.

### b. Jenis-Jenis Metode Konvensional

Jenis-jenis metode konvensional dalam pengumpulan data mencakup berbagai pendekatan yang menekankan interaksi langsung antara peneliti dan responden, serta pengamatan langsung terhadap fenomena yang diteliti, sehingga memungkinkan pengumpulan data yang mendalam dan kontekstual sesuai tujuan penelitian (Creswell & Creswell, 2018). Metode ini mencakup wawancara tatap muka yang dapat berupa wawancara terstruktur, semi-terstruktur, atau tidak terstruktur, di mana pertanyaan dapat disesuaikan secara fleksibel selama proses interaksi untuk memperoleh informasi yang lebih akurat dan relevan dari responden.

Kuesioner cetak atau formulir tertulis juga menjadi metode konvensional yang umum digunakan, di mana responden mengisi pertanyaan secara manual, sehingga data yang diperoleh dapat dianalisis secara kuantitatif maupun kualitatif, dengan catatan bahwa desain instrumen harus jelas dan terstruktur untuk mengurangi risiko ambiguitas atau kesalahan pengisian. Observasi lapangan juga termasuk metode ini, baik melalui observasi partisipatif, di mana peneliti terlibat langsung dalam kegiatan yang diamati, maupun non-partisipatif, di mana peneliti hanya mengamati perilaku atau kondisi secara objektif, memungkinkan identifikasi pola, kebiasaan, atau fenomena yang sulit dijelaskan melalui wawancara saja.

Metode tambahan yang sering digunakan dalam pengumpulan data konvensional adalah *Focus Group Discussion* (FGD), yang melibatkan diskusi kelompok kecil untuk menggali persepsi kolektif, serta studi dokumen atau arsip yang menganalisis laporan, catatan historis, notulen, atau dokumen resmi untuk memperoleh data sekunder yang relevan. Teknik ini juga mencakup studi kasus, di mana data dikumpulkan secara mendalam melalui kombinasi wawancara, observasi, dan analisis dokumen, sehingga memungkinkan pemahaman holistik terhadap suatu entitas atau fenomena tertentu, memberikan konteks yang lebih kaya dibanding metode digital.

c. Langkah Pelaksanaan Metode Konvensional

Langkah pelaksanaan metode konvensional dimulai dengan perencanaan yang matang, di mana peneliti menetapkan tujuan penelitian, merumuskan pertanyaan yang relevan, dan

menentukan populasi serta sampel yang akan diteliti, sehingga seluruh proses pengumpulan data dapat dilakukan secara sistematis dan terarah (Saunders *et al.*, 2019). Selanjutnya, desain instrumen pengumpulan data seperti kuesioner, panduan wawancara, atau lembar observasi dibuat dengan memperhatikan kejelasan, konsistensi, dan kemampuan instrumen untuk menangkap informasi yang dibutuhkan secara akurat.

Tahap berikutnya adalah uji coba instrumen pada sampel kecil untuk mengidentifikasi kekurangan atau ambiguitas, diikuti dengan pelatihan enumerator atau peneliti lapangan agar memahami prosedur, etika penelitian, dan teknik pencatatan data yang benar, sehingga proses pengumpulan data di lapangan dapat berjalan lancar dan konsisten. Pengumpulan data kemudian dilakukan melalui wawancara, observasi, studi dokumen, atau survei lapangan sesuai protokol yang telah disiapkan, dengan pencatatan rinci yang mencakup metadata seperti waktu, lokasi, dan kondisi pengumpulan data.

Dilakukan proses validasi dan pembersihan untuk memeriksa konsistensi, kelengkapan, dan akurasi data, termasuk menghapus duplikasi, menangani *missing values*, dan memastikan data siap dianalisis. Data kemudian diorganisasi dan disimpan secara sistematis agar memudahkan proses entri, analisis, dan pelaporan, sekaligus menjaga integritas dan keamanan informasi untuk penelitian selanjutnya. Langkah-langkah ini memastikan bahwa metode konvensional tidak hanya menghasilkan data yang valid, tetapi juga meminimalkan kesalahan manusia dan bias yang mungkin terjadi selama pengumpulan data.

d. Kelebihan Metode Konvensional

Metode konvensional dalam pengumpulan data memiliki kelebihan signifikan karena memungkinkan peneliti untuk memperoleh informasi yang kaya, mendalam, dan kontekstual melalui interaksi langsung dengan responden maupun objek penelitian, sehingga data yang dikumpulkan mampu mencerminkan realitas sosial atau perilaku secara komprehensif dan akurat (Turner *et al.*, 2022). Interaksi semacam ini sangat penting dalam memahami motivasi, sikap, pengalaman, serta nuansa yang tidak mudah tertangkap oleh metode otomatis atau

digital, terutama ketika penelitian bertujuan mengungkap makna, konteks budaya, atau fenomena kompleks dalam kehidupan nyata.

Kelebihan metode konvensional terletak pada kemampuan peneliti untuk mengadaptasi instrumen pengumpulan data secara langsung, melakukan klarifikasi jika responden tidak memahami pertanyaan, serta menjalin hubungan yang dapat meningkatkan kualitas respons, yang pada akhirnya meningkatkan validitas internal data dan membantu menghasilkan kesimpulan yang lebih dapat dipertanggungjawabkan dalam studi ilmiah.

e. Keterbatasan Metode Konvensional

Metode konvensional memiliki keterbatasan signifikan berupa kebutuhan waktu, tenaga, dan biaya yang besar untuk melaksanakan proses pengumpulan data karena wawancara tatap muka, observasi lapangan, atau survei kuesioner cetak memerlukan koordinasi personal, perjalanan ke lokasi, pencetakan instrumen, serta pencatatan manual yang memakan waktu lama sehingga skalabilitasnya rendah terutama untuk sampel yang besar (Creswell & Creswell, 2018). Keterbatasan ini sering menghasilkan proses yang lambat dan kurang efisien jika dibandingkan dengan metode berbasis teknologi, yang dapat mengumpulkan data dalam jumlah besar secara otomatis dalam waktu singkat tanpa memerlukan banyak sumber daya manusia di lapangan.

Metode konvensional rentan terhadap bias manusia seperti *interviewer* bias, kesalahan pencatatan, serta *desirability* bias dari responden yang cenderung memberikan jawaban yang dianggap “diinginkan”, sehingga mengurangi objektivitas data dan konsistensi antar pengumpul data. Risiko tersebut mempengaruhi kualitas data dan dapat menimbulkan kesalahan interpretasi jika tidak ditangani dengan prosedur kontrol kualitas yang ketat, yang pada akhirnya berdampak pada validitas hasil penelitian secara keseluruhan.

## 2. Metode Berbasis Teknologi

Metode berbasis teknologi adalah pendekatan pengumpulan data yang memanfaatkan perangkat digital, sistem otomatis, dan teknologi informasi untuk memperoleh, menyimpan, dan mengelola data. Metode

ini memungkinkan pengumpulan data dalam jumlah besar secara cepat, efisien, dan sering kali *real-time*, sehingga memudahkan analisis lanjutan dalam konteks sains data dan penelitian modern.

a. Karakteristik Metode Berbasis Teknologi

Metode berbasis teknologi memiliki karakteristik utama berupa kemampuan mengumpulkan dan memproses data secara otomatis dan *real-time* menggunakan perangkat digital seperti aplikasi *mobile*, sensor, alat IoT, *web scraping*, dan sistem *database* terintegrasi yang memungkinkan pengambilan data dalam volume besar dengan efisiensi tinggi dan minim keterlibatan manual (Kitchin, 2018). Karakteristik ini mendukung skalabilitas pengumpulan data yang jauh lebih cepat dibandingkan metode tradisional serta memungkinkan analisis data yang langsung terhubung dengan sistem pemrosesan seperti *cloud computing*, *big data analytics*, dan *machine learning* untuk menghasilkan *insight* yang lebih dinamis.

Metode berbasis teknologi memiliki akurasi dan konsistensi yang lebih tinggi karena penggunaan algoritma, format digital baku, dan integrasi sistem yang mengurangi risiko kesalahan manusia seperti kesalahan entri atau bias pencatat, serta memungkinkan pengelolaan data secara terpusat, penyimpanan aman, dan kemudahan akses data lintas platform bagi peneliti atau tim analisis. Integrasi teknologi juga memfasilitasi pemantauan berkelanjutan, pengambilan keputusan berbasis data secara cepat, serta kemampuan menyimpan metadata yang komprehensif untuk keperluan audit dan pelacakan perubahan data dari waktu ke waktu.

b. Jenis-Jenis Metode Berbasis Teknologi

Jenis-jenis metode berbasis teknologi dalam pengumpulan data mencakup berbagai pendekatan digital yang digunakan untuk memperoleh, mengekstraksi, dan merekam informasi secara otomatis dari sumber elektronik seperti internet, aplikasi, perangkat sensor, atau perangkat *mobile*, di mana teknik seperti *web scraping* untuk mengekstrak konten dari halaman web, pemanfaatan API (*Application Programming Interface*) untuk mengakses data terstruktur dari layanan digital, serta pengambilan data sensor dari perangkat IoT (*Internet of Things*) merupakan contoh nyata dari metode ini yang memungkinkan

pengumpulan data dalam volume besar secara efisien untuk analisis lebih lanjut (Nasution, 2024). Metode-metode tersebut juga sering melibatkan pengumpulan data dari media sosial, platform *online*, dan aplikasi *mobile* melalui teknik otomatis yang mempercepat proses pengambilan serta mengurangi keterlibatan manual, sedangkan sensor dan perangkat IoT dapat menangkap keadaan lingkungan atau aktivitas pengguna secara *real-time* dan mentransmisikan data terus-menerus ke sistem pusat untuk penyimpanan dan pemrosesan.

Jenis lain dari metode berbasis teknologi termasuk crowdsensing yang memanfaatkan sensor dan perangkat *mobile* dari sekelompok besar pengguna untuk berbagi data terkait fenomena tertentu di lingkungan serta teknik pengumpulan data melalui aplikasi khusus yang mengintegrasikan antarmuka digital dan perangkat keras untuk merekam *input* pengguna secara langsung, sehingga penelitian dan analisis dapat dilakukan dengan cakupan yang lebih luas dan granular tanpa batasan fisik serta dapat menyokong kebutuhan penelitian modern seperti *big data* dan *machine learning*.

c. Langkah Pelaksanaan Metode Berbasis Teknologi

Langkah pelaksanaan metode berbasis teknologi dimulai dengan perencanaan yang matang, di mana peneliti menentukan jenis data yang diperlukan, sumber data digital, serta platform atau perangkat teknologi yang akan digunakan, termasuk sistem sensor, aplikasi *mobile*, *database*, atau API, sehingga proses pengumpulan data dapat dilakukan secara otomatis dan efisien sesuai tujuan penelitian (Kitchin, 2018). Tahap ini juga mencakup desain instrumen digital dan pengaturan sistem integrasi agar data dapat dikumpulkan secara *real-time*, tersimpan secara aman, dan siap diproses untuk analisis lebih lanjut menggunakan tools analitik atau *big data*.

Tahap implementasi dilakukan dengan memasang sensor, mengaktifkan aplikasi, atau menjalankan program otomatis untuk mengekstrak data dari berbagai sumber, kemudian data yang terkumpul divalidasi dan dibersihkan untuk memastikan konsistensi, akurasi, dan kelengkapan sebelum dianalisis. Proses ini memungkinkan pemantauan berkelanjutan dan pemrosesan data yang cepat, sekaligus meminimalkan keterlibatan manual,

mengurangi risiko kesalahan manusia, dan memastikan bahwa data digital yang diperoleh dapat mendukung pengambilan keputusan berbasis bukti secara efektif.

d. Kelebihan Metode Berbasis Teknologi

Metode berbasis teknologi memiliki kelebihan signifikan karena memungkinkan pengumpulan data dalam volume besar dengan kecepatan tinggi melalui otomatisasi, sensor, aplikasi, dan sistem digital, sehingga proses penelitian menjadi lebih efisien dibanding metode manual, serta data yang diperoleh dapat langsung diintegrasikan untuk analisis *real-time* dan pemodelan prediktif (Hashem *et al.*, 2018). Kelebihan ini juga mencakup kemampuan untuk meminimalkan kesalahan manusia, meningkatkan konsistensi data, serta menyimpan informasi dengan aman di *cloud* atau *database* terpusat yang mendukung akses lintas platform bagi tim peneliti atau analis.

Metode berbasis teknologi memungkinkan monitoring dan pengumpulan data secara berkelanjutan, termasuk pemantauan *real-time* terhadap aktivitas, kondisi lingkungan, atau interaksi pengguna, sehingga penelitian dapat memperoleh wawasan yang lebih akurat dan relevan dengan konteks, serta mendukung pengambilan keputusan berbasis data yang cepat dan berbasis bukti. Kelebihan ini membuat metode berbasis teknologi sangat cocok digunakan dalam penelitian modern yang membutuhkan data besar (*big data*), analitik lanjutan, serta integrasi dengan *machine learning* atau sistem otomatis lainnya.

e. Keterbatasan dan Tantangan

Metode berbasis teknologi memiliki keterbatasan dan tantangan, terutama terkait biaya awal yang tinggi untuk perangkat keras, perangkat lunak, dan infrastruktur digital, serta kebutuhan sumber daya manusia yang memiliki kompetensi teknis untuk mengoperasikan dan memelihara sistem, sehingga implementasinya tidak selalu mudah bagi organisasi atau peneliti dengan sumber daya terbatas (Chen *et al.*, 2019). Selain itu, risiko gangguan teknis atau kerusakan sistem dapat menyebabkan hilangnya data atau interupsi pengumpulan, sehingga memerlukan mekanisme backup dan pemeliharaan yang ketat untuk menjaga kontinuitas pengumpulan data.

Tantangan lain yang muncul adalah masalah keamanan dan privasi data, terutama ketika data yang dikumpulkan bersifat sensitif, sehingga memerlukan protokol enkripsi, kontrol akses, dan kepatuhan terhadap regulasi perlindungan data, sementara sebagian data yang diperoleh dari sumber eksternal dapat memiliki kualitas yang bervariasi dan memerlukan validasi sebelum dianalisis. Selain itu, ketergantungan pada teknologi dapat membatasi pengumpulan informasi yang memerlukan konteks sosial atau nuansa manusia, sehingga kadang perlu dikombinasikan dengan metode manual untuk memperoleh data yang lebih komprehensif dan mendalam.

#### **D. Kualitas Data: Validitas, Reliabilitas, dan Bias**

Kualitas data adalah faktor utama yang menentukan keandalan hasil analisis dalam sains data. Data yang berkualitas tinggi memastikan informasi yang diperoleh akurat, konsisten, dan dapat dipertanggungjawabkan. Tiga aspek penting yang menjadi fokus penilaian kualitas data adalah validitas, reliabilitas, dan bias, karena ketiganya saling berkaitan dan memengaruhi kesimpulan yang ditarik dari data.

##### **1. Validitas**

Validitas adalah ukuran sejauh mana data benar-benar merepresentasikan konsep atau fenomena yang ingin diukur. Data yang valid berarti hasil pengukuran akurat, tepat, dan relevan dengan tujuan penelitian atau analisis yang dilakukan. Tanpa validitas yang tinggi, kesimpulan yang ditarik dari data bisa menyesatkan atau tidak sesuai dengan kenyataan. Validitas merupakan aspek fundamental dalam kualitas data karena menentukan apakah data dapat dipercaya untuk pengambilan keputusan, perencanaan strategi, atau penelitian ilmiah.

###### **a. Jenis-jenis Validitas**

Validitas merupakan ukuran penting dalam penelitian untuk memastikan bahwa instrumen atau data benar-benar mengukur apa yang dimaksudkan. Jenis-jenis validitas mencakup validitas konten, validitas kriteria, dan validitas konstruk, yang masing-masing memiliki peran dalam menilai kesesuaian data dengan konsep yang ingin diukur. Validitas konten memastikan bahwa

seluruh aspek konsep tercakup oleh indikator yang digunakan, validitas kriteria menilai hubungan antara hasil pengukuran dengan standar eksternal, sedangkan validitas konstruk menekankan kesesuaian data dengan teori atau konsep abstrak yang mendasari penelitian (Heale & Twycross, 2018).

Penerapan validitas yang tepat membutuhkan perancangan instrumen yang cermat, pengujian awal melalui pilot study, serta evaluasi oleh ahli yang memahami konteks konsep yang diukur. Kesalahan dalam menentukan validitas dapat menyebabkan interpretasi data yang keliru dan keputusan yang tidak akurat dalam penelitian maupun analisis data. Oleh karena itu, validitas menjadi fondasi yang menentukan kualitas dan keandalan seluruh proses penelitian serta analisis data yang dilakukan.

b. Faktor yang Mempengaruhi Validitas

Faktor yang mempengaruhi validitas data sangat penting untuk diperhatikan agar hasil penelitian atau analisis sains data dapat dipercaya. Salah satu faktor utama adalah kejelasan definisi konsep yang akan diukur, karena ketidakjelasan dapat menyebabkan data tidak sesuai dengan tujuan penelitian, sehingga interpretasi hasil menjadi keliru, selain itu, desain instrumen pengukuran yang tepat juga berperan besar, karena pertanyaan atau alat ukur yang ambigu dapat menimbulkan kesalahan sistematis dalam pengumpulan data, dan penggunaan data proxy yang tidak representatif dapat semakin menurunkan akurasi pengukuran (Heale & Twycross, 2018).

Faktor prosedur pengumpulan data juga memengaruhi validitas, karena perbedaan pelaksanaan atau kondisi lingkungan saat pengumpulan dapat menimbulkan variasi hasil yang tidak diinginkan, kesalahan manusia dalam pencatatan atau *input* data dapat menambah distorsi, dan pemrosesan data yang tidak cermat, seperti transformasi atau penghapusan data yang tidak tepat, dapat mengubah makna asli data sehingga menghasilkan informasi yang tidak valid. Upaya mitigasi terhadap faktor-faktor ini termasuk standarisasi instrumen, pelatihan pengumpul data, validasi awal melalui pilot study, dan peninjauan oleh ahli untuk memastikan kesesuaian dengan konsep yang dimaksud.

c. Cara Meningkatkan Validitas

Meningkatkan validitas data merupakan langkah penting agar hasil penelitian atau analisis sains data akurat dan dapat diandalkan. Salah satu cara utama adalah dengan menyusun instrumen pengukuran yang jelas dan sesuai dengan konsep yang ingin diukur, termasuk membuat definisi variabel yang spesifik serta indikator yang komprehensif, kemudian melakukan pilot testing untuk mengevaluasi apakah instrumen dapat menangkap data secara tepat, dan melibatkan pakar atau reviewer untuk menilai kesesuaian instrumen terhadap konsep yang diukur sehingga potensi kesalahan dapat diminimalkan.

Penerapan triangulasi data juga dapat meningkatkan validitas, yaitu dengan membandingkan dan memverifikasi data dari berbagai sumber atau metode untuk memastikan konsistensi hasil, penerapan kontrol kualitas dalam pengumpulan data seperti pengecekan *input* dan range check dapat mengurangi kesalahan sistematis, serta dokumentasi yang baik dalam bentuk data dictionary atau catatan metodologi membantu menjaga kejelasan dan konsistensi interpretasi data sepanjang proses penelitian. Dengan menerapkan langkah-langkah ini, validitas data meningkat dan hasil analisis dapat memberikan *insight* yang lebih akurat dan dapat dipercaya.

## 2. Reliabilitas

Reliabilitas adalah ukuran konsistensi dan stabilitas data ketika pengukuran dilakukan berulang kali dalam kondisi yang sama. Data yang reliabel memberikan hasil yang serupa pada pengukuran yang berulang, sehingga analisis yang dilakukan menjadi lebih dapat dipercaya. Reliabilitas tidak menjamin keakuratan mutlak, tetapi memastikan bahwa data tidak terpengaruh oleh fluktuasi acak, kesalahan pengukuran sementara, atau variasi subjektif dari pengumpul data. Dalam sains data, reliabilitas sangat penting karena data yang konsisten memudahkan interpretasi tren, pola, dan hubungan antar variabel.

### a. Jenis-jenis Reliabilitas

Reliabilitas merupakan ukuran konsistensi data dalam penelitian yang menunjukkan sejauh mana hasil pengukuran dapat direproduksi dalam kondisi yang sama, sehingga data yang reliabel akan memberikan hasil yang stabil dan dapat dipercaya meskipun diukur berulang kali. Jenis-jenis reliabilitas mencakup

*test-retest reliability*, yang menilai konsistensi hasil pengukuran pada waktu yang berbeda dengan sampel yang sama, *inter-rater reliability* yang mengukur kesepakatan antar-pengamat dalam menilai fenomena yang sama, serta *internal consistency reliability* yang menilai konsistensi antar-item dalam suatu instrumen multi-indikator, dan penerapan semua jenis reliabilitas ini penting untuk memastikan bahwa instrumen dan metode pengumpulan data menghasilkan informasi yang stabil dan akurat.

Penerapan reliabilitas yang tepat membutuhkan prosedur pengumpulan data yang terstandarisasi, pelatihan pengumpul data agar mengurangi variasi subjektif, serta penggunaan instrumen yang telah diuji konsistensinya melalui pilot testing atau analisis statistik seperti Cronbach's alpha. Selain itu, automasi pengukuran dan audit kualitas data secara rutin dapat membantu meminimalkan kesalahan manusia dan fluktuasi acak, sehingga hasil penelitian tetap konsisten dan valid. Dengan memastikan berbagai jenis reliabilitas diterapkan dengan tepat, data yang diperoleh dapat mendukung analisis yang akurat dan keputusan berbasis bukti yang lebih kuat.

b. Faktor yang Mempengaruhi Reliabilitas

Reliabilitas data sangat dipengaruhi oleh kualitas instrumen dan konsistensi prosedur pengukuran, karena alat ukur yang tidak standar atau ambigu dapat menghasilkan variasi yang signifikan dalam hasil, selain itu perbedaan teknik pengumpulan data antar pengumpul atau perbedaan waktu pengukuran juga dapat menurunkan konsistensi, dan faktor eksternal seperti lingkungan, kondisi fisik, maupun psikologis responden turut memengaruhi stabilitas hasil pengukuran sehingga mengurangi reliabilitas (Kimberlin & Winterstein, 2018).

Kesalahan manusia dalam *input* atau pencatatan data menjadi sumber ketidakstabilan, sehingga pelatihan pengumpul data yang cermat dan penerapan prosedur standar sangat diperlukan untuk menjaga konsistensi, penggunaan teknologi atau automasi pengukuran dapat meminimalkan fluktuasi acak, dan audit kualitas data secara rutin membantu memastikan bahwa hasil pengukuran tetap stabil serta dapat diandalkan untuk analisis lebih lanjut. Dengan memperhatikan faktor-faktor ini,

reliabilitas data meningkat sehingga hasil penelitian menjadi lebih konsisten, valid, dan dapat dipertanggungjawabkan.

c. Cara Meningkatkan Reliabilitas

Meningkatkan reliabilitas data merupakan langkah penting untuk memastikan hasil penelitian konsisten dan dapat dipercaya, karena data yang stabil akan mendukung analisis yang akurat dan pengambilan keputusan berbasis bukti. Salah satu cara utama adalah dengan menggunakan instrumen pengukuran yang telah terstandarisasi dan diuji sebelumnya sehingga menghasilkan hasil yang konsisten, diikuti dengan pelatihan pengumpul data agar prosedur pengukuran dilakukan seragam, serta penerapan pilot testing atau uji coba awal untuk mengevaluasi konsistensi alat ukur sebelum pengumpulan data penuh (Polit & Beck, 2018).

Penggunaan automasi atau perangkat digital dalam pengumpulan data dapat mengurangi kesalahan manusia dan variasi acak yang tidak diinginkan, audit kualitas data secara rutin membantu memverifikasi konsistensi hasil, dan dokumentasi prosedur pengumpulan data yang jelas akan menjaga standarisasi sepanjang proses penelitian. Dengan menerapkan langkah-langkah ini secara sistematis, reliabilitas data dapat ditingkatkan, sehingga informasi yang diperoleh lebih akurat, stabil, dan dapat dipertanggungjawabkan dalam analisis maupun laporan penelitian.

### 3. Bias

Bias adalah distorsi sistematis yang menyebabkan data tidak merepresentasikan kondisi atau populasi yang sebenarnya. Data yang mengandung bias menghasilkan kesimpulan yang tidak akurat atau menyesatkan, karena informasi yang diperoleh cenderung condong ke arah tertentu. Bias berbeda dengan kesalahan acak; ia bersifat sistematis dan memengaruhi seluruh dataset atau sebagian besar pengukuran. Dalam sains data, identifikasi dan mitigasi bias sangat penting agar analisis, prediksi, dan keputusan yang diambil berbasis data tetap sah dan dapat dipercaya.

a. Jenis-jenis Bias

Bias dalam data adalah distorsi sistematis yang menyebabkan hasil penelitian atau analisis tidak

merepresentasikan kondisi atau populasi yang sebenarnya, sehingga informasi yang diperoleh dapat menyesatkan atau tidak akurat. Jenis-jenis bias yang umum ditemui meliputi selection bias, yang terjadi ketika sampel tidak representatif dari populasi target; measurement bias, yang timbul akibat instrumen atau metode pengukuran menghasilkan kesalahan sistematis; reporting bias, ketika responden memberikan jawaban yang dianggap lebih pantas atau diinginkan; survivorship bias, yang muncul ketika hanya kasus yang bertahan atau sukses yang dianalisis; dan omitted variable bias, yang terjadi jika variabel penting tidak dimasukkan dalam model analisis sehingga hubungan antar variabel lain tampak berbeda dari kenyataan (Suresh & Chandrashekhara, 2018).

Penyebab bias bisa berasal dari desain sampling yang buruk, instrumen pengukuran yang tidak tepat, data hilang yang tidak acak, atau praktik pembersihan data yang menghapus informasi penting. Mitigasi bias dapat dilakukan melalui penggunaan sampling probabilistik, penerapan instrumen yang netral, analisis pola missing data, penerapan post-stratification atau weighting, dan sensitivity analysis untuk menguji sejauh mana hasil berubah bila asumsi atau metode berbeda diterapkan. Dengan memahami dan mengurangi bias, kualitas data meningkat sehingga analisis dapat menghasilkan kesimpulan yang lebih akurat dan dapat dipercaya.

b. Sumber Bias

Sumber bias dalam data muncul dari berbagai aspek yang dapat memengaruhi representativitas dan keakuratan informasi, sehingga hasil penelitian berpotensi menyimpang dari kenyataan. Salah satu sumber utama adalah desain sampling yang tidak tepat, di mana sampel yang dipilih tidak mencerminkan populasi target, diikuti oleh instrumen pengukuran yang tidak sesuai atau pertanyaan yang memimpin responden sehingga menghasilkan jawaban yang sistematis berbeda dari kondisi sebenarnya, serta data hilang atau tidak lengkap yang tidak acak yang dapat menyebabkan distorsi interpretasi hasil (Suresh & Chandrashekhara, 2018).

Praktik pembersihan data yang tidak hati-hati, seperti penghapusan *outlier* tanpa analisis yang memadai, atau

transformasi data yang salah dapat menjadi sumber bias tambahan yang memengaruhi kualitas data. Faktor manusia, seperti kesalahan pencatatan atau interpretasi, juga turut berkontribusi terhadap bias, sehingga pelatihan pengumpul data, penggunaan instrumen yang terstandarisasi, dan prosedur audit kualitas menjadi langkah penting untuk meminimalkan dampak bias. Dengan mengenali dan mengelola sumber bias, data yang dikumpulkan menjadi lebih representatif, akurat, dan dapat diandalkan untuk analisis maupun pengambilan keputusan.

c. Cara Mengidentifikasi dan Mengurangi Bias

Mengidentifikasi dan mengurangi bias dalam data merupakan langkah penting agar hasil penelitian atau analisis dapat dipercaya dan akurat, karena bias dapat menyesatkan interpretasi dan keputusan yang diambil. Salah satu pendekatan utama adalah memeriksa representativitas sampel terhadap populasi target, mengevaluasi apakah metode pengumpulan data memungkinkan distorsi sistematis, serta menganalisis pola missing data untuk menentukan apakah data hilang secara acak atau terkait variabel tertentu, sehingga langkah-langkah mitigasi dapat diterapkan secara tepat (Grove *et al.*, 2019).

Penggunaan instrumen pengukuran yang netral dan standar sangat penting untuk mengurangi measurement bias, penerapan weighting atau post-stratification dapat menyesuaikan ketidakseimbangan dalam sampel, dan sensitivity analysis memungkinkan peneliti mengevaluasi sejauh mana hasil berubah bila asumsi atau metode berbeda diterapkan. Dengan menerapkan strategi-strategi ini secara sistematis, bias dapat diminimalkan sehingga data menjadi lebih representatif, konsisten, dan mendukung kesimpulan yang akurat serta dapat dipertanggungjawabkan.

## **E. Reprodusibilitas dan Dokumentasi Proses Pengumpulan Data**

Reprodusibilitas dan dokumentasi proses pengumpulan data merupakan aspek fundamental dalam sains data yang menentukan tingkat keandalan, transparansi, dan akuntabilitas data yang dihasilkan. Kedua aspek ini memastikan bahwa proses pengumpulan data dapat

dipahami, ditelusuri, dan diulang secara sistematis sehingga data yang digunakan dalam analisis memiliki dasar metodologis yang kuat dan dapat dipertanggungjawabkan.

## 1. Reprodusibilitas Proses Pengumpulan Data

Reprodusibilitas dalam proses pengumpulan data merujuk pada kemampuan suatu prosedur pengumpulan data untuk diulang kembali oleh pihak lain dengan langkah, metode, dan kondisi yang sama sehingga menghasilkan data yang konsisten dan dapat dipertanggungjawabkan. Dalam konteks sains data, reprodusibilitas menjadi indikator penting kualitas metodologi karena memastikan bahwa data tidak bersifat kebetulan, bias, atau bergantung pada individu tertentu.

### a. Tujuan dan Peran Reprodusibilitas

Reprodusibilitas memiliki tujuan utama untuk memastikan bahwa proses pengumpulan data dapat diulang dengan prosedur, kondisi, dan parameter yang sama sehingga menghasilkan data yang konsisten dan dapat diverifikasi secara metodologis oleh pihak lain. Dalam konteks sains data, reprodusibilitas berperan penting dalam meningkatkan kepercayaan terhadap data dan hasil analisis, karena data yang diperoleh melalui proses yang dapat direproduksi memungkinkan pengujian ulang, penelusuran sumber kesalahan, serta evaluasi objektif terhadap kualitas metode yang digunakan (*National Academies of Sciences, Engineering, and Medicine*, 2019).

Peran reprodusibilitas juga berkaitan erat dengan transparansi dan akuntabilitas dalam pengelolaan data, karena setiap tahapan pengumpulan data harus dapat dijelaskan dan dipertanggungjawabkan secara sistematis. Dengan adanya reprodusibilitas, data tidak hanya berfungsi sebagai bahan analisis sesaat, tetapi juga sebagai aset ilmiah yang dapat digunakan kembali, dikembangkan lebih lanjut, serta menjadi dasar pengambilan keputusan yang rasional dan berbasis bukti dalam jangka panjang.

### b. Unsur-Unsur Utama Reprodusibilitas

Reprodusibilitas dalam pengumpulan data mencakup sejumlah unsur utama yang memungkinkan proses tersebut dijalankan kembali dengan hasil yang konsisten dan dapat diverifikasi secara independen, seperti ketersediaan data dan

materi, transparansi metode, serta dokumentasi lengkap dari semua langkah yang dilakukan. Menurut kerangka standar *reproducibility*, unsur-unsur utama ini mencakup data *availability* (ketersediaan data), *methods transparency* (transparansi metode), *materials accessibility* (aksesibilitas materi/alat), dan *analysis reproducibility* (kemampuan memutar ulang analisis) yang semuanya saling terkait untuk menjamin bahwa pihak lain dapat mengikuti alur kerja yang sama dan menghasilkan keluaran yang setara tanpa ambiguitas (CASRAI, 2025).

Setiap unsur tersebut saling memperkuat: data yang tersedia dan terdokumentasi dengan jelas membuat peneliti lain dapat mengakses sumber *input* yang sama, transparansi metode memastikan semua langkah pengumpulan dicatat secara detail, akses terhadap materi atau konfigurasi alat memungkinkan prosedur dijalankan kembali, dan *reproducibility* analisis menjamin bahwa pemrosesan data serta analisis statistik atau komputasi dapat diulang dengan hasil yang konsisten. Penerapan semua unsur ini secara konsisten dapat meningkatkan kredibilitas, keterlacakan, dan keandalan data serta hasil analisis yang dihasilkan dari proses pengumpulan data yang direproduksi.

c. Tantangan dalam Mewujudkan Reprodusibilitas

Reprodusibilitas sering kali menghadapi berbagai tantangan signifikan yang menghambat kemampuan peneliti lain untuk mengulangi proses pengumpulan data dan menghasilkan hasil yang konsisten karena faktor-faktor seperti keterbatasan akses terhadap data asli, kurangnya transparansi dokumentasi metode, dan kompleksitas pengaturan lingkungan komputasi yang digunakan dalam penelitian. Menurut studi tentang *reproducibility* di bidang komputasi dan ilmu data, hambatan utama termasuk *restricted access to underlying data and code*, *incomplete documentation*, and *methodological opacity* yang secara langsung mengurangi kemungkinan pihak lain mengulangi eksperimen atau pengumpulan data dengan akurat (AL Beam *et al.*, 2020).

Tantangan dalam mewujudkan reprodusibilitas juga muncul dari berbagai aspek praktis dan struktural seperti penggunaan perangkat lunak dan alat yang tidak umum, kebijakan privasi

yang membatasi berbagi data, serta beban dokumentasi yang tinggi yang sering kali memerlukan waktu dan sumber daya lebih banyak dari yang tersedia bagi peneliti. Kompleksitas ini memperlihatkan bahwa mencapai *reproducibility* bukan hanya masalah teknis, tetapi juga berkaitan dengan budaya penelitian, kebijakan akses data, dan standar pelaporan yang diterapkan di komunitas ilmiah global.

d. Implikasi Reprodusibilitas terhadap Kualitas Data

Reprodusibilitas memiliki implikasi yang signifikan terhadap kualitas data, karena ketika proses pengumpulan dan pengolahan data dilakukan dengan cara yang dapat direproduksi oleh peneliti lain, tingkat kepercayaan, konsistensi, dan validitas dataset meningkat secara langsung. Hal ini memperkuat bahwa data tidak hanya lengkap dan akurat pada saat dikumpulkan, tetapi juga dapat dipertanggungjawabkan saat di-uji ulang oleh pihak lain, sehingga memperkecil risiko bias, kesalahan sistematis, atau data artefak yang merusak hasil penelitian (Condon, Simpson & Emanuel, 2022).

## 2. Dokumentasi Proses Pengumpulan Data

Dokumentasi proses pengumpulan data adalah kegiatan pencatatan dan penyimpanan informasi secara sistematis mengenai seluruh tahapan, metode, dan keputusan yang diambil selama proses pengumpulan data. Dokumentasi ini berfungsi sebagai sumber informasi resmi yang menjelaskan bagaimana, dari mana, dan dalam kondisi apa data diperoleh, sehingga data dapat dipahami, dievaluasi, dan digunakan kembali secara tepat.

a. Tujuan dan Fungsi Dokumentasi

Dokumentasi proses pengumpulan data memiliki tujuan dan fungsi yang sangat penting dalam memastikan bahwa data yang dihasilkan tidak hanya tersimpan secara sistematis, tetapi juga dapat dipahami, diakses, serta digunakan kembali oleh peneliti lain untuk kebutuhan verifikasi atau analisis lanjutan. Sebagaimana dijelaskan dalam studi ilmiah tentang peran profesi informasi dan dokumentasi dalam manajemen data penelitian, dokumentasi tidak sekadar mencatat data, tetapi juga menyediakan konteks deskriptif dan struktur yang jelas sehingga data dapat ditemukan, dipahami, dan digunakan dengan tepat

oleh komunitas ilmiah lainnya, terutama ketika dataset menjadi semakin besar dan kompleks seiring perkembangan *big data* (Ávila & Sulé, 2025).

Dokumentasi dalam pengumpulan data berfungsi sebagai fondasi transparansi dan keterlacakan ilmu pengetahuan, karena informasi mengenai asal, metode, struktur, dan aturan penggunaan data direkam secara menyeluruh sehingga meminimalkan risiko kesalahpahaman atau interpretasi yang keliru di kemudian hari. Selain itu, dokumentasi berperan sebagai sarana komunikasi ilmiah antarpeneliti dan sebagai arsip digital yang mendukung keberlanjutan penelitian di masa mendatang, sehingga data tidak hanya tersimpan tetapi juga memiliki nilai guna yang tinggi dalam jangka panjang.

b. Komponen Utama dalam Dokumentasi Pengumpulan Data

Dokumentasi pengumpulan data mencakup beragam komponen utama yang harus dicatat secara sistematis sehingga data dapat dipahami, digunakan kembali, dan dievaluasi oleh peneliti lain secara efektif; komponen tersebut antara lain meliputi informasi kontekstual tentang asal data, metodologi pengumpulan, struktur dan format data, definisi variabel, serta prosedur pemrosesan yang diterapkan, karena tanpa informasi-informasi ini dataset akan sulit dipahami dan digunakan secara benar oleh pihak lain. Dalam penjelasan akademis tentang dokumentasi dataset, aspek-aspek seperti tujuan pembuatan dataset, komposisi data, proses pengumpulan, langkah-langkah pemrosesan, penggunaan yang dimaksudkan, dan pemeliharaan (termasuk versi dan pembaruan) dikemukakan sebagai komponen dokumentasi yang esensial untuk menjamin transparansi, keterbacaan, dan relevansi data bagi berbagai pemangku kepentingan yang mengakses atau menerapkan dataset tersebut (Pushkarna, Zaldivar & Kjartansson, 2022).

Dokumentasi yang lengkap tidak hanya mencantumkan data mentah saja, tetapi juga catatan terperinci mengenai setiap tahap dalam siklus hidup data dari bagaimana data dikumpulkan hingga bagaimana interpretasi dan transformasi dilakukan sehingga pengguna lain dapat memverifikasi dan mereplikasi penggunaan dataset dengan benar tanpa kehilangan konteks penting yang memengaruhi kualitas dan relevansinya.

c. Peran Dokumentasi terhadap Kualitas Data

Dokumentasi memiliki peran yang sangat krusial terhadap kualitas data karena melalui pencatatan yang sistematis, jelas, dan terstruktur, setiap tahapan pengumpulan data dapat ditelusuri kembali sehingga sumber kesalahan, inkonsistensi, atau bias dapat diidentifikasi secara lebih akurat. Dokumentasi yang baik memastikan bahwa informasi mengenai asal data, metode pengumpulan, definisi variabel, serta kondisi teknis dan kontekstual saat data dikumpulkan tersedia secara lengkap, sehingga kualitas data dapat dinilai dan diverifikasi secara objektif oleh pihak lain.

Dokumentasi yang memadai berkontribusi langsung pada peningkatan konsistensi, keandalan, dan validitas data, karena data tidak berdiri sendiri tanpa konteks, melainkan disertai penjelasan yang memandu interpretasi dan penggunaannya. Dengan adanya dokumentasi yang jelas, data menjadi lebih mudah digunakan kembali, diaudit, dan dikombinasikan dengan sumber data lain, sehingga kualitas data tetap terjaga meskipun digunakan dalam analisis lanjutan atau pada konteks penelitian yang berbeda.

d. Implikasi Dokumentasi dalam Sains Data

Dokumentasi dalam sains data memiliki implikasi yang sangat luas karena menentukan sejauh mana data dapat dipahami, digunakan kembali, dan dikembangkan secara berkelanjutan dalam berbagai konteks analisis dan pengambilan keputusan. Dokumentasi yang lengkap dan terstruktur memungkinkan proses sains data berjalan secara transparan dan akuntabel, karena seluruh tahapan mulai dari pengumpulan, pemrosesan, hingga pemanfaatan data dapat ditelusuri dan dievaluasi secara sistematis oleh peneliti atau praktisi lain.

Implikasi dokumentasi juga terlihat pada peningkatan kolaborasi dan efisiensi kerja dalam proyek sains data, karena data yang terdokumentasi dengan baik dapat dipahami dengan cepat tanpa ketergantungan pada individu tertentu. Dokumentasi yang kuat menjadikan data sebagai aset jangka panjang yang bernilai tinggi, mendukung reproduktibilitas, menjaga kualitas analisis, serta memperkuat kepercayaan terhadap hasil sains data dalam konteks ilmiah maupun praktis.

# BAB III

## DATA PRIMER

---

Data primer dalam metode pengumpulan data untuk sains data merupakan data yang diperoleh secara langsung dari sumber pertama melalui proses terencana seperti observasi, survei, eksperimen, atau pengukuran sistem digital, sehingga data yang dihasilkan bersifat spesifik sesuai tujuan analisis. Keunggulan data primer terletak pada tingkat relevansi, akurasi, dan kendali kualitas yang lebih tinggi, karena proses pengumpulan dapat disesuaikan dengan kebutuhan model, konteks permasalahan, serta karakteristik variabel yang dianalisis dalam pengembangan solusi berbasis sains data.

### A. Pengertian dan Karakteristik Data Primer

Pada kegiatan penelitian dan analisis berbasis data, pemahaman mengenai jenis data yang digunakan menjadi aspek fundamental dalam menentukan kualitas dan ketepatan hasil analisis. Salah satu jenis data yang memiliki peran penting dalam proses tersebut adalah data primer, karena data ini diperoleh langsung dari sumber utama dan dirancang secara khusus untuk menjawab tujuan pengumpulan data.

#### 1. Pengertian Data Primer

Data primer merupakan data yang diperoleh secara langsung oleh peneliti dari sumber utama melalui proses pengumpulan yang dirancang secara khusus untuk menjawab tujuan penelitian atau analisis tertentu. Data ini dikumpulkan secara langsung dari objek, individu, atau sistem yang menjadi fokus kajian, sehingga informasi yang diperoleh masih bersifat asli dan belum mengalami pengolahan atau interpretasi dari pihak lain. Menurut Sugiyono (2019), data primer adalah sumber data yang langsung memberikan data kepada pengumpul data, yang diperoleh melalui teknik pengumpulan seperti observasi, wawancara, dan kuesioner yang disesuaikan dengan kebutuhan penelitian.

Pada konteks penelitian ilmiah dan analisis data modern, data primer memiliki peran strategis karena mampu merepresentasikan kondisi aktual sesuai dengan ruang lingkup dan waktu penelitian yang ditetapkan. Penggunaan data primer memungkinkan peneliti untuk merancang variabel dan indikator pengukuran secara tepat sehingga data yang diperoleh selaras dengan permasalahan yang dianalisis. Oleh karena itu, data primer sering digunakan sebagai dasar utama dalam pengambilan keputusan, pengujian hipotesis, serta pengembangan model analisis karena tingkat relevansinya yang tinggi terhadap tujuan penelitian.

## **2. Karakteristik Data Primer**

### **a. Dikumpulkan Langsung dari Sumber Asli**

Karakteristik data primer yang dikumpulkan langsung dari sumber asli menunjukkan bahwa data diperoleh secara langsung dari responden, objek, atau fenomena yang sedang diteliti tanpa melalui proses perantara atau pengolahan oleh pihak lain. Proses ini memungkinkan peneliti memperoleh informasi yang masih mentah dan autentik, sehingga data benar-benar mencerminkan kondisi nyata sesuai dengan konteks waktu dan situasi penelitian berlangsung. Menurut Sekaran dan Bougie (2020), data primer adalah data yang dikumpulkan secara langsung oleh peneliti untuk tujuan penelitian tertentu, sehingga relevansinya sangat tinggi terhadap permasalahan yang sedang dikaji.

Pengumpulan data langsung dari sumber asli memberikan keleluasaan bagi peneliti untuk menentukan instrumen, teknik, dan variabel pengukuran yang paling sesuai dengan tujuan analisis. Kondisi ini penting karena data yang tidak melalui pihak ketiga cenderung memiliki tingkat distorsi informasi yang lebih rendah dan mampu menggambarkan fenomena secara lebih objektif. Dalam konteks penelitian dan sains data, karakteristik ini menjadikan data primer sebagai dasar yang kuat untuk analisis mendalam, pengambilan keputusan, serta pengembangan model yang membutuhkan representasi kondisi aktual secara akurat.

### **b. Spesifik terhadap Tujuan Analisis**

Karakteristik data primer yang bersifat spesifik terhadap tujuan analisis menunjukkan bahwa data dikumpulkan berdasarkan kebutuhan penelitian yang telah dirumuskan secara

jelas sejak awal, sehingga variabel, indikator, dan teknik pengumpulan disesuaikan secara langsung dengan permasalahan yang ingin dianalisis. Pendekatan ini memungkinkan peneliti memperoleh data yang relevan dan terfokus, karena proses pengumpulan tidak bersifat umum, melainkan diarahkan untuk menjawab pertanyaan penelitian atau mendukung tujuan analisis tertentu secara tepat dan mendalam.

Pada konteks metodologi penelitian, spesifisitas data primer memberikan keunggulan karena data dirancang untuk memenuhi kebutuhan analisis yang unik dan kontekstual, bukan sekadar memanfaatkan data yang sudah tersedia. Creswell (2018) menjelaskan bahwa data primer dikumpulkan secara sengaja untuk menjawab tujuan penelitian tertentu, sehingga peneliti memiliki kontrol penuh terhadap jenis informasi yang dikumpulkan dan tingkat kesesuaiannya dengan fokus analisis yang dilakukan.

c. Tingkat Akurasi dan Validitas Lebih Tinggi

Karakteristik data primer yang memiliki tingkat akurasi dan validitas lebih tinggi muncul karena data dikumpulkan langsung dari sumber asli dengan metode yang dirancang secara sistematis sesuai tujuan penelitian, sehingga risiko kesalahan, bias, atau distorsi informasi dapat diminimalkan. Pendekatan ini memungkinkan peneliti memastikan bahwa data yang diperoleh benar-benar merepresentasikan fenomena yang diamati, serta dapat digunakan untuk analisis yang membutuhkan presisi tinggi dalam pengambilan keputusan atau pengembangan model.

Kontrol penuh terhadap proses pengumpulan data memungkinkan peneliti untuk menyesuaikan instrumen dan prosedur agar relevan dengan variabel yang sedang dianalisis, sehingga validitas internal dan eksternal data meningkat. Menurut Saunders *et al.* (2019), data primer memiliki keunggulan utama dalam hal keakuratan dan relevansi karena dikumpulkan langsung oleh peneliti untuk tujuan penelitian tertentu, sehingga informasi yang diperoleh lebih dapat dipercaya dibandingkan data sekunder.

d. Fleksibel dalam Desain Pengumpulan

Karakteristik data primer yang fleksibel dalam desain pengumpulan menunjukkan bahwa peneliti memiliki kebebasan

untuk menyesuaikan metode, instrumen, dan variabel pengukuran sesuai dengan kebutuhan penelitian, sehingga data yang diperoleh benar-benar relevan dengan tujuan analisis. Fleksibilitas ini memungkinkan penyesuaian terhadap konteks, kondisi lapangan, atau dinamika responden, sehingga proses pengumpulan data dapat dilakukan secara adaptif tanpa mengorbankan kualitas dan keakuratan informasi yang dikumpulkan.

Pendekatan fleksibel dalam desain pengumpulan data primer sangat penting dalam penelitian modern, karena memudahkan peneliti untuk menanggapi perubahan atau kendala yang muncul selama proses pengumpulan, termasuk menyesuaikan skala, format kuesioner, atau metode observasi. Menurut Bryman dan Bell (2019), data primer menawarkan fleksibilitas tinggi dalam desain pengumpulan karena peneliti dapat mengubah dan menyesuaikan instrumen sesuai dengan kebutuhan analisis, sehingga data yang diperoleh tetap valid dan sesuai dengan tujuan penelitian.

## **B. Kelebihan dan Keterbatasan Data Primer**

Data primer memiliki peran penting dalam sains data karena memberikan informasi yang langsung diperoleh dari sumber asli. Keunggulan dan keterbatasan data primer perlu dipahami agar penggunaan data ini lebih efektif dalam penelitian dan pengambilan keputusan.

### **1. Kelebihan Data Primer**

Secara umum, data primer menawarkan beberapa kelebihan yang membuatnya sangat bernilai dalam penelitian:

#### **a. Keakuratan Tinggi**

Data primer memiliki keunggulan utama berupa keakuratan tinggi karena informasi dikumpulkan langsung dari sumber asli, sehingga meminimalkan risiko distorsi atau kesalahan interpretasi yang umum terjadi pada data sekunder. Proses pengumpulan yang dikontrol secara langsung memungkinkan peneliti menyesuaikan instrumen pengukuran dan metode pengumpulan sesuai dengan kebutuhan spesifik penelitian,

sehingga hasil analisis lebih dapat dipercaya dan relevan. Selain itu, kualitas data primer yang tinggi mendukung pengambilan keputusan berbasis bukti yang lebih tepat, karena informasi yang diperoleh merefleksikan kondisi nyata dari objek atau responden yang diteliti (Saunders *et al.*, 2019).

Keakuratan ini juga menjadi faktor penting dalam konteks sains data, di mana ketelitian informasi sangat mempengaruhi hasil model prediktif atau analisis statistik. Dengan memanfaatkan data primer, peneliti dapat memperoleh data yang tidak hanya lengkap tetapi juga memiliki validitas internal yang tinggi, sehingga kesalahan prediksi atau bias dapat diminimalkan. Hal ini membuat data primer menjadi pilihan utama ketika penelitian membutuhkan *insight* yang spesifik, terpercaya, dan relevan dengan konteks yang sedang dianalisis.

b. Relevansi dan Spesifik

Data primer unggul dalam hal relevansi dan spesifik karena data yang dikumpulkan memang dirancang khusus untuk menjawab pertanyaan penelitian atau tujuan penelitian tertentu, sehingga semua informasi yang diperoleh benar-benar sesuai dengan kebutuhan studi tanpa adanya potongan data yang tidak diperlukan atau tidak terkait. Dengan pengumpulan yang disesuaikan secara langsung oleh peneliti, variabel, indikator, dan instrumen pengukuran dapat ditetapkan sedemikian rupa sehingga mencerminkan fokus penelitian secara akurat dan mendalam, yang tidak mungkin diperoleh dari data yang dikumpulkan untuk tujuan lain. Hal ini didukung oleh penelitian yang menjelaskan bahwa data primer dihasilkan dan dikumpulkan secara langsung oleh peneliti untuk kebutuhan studi tertentu, sehingga informasi yang diperoleh lebih relevan dan sesuai dengan tujuan penelitian dibandingkan data lain yang sudah ada sebelumnya (U Sulung, 2024).

Kualitas relevansi dan spesifikasi ini menjadikan data primer sangat penting dalam konteks sains data, di mana model analitik dan pengambilan keputusan bergantung pada kesesuaian dan kedalaman informasi yang dianalisis. Ketika data yang masuk benar-benar mencerminkan fenomena atau perilaku yang diteliti, proses pemodelan, validasi, dan interpretasi hasil menjadi lebih kuat dan dapat dipertanggungjawabkan secara ilmiah. Tanpa

relevansi dan spesifik seperti ini, hasil riset bisa menjadi bias atau kurang menggambarkan realitas yang sebenarnya, sehingga keputusan yang diambil berdasarkan data tersebut kurang efektif atau bahkan menyesatkan.

c. Kontrol Penuh atas Proses Pengumpulan

Data primer memberikan keunggulan berupa kontrol penuh atas proses pengumpulan, karena peneliti memiliki kendali langsung terhadap metode, instrumen, dan teknik pengambilan data yang digunakan. Hal ini memungkinkan penyesuaian setiap aspek pengumpulan data agar sesuai dengan tujuan penelitian dan karakteristik responden atau fenomena yang diamati, sehingga risiko kesalahan pengukuran atau bias dapat diminimalkan. Peneliti juga dapat memonitor kualitas data secara *real-time*, memastikan bahwa setiap informasi yang dikumpulkan memenuhi standar validitas dan reliabilitas yang diinginkan (Creswell & Creswell, 2018).

Keuntungan kontrol ini sangat penting dalam konteks sains data, di mana kualitas *input* data sangat menentukan keakuratan analisis, model prediktif, dan hasil keputusan yang diambil. Dengan kontrol penuh, peneliti dapat menyesuaikan strategi pengumpulan jika ditemukan kendala di lapangan, sehingga hasil penelitian tetap sesuai rencana dan tetap relevan. Selain itu, pengawasan yang langsung memungkinkan identifikasi dini terhadap inkonsistensi atau anomali data, sehingga langkah korektif dapat segera dilakukan sebelum data digunakan dalam analisis lebih lanjut.

d. Ketersediaan Informasi Unik

Data primer memberikan keunggulan berupa ketersediaan informasi yang unik karena informasi tersebut dikumpulkan secara langsung oleh peneliti sesuai kebutuhan studi, sehingga dapat menangkap aspek-aspek fenomena yang belum pernah diobservasi sebelumnya atau tidak tersedia dalam sumber lain. Dengan mengumpulkan data dari responden, eksperimen, atau observasi yang dirancang khusus untuk menjawab pertanyaan penelitian, peneliti dapat memperoleh informasi baru yang relevan dengan konteks riset serta tidak terpengaruh oleh interpretasi atau pemrosesan pihak lain. Oleh karena itu, data primer sering kali menawarkan wawasan asli dan mendalam yang

dapat memperluas pengetahuan ilmiah atau memberikan kontribusi baru terhadap bidang studi yang diteliti, karena data primer merupakan data asli dan unik yang dikumpulkan langsung oleh peneliti dari sumbernya untuk memenuhi kebutuhan riset tertentu (Ajayi, 2023).

## **2. Keterbatasan Data Primer**

Meskipun memiliki banyak kelebihan, data primer juga memiliki keterbatasan yang perlu diperhatikan:

### **a. Biaya Tinggi**

Pengumpulan data primer sering kali menghadirkan keterbatasan berupa biaya yang tinggi karena peneliti membutuhkan sumber daya finansial yang signifikan untuk merancang instrumen penelitian, membayar tenaga pengumpul data, dan mencapai responden di lokasi yang beragam. Biaya ini mencakup berbagai komponen seperti upah interviewer, transportasi, penyediaan alat ukur, hingga kompensasi untuk partisipan, sehingga dapat membebani anggaran penelitian terutama ketika dilakukan secara luas atau melibatkan metode yang kompleks. Dengan demikian, keterbatasan biaya menjadi salah satu hambatan utama dalam penggunaan data primer, karena pengumpulan data primer membutuhkan investasi sumber daya yang besar yang mencakup waktu, tenaga, dan modal sehingga sering kali menjadi lebih mahal dibandingkan dengan penggunaan data sekunder yang sudah tersedia (Kumar, 2024).

Biaya tinggi dalam pengumpulan data primer juga dapat membatasi skala penelitian yang dapat dilakukan, karena anggaran yang terbatas sering kali memaksa peneliti untuk mengecilkan sampel atau area studi. Hal ini pada gilirannya berpotensi memengaruhi kemampuan penelitian untuk menghasilkan temuan yang representatif atau generalisasi yang luas, sehingga meskipun data primer sangat bermanfaat, keterbatasan biaya harus dipertimbangkan secara matang dalam perencanaan riset.

### **b. Waktu Pengumpulan Lama**

Data primer sering kali menghadirkan keterbatasan berupa waktu pengumpulan yang lama karena proses pengambilan data yang harus dilakukan secara langsung dari responden atau

fenomena tertentu memerlukan tahapan yang panjang mulai dari perancangan instrumen, penjadwalan wawancara atau survei, hingga pengumpulan dan verifikasi setiap jawaban. Hal ini berbeda dengan data sekunder yang sudah tersedia dan dapat diakses dengan cepat, sehingga penggunaan data primer akan memperpanjang jadwal penelitian secara keseluruhan, terutama ketika skala sampel besar atau responden tersebar di area geografis luas. Karena itu, penelitian yang mengandalkan data primer sering kali menghadapi keterlambatan signifikan dalam memperoleh hasil karena waktu yang dibutuhkan untuk mengumpulkan dan memproses data asli sering jauh lebih panjang dibandingkan dengan metode lain sehingga mempengaruhi kecepatan analisis dan pengambilan keputusan berbasis data (Qualtrics, 2023).

Waktu pengumpulan yang lama ini juga dapat berdampak pada relevansi temuan, terutama dalam konteks sains data di mana tren atau kondisi lapangan berubah cepat; data yang berhasil dikumpulkan setelah berbulan-bulan bisa saja tidak lagi mencerminkan kondisi terkini. Selain itu, durasi panjang ini dapat menambah tekanan pada sumber daya manusia dan logistik peneliti, memperbesar risiko keterlambatan jadwal penelitian lainnya, sehingga penyusunan perencanaan dan manajemen waktu menjadi sangat krusial dalam riset dengan data primer.

c. Kesulitan Akses dan Partisipasi

Data primer memiliki keterbatasan berupa kesulitan akses dan partisipasi, karena peneliti sering kali harus menghadapi tantangan dalam menjangkau responden yang relevan atau lokasi yang sulit dijangkau secara geografis atau administratif. Kesulitan ini dapat mencakup keterbatasan izin akses, respons rendah dari partisipan, atau kendala logistik yang memperlambat proses pengumpulan data dan mengurangi cakupan sampel yang seharusnya representatif. Hal ini menjadikan data primer seringkali sulit diperoleh secara lengkap karena hambatan dalam mendapatkan akses dan tingkat partisipasi yang memadai dari responden yang relevan, sehingga mempengaruhi kualitas dan kelengkapan informasi yang dikumpulkan (Saunders *et al.*, 2019).

Rendahnya partisipasi atau keterbatasan akses dapat menimbulkan bias non-respons, di mana kelompok tertentu mungkin kurang terwakili dalam data yang dikumpulkan. Dampaknya, hasil analisis bisa kurang mencerminkan kondisi nyata atau populasi sasaran, sehingga strategi pengumpulan data, pendekatan komunikasi, dan motivasi partisipan perlu direncanakan secara matang untuk memaksimalkan respons dan kualitas data primer.

d. Risiko Kesalahan Pengumpulan

Data primer memiliki keterbatasan berupa risiko kesalahan pengumpulan, karena proses pengambilan data yang dilakukan langsung oleh peneliti atau enumerator rentan terhadap kesalahan manusia, bias pengamat, atau ketidakkonsistenan dalam mencatat jawaban responden. Kesalahan ini dapat muncul dari berbagai faktor, termasuk ketidaktepatan instrumen pengukuran, kurangnya pelatihan pengumpul data, atau kesalahan dalam proses *input* dan validasi data, sehingga memengaruhi kualitas dan reliabilitas informasi yang diperoleh. Oleh karena itu, meskipun data primer bersifat asli dan relevan, risiko kesalahan dalam pengumpulan data primer tetap ada dan memerlukan perhatian khusus dalam perencanaan, pelaksanaan, dan validasi data agar hasil penelitian tetap akurat dan dapat dipercaya (Flick, 2018).

Risiko kesalahan pengumpulan ini dapat berdampak signifikan pada analisis dan interpretasi data, karena kesalahan kecil yang tidak terdeteksi dapat menyebabkan bias, distorsi, atau kesimpulan yang menyesatkan. Untuk mengurangi risiko tersebut, peneliti harus menerapkan prosedur kontrol kualitas, pelatihan pengumpul data, dan mekanisme verifikasi silang agar data yang dikumpulkan tetap valid dan dapat diandalkan dalam konteks penelitian maupun pengambilan keputusan berbasis sains data.

### C. Peran Data Primer dalam Analisis Sains Data

Data primer memiliki peran yang sangat penting dalam analisis sains data karena menyediakan informasi yang asli, akurat, dan sesuai dengan konteks penelitian. Data ini menjadi fondasi untuk membangun

model, melakukan prediksi, dan mendukung pengambilan keputusan berbasis bukti. Beberapa peran utama data primer dalam sains data dapat dijabarkan sebagai berikut:

### **1. Menjamin Keakuratan dan Validitas Data**

Data primer berperan penting dalam menjamin keakuratan dan validitas data karena sumber ini dikumpulkan langsung oleh peneliti dengan desain dan instrumen yang secara khusus disesuaikan untuk menjawab pertanyaan riset, sehingga setiap informasi yang diperoleh benar-benar mencerminkan fenomena yang diteliti tanpa distorsi dari pihak lain. Keakuratan di sini mencakup sejauh mana data mencerminkan nilai sebenarnya dari variabel yang diukur, sementara validitas mengacu pada kemampuan data tersebut untuk mengukur konsep atau fenomena yang dimaksud secara tepat, sehingga hasil analisis dan model yang dibangun memiliki dasar yang kuat untuk interpretasi dan pengambilan keputusan. Dalam penelitian ilmiah, validitas dan reliabilitas data sangat penting untuk menghasilkan temuan penelitian yang dapat dipercaya dan akurat, dan metode pengumpulan serta pengelolaan data yang tepat akan sangat memengaruhi kualitas ini (Karnia, 2024).

Keunggulan ini membuat data primer menjadi sumber yang sangat penting dalam sains data karena setiap kesalahan atau bias dalam pengukuran dapat berdampak langsung pada *output* model, prediksi, atau *insight* yang dihasilkan. Dengan kontrol penuh terhadap cara data dikumpulkan, peneliti dapat menerapkan langkah-langkah validasi internal seperti uji coba instrumen, pemantauan lapangan, dan verifikasi data sehingga risiko kesalahan dan bias dapat diminimalkan sebelum data digunakan dalam analisis lebih lanjut. Akhirnya, data yang akurat dan valid tidak hanya meningkatkan kualitas hasil penelitian tetapi juga memperkuat kepercayaan terhadap rekomendasi atau keputusan yang diambil berdasarkan analisis data tersebut.

### **2. Menyediakan Informasi Relevan dan Spesifik**

Data primer memiliki keunggulan dalam menyediakan informasi yang relevan dan spesifik karena dikumpulkan dengan tujuan penelitian yang jelas, sehingga setiap variabel, indikator, dan instrumen pengukuran dapat disesuaikan untuk menangkap aspek-aspek yang paling penting dari fenomena yang sedang dianalisis. Informasi yang relevan ini memungkinkan peneliti atau analis sains data untuk fokus

pada data yang benar-benar mendukung hipotesis, model, atau keputusan, tanpa gangguan dari informasi yang tidak terkait. Dengan demikian, data primer menjadi sangat berharga karena penelitian yang menggunakan data primer dapat menghasilkan informasi yang secara langsung relevan dengan tujuan studi dan lebih spesifik dibandingkan data sekunder yang bersifat umum (Bryman, 2021).

Ketersediaan data yang relevan dan spesifik juga meningkatkan efektivitas proses analisis, karena model statistik atau algoritma *machine learning* dapat dibangun berdasarkan data yang sesuai konteks, mengurangi risiko bias atau kesalahan interpretasi. Selain itu, data yang spesifik memungkinkan pengambilan keputusan yang lebih tepat dan strategi intervensi yang lebih efektif, karena *insight* yang diperoleh benar-benar mencerminkan kebutuhan penelitian atau kondisi nyata dari populasi yang diteliti. Dengan demikian, relevansi dan spesifikasi data primer berperan krusial dalam menghasilkan temuan yang valid, terukur, dan aplikatif dalam praktik sains data.

### 3. Mendukung Pengembangan Model dan Algoritma

Data primer memiliki peran penting dalam mendukung pengembangan model dan algoritma dalam sains data karena data yang dikumpulkan secara langsung dari sumber asli memungkinkan pembuatan model yang sesuai dengan kondisi nyata dan variabel yang relevan. Kualitas data yang tinggi dan spesifik memungkinkan algoritma *machine learning* atau model statistik untuk belajar dari informasi yang benar-benar mencerminkan fenomena yang sedang diteliti, sehingga hasil prediksi atau klasifikasi menjadi lebih akurat dan dapat diandalkan. Hal ini karena data primer menyediakan *input* yang autentik dan sesuai konteks, yang sangat penting untuk membangun model analitik dan algoritma yang efektif dalam menghasilkan *insight* yang valid (Shmueli *et al.*, 2020).

Penggunaan data primer memfasilitasi iterasi model yang lebih baik, karena peneliti dapat menyesuaikan *input*, memperbaiki variabel, dan menambahkan data tambahan jika diperlukan untuk meningkatkan kinerja algoritma. Hal ini sangat penting dalam pengembangan model prediktif, klasifikasi, dan segmentasi, di mana kualitas *input* langsung memengaruhi akurasi *output*, interpretasi, dan relevansi *insight* yang diperoleh. Dengan data primer, proses pengembangan model dan algoritma menjadi lebih adaptif, terkontrol, dan dapat menghasilkan

temuan yang lebih valid serta implementasi yang lebih efektif dalam praktik sains data.

#### **4. Menyediakan Wawasan Baru atau Informasi Unik**

Data primer memiliki keunggulan dalam menyediakan wawasan baru atau informasi unik karena data dikumpulkan langsung dari sumber asli dengan fokus pada tujuan penelitian yang spesifik, sehingga memungkinkan penemuan fenomena, pola, atau *insight* yang belum pernah diidentifikasi sebelumnya. Informasi ini sangat berharga dalam sains data, di mana *insight* baru dapat digunakan untuk inovasi, pengembangan strategi, atau pengambilan keputusan yang lebih efektif dan berbasis bukti. Dengan demikian, penggunaan data primer memungkinkan peneliti untuk memperoleh informasi unik dan mendalam yang tidak tersedia dari sumber sekunder, sehingga mendukung penciptaan pengetahuan baru dan analisis yang lebih kaya (Hair *et al.*, 2019).

Ketersediaan informasi unik ini juga memungkinkan organisasi atau peneliti untuk mengidentifikasi peluang tersembunyi, memahami perilaku responden secara lebih mendalam, dan menyesuaikan strategi berdasarkan data asli yang relevan. Selain itu, wawasan yang diperoleh dari data primer dapat digunakan untuk memvalidasi atau menantang temuan dari data sekunder, meningkatkan akurasi dan keandalan hasil analisis. Dengan demikian, data primer menjadi komponen strategis dalam sains data karena menyediakan fondasi informasi yang eksklusif, spesifik, dan relevan bagi pengambilan keputusan maupun penelitian lebih lanjut.

#### **5. Memfasilitasi Validasi dan Verifikasi Temuan**

Data primer memiliki peran penting dalam memfasilitasi validasi dan verifikasi temuan karena data dikumpulkan langsung oleh peneliti sesuai dengan tujuan penelitian, sehingga memungkinkan pemeriksaan ulang atas hasil analisis dan memastikan kesesuaian antara data yang diperoleh dengan fenomena yang diteliti. Proses validasi ini penting untuk menilai keandalan dan konsistensi hasil, serta untuk mendeteksi adanya bias atau kesalahan pengukuran sebelum data digunakan dalam pengambilan keputusan atau publikasi ilmiah. Dengan demikian, penggunaan data primer memungkinkan peneliti untuk melakukan

verifikasi temuan secara sistematis, sehingga meningkatkan keakuratan, kredibilitas, dan validitas hasil penelitian (Flick, 2020).

Kemampuan memverifikasi dan memvalidasi temuan dari data primer mendukung transparansi dan akuntabilitas dalam penelitian, karena setiap langkah pengumpulan dan analisis dapat ditelusuri kembali. Hal ini juga memperkuat interpretasi hasil dan memberikan dasar yang lebih kokoh untuk pengembangan teori atau penerapan praktis, karena keputusan dan strategi yang diambil didasarkan pada data yang telah teruji validitasnya. Dengan demikian, data primer tidak hanya berfungsi sebagai sumber informasi, tetapi juga sebagai alat penting untuk memastikan kualitas, reliabilitas, dan ketepatan hasil analisis dalam konteks sains data.



# BAB IV

## METODE PENGUMPULAN DATA PRIMER KONVENSIONAL

---

Metode pengumpulan data primer konvensional merupakan pendekatan tradisional untuk memperoleh informasi langsung dari sumbernya melalui wawancara, kuesioner, atau observasi. Metode ini memungkinkan peneliti mendapatkan data yang spesifik dan relevan, serta menyesuaikan pertanyaan secara fleksibel, meskipun prosesnya cenderung memakan waktu dan memerlukan tenaga lebih.

### A. Wawancara

Wawancara adalah metode pengumpulan data primer yang dilakukan melalui interaksi langsung antara peneliti dan responden untuk memperoleh informasi yang mendalam dan akurat sesuai konteks penelitian. Tujuan utama dari wawancara adalah menggali pandangan, pengalaman, dan persepsi responden secara rinci, sehingga peneliti dapat memahami fenomena yang diteliti secara komprehensif dan menangkap nuansa yang mungkin tidak muncul melalui metode tertulis. Menurut Kvale dan Brinkmann (2018), wawancara memungkinkan peneliti untuk memperoleh data kontekstual dan interpretatif yang relevan, sehingga hasil penelitian dapat mencerminkan realitas subjektif responden secara lebih autentik.

Proses wawancara dirancang agar fleksibel namun tetap sistematis, dimulai dari persiapan pertanyaan, pemilihan responden, hingga pencatatan dan analisis data secara cermat. Melalui wawancara, peneliti tidak hanya mengumpulkan jawaban, tetapi juga memperhatikan ekspresi, nada suara, dan reaksi non-verbal yang menambah kedalaman informasi. Dengan demikian, wawancara menjadi alat penting dalam penelitian sosial dan kualitatif, karena mampu menghadirkan pemahaman yang lebih kaya dibandingkan sekadar data kuantitatif.

## 1. Jenis Wawancara

### a. Wawancara Terstruktur

Wawancara terstruktur adalah bentuk wawancara yang dirancang dengan daftar pertanyaan baku yang disusun secara sistematis sehingga setiap responden menjawab pertanyaan yang sama dengan urutan yang konsisten, bertujuan untuk meminimalkan variasi dalam pengumpulan data agar hasilnya lebih mudah dibandingkan dan dianalisis secara kuantitatif. Dengan pendekatan ini, peneliti berfokus pada pertanyaan tertutup atau semi-tertutup yang telah ditentukan sebelumnya sehingga responden memberikan jawaban yang spesifik sesuai dengan pilihan yang tersedia, memungkinkan validitas dan reliabilitas data lebih terjaga karena batasan pada interpretasi bebas responden; menurut Bryman (2018), wawancara terstruktur memfasilitasi pengumpulan data yang seragam yang mendukung analisis statistik dan generalisasi hasil penelitian.

Pada pelaksanaan wawancara terstruktur, pewawancara berfungsi sebagai pengumpul data yang mengikuti skrip pertanyaan tanpa menambahkan atau mengubah isi pertanyaan, sehingga interaksi antara pewawancara dan responden tetap terkontrol dan berfokus pada tujuan pengumpulan data. Metode ini sangat efektif digunakan dalam penelitian besar yang membutuhkan standarisasi tinggi dan pengolahan data yang cepat, meskipun keterbatasannya mencakup kurangnya fleksibilitas untuk menggali jawaban lebih dalam karena responden hanya terbatas pada pertanyaan dan opsi jawaban yang telah ditetapkan.

### b. Wawancara Semi-terstruktur

Wawancara semi-terstruktur adalah metode pengumpulan data yang memadukan fleksibilitas wawancara tidak terstruktur dengan panduan pertanyaan dasar yang telah ditentukan sebelumnya, sehingga peneliti dapat melihat jawaban responden lebih mendalam sambil tetap menjaga fokus penelitian. Metode ini memungkinkan peneliti menyesuaikan pertanyaan lanjutan secara spontan sesuai respons yang diberikan, sehingga informasi yang diperoleh tidak hanya relevan tetapi juga kontekstual dan kaya nuansa; menurut DiCicco-Bloom dan Crabtree (2019), wawancara semi-terstruktur efektif untuk memperoleh data

kualitatif yang mendalam dan memperkuat pemahaman terhadap fenomena yang diteliti.

Pada praktiknya, wawancara semi-terstruktur memerlukan persiapan panduan pertanyaan yang fleksibel, di mana pewawancara dapat menambahkan pertanyaan tambahan untuk menggali informasi lebih detail. Pendekatan ini sangat berguna untuk penelitian sosial dan perilaku, karena memungkinkan respons yang luas dan mendalam dari peserta, sekaligus menjaga relevansi data dengan tujuan penelitian.

c. Wawancara Tidak Terstruktur

Wawancara tidak terstruktur adalah teknik pengumpulan data primer yang mengandalkan percakapan bebas tanpa daftar pertanyaan baku, di mana peneliti mengajukan pertanyaan secara spontan berdasarkan jawaban responden untuk melihat pengalaman, pandangan, dan realitas subjektif secara mendalam; pendekatan ini sangat fleksibel sehingga alur wawancara mengikuti respons peserta dan memungkinkan munculnya informasi yang tidak terduga, sehingga sangat cocok untuk penelitian kualitatif yang bertujuan memahami fenomena sosial secara intuitif dan kontekstual (George, 2025).

Pada praktiknya, wawancara tidak terstruktur menyerupai percakapan alami antara peneliti dan responden, di mana pertanyaan dapat berkembang seiring berjalannya dialog sehingga peneliti dapat menggali nuansa, makna, serta rincian pengalaman yang mungkin tidak akan terungkap melalui format wawancara yang lebih kaku; sifat ini menjadikannya metode yang efektif ketika tujuan penelitian adalah memperoleh wawasan yang kaya dan mendalam mengenai realitas hidup responden.

## 2. Proses Wawancara

Proses wawancara merupakan rangkaian kegiatan sistematis yang dimulai dari perencanaan, pemilihan responden, penyusunan pertanyaan, pelaksanaan wawancara, hingga pencatatan dan analisis data, yang dirancang untuk memastikan data yang diperoleh relevan, lengkap, dan akurat. Tahapan ini menuntut perencanaan matang agar interaksi antara peneliti dan responden berlangsung efektif, sehingga informasi yang dikumpulkan dapat mencerminkan perspektif responden

secara autentik dan mendukung tujuan penelitian secara keseluruhan (Qu & Dumay, 2018).

Pada tahap persiapan, peneliti menyusun panduan wawancara yang menyesuaikan dengan jenis wawancara yang digunakan, baik terstruktur, semi-terstruktur, maupun tidak terstruktur, sambil mempertimbangkan konteks penelitian dan karakteristik responden. Persiapan ini juga meliputi pelatihan pewawancara agar mampu mengelola komunikasi dengan responden secara profesional, menjaga objektivitas, dan meminimalkan bias dalam proses pengumpulan data.

Pelaksanaan wawancara menekankan interaksi langsung yang adaptif, di mana peneliti menyesuaikan pertanyaan dan pendekatan komunikasi sesuai respons yang muncul, sambil tetap menjaga fokus penelitian. Selama proses ini, peneliti mencatat jawaban, ekspresi non-verbal, dan reaksi emosional responden secara cermat untuk memastikan data yang dikumpulkan kaya konteks dan dapat dianalisis secara komprehensif.

Tahap terakhir dalam proses wawancara adalah analisis data yang melibatkan transkripsi, pengkodean, dan interpretasi hasil wawancara untuk menemukan pola, tema, dan *insight* yang relevan dengan tujuan penelitian. Proses ini menuntut ketelitian tinggi agar data yang diperoleh tidak hanya deskriptif tetapi juga dapat digunakan untuk menghasilkan kesimpulan yang valid dan dapat dipertanggungjawabkan secara ilmiah.

### **3. Keunggulan Wawancara**

Wawancara memiliki keunggulan utama dalam menghasilkan data yang kaya, mendalam, dan kontekstual karena peneliti dapat menggali informasi langsung dari responden serta menyesuaikan pertanyaan secara fleksibel berdasarkan jawaban yang muncul, sehingga memungkinkan pemahaman yang lebih komprehensif terhadap fenomena yang diteliti. Metode ini juga memungkinkan peneliti menangkap nuansa non-verbal, ekspresi, dan intonasi suara yang menambah kualitas dan validitas data, serta memfasilitasi klarifikasi jawaban secara langsung untuk mengurangi ambiguitas. Menurut Barriball dan While (2019), wawancara memberikan akses pada wawasan subjektif responden yang sulit diperoleh melalui instrumen tertulis, sehingga menjadikannya metode yang efektif untuk penelitian kualitatif yang menekankan kedalaman dan konteks data.

#### **4. Keterbatasan Wawancara**

Wawancara memiliki keterbatasan yang signifikan terkait waktu dan sumber daya, karena proses pengumpulan data memerlukan interaksi langsung dengan responden, pencatatan, dan analisis yang cermat, sehingga penelitian dapat menjadi lebih memakan waktu dan mahal dibandingkan metode lain. Selain itu, kualitas data sangat bergantung pada keterampilan pewawancara dalam mengajukan pertanyaan dan mengelola interaksi, sehingga adanya bias pewawancara atau responden dapat memengaruhi validitas hasil penelitian. Menurut Opdenakker (2019), wawancara juga rentan terhadap kesalahan subjektivitas dan variabilitas jawaban antarresponden, sehingga membutuhkan strategi pengendalian yang hati-hati untuk memastikan reliabilitas dan konsistensi data.

### **B. Kuesioner**

Kuesioner adalah instrumen pengumpulan data yang disusun dalam bentuk pertanyaan tertulis yang dirancang untuk dijawab langsung oleh responden, baik secara manual maupun daring, sehingga memungkinkan pengumpulan informasi secara sistematis dan terstandarisasi. Tujuan utama penggunaan kuesioner adalah untuk memperoleh data yang dapat diukur, dianalisis, dan dibandingkan dengan mudah, termasuk informasi mengenai sikap, opini, pengetahuan, atau perilaku responden sesuai fokus penelitian. Menurut Fink (2019), kuesioner memberikan metode yang efisien dan konsisten untuk mengumpulkan data dari sejumlah besar responden, sehingga mempermudah analisis kuantitatif maupun kualitatif dengan tingkat reliabilitas yang tinggi.

Pada praktiknya, kuesioner memungkinkan peneliti untuk menjangkau responden yang luas tanpa interaksi tatap muka, sekaligus memberi waktu bagi responden untuk mempertimbangkan jawabannya secara lebih reflektif dan mendetail. Dengan demikian, kuesioner tidak hanya menghemat waktu dan biaya, tetapi juga membantu menghasilkan data yang lebih terstruktur dan mudah diolah, menjadikannya salah satu metode paling populer dalam penelitian sosial dan pendidikan.

## 1. Jenis Kuesioner

### a. Kuesioner Tertutup

Kuesioner tertutup adalah jenis kuesioner di mana setiap pertanyaan disertai dengan pilihan jawaban yang telah ditentukan sebelumnya, sehingga responden hanya memilih jawaban yang tersedia, memungkinkan pengumpulan data yang lebih mudah dianalisis secara kuantitatif dan dibandingkan antarresponden. Metode ini efektif untuk penelitian yang memerlukan standarisasi tinggi dan pengolahan data yang cepat, karena meminimalkan ambiguitas jawaban dan meningkatkan konsistensi data; menurut Creswell dan Creswell (2018), kuesioner tertutup sangat berguna dalam penelitian kuantitatif karena menyediakan informasi yang terstruktur dan mempermudah analisis statistik.

Penggunaan kuesioner tertutup memungkinkan peneliti memperoleh data dalam jumlah besar dengan cepat, karena responden dapat menjawab pertanyaan tanpa memikirkan jawaban bebas, sehingga efisiensi waktu dan biaya meningkat. Selain itu, kuesioner tertutup mempermudah pengkodean jawaban untuk keperluan analisis data, mengurangi risiko kesalahan interpretasi, dan memfasilitasi perbandingan lintas sampel secara sistematis.

### b. Kuesioner Terbuka

Kuesioner terbuka adalah jenis instrumen pengumpulan data yang memungkinkan responden menjawab pertanyaan dengan kata-katanya sendiri tanpa dibatasi opsi jawaban, sehingga peneliti dapat memperoleh informasi yang lebih rinci, deskriptif, dan kontekstual mengenai pandangan, pengalaman, atau persepsi peserta. Metode ini sangat bermanfaat dalam penelitian kualitatif karena memberikan fleksibilitas untuk melihat jawaban yang tidak terduga, memperkaya pemahaman terhadap fenomena yang diteliti, dan memungkinkan analisis tematik yang mendalam; menurut Saunders *et al.* (2019), kuesioner terbuka memungkinkan responden mengekspresikan opini dan pengalaman secara luas, sehingga peneliti dapat menemukan wawasan baru yang tidak mungkin diperoleh melalui pertanyaan tertutup.

Pada praktiknya, kuesioner terbuka memberikan kesempatan kepada responden untuk menyampaikan pemikiran dengan

bebas, sehingga hasil jawaban cenderung lebih reflektif dan autentik. Pendekatan ini mendukung penggalan informasi yang kompleks dan nuansa konteks yang tidak dapat ditangkap melalui pertanyaan pilihan ganda, sehingga mempermudah identifikasi pola, tema, atau fenomena yang sebelumnya tidak terduga.

c. Kuesioner Kombinasi

Kuesioner kombinasi adalah jenis kuesioner yang menggabungkan pertanyaan tertutup dan terbuka dalam satu instrumen, sehingga memungkinkan peneliti memperoleh data yang terstruktur untuk analisis kuantitatif sekaligus data deskriptif yang kaya konteks untuk analisis kualitatif. Metode ini efektif karena memadukan kemudahan pengolahan data dari pertanyaan tertutup dengan fleksibilitas dan kedalaman informasi dari pertanyaan terbuka, sehingga hasil penelitian menjadi lebih komprehensif dan menyeluruh; menurut Bryman dan Bell (2019), kuesioner kombinasi memungkinkan peneliti memanfaatkan keunggulan masing-masing jenis pertanyaan untuk memperoleh gambaran yang lebih utuh mengenai fenomena yang diteliti.

Pada praktiknya, kuesioner kombinasi dirancang agar pertanyaan tertutup dapat memberikan jawaban yang mudah diolah dan dibandingkan, sementara pertanyaan terbuka memberikan ruang bagi responden untuk menjelaskan alasan, pengalaman, atau pandangannya secara bebas. Pendekatan ini mendukung penggalan data yang lebih kaya dan relevan, serta memungkinkan peneliti menyesuaikan analisis dengan tujuan penelitian baik dari sisi kuantitatif maupun kualitatif.

## 2. Proses Penggunaan Kuesioner

Proses penggunaan kuesioner dimulai dengan perencanaan yang matang, termasuk penentuan tujuan penelitian, identifikasi variabel yang akan diukur, dan pemilihan jenis pertanyaan yang sesuai, baik tertutup, terbuka, maupun kombinasi, sehingga data yang dikumpulkan relevan dan dapat dianalisis dengan tepat. Tahap ini juga mencakup penentuan sampel responden yang representatif dan perancangan instrumen yang jelas, singkat, serta bebas bias untuk memastikan validitas dan reliabilitas data; menurut Zikmund *et al.* (2019), perencanaan dan desain

kuesioner yang cermat sangat penting untuk menghasilkan data yang akurat dan dapat dipertanggungjawabkan secara ilmiah.

Setelah tahap perencanaan, kuesioner didistribusikan kepada responden melalui berbagai media, seperti tatap muka, pos, atau platform daring, dengan instruksi yang jelas untuk memastikan responden memahami pertanyaan dan menjawab dengan jujur. Pendekatan ini membantu meminimalkan kesalahan pengisian dan meningkatkan tingkat respons, sehingga data yang dikumpulkan lebih representatif dan dapat digunakan untuk analisis selanjutnya.

Tahap berikutnya adalah pengumpulan dan pengecekan jawaban, di mana peneliti meninjau setiap kuesioner untuk mendeteksi ketidakkonsistenan, kekosongan, atau jawaban yang ambigu, sehingga kualitas data tetap terjaga. Proses ini penting untuk memastikan bahwa semua jawaban dapat dianalisis secara sistematis, baik melalui pengkodean jawaban tertutup maupun kategorisasi jawaban terbuka, sehingga data siap digunakan untuk menghasilkan temuan yang valid.

Tahap terakhir meliputi analisis data, yang mencakup pengolahan jawaban tertutup secara statistik dan interpretasi jawaban terbuka secara tematik, sehingga penelitian dapat menghasilkan *insight* yang mendalam dan komprehensif. Dengan proses yang terstruktur ini, kuesioner menjadi alat yang efektif untuk mengumpulkan data dalam jumlah besar secara efisien, sambil tetap mempertahankan kualitas dan relevansi informasi yang diperoleh.

### **3. Keunggulan Kuesioner**

Kuesioner memiliki keunggulan karena memungkinkan pengumpulan data dari banyak responden secara efisien, baik secara langsung maupun melalui media daring, sehingga waktu dan biaya penelitian dapat dihemat secara signifikan dibandingkan metode wawancara tatap muka. Selain itu, kuesioner menyediakan format yang terstruktur dan konsisten, sehingga memudahkan pengkodean, pengolahan, dan analisis data secara kuantitatif maupun kualitatif, sekaligus meminimalkan kesalahan interpretasi jawaban. Menurut Saunders, Lewis, dan Thornhill (2019), kuesioner memberikan metode yang sistematis dan fleksibel untuk memperoleh data yang dapat diandalkan dari sampel besar, sambil tetap memungkinkan eksplorasi opini dan persepsi responden melalui pertanyaan terbuka.

#### 4. Keterbatasan Kuesioner

Kuesioner memiliki keterbatasan karena jawaban yang diperoleh sangat bergantung pada pemahaman dan kesungguhan responden dalam mengisi pertanyaan, sehingga risiko kesalahan atau bias responden dapat memengaruhi validitas data. Selain itu, kuesioner terbatas dalam menggali informasi yang mendalam karena interaksi langsung antara peneliti dan responden minim, sehingga nuansa, ekspresi, dan konteks jawaban sering kali tidak tertangkap. Menurut Zikmund *et al.* (2019), kuesioner rentan terhadap tingkat respons yang rendah, interpretasi yang berbeda antarresponden, serta kesulitan dalam memperoleh data yang bersifat kompleks atau emosional, sehingga peneliti harus berhati-hati dalam perancangan dan analisis data.

### C. Observasi

Observasi adalah metode pengumpulan data primer di mana peneliti secara langsung mengamati perilaku, interaksi, atau fenomena dalam konteks alaminya tanpa bergantung pada laporan atau opini dari subjek penelitian, sehingga memungkinkan data yang diperoleh lebih objektif dan autentik. Tujuan utama observasi adalah untuk memahami perilaku atau situasi secara langsung, menangkap dinamika interaksi, dan memperoleh informasi kontekstual yang sulit dijangkau melalui metode lain seperti wawancara atau kuesioner. Menurut Flick (2018), observasi memungkinkan peneliti mendapatkan data yang nyata dan relevan dengan kondisi lapangan, sehingga mendukung analisis yang akurat dan pemahaman mendalam terhadap fenomena yang diteliti.

Pada praktiknya, observasi digunakan untuk mempelajari perilaku manusia, proses organisasi, atau kejadian tertentu dengan tujuan menghasilkan pemahaman yang lebih lengkap dan holistik. Metode ini memberikan kesempatan bagi peneliti untuk mencatat interaksi, ekspresi, dan kondisi lingkungan secara langsung, sehingga informasi yang diperoleh mencerminkan realitas yang sebenarnya dan dapat digunakan untuk mendukung kesimpulan penelitian secara ilmiah.

#### 1. Jenis Observasi

Observasi memiliki berbagai jenis yang disesuaikan dengan tujuan dan konteks penelitian, termasuk observasi partisipatif di mana peneliti terlibat langsung dalam kegiatan yang diamati untuk

memperoleh pemahaman mendalam, serta observasi non-partisipatif di mana peneliti hanya mengamati dari luar tanpa ikut serta agar perilaku subjek tetap alami dan tidak terpengaruh kehadiran peneliti. Jenis lain mencakup observasi terstruktur, yang menggunakan pedoman atau checklist untuk mencatat perilaku secara sistematis sehingga mempermudah analisis data, dan observasi tidak terstruktur, yang lebih fleksibel dan memungkinkan penemuan informasi tak terduga; menurut Cohen *et al.* (2018), pemilihan jenis observasi yang tepat sangat penting untuk memperoleh data yang relevan, valid, dan mencerminkan fenomena yang sebenarnya dalam konteks penelitian.

Observasi partisipatif memungkinkan peneliti mendapatkan wawasan subjektif dari pengalaman langsung, sehingga dapat memahami dinamika dan interaksi secara lebih lengkap. Sementara observasi non-partisipatif membantu menjaga objektivitas, karena subjek yang diamati tidak merasa diawasi secara aktif, sehingga perilaku yang muncul lebih natural dan dapat dijadikan data yang representatif.

Observasi terstruktur ideal untuk penelitian yang membutuhkan pengumpulan data sistematis dan analisis kuantitatif, karena pedoman atau checklist memastikan setiap kejadian atau perilaku dicatat secara konsisten. Sebaliknya, observasi tidak terstruktur lebih cocok untuk penelitian eksploratif atau kualitatif yang bertujuan menggali fenomena kompleks, karena fleksibilitasnya memungkinkan peneliti menangkap detail, pola, dan informasi kontekstual yang tidak terduga.

## **2. Proses Observasi**

Proses observasi dimulai dengan perencanaan yang matang, termasuk penentuan tujuan, fokus pengamatan, dan pemilihan jenis observasi yang sesuai, baik partisipatif maupun non-partisipatif, serta terstruktur atau tidak terstruktur, agar data yang diperoleh relevan dan dapat dianalisis secara sistematis. Tahap ini juga melibatkan penyusunan pedoman atau checklist jika diperlukan, sehingga setiap perilaku atau kejadian yang diamati dapat dicatat secara konsisten; menurut Flick (2018), perencanaan dan desain observasi yang cermat sangat penting untuk memastikan bahwa pengumpulan data menghasilkan informasi yang akurat dan mendukung validitas penelitian.

Tahap pelaksanaan observasi menekankan pencatatan langsung terhadap perilaku, interaksi, atau fenomena di lapangan, baik secara manual maupun menggunakan alat rekam, sambil meminimalkan

gangguan terhadap subjek penelitian agar perilaku tetap alami. Proses ini menuntut kewaspadaan dan konsentrasi tinggi dari peneliti untuk menangkap detail, ekspresi, dan konteks lingkungan yang relevan, sehingga data yang dikumpulkan dapat mencerminkan realitas secara autentik.

Tahap terakhir adalah analisis data observasi, yang meliputi pengolahan catatan, pengkodean perilaku, dan interpretasi hasil untuk menemukan pola, tema, dan *insight* yang mendukung tujuan penelitian. Dengan proses yang terstruktur ini, observasi menjadi metode yang efektif untuk menghasilkan data empiris yang kaya konteks, memungkinkan pemahaman fenomena secara mendalam dan holistik.

### **3. Keunggulan Observasi**

Observasi memiliki keunggulan utama karena memungkinkan peneliti mengumpulkan data yang nyata dan autentik secara langsung dari objek penelitian, sehingga informasi yang diperoleh mencerminkan perilaku, interaksi, dan konteks lingkungan secara akurat tanpa bergantung pada persepsi atau ingatan responden. Metode ini juga memungkinkan peneliti menangkap detail non-verbal, ekspresi, dan dinamika situasional yang sulit diakses melalui kuesioner atau wawancara, sehingga kualitas data lebih kaya dan mendalam. Menurut Creswell dan Creswell (2018), observasi memberikan peluang untuk memperoleh wawasan empiris yang konkret dan kontekstual, yang dapat digunakan untuk analisis mendalam dan pengembangan teori yang lebih valid dalam penelitian sosial dan perilaku.

### **4. Keterbatasan Observasi**

Observasi memiliki keterbatasan karena proses pengumpulan data sering kali memerlukan waktu dan tenaga yang cukup besar, terutama jika dilakukan dalam jangka panjang atau melibatkan banyak subjek, sehingga penelitian menjadi lebih mahal dan kompleks. Selain itu, kualitas data sangat tergantung pada kemampuan dan objektivitas peneliti dalam mengamati dan mencatat perilaku atau fenomena yang terjadi, sehingga risiko bias subjektif tidak dapat diabaikan. Menurut Cohen, Manion, dan Morrison (2018), observasi juga menghadapi tantangan berupa kemungkinan perilaku subjek berubah jika menyadari sedang diamati, serta keterbatasan dalam menangkap fenomena internal

atau motivasi yang tidak tampak secara langsung, sehingga peneliti harus berhati-hati dalam interpretasi data.

#### **D. Focus Group Discussion (FGD)**

*Focus Group Discussion* (FGD) adalah metode pengumpulan data kualitatif yang melibatkan sekelompok peserta, biasanya antara enam hingga dua belas orang, yang berkumpul untuk berdiskusi mengenai topik tertentu di bawah fasilitasi seorang moderator, sehingga interaksi kelompok dapat memunculkan pandangan, pengalaman, dan opini yang beragam. Tujuan FGD adalah untuk memperoleh wawasan mendalam mengenai persepsi, sikap, dan motivasi peserta serta memahami dinamika sosial yang muncul selama diskusi, yang seringkali sulit diperoleh melalui wawancara individu atau kuesioner. Menurut Krueger dan Casey (2019), FGD efektif dalam melihat opini kolektif dan interaksi sosial, sehingga memungkinkan peneliti mengidentifikasi tema-tema penting, pola pemikiran, dan isu yang relevan dengan konteks penelitian.

Pada praktiknya, FGD digunakan untuk menggali informasi yang bersifat kompleks dan kontekstual, termasuk persepsi peserta terhadap fenomena sosial, budaya, atau organisasi tertentu. Metode ini memberikan kesempatan bagi peserta untuk saling bertukar pendapat, memunculkan ide baru, dan memperluas wawasan melalui diskusi kelompok, sehingga hasil penelitian menjadi lebih kaya, beragam, dan mendalam.

##### **1. Karakteristik FGD**

*Focus Group Discussion* (FGD) memiliki karakteristik khas, yaitu dilakukan dalam kelompok kecil yang memungkinkan interaksi dinamis antara peserta sehingga berbagai pendapat, pengalaman, dan persepsi dapat muncul secara spontan dan beragam. Diskusi dipandu oleh seorang moderator yang berperan menjaga fokus topik, menyeimbangkan partisipasi, dan memastikan setiap peserta berkesempatan menyampaikan pandangan, sehingga data yang diperoleh relevan dan dapat dianalisis secara mendalam; menurut Morgan (2018), FGD menekankan interaksi sosial sebagai sumber informasi, memungkinkan peneliti memahami dinamika kelompok serta pola pemikiran kolektif yang muncul selama diskusi.

Peserta FGD dipilih secara purposive berdasarkan kesamaan karakteristik atau relevansi dengan topik penelitian, sehingga diskusi menghasilkan informasi yang kontekstual dan representatif terhadap fenomena yang dikaji. Metode ini juga fleksibel dalam pelaksanaan, memungkinkan moderator menyesuaikan pertanyaan atau topik tambahan berdasarkan respons kelompok, sehingga wawasan yang diperoleh lebih kaya dan mendalam.

## **2. Proses Pelaksanaan FGD**

Proses pelaksanaan *Focus Group Discussion* (FGD) dimulai dengan perencanaan yang matang, termasuk penentuan tujuan, fokus diskusi, dan pemilihan peserta yang relevan serta representatif terhadap topik penelitian, agar informasi yang diperoleh mencerminkan pandangan dan pengalaman yang tepat. Tahap ini juga mencakup penyusunan panduan pertanyaan atau topik diskusi yang jelas, sehingga moderator dapat memandu diskusi secara sistematis dan menjaga jalannya interaksi tetap fokus; menurut Krueger dan Casey (2019), perencanaan yang cermat merupakan kunci untuk memastikan FGD menghasilkan data yang kaya, relevan, dan dapat dianalisis secara mendalam.

Tahap pelaksanaan FGD menekankan peran moderator dalam memfasilitasi diskusi, menjaga keseimbangan partisipasi antar peserta, dan mendorong interaksi yang spontan serta konstruktif, sehingga setiap peserta memiliki kesempatan untuk mengekspresikan opini dan pengalaman secara terbuka. Moderator juga memonitor dinamika kelompok untuk mengidentifikasi pandangan dominan maupun minoritas yang dapat memberikan wawasan berbeda, sehingga hasil diskusi mencerminkan variasi perspektif secara komprehensif.

Selama diskusi, proses pencatatan dan/atau perekaman sangat penting untuk memastikan seluruh informasi terekam dengan baik, termasuk interaksi verbal maupun non-verbal peserta, sehingga memudahkan analisis data secara tematik. Pendekatan ini memungkinkan peneliti mengolah data yang diperoleh menjadi kategori, tema, dan pola yang relevan dengan tujuan penelitian, sambil tetap mempertahankan konteks alami diskusi.

Tahap terakhir adalah analisis hasil FGD, yang melibatkan transkripsi percakapan, pengkodean data, dan interpretasi tematik untuk mengidentifikasi isu, persepsi, dan pola pemikiran peserta. Dengan

mengikuti proses yang sistematis ini, FGD menjadi metode efektif untuk memperoleh data kualitatif yang kaya, beragam, dan kontekstual, mendukung pemahaman yang mendalam terhadap fenomena yang diteliti.

### **3. Keunggulan FGD**

*Focus Group Discussion* (FGD) memiliki keunggulan karena memungkinkan peneliti mengumpulkan data kualitatif yang kaya dan mendalam melalui interaksi kelompok, di mana peserta dapat saling bertukar opini, pengalaman, dan pandangan yang sering kali memunculkan wawasan baru yang tidak diperoleh melalui wawancara individu. Metode ini juga efektif untuk mengidentifikasi persepsi kolektif, dinamika sosial, dan isu yang relevan dengan konteks penelitian, sehingga peneliti dapat menangkap nuansa serta variasi perspektif yang kompleks. Menurut Krueger dan Casey (2019), FGD memberikan peluang unik untuk melihat tema dan pola pemikiran secara lebih komprehensif melalui diskusi yang interaktif dan terarah, sehingga hasil penelitian menjadi lebih kaya, autentik, dan relevan.

### **4. Keterbatasan FGD**

*Focus Group Discussion* (FGD) memiliki keterbatasan karena hasil diskusi sangat dipengaruhi oleh dinamika kelompok, di mana peserta yang dominan dapat memengaruhi pandangan peserta lain, sehingga data yang diperoleh terkadang kurang mencerminkan opini individu secara independen. Selain itu, efektivitas FGD sangat bergantung pada keterampilan moderator dalam memfasilitasi diskusi dan menjaga fokus, sehingga kesalahan atau bias moderator dapat memengaruhi kualitas data. Menurut Morgan (2018), FGD juga menghadapi tantangan berupa keterbatasan waktu, biaya, dan kesulitan dalam menganalisis interaksi verbal maupun non-verbal secara sistematis, sehingga peneliti perlu berhati-hati dalam interpretasi dan generalisasi hasil diskusi.

## **E. Eksperimen**

Eksperimen adalah metode penelitian yang dilakukan dengan mengatur, memanipulasi, dan mengendalikan satu atau lebih variabel independen untuk mengamati pengaruhnya terhadap variabel dependen,

sehingga memungkinkan peneliti mengidentifikasi hubungan sebab-akibat secara sistematis dan terukur. Tujuan eksperimen adalah untuk menguji hipotesis secara ilmiah, memperoleh data yang valid dan reliabel, serta memahami bagaimana perubahan pada satu variabel memengaruhi variabel lain dalam kondisi yang terkendali. Menurut Creswell dan Creswell (2018), eksperimen memungkinkan peneliti untuk menilai efek perlakuan atau intervensi secara langsung, sehingga kesimpulan yang dihasilkan dapat dipercaya dan mendukung pengembangan teori yang akurat.

Pada praktiknya, eksperimen digunakan untuk mempelajari fenomena yang dapat diukur secara kuantitatif, termasuk perilaku manusia, respons terhadap perlakuan tertentu, atau efek perubahan lingkungan, dengan tujuan menghasilkan data yang dapat diuji ulang (replikasi) dan dianalisis secara objektif. Metode ini sangat berguna dalam penelitian ilmiah karena memungkinkan kontrol terhadap variabel pengganggu, meminimalkan bias, dan menghasilkan informasi yang dapat dijadikan dasar pengambilan keputusan atau perumusan kebijakan berbasis bukti.

## 1. Jenis Eksperimen

Eksperimen memiliki berbagai jenis yang disesuaikan dengan tujuan penelitian dan kondisi subjek, termasuk eksperimen laboratorium yang dilakukan dalam lingkungan terkendali penuh untuk meminimalkan pengaruh variabel luar dan memfokuskan pada hubungan sebab-akibat antarvariabel, serta eksperimen lapangan yang dilaksanakan di lingkungan alami subjek agar hasil lebih mencerminkan realitas sehari-hari. Jenis lain mencakup eksperimen murni (*true experiment*) yang melibatkan penempatan acak subjek ke kelompok eksperimen dan kontrol untuk mengurangi bias, dan eksperimen semiterkontrol (*quasi-experiment*) yang tetap membandingkan kelompok eksperimen dan kontrol meskipun penempatan subjek tidak acak; menurut Shadish, Cook, dan Campbell (2018), pemilihan jenis eksperimen yang tepat sangat penting untuk memastikan validitas internal dan eksternal, sehingga kesimpulan tentang hubungan sebab-akibat dapat dipercaya.

Eksperimen laboratorium unggul dalam kontrol variabel dan pengulangan percobaan, memungkinkan peneliti memperoleh data yang reliabel dan lebih mudah dianalisis secara statistik. Sebaliknya,

eksperimen lapangan lebih realistis dan relevan dengan kondisi nyata, namun rentan terhadap gangguan variabel luar, sehingga peneliti harus berhati-hati dalam interpretasi hasil.

Eksperimen semiterkontrol dan murni memungkinkan fleksibilitas dalam desain penelitian sesuai keterbatasan praktis dan etika, sekaligus tetap menyediakan informasi mengenai efek perlakuan. Kedua jenis ini memungkinkan perbandingan antara kelompok eksperimen dan kontrol untuk mengidentifikasi efek intervensi secara jelas, sehingga peneliti dapat membuat inferensi kausal yang lebih akurat.

## **2. Proses Pelaksanaan Eksperimen**

Peneliti dalam eksperimen memulai proses pelaksanaan dengan merumuskan pertanyaan penelitian yang jelas dan hipotesis yang dapat diuji, lalu menentukan desain eksperimen yang tepat, termasuk variabel bebas yang akan dimanipulasi dan variabel dependen yang akan diukur secara sistematis; langkah ini juga mencakup pemilihan atau penetapan kelompok perlakuan dan kontrol serta strategi untuk mengendalikan variabel pengganggu agar hasil pengukuran lebih valid. Perencanaan ini menjadi dasar bagi pengambilan sampel atau pembagian subjek ke dalam kelompok yang relevan, yang selanjutnya eksperimen dilaksanakan melalui pemberian perlakuan (treatment) kepada kelompok eksperimen sedangkan kelompok kontrol tetap dalam kondisi standar. Setelah perlakuan diberikan, peneliti melakukan pengukuran variabel dependen sesuai protokol, mencatat respons secara akurat, lalu mengumpulkan data empiris yang diperlukan untuk mengevaluasi apakah hipotesis telah didukung oleh hasil eksperimen.

Pada seluruh proses pelaksanaan eksperimen, peneliti harus memantau dan mencatat setiap tahap dengan teliti, termasuk kontrol terhadap kondisi lingkungan dan konsistensi pemberian perlakuan, guna meminimalkan kemungkinan adanya bias atau faktor luar yang dapat memengaruhi hasil akhir; proses ini sering dilakukan dengan prosedur standar yang tertulis dalam rancangan penelitian. Tahap berikutnya adalah analisis data, di mana data yang dikumpulkan diproses dengan teknik statistik yang sesuai guna menguji hubungan antara variabel independen dan dependen serta menjawab pertanyaan penelitian yang diajukan. Pada akhirnya, peneliti menarik kesimpulan berdasarkan hasil analisis tersebut serta mempertimbangkan implikasi temuan untuk teori

atau praktik ilmiah yang lebih luas, kemudian menyusun laporan penelitian agar proses dan temuan eksperimen dapat dipahami dan direplikasi oleh peneliti lain.

Proses yang sistematis dalam eksperimen membantu memastikan bahwa hubungan sebab-akibat antara variabel dapat diidentifikasi secara jelas dan logis, serta memungkinkan peneliti memprediksi fenomena berdasarkan bukti empiris yang diperoleh dari percobaan yang dilakukan. Menurut Creswell dan Guetterman (2025), penelitian eksperimental melibatkan urutan langkah yang jelas dalam perencanaan, pelaksanaan, dan evaluasi, sehingga struktur proses penelitian sangat penting untuk menghasilkan temuan yang valid dan dapat dipercaya. Prosedur yang tertata rapi tersebut menjadikan eksperimen sebagai metode yang kuat dalam menguji hipotesis dan memahami dinamika sebab-akibat dalam berbagai disiplin ilmu.

### **3. Keunggulan Eksperimen**

Eksperimen memiliki keunggulan karena memungkinkan peneliti mengendalikan kondisi penelitian secara ketat, sehingga pengaruh variabel luar dapat diminimalkan dan hasil yang diperoleh lebih mencerminkan dampak langsung dari perlakuan yang diberikan. Keunggulan lainnya adalah kemampuan menghasilkan bukti empiris yang kuat dan terukur, sebab data dikumpulkan melalui prosedur yang terstandar, terkontrol, dan dapat direplikasi, sehingga mendukung keandalan serta konsistensi temuan penelitian. Menurut Montgomery (2019), metode eksperimen unggul dalam mengevaluasi pengaruh perlakuan secara sistematis melalui manipulasi variabel dan analisis statistik yang tepat, sehingga sangat efektif digunakan untuk pengujian hipotesis dan pengambilan keputusan berbasis data.

### **4. Keterbatasan Eksperimen**

Eksperimen memiliki keterbatasan karena kondisi penelitian yang sangat terkontrol sering kali tidak sepenuhnya mencerminkan situasi nyata, sehingga hasil eksperimen berpotensi memiliki validitas eksternal yang rendah dan sulit digeneralisasikan ke konteks kehidupan sehari-hari. Keterbatasan lainnya terletak pada kendala etis dan praktis, karena tidak semua variabel dapat dimanipulasi secara bebas, terutama ketika menyangkut perilaku manusia, nilai sosial, atau dampak yang berisiko merugikan subjek penelitian. Menurut Gravetter dan Forzano

(2018), penelitian eksperimen juga dapat menghadapi masalah keterbatasan sumber daya, seperti waktu, biaya, serta kesulitan dalam mengontrol seluruh variabel pengganggu, yang pada akhirnya dapat memengaruhi interpretasi dan penerapan hasil penelitian.

## **F. Digitalisasi Data Primer Konvensional**

Digitalisasi Data Primer Konvensional merupakan proses sistematis untuk mengubah data primer yang dikumpulkan melalui cara-cara tradisional atau manual ke dalam format digital yang terstruktur, terdokumentasi, dan siap diolah lebih lanjut. Data primer konvensional umumnya diperoleh secara langsung dari sumber pertama melalui metode seperti kuesioner kertas, wawancara tatap muka, observasi lapangan, pencatatan manual, serta dokumentasi fisik lainnya. Digitalisasi bukan sekadar pemindahan bentuk fisik ke digital, tetapi juga mencakup penataan, pengelolaan, dan peningkatan kualitas data agar bernilai guna tinggi dalam analisis, pengambilan keputusan, dan pengembangan pengetahuan.

Secara konseptual, digitalisasi data primer konvensional bertujuan untuk menjembatani keterbatasan metode manual dengan kebutuhan pengolahan data modern yang menuntut kecepatan, ketepatan, dan efisiensi. Dalam konteks penelitian, organisasi, maupun pengembangan sistem informasi, digitalisasi memungkinkan data yang sebelumnya tersebar dalam bentuk fisik menjadi satu kesatuan data yang mudah diakses, dianalisis, serta diverifikasi. Digitalisasi data primer konvensional dilakukan melalui beberapa tahapan yang saling berkaitan dan harus dirancang secara cermat agar kualitas data tetap terjaga:

### **1. Identifikasi dan Inventarisasi Data**

Identifikasi dan inventarisasi data merupakan tahap awal yang sangat krusial dalam pengelolaan data karena berfungsi untuk mengenali secara sistematis seluruh data yang dimiliki, mencakup sumber data, jenis data, format, volume, waktu pengumpulan, serta pihak yang bertanggung jawab terhadap data tersebut. Proses ini membantu memastikan bahwa tidak ada data penting yang terlewat, terduplikasi, atau tidak terdokumentasi, sehingga organisasi atau peneliti memiliki gambaran menyeluruh mengenai aset data yang tersedia sebelum melangkah ke tahap digitalisasi, pengolahan, maupun analisis lanjutan.

Melalui identifikasi dan inventarisasi yang baik, data dapat diklasifikasikan secara tepat berdasarkan karakteristik dan tingkat kepentingannya, yang pada akhirnya mendukung efisiensi pengelolaan dan kualitas pengambilan keputusan berbasis data.

Inventarisasi data juga berperan penting dalam pengendalian kualitas, keamanan, dan keberlanjutan data, karena setiap data yang terdaftar dapat dilacak asal-usulnya serta status penggunaannya sepanjang siklus hidup data. Pendekatan ini memungkinkan penetapan kebijakan pengelolaan data yang lebih terarah, termasuk pengaturan akses, perlindungan data sensitif, serta perencanaan integrasi data antar sistem, sehingga risiko kehilangan atau penyalahgunaan data dapat diminimalkan. Pentingnya identifikasi dan inventarisasi data sebagai fondasi tata kelola data ditegaskan oleh ISO/IEC 27001 (2019) yang menyatakan bahwa pengenalan dan pencatatan aset informasi secara sistematis merupakan prasyarat utama untuk pengelolaan data yang efektif, aman, dan berkelanjutan.

## **2. Konversi Fisik ke Digital**

Konversi fisik ke digital merupakan proses inti dalam digitalisasi data yang bertujuan mengubah data primer berbentuk analog, seperti dokumen kertas, catatan manual, foto cetak, serta rekaman audio dan video, menjadi format digital yang dapat disimpan dan diproses menggunakan sistem komputer. Proses ini mencakup berbagai teknik, antara lain pemindaian dokumen, pengetikan ulang data, penggunaan teknologi pengenalan karakter optik, serta digitalisasi media audio-visual, yang semuanya harus dilakukan dengan menjaga kesesuaian antara data digital dan sumber aslinya agar tidak terjadi distorsi informasi. Pentingnya konversi fisik ke digital dalam konteks pengelolaan data modern ditegaskan oleh Borgman (2018) yang menyatakan bahwa transformasi data analog ke format digital merupakan langkah fundamental untuk memastikan data dapat diakses, dianalisis, dan dimanfaatkan secara berkelanjutan dalam ekosistem informasi digital.

Konversi fisik ke digital juga berperan penting dalam meningkatkan efisiensi, ketertelusuran, dan daya guna data primer dalam jangka panjang. Data yang telah dikonversi ke bentuk digital lebih mudah diintegrasikan dengan sistem informasi, dianalisis menggunakan perangkat lunak analitik, serta diamankan melalui mekanisme

rencana dan pengendalian akses yang lebih andal dibandingkan penyimpanan fisik. Dengan demikian, konversi fisik ke digital tidak hanya bersifat teknis, tetapi juga strategis karena menjadi fondasi bagi pengelolaan data yang sistematis, transparan, dan mendukung pengambilan keputusan berbasis data.

### **3. Strukturisasi Data**

Strukturisasi data merupakan proses penataan data hasil pengumpulan atau digitalisasi ke dalam format yang terorganisasi, konsisten, dan mudah dipahami sehingga data dapat digunakan secara efektif dalam pengolahan dan analisis. Proses ini mencakup penentuan atribut data, tipe data, relasi antar data, serta penetapan aturan pengisian yang bertujuan untuk mengurangi ambiguitas dan meningkatkan keterbacaan data. Pentingnya strukturisasi data dalam sistem pengelolaan informasi ditegaskan oleh Silberschatz *et al.* (2019) yang menjelaskan bahwa data yang disusun dalam struktur yang jelas dan terdefinisi dengan baik akan lebih mudah dikelola, diintegrasikan, serta diproses oleh sistem basis data secara efisien.

Strukturisasi data juga berperan penting dalam menjaga kualitas dan konsistensi data sepanjang siklus hidupnya. Data yang terstruktur memungkinkan penerapan validasi, pengendalian redundansi, serta pemeliharaan integritas data, sehingga kesalahan logis dan inkonsistensi antar entitas dapat diminimalkan. Dengan demikian, strukturisasi data tidak hanya mendukung kebutuhan teknis pengolahan data, tetapi juga menjadi fondasi strategis bagi analisis yang akurat, pengambilan keputusan berbasis data, serta pengembangan sistem informasi yang berkelanjutan.

### **4. Pemeriksaan dan Validasi**

Pemeriksaan dan validasi merupakan tahap krusial dalam pengelolaan data yang bertujuan untuk memastikan bahwa data yang dikumpulkan, dikonversi, dan disusun telah sesuai dengan sumber aslinya serta memenuhi standar kualitas yang ditetapkan. Proses ini melibatkan pengecekan kelengkapan, akurasi, konsistensi, dan kewajaran nilai data melalui teknik seperti pengecekan ulang terhadap sumber awal, pengujian logika antar variabel, serta deteksi kesalahan entri atau inkonsistensi format. Pentingnya pemeriksaan dan validasi data ditegaskan dalam ISO/IEC 25012 (2019) yang menyatakan bahwa

kualitas data hanya dapat dijamin apabila data tersebut melalui proses verifikasi dan validasi yang sistematis berdasarkan karakteristik kualitas yang terdefinisi dengan jelas.

Pemeriksaan dan validasi juga berperan dalam meningkatkan keandalan data sebagai dasar analisis dan pengambilan keputusan. Data yang telah divalidasi memberikan tingkat kepercayaan yang lebih tinggi karena risiko kesalahan, bias, dan ketidaksesuaian konteks dapat diminimalkan sejak tahap awal pengelolaan data. Dengan demikian, pemeriksaan dan validasi tidak hanya bersifat teknis, tetapi juga strategis karena menentukan sejauh mana data dapat digunakan secara bertanggung jawab, akurat, dan berkelanjutan dalam berbagai proses analitik maupun operasional.

## **5. Penyimpanan dan Pengelolaan**

Penyimpanan dan pengelolaan data merupakan tahap lanjutan yang sangat penting setelah data melalui proses pemeriksaan dan validasi, karena tahap ini menentukan bagaimana data disimpan secara aman, terorganisasi, dan siap digunakan kembali dalam jangka pendek maupun jangka panjang. Proses ini mencakup pemilihan media penyimpanan yang sesuai, penataan struktur folder atau basis data, penetapan mekanisme pencadangan, serta pengaturan hak akses agar data terlindungi dari kehilangan, kerusakan, maupun penggunaan yang tidak sah. Pentingnya penyimpanan dan pengelolaan data yang sistematis ditegaskan oleh Laudon dan Laudon (2020) yang menyatakan bahwa pengelolaan data yang efektif memerlukan sistem penyimpanan terstruktur dan kebijakan pengendalian yang jelas agar data dapat mendukung operasi organisasi dan pengambilan keputusan secara optimal.

Penyimpanan dan pengelolaan data juga memiliki dimensi strategis karena berkaitan langsung dengan keberlanjutan dan nilai guna data. Data yang dikelola dengan baik akan lebih mudah ditelusuri, diperbarui, serta diintegrasikan dengan sistem lain, sehingga mendukung konsistensi dan kontinuitas informasi dalam jangka panjang. Dengan demikian, penyimpanan dan pengelolaan data tidak hanya berfungsi sebagai tempat penampungan data, tetapi juga sebagai fondasi penting bagi tata kelola data yang andal, efisien, dan berorientasi pada pemanfaatan data secara maksimal.

## G. Pengolahan Awal Data Primer Menggunakan Python

Pengolahan awal data primer menggunakan Python merupakan tahapan krusial dalam siklus analisis data, khususnya ketika data diperoleh langsung dari sumber lapangan seperti survei, wawancara, observasi, atau sensor. Data primer umumnya masih mentah, tidak terstruktur dengan baik, serta berpotensi mengandung kesalahan pencatatan, nilai kosong, dan inkonsistensi format. Python banyak digunakan pada tahap ini karena memiliki pustaka yang kuat, fleksibel, dan efisien untuk mempersiapkan data agar siap dianalisis lebih lanjut. Secara rinci, pengolahan awal data primer dengan Python mencakup beberapa aktivitas utama berikut.

### 1. *Input dan Encoding Data Wawancara/Kuesioner*

*Input dan encoding data* wawancara/kuesioner dalam pengolahan awal data primer menggunakan Python mencakup proses penerimaan respons dari responden, pembersihan awal teks jawaban, serta transformasi jawaban tersebut menjadi format yang terstruktur dan dapat dianalisis. Dalam praktiknya, data wawancara sering kali memiliki bentuk teks naratif atau jawaban terbuka yang harus dibersihkan dari karakter tidak relevan dan distandarisasi sebelum dilakukan coding, sedangkan tanggapan kuesioner pilihan ganda perlu diberi kode numerik terlebih dahulu agar bisa diolah secara efisien dalam Python menggunakan pustaka seperti pandas untuk manipulasi data tabular. Proses pemberian kode atau encoding ini bertujuan mengonversi nilai kategorikal ke representasi numerik yang konsisten sehingga Python dapat melakukan operasi statistik, filtering, dan visualisasi data tanpa ambiguitas pada nilai string atau teks jawaban.

Encoding jawaban kuesioner bisa dilakukan melalui berbagai pendekatan di Python seperti label encoding untuk skala berurutan (misalnya “sangat setuju” ke 5 dan “tidak setuju” ke 1) atau *one-hot encoding* untuk variabel nominal, sementara teks naratif dari wawancara dapat melalui *preprocessing* teks dasar termasuk normalisasi huruf, penghapusan tanda baca serta transformasi ke token sebelum dianalisis lebih lanjut misalnya untuk tema atau sentimen. Menurut Koukaras dan Tjortjis (2025), penyiapan data mentah berupa pembersihan, normalisasi, dan encoding merupakan langkah penting dalam mempersiapkan data *input* agar siap dan optimal untuk analisis atau

pemodelan lebih lanjut, yang mencakup pemberian kode kategori dan penanganan nilai yang tidak konsisten dalam dataset primer.

## 2. Struktur Dataset (CSV, Excel, JSON)

Struktur dataset merupakan cara data disusun, disimpan, dan direpresentasikan dalam suatu berkas sehingga dapat dibaca, diproses, dan dianalisis dengan mudah. Dalam konteks pengolahan data primer dan sains data, pemahaman terhadap struktur dataset sangat penting karena setiap format penyimpanan data memiliki karakteristik, kelebihan, serta keterbatasan yang berbeda. Tiga format dataset yang paling umum digunakan adalah CSV, Excel, dan JSON. Perbedaan struktur pada masing-masing format memengaruhi cara data diakses, dibersihkan, dan diolah dalam proses analisis. Secara rinci, struktur dari masing-masing dataset dapat dijelaskan sebagai berikut.

### a. Struktur Dataset CSV (*Comma-Separated Values*)

Struktur dataset CSV (*Comma-Separated Values*) merupakan bentuk penyimpanan data tabular yang paling sederhana dan banyak digunakan dalam pengolahan data primer karena bersifat ringan, terbuka, dan mudah dipertukarkan antar sistem. Dalam struktur CSV, setiap baris merepresentasikan satu unit observasi, sedangkan setiap kolom menunjukkan variabel yang dipisahkan oleh tanda koma atau delimiter tertentu, dengan baris pertama umumnya berisi header sebagai penanda nama variabel agar data dapat dikenali secara semantik oleh perangkat lunak analisis. Menurut McKinney (2018), format CSV dirancang sebagai representasi teks sederhana untuk data terstruktur yang memungkinkan pembacaan cepat, interoperabilitas lintas platform, serta kemudahan integrasi dengan berbagai bahasa pemrograman dan alat analitik.

Struktur dataset CSV memiliki karakteristik datar tanpa dukungan relasi hierarkis maupun metadata kompleks, sehingga setiap nilai data disimpan secara eksplisit sebagai teks yang bergantung pada interpretasi sistem pemrosesan. Kondisi ini menjadikan CSV sangat efisien untuk menyimpan data numerik dan kategorikal dalam skala besar, namun sekaligus menuntut konsistensi penulisan data, terutama terkait tipe data, penanganan nilai kosong, dan penggunaan delimiter yang seragam. Oleh karena itu, dalam praktik pengolahan awal data primer,

pemahaman yang baik terhadap struktur CSV sangat penting agar proses pembacaan, pembersihan, dan analisis data dapat dilakukan secara akurat dan bebas dari kesalahan interpretasi struktural.

b. Struktur Dataset Excel

Struktur dataset Excel merupakan bentuk penyimpanan data tabular yang disusun dalam baris dan kolom pada lembar kerja (*worksheet*) di dalam satu file, di mana setiap sel dapat memuat nilai data, rumus, atau format tertentu yang memberikan konteks tambahan terhadap data tersebut. Keunikan struktur Excel terletak pada kemampuannya menampung banyak *worksheet* dalam satu file, sehingga memungkinkan pengelompokan data berdasarkan topik, periode, atau unit analisis tanpa harus memisahkannya ke dalam banyak berkas. Menurut McKinney (2022), format Excel mendukung penyimpanan data tabular yang kaya dengan metadata seperti formula dan format sel, sehingga sering digunakan dalam tahap awal pengumpulan dan eksplorasi data sebelum diproses lebih lanjut secara komputasional.

Struktur dataset Excel bersifat fleksibel karena memungkinkan kombinasi antara data mentah dan hasil perhitungan dalam satu lembar kerja, namun fleksibilitas ini juga berpotensi menimbulkan inkonsistensi struktur apabila tidak dikelola dengan standar yang jelas. Dataset Excel sering mengandung sel kosong, header bertingkat, atau penggabungan sel yang dapat menyulitkan proses pembacaan otomatis oleh bahasa pemrograman seperti Python jika tidak dilakukan pengolahan awal terlebih dahulu. Oleh karena itu, dalam konteks pengolahan data primer dan sains data, pemahaman terhadap struktur dataset Excel sangat penting agar data dapat dikonversi ke format yang lebih terstruktur dan konsisten untuk mendukung analisis yang akurat dan dapat direproduksi.

c. Struktur Dataset JSON (*JavaScript Object Notation*)

Struktur dataset JSON (*JavaScript Object Notation*) merupakan format penyimpanan data berbasis teks yang menggunakan pasangan kunci dan nilai (*key-value*) untuk merepresentasikan informasi secara terstruktur dan mudah dibaca oleh manusia maupun mesin. Berbeda dengan format tabular seperti CSV dan Excel, JSON memiliki struktur hierarkis yang

memungkinkan data disusun dalam bentuk objek dan array yang dapat bersarang, sehingga mampu merepresentasikan hubungan data yang kompleks dan tidak seragam dalam satu dokumen. Menurut McKinney (2022), JSON dirancang sebagai format pertukaran data yang ringan dan fleksibel, sangat sesuai untuk merepresentasikan data semi-terstruktur yang sering dihasilkan oleh aplikasi web, layanan API, dan sistem berbasis jaringan.

Dataset JSON memungkinkan setiap entri data memiliki atribut yang berbeda tanpa harus mengikuti skema kolom yang kaku, sehingga sangat efektif untuk menyimpan data dinamis atau berkembang. Namun, struktur bertingkat dan bersarang dalam JSON sering kali menyulitkan proses analisis langsung karena data perlu diratakan atau dinormalisasi terlebih dahulu agar dapat diproses dalam bentuk tabel. Oleh karena itu, dalam pengolahan awal data primer, pemahaman terhadap struktur dataset JSON menjadi penting agar proses transformasi ke format analitis dapat dilakukan secara akurat, konsisten, dan tetap menjaga makna hubungan antar data.

### 3. Contoh Script Python untuk Data Primer

Berikut contoh script Python untuk data primer mulai dari *input*–pembersihan–uji reliabilitas–analisis deskriptif–uji asumsi–regresi.

#### a. Struktur data primer

Misal file **data\_primer.csv** berisi kolom:

- 1) id, gender, usia, pendidikan
- 2) Item Likert (1–5): X1\_1, X1\_2, X1\_3, X1\_4 (konstruk X1) dan Y\_1, Y\_2, Y\_3 (konstruk Y)

#### b. Script lengkap

```
import pandas as pd
import numpy as np
from scipy import stats
import statsmodels.api as sm
from statsmodels.stats.outliers_influence import
variance_inflation_factor

# =====
# A. BACA DATA PRIMER
# =====
```

```

df = pd.read_csv("data_primer.csv") # ganti path jika perlu
print("Jumlah responden:", len(df))
print(df.head())

# =====
# B. PEMBERSIHAN DATA
# =====
# 1) Standarisasi nama kolom (opsional)
df.columns = [c.strip().replace(" ", "_") for c in df.columns]

# 2) Pastikan kolom Likert numerik
likert_cols = [c for c in df.columns if c.startswith(("X1_", "Y_"))]
for c in likert_cols:
    df[c] = pd.to_numeric(df[c], errors="coerce")

# 3) Tangani missing value: contoh isi dengan median per kolom
df[likert_cols] = df[likert_cols].apply(lambda s:
s.fillna(s.median()))

# 4) Filter nilai tidak wajar (contoh Likert harus 1-5)
for c in likert_cols:
    df = df[df[c].between(1, 5)]

print("Setelah cleaning, responden:", len(df))

# =====
# C. BUAT SKOR KONSTRUK (RATA-RATA ITEM)
# =====
X1_items = ["X1_1", "X1_2", "X1_3", "X1_4"]
Y_items = ["Y_1", "Y_2", "Y_3"]

df["X1"] = df[X1_items].mean(axis=1)
df["Y"] = df[Y_items].mean(axis=1)

# =====
# D. CRONBACH'S ALPHA (UJI RELIABILITAS)
# =====
def cronbach_alpha(items_df: pd.DataFrame) -> float:

```

```

items = items_df.dropna()
k = items.shape[1]
if k < 2:
    return np.nan
item_var = items.var(axis=0, ddof=1)
total_var = items.sum(axis=1).var(ddof=1)
return (k / (k - 1)) * (1 - (item_var.sum() / total_var))

alpha_X1 = cronbach_alpha(df[X1_items])
alpha_Y = cronbach_alpha(df[Y_items])

print(f'Cronbach Alpha X1: {alpha_X1:.3f}')
print(f'Cronbach Alpha Y : {alpha_Y:.3f}')

# =====
# E. DESKRIPTIF
# =====
print("\nDeskriptif X1 dan Y:")
print(df[["X1", "Y"]].describe())

# =====
# F. UJI NORMALITAS (Shapiro-Wilk; baik untuk n kecil-sedang)
# =====
sh_x1 = stats.shapiro(df["X1"])
sh_y = stats.shapiro(df["Y"])
print("\nShapiro X1:", sh_x1)
print("Shapiro Y :", sh_y)

# =====
# G. KORELASI (Pearson jika normal; Spearman jika tidak)
# =====
corr_pearson = stats.pearsonr(df["X1"], df["Y"])
corr_spearman = stats.spearmanr(df["X1"], df["Y"])
print("\nKorelasi Pearson (X1,Y):", corr_pearson)
print("Korelasi Spearman (X1,Y):", corr_spearman)

# =====

```

```

# H. REGRESI LINIER SEDERHANA:  $Y \sim X_1$ 
# =====
X = sm.add_constant(df["X1"]) # tambah intercept
model = sm.OLS(df["Y"], X).fit()
print("\nHasil Regresi:")
print(model.summary())

# =====
# I. DIAGNOSTIK SEDERHANA
# =====
# 1) Residual normality (opsional)
resid = model.resid
print("\nShapiro residual:", stats.shapiro(resid))

# 2) Outlier sederhana pakai z-score pada residual
df["resid_z"] = stats.zscore(resid)
outliers = df[np.abs(df["resid_z"]) > 3]
print("\nJumlah outlier ( $|z|>3$ ):", len(outliers))

# =====
# J. SIMPAN OUTPUT (opsional)
# =====
df.to_csv("data_primer_clean_scored.csv", index=False)
print("\nFile tersimpan: data_primer_clean_scored.csv")

```

### 3) Kalau Data Primer Kamu Bentuknya Wawancara (Teks)

Kalau yang kamu maksud data primer adalah transkrip wawancara, alurnya beda (cleaning teks, coding tema, word frequency, dll). Tapi script di atas khusus untuk data primer kuantitatif (kuesioner/observasi).

Kalau kamu kasih info singkat:

1. datanya kuesioner atau wawancara,
2. format file (Excel/CSV), dan
3. contoh nama kolom (misal: X1\_1...X1\_5, Y\_1...)

# BAB V

## METODE PENGUMPULAN DATA PRIMER BERBASIS TEKNOLOGI

---

Metode pengumpulan data primer berbasis teknologi merupakan pendekatan modern yang memanfaatkan perangkat digital, aplikasi, dan sistem informasi untuk memperoleh data langsung dari sumber pertama secara lebih efisien dan akurat. Pemanfaatan teknologi dalam proses ini memungkinkan pengumpulan data dilakukan secara real time, terstruktur, serta mampu menjangkau responden dalam skala yang lebih luas dengan tetap menjaga kualitas dan konsistensi data yang dihasilkan.

### A. Sensor dan Internet of Things (IoT)

Sensor dan *Internet of Things* (IoT) merupakan fondasi utama dalam sistem pengumpulan data primer berbasis teknologi modern. Sensor berfungsi sebagai alat pengindra yang menangkap fenomena fisik atau lingkungan tertentu dan mengubahnya menjadi data yang dapat diproses secara digital, sedangkan IoT berperan sebagai kerangka sistem yang menghubungkan sensor tersebut ke jaringan, sistem pemrosesan, dan pengguna secara otomatis dan berkelanjutan.

#### 1. Pengertian dan Peran Sensor

Sensor merupakan perangkat elektronik yang dirancang untuk mendeteksi perubahan atau fenomena fisik di lingkungan sekitarnya, seperti cahaya, suhu, tekanan, atau gerakan, kemudian mengubahnya menjadi sinyal listrik yang dapat diproses dan dianalisis oleh sistem komputer atau jaringan. Menurut Robert Sheldon (2022), sensor adalah perangkat yang mendeteksi dan merespons *input* dari lingkungan fisik, menghasilkan sinyal yang bisa dibaca atau ditransmisikan untuk diproses lebih lanjut sehingga memungkinkan interpretasi dan kontrol atas fenomena tersebut dalam berbagai aplikasi teknologi modern, termasuk

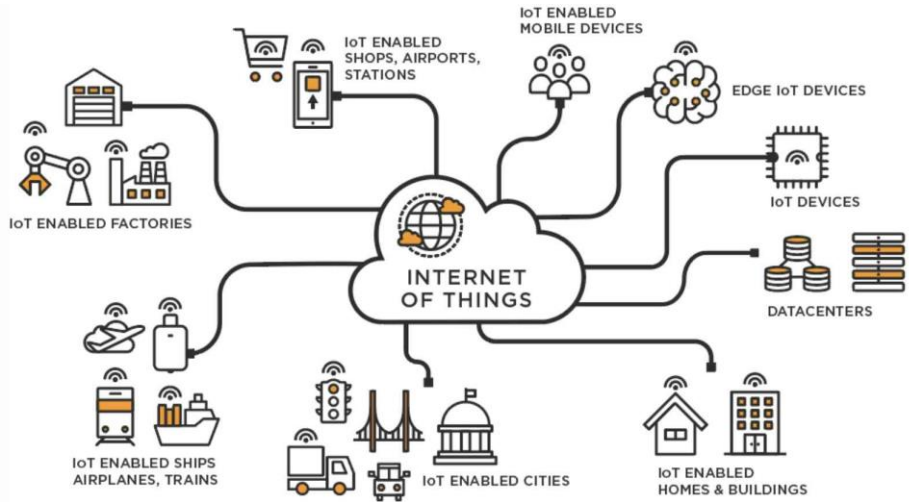
sistem otomatisasi dan IoT (*Internet of Things*). Peran sensor dalam teknologi sangat krusial karena sensor bertindak sebagai “mata dan telinga” dari sistem digital, menjembatani dunia fisik dengan dunia digital sehingga memungkinkan pemantauan, pengambilan keputusan, dan otomatisasi proses dengan akurasi tinggi dan respon *real-time*.

Sensor tidak hanya menghasilkan data mentah; juga memungkinkan sistem untuk merespons perubahan secara otomatis dan cepat tanpa campur tangan manusia langsung, seperti mengatur suhu ruangan, memicu alarm keselamatan, atau menyesuaikan operasi mesin industri. Dengan kemampuan ini, sensor menjadi komponen fundamental dalam pengembangan teknologi cerdas yang semakin otonom dan terhubung, menjadikan data yang dikumpulkan lebih bermakna dan berguna dalam berbagai konteks aplikasi teknologi saat ini.

## **2. Konsep Dasar *Internet of Things* (IoT)**

*Internet of Things* (IoT) adalah konsep teknologi di mana berbagai perangkat fisik yang dilengkapi dengan sensor, perangkat lunak, dan kemampuan komunikasi jaringan saling terhubung untuk mengumpulkan, bertukar, dan memproses data tanpa bergantung pada interaksi manusia secara langsung, sehingga mengubah objek biasa menjadi “perangkat pintar” yang mampu beroperasi secara otomatis dalam sistem yang lebih luas. Menurut definisi yang dimuat oleh TechTarget, IoT merupakan jaringan perangkat yang saling terhubung yang dapat mengirim dan menerima data melalui internet dengan kemampuan tersemat seperti sensor dan perangkat lunak untuk memungkinkan pengambilan keputusan serta otomatisasi proses (Gillis & Yasar, 2025). Dengan kerangka ini, IoT memungkinkan terjadinya aliran informasi secara *real-time* antara dunia fisik dan digital, sehingga aplikasi teknologi IoT dapat ditemukan dalam berbagai bidang seperti sistem otomasi rumah, pemantauan kesehatan, transportasi cerdas, serta manajemen energi yang lebih efisien dan responsif terhadap perubahan kondisi lingkungan atau kebutuhan pengguna.

Gambar 3. *Internet of Things*



Sumber: Dicoding

Pengembangan IoT mengintegrasikan elemen hardware, jaringan komunikasi, dan sistem pemrosesan data yang terpusat atau terdistribusi di *cloud*, yang kemudian dianalisis untuk menghasilkan wawasan operasional yang mendalam serta respon otomatis terhadap situasi tertentu. Perangkat IoT tidak hanya berfungsi sebagai pengirim data; juga dapat saling berinteraksi dan berkoordinasi untuk menyelesaikan tugas kompleks, seperti optimasi produksi industri, pengaturan lalu lintas kota, atau adaptasi layanan kesehatan berdasarkan data pasien terkini, sehingga menjadikan IoT sebagai infrastruktur penting dalam revolusi teknologi saat ini.

### 3. Komponen Utama Sistem Sensor dan IoT

Sistem IoT berbasis sensor umumnya terdiri dari beberapa komponen yang saling terintegrasi, yaitu:

#### a. Sensor

Sensor merupakan komponen utama dalam sistem sensor dan *Internet of Things* (IoT) karena berfungsi sebagai titik awal pengumpulan data primer yang menangkap kondisi fisik atau lingkungan secara langsung dan mengubahnya menjadi sinyal digital yang dapat diproses oleh sistem. Dalam konteks IoT, sensor memungkinkan objek fisik berperan aktif dalam sistem cerdas dengan menyediakan data real time yang menjadi dasar pemantauan, analisis, serta pengambilan keputusan otomatis dalam berbagai aplikasi seperti industri, lingkungan, dan

kesehatan. Menurut Sheldon (2022), sensor adalah perangkat yang mendeteksi dan merespons masukan dari lingkungan fisik dengan menghasilkan data yang dapat diproses dan ditransmisikan oleh sistem digital, sehingga keberadaan sensor menjadi elemen krusial yang menghubungkan dunia fisik dengan infrastruktur komputasi dan jaringan IoT modern.

b. Perangkat Pengolah Awal (Mikrokontroler atau *Edge Device*)

Perangkat pengolah awal berupa mikrokontroler atau *edge device* merupakan komponen utama dalam sistem sensor dan *Internet of Things* (IoT) yang berfungsi menerima, mengelola, dan memproses data mentah dari sensor sebelum dikirim ke sistem yang lebih tinggi seperti server atau *cloud*. Perangkat ini memungkinkan dilakukannya pemrosesan awal, seperti penyaringan data, pengurangan *noise*, dan pengambilan keputusan sederhana secara lokal sehingga dapat mengurangi beban jaringan serta latensi pengiriman data. Menurut Gillis dan Yasar (2023), *edge device* adalah perangkat yang melakukan pemrosesan data di dekat sumber data untuk meningkatkan efisiensi, kecepatan respons, dan keandalan sistem IoT, yang menegaskan perannya sebagai penghubung penting antara sensor dan infrastruktur komputasi jaringan.

c. Media Komunikasi

Media komunikasi merupakan komponen utama dalam sistem sensor dan *Internet of Things* (IoT) yang berfungsi sebagai sarana pengiriman data dari sensor atau perangkat pengolah awal menuju gateway, server, atau platform *cloud* secara andal dan berkelanjutan. Media ini mencakup berbagai teknologi jaringan seperti Wi-Fi, Bluetooth, jaringan seluler, maupun jaringan khusus IoT yang memungkinkan perangkat saling terhubung dan bertukar data meskipun berada pada lokasi yang berjauhan. Menurut Gillis dan Yasar (2023), media komunikasi dalam IoT mendukung penggunaan *communication protocols that enable devices to exchange data efficiently and reliably over a network*, sehingga perannya sangat krusial dalam memastikan data sensor dapat diterima tepat waktu dan digunakan secara efektif dalam proses pemantauan, analisis, serta pengambilan keputusan otomatis.

d. Server atau Platform *Cloud*

Server atau platform *cloud* merupakan komponen utama dalam sistem sensor dan *Internet of Things* (IoT) yang berfungsi sebagai pusat penyimpanan, pengolahan, dan analisis data yang dikirim dari sensor maupun perangkat pengolah awal, sehingga memungkinkan pengambilan keputusan berbasis data secara efisien dan real time. Platform *cloud* juga menyediakan kemampuan skalabilitas, keamanan, dan integrasi layanan analitik yang mendukung visualisasi data, pemantauan performa perangkat, serta implementasi algoritma *machine learning* untuk meningkatkan nilai informasi dari data yang dikumpulkan. Menurut Gillis dan Yasar (2024), *cloud platforms in IoT serve as centralized hubs for collecting, storing, and processing data from connected devices, enabling real-time analytics, system integration, and scalable data management*, yang menegaskan peran krusial *cloud* sebagai tulang punggung operasional dan manajemen data dalam ekosistem IoT.

## B. Mobile Data Collection Applications

*Mobile Data Collection Applications* adalah perangkat lunak atau aplikasi yang berjalan pada perangkat *mobile*, seperti smartphone atau tablet, yang digunakan untuk mengumpulkan, menyimpan, dan mengirimkan data secara digital dari lokasi lapangan atau pengguna langsung. Aplikasi ini menggantikan metode pengumpulan data tradisional berbasis kertas, sehingga memudahkan pencatatan, pengolahan, dan analisis data secara lebih cepat, akurat, dan terstruktur.

### 1. Fungsi dan Peran Utama

*Mobile Data Collection Applications* memiliki peran dan fungsi utama yang sangat penting dalam pengumpulan data modern karena memungkinkan data dikumpulkan langsung dari sumbernya dengan cepat, akurat, dan efisien serta mendukung proses pengambilan keputusan yang lebih baik. Aplikasi jenis ini tidak hanya menggantikan metode manual berbasis kertas, tetapi juga menyediakan mekanisme validasi otomatis, perekaman data *real-time*, hingga integrasi sensor perangkat *mobile* (GPS, kamera, dan sebagainya) sehingga informasi yang dikumpulkan lebih komprehensif dan siap untuk dianalisis secara

langsung dalam berbagai konteks pemantauan dan penelitian. Menurut Seifert (2018), *mobile data collection with smartphones offers unique and innovative opportunities for studying human beings and processes in real life and real time*, yang menunjukkan bahwa aplikasi pengumpulan data *mobile* membuka peluang baru bagi riset dan evaluasi lapangan yang responsif terhadap dinamika kondisi di lingkungan nyata dan waktunya sesuai kebutuhan penelitian atau operasional organisasi.

Dengan kemampuan ini, aplikasi pengumpulan data *mobile* secara signifikan mengurangi kesalahan *input* yang sering muncul pada pencatatan manual, mempercepat proses sinkronisasi data ke server atau *cloud*, serta memungkinkan tim lapangan dan pengambil keputusan pusat untuk memantau progres secara simultan tanpa perlu kembali ke lokasi pengumpulan data. Hal ini menjadikan aplikasi *mobile* sebagai alat strategis dalam pengelolaan data di berbagai bidang seperti survei sosial, evaluasi layanan publik, pemantauan kesehatan masyarakat, serta riset pasar, karena data yang dihasilkan dapat diproses dan dianalisis dengan cepat untuk menghasilkan *insight* operasional yang relevan dan dapat ditindaklanjuti secara efektif.

## **2. Keunggulan *Mobile Data Collection Applications***

Keunggulan *Mobile Data Collection Applications* terletak pada kemampuannya untuk mengumpulkan data secara *real-time* dan akurat langsung dari lapangan, sekaligus meminimalkan kesalahan manusia yang sering terjadi pada metode pencatatan manual, serta memungkinkan integrasi berbagai jenis *input* seperti teks, angka, gambar, audio, dan lokasi GPS sehingga data yang dikumpulkan menjadi lebih kaya dan mudah dianalisis. Menurut Mutka *et al.* (2020), *mobile data collection systems improve data accuracy, reduce time delays, and enhance operational efficiency in field research and monitoring activities*, yang menegaskan bahwa aplikasi *mobile* secara signifikan meningkatkan kualitas, kecepatan, dan efektivitas pengumpulan data dalam berbagai konteks penelitian dan operasional.

Keunggulan lain termasuk efisiensi biaya dan waktu karena mengurangi kebutuhan kertas, perjalanan, dan pekerjaan administratif, serta fleksibilitas penggunaan di berbagai kondisi lapangan termasuk area terpencil atau situasi dengan konektivitas terbatas melalui fitur penyimpanan offline yang dapat disinkronkan kemudian. Dengan kemampuan ini, *Mobile Data Collection Applications* menjadi alat

strategis bagi organisasi dan peneliti untuk memperoleh data yang valid, cepat, dan siap diproses untuk pengambilan keputusan atau evaluasi program secara lebih efektif dan terukur.

### 3. Fitur Umum yang Tersedia

Fitur umum yang tersedia pada *Mobile Data Collection Applications* mencakup formulir digital dengan logika percabangan, kemampuan pengumpulan data multimedia seperti foto, video, dan audio, serta integrasi GPS untuk pelacakan lokasi, yang semuanya memungkinkan pengumpulan data yang lebih komprehensif dan terstruktur langsung dari lapangan. Selain itu, aplikasi modern sering dilengkapi dengan sinkronisasi *cloud* otomatis, validasi data secara *real-time*, dan notifikasi atau alert, sehingga proses pengumpulan dan pemantauan data menjadi lebih cepat, akurat, dan efisien tanpa memerlukan interaksi manual yang berulang. Menurut Mutka *et al.* (2020), *modern mobile data collection platforms offer features such as multimedia capture, GPS integration, real-time validation, and cloud synchronization, enhancing both the accuracy and efficiency of field data collection*, yang menekankan pentingnya fitur-fitur tersebut dalam mendukung pengumpulan data yang efektif dan responsif.

Fitur-fitur ini tidak hanya meningkatkan kualitas data dan kecepatan pemrosesan, tetapi juga memberikan fleksibilitas bagi pengguna untuk bekerja dalam kondisi lapangan yang beragam, termasuk area terpencil dengan keterbatasan konektivitas, melalui kemampuan penyimpanan offline yang dapat disinkronkan kemudian. Dengan demikian, keberadaan fitur-fitur tersebut menjadikan *Mobile Data Collection Applications* sebagai alat yang sangat strategis dalam berbagai bidang, mulai dari penelitian sosial, evaluasi program, hingga pemantauan operasional organisasi, karena memungkinkan data dikumpulkan secara *real-time*, akurat, dan siap digunakan untuk analisis maupun pengambilan keputusan.

### 4. Contoh Aplikasi Populer

Contoh aplikasi populer yang digunakan untuk *Mobile Data Collection Applications* mencakup platform-platform seperti Open Data Kit (ODK), KoBoToolbox, SurveyCTO Collect, dan CommCare, yang masing-masing menyediakan antarmuka pengumpulan data lapangan melalui perangkat *mobile* dengan kemampuan offline, *input* multimedia,

dan sinkronisasi ke server atau *cloud* untuk analisis lebih lanjut. Aplikasi-aplikasi ini sering dipilih oleh peneliti, organisasi kemanusiaan, maupun tim operasional karena fleksibilitasnya dalam menyusun formulir digital, menangkap data berbasis teks, foto, GPS, serta fitur logika survei yang kompleks tanpa ketergantungan penuh pada koneksi internet, sehingga mendukung pengumpulan data primer dalam berbagai konteks penelitian dan monitoring yang dinamis. Dalam studi yang menilai berbagai solusi pengumpulan data *mobile*, disebutkan bahwa koBoToolbox dan ODK termasuk di antara solusi open-source *mobile* data collection yang lengkap dan banyak digunakan oleh organisasi di seluruh dunia, menunjukkan popularitas dan relevansi aplikasi-aplikasi tersebut dalam praktik pengumpulan data digital modern (Loola Bokonda, OUazanni-Touhami & Souissi, 2020)

### C. Digital Tracking dan Log Data

*Digital tracking* dan *log data* merupakan dua komponen penting dalam pengumpulan, pemantauan, dan analisis aktivitas digital. Digital tracking mengacu pada proses pengamatan dan pencatatan aktivitas pengguna, perangkat, atau sistem dalam lingkungan digital, termasuk klik pada *website*, interaksi dengan aplikasi, penggunaan sensor IoT, hingga perilaku pengguna dalam platform digital. Sementara itu, log data adalah catatan terstruktur atau semi-terstruktur yang dihasilkan oleh sistem komputer, aplikasi, server, atau perangkat jaringan, yang merekam peristiwa (*events*) beserta informasi pendukung seperti waktu kejadian, identitas pengguna, status, dan metadata tambahan.

#### 1. Jenis *Digital Tracking* dan *Log Data*

##### a. *Web / Clickstream Logs*

*Web/Clickstream logs* adalah rekaman terperinci dari jejak navigasi pengguna di dunia daring yang mencatat urutan klik, halaman yang dikunjungi, serta interaksi lainnya dalam web server log untuk memahami bagaimana pengguna menelusuri dan berinteraksi dengan situs web menurut perilaku penggunaan yang sebenarnya dalam konteks browsing. Menurut penelitian *You, the Web and Your Device: Longitudinal Characterization of Browsing Habits* (2018), clickstream direpresentasikan sebagai jaringan jalur yang diikuti oleh pengguna di web, di mana setiap

simpul mewakili halaman web dan setiap tepi menunjukkan perjalanan pengguna dari satu halaman ke halaman lain, sehingga clickstream dapat digunakan untuk menggambarkan pola penggunaan, tren perangkat yang digunakan, dan perbedaan perilaku browsing dari waktu ke waktu.

b. *Application Logs*

*Application logs* adalah rekaman digital yang dihasilkan oleh perangkat lunak untuk mencatat peristiwa, status internal, kesalahan, serta interaksi fungsi selama runtime aplikasi dan digunakan untuk memahami bagaimana suatu program berjalan serta mendukung pemecahan masalah dan optimasi sistem, di mana data log ini berguna dalam monitoring dan diagnosis perilaku perangkat lunak “*because the data provided by application logs and runtime environment are essential to detect and diagnose undesired behavior and improve system reliability*” (Cândido, Aniche & van Deursen, 2021). Dengan demikian, log aplikasi tidak hanya berfungsi sebagai catatan historis aktivitas program tetapi juga menjadi dasar analisis performa, identifikasi bug, pelacakan transaksi penting, serta pengambilan keputusan teknis dan operasional berdasarkan data runtime yang terekam.

c. *Server / Access Logs*

*Server/Access logs* adalah file catatan yang dihasilkan secara otomatis oleh web server setiap kali server menerima permintaan (*request*) dari klien, yang mencatat secara rinci semua interaksi dengan server termasuk alamat IP pengguna, waktu permintaan, metode HTTP yang digunakan, URL yang diminta, status respon, dan atribut lain yang berkaitan dengan komunikasi tersebut “*Web server access logs ... provide a rich source of data for system monitoring, security assessment, and user behavior analysis*” (Chodak *et al.*, 2020). Dengan adanya informasi tersebut, file log akses server tidak hanya berfungsi sebagai sejarah teknis dari semua permintaan yang diproses, tetapi juga menjadi basis penting untuk melakukan pemantauan kinerja server, mendeteksi anomali atau serangan siber, memahami pola lalu lintas pengguna, dan mendukung optimasi layanan digital dengan cara menganalisis volume lalu lintas, frekuensi permintaan, pola navigasi, serta karakteristik klien yang berinteraksi dengan aplikasi atau situs web secara *real-time*.

d. *Event Tracking*

*Event tracking* adalah proses pemantauan dan pencatatan interaksi spesifik yang dilakukan pengguna dengan sebuah aplikasi atau situs web di mana setiap tindakan seperti klik tombol, pengisian formulir, pemutaran video hingga transaksi diperlakukan sebagai peristiwa (*event*) yang terekam untuk dianalisis sehingga perilaku pengguna dapat dipahami dengan lebih detail “*Event tracking gives businesses a competitive advantage by enabling a deeper understanding of customer movements across marketing channels*” (Patterson, 2024) . Dengan mengumpulkan dan menganalisis data event tersebut dalam urutan kronologis berdasarkan atribut seperti jenis aksi, waktu, perangkat, dan konteks, organisasi bisa mengukur keterlibatan, mengidentifikasi hambatan dalam pengalaman pengguna, memetakan konversi dari awareness hingga pembelian, serta mengoptimalkan fitur dan strategi produk atau pemasaran secara berbasis data.

e. *Telemetry / Metrics*

*Telemetry/metrics* dalam konteks sistem digital adalah kumpulan ukuran numerik *real-time* yang mencerminkan kondisi, kinerja, dan kesehatan suatu sistem atau aplikasi, seperti waktu respons, penggunaan CPU, throughput permintaan, serta tingkat error, yang dikumpulkan secara otomatis untuk memantau dan menganalisis operasi teknis secara kontinu “*Metrics are numerical measurements that track the performance and health of different components, such as response times, error rates, and resource usage*” (Zhou et al., 2024) dan menyediakan landasan objektif bagi tim operasi atau pengembang dalam mengidentifikasi tren, mendeteksi anomali, serta merencanakan kapasitas dan pengoptimalan sistem. Dengan menggabungkan telemetry yang secara otomatis menangkap data dari sumber terdistribusi dengan alat analitik yang menyajikan informasi tersebut dalam bentuk yang mudah dipahami, organisasi dapat meningkatkan visibilitas terhadap perilaku sistem, mempercepat deteksi masalah, dan mendukung keputusan berbasis data untuk pengelolaan performa serta peningkatan kualitas layanan secara efisien.

f. *Security / Audit Logs*

*Security/Audit logs* adalah catatan terperinci yang merekam seluruh aktivitas terkait keamanan dan akses dalam sistem digital, termasuk autentikasi pengguna, perubahan hak akses, percobaan login gagal, aktivitas file sensitif, dan konfigurasi sistem, yang berfungsi sebagai alat utama untuk mendeteksi ancaman, meninjau kepatuhan terhadap regulasi, serta mendukung investigasi insiden keamanan “*Audit logs serve as a critical mechanism for detecting, investigating, and preventing security incidents by providing a detailed chronological record of system activities*” (Sharma & Gupta, 2022). Dengan memanfaatkan informasi ini secara sistematis, organisasi dapat mengidentifikasi pola serangan, melakukan *forensic analysis*, memastikan kepatuhan terhadap kebijakan keamanan dan regulasi seperti GDPR atau ISO 27001, serta memperkuat pertahanan sistem dengan melakukan koreksi proaktif terhadap celah atau penyimpangan yang terdeteksi dari aktivitas log.

g. *IoT / Sensor Logs*

*IoT/Sensor logs* adalah kumpulan data yang ditangkap secara terus-menerus oleh jaringan sensor yang tertanam dalam perangkat *Internet of Things* (IoT), di mana setiap entri log mencerminkan pembacaan sensor waktu-nyata seperti suhu, kelembapan, tekanan, gerakan atau parameter lingkungan lain yang kemudian direkam dan disimpan untuk tujuan monitoring, analisis tren, dan pengambilan keputusan otomatis dalam sistem yang terhubung. Dalam konteks aplikasi IoT yang kompleks, sensor logs menyediakan “*real-time data streams generated by interconnected sensors that reflect the current state or condition of physical systems, enabling advanced monitoring and analytics for operational insights*” (Qiu, 2025), sehingga informasi tersebut sangat penting untuk memodelkan kondisi fisik secara akurat, mendeteksi perubahan atau anomali sejak dini, serta mendukung sistem otomatisasi, pemeliharaan prediktif, dan optimasi proses berbasis data yang terus berjalan.

## 2. Struktur Log Data

Struktur log data adalah kerangka terorganisir yang mendefinisikan bagaimana informasi dari berbagai sumber sistem digital

direkam, disimpan, dan dikategorikan, mencakup elemen seperti timestamp, tipe event, identitas entitas, konteks atau payload, status hasil, serta metadata sumber, sehingga memudahkan pemrosesan, analisis, dan korelasi antar peristiwa yang terjadi secara bersamaan atau berurutan. Menurut penelitian oleh Al-Fedaghi (2019), struktur log yang baik memungkinkan konsistensi data lintas sistem, mendukung validasi, deduplikasi, dan transformasi informasi menjadi *insight* yang dapat diandalkan, serta memfasilitasi audit dan monitoring sistem secara efektif.

Implementasi struktur log biasanya menggunakan format terstruktur seperti JSON, XML, atau CSV untuk menyimpan data log agar mudah di-parse dan dianalisis, atau format kolumnar seperti Parquet untuk penyimpanan di data lake skala besar, sementara untuk telemetry atau metrik waktu-nyata, *time-series database* digunakan agar query cepat terhadap data kontinu dapat dilakukan. Struktur ini membantu sistem untuk mengidentifikasi event unik, mengelompokkan peristiwa berdasarkan sesi atau user, dan mengekstrak pola dari volume log yang besar, sehingga manajemen dan analitik data menjadi lebih efisien.

## D. Drones dan Remote Sensing

*Drones*, atau UAV (*Unmanned Aerial Vehicles*), adalah pesawat tanpa awak yang dapat dikendalikan secara manual atau otomatis untuk melakukan pengamatan dan pengumpulan data dari udara. *Remote sensing* adalah teknik penginderaan jarak jauh yang menggunakan sensor untuk mengamati objek atau permukaan bumi tanpa kontak langsung, baik melalui cahaya, gelombang elektromagnetik, atau sensor khusus seperti LiDAR dan radar.

### 1. Komponen Utama Sistem *Drone* untuk *Remote Sensing*

Komponen utama sistem *drone* untuk *remote sensing* melibatkan integrasi platform udara tanpa awak dengan berbagai sensor, perangkat navigasi, dan perangkat lunak pemrosesan data sehingga memungkinkan pengambilan informasi dari permukaan bumi secara akurat dan efisien. Platform *drone* seperti multi-rotor dan *fixed-wing* dipilih berdasarkan kebutuhan misi, sementara sensor optik, multispektral, LiDAR, dan termal memungkinkan pengukuran spektrum elektromagnetik yang

berbeda untuk aplikasi seperti pemetaan topografi, monitoring vegetasi, dan inspeksi infrastruktur (Colomina & Molina, 2018).

Sistem navigasi dan penentuan posisi menggunakan GNSS dan IMU berperan penting dalam georeferensi data yang dikumpulkan oleh *drone*, sehingga setiap titik data dapat ditentukan koordinatnya dengan presisi tinggi. *Ground Control Points* (GCP) atau teknologi RTK/PPK juga sering digunakan untuk meningkatkan akurasi spasial orthomosaic dan model 3D, sehingga hasil penginderaan jauh dapat digunakan untuk analisis presisi tinggi dan perencanaan berbasis data.

## **2. Alur Kerja Penginderaan Jauh dengan *Drone***

Alur kerja penginderaan jauh dengan *drone* dimulai dari perencanaan misi yang matang, termasuk penentuan area cakupan, ketinggian terbang, jalur penerbangan, dan overlap foto, sehingga pengumpulan data dapat dilakukan secara sistematis dan konsisten. Proses ini juga mencakup kalibrasi sensor dan pengaturan parameter kamera atau LiDAR sebelum penerbangan untuk memastikan kualitas data yang tinggi dan dapat diandalkan (Nex & Remondino, 2018).

Selama penerbangan, *drone* mengumpulkan data citra optik, multispektral, termal, atau LiDAR, yang kemudian melalui pra-pemrosesan untuk koreksi radiometrik, distorsi lensa, dan georeferensi kasar sehingga data mentah siap untuk analisis lebih lanjut. Penggunaan GCP atau sistem RTK/PPK membantu meningkatkan akurasi spasial, sementara metadata penerbangan dicatat untuk keperluan audit dan validasi kualitas data.

Tahap pemrosesan berikutnya melibatkan fotogrametri atau pemrosesan LiDAR untuk menghasilkan orthomosaic, *point cloud*, DSM/DTM, serta indeks vegetasi atau klasifikasi tutupan lahan, yang kemudian dievaluasi melalui validasi lapangan agar hasil analisis dapat dipercaya. Data akhir yang diperoleh kemudian dapat digunakan untuk berbagai aplikasi praktis, mulai dari pertanian presisi hingga pemetaan dan monitoring lingkungan, dengan efisiensi waktu dan biaya yang tinggi dibanding metode konvensional.

## **3. Jenis Keluaran Data dan Aplikasinya**

Jenis keluaran data dari penginderaan jauh menggunakan *drone* mencakup orthomosaic, model elevasi digital (DSM/DTM), *point cloud* 3D, serta indeks vegetasi seperti NDVI, yang memungkinkan analisis

spasial dan spektral secara detail untuk berbagai kebutuhan. Data ini dapat digunakan untuk pemetaan topografi, monitoring kesehatan tanaman, inventarisasi vegetasi, dan perencanaan wilayah secara akurat karena resolusi tinggi dan fleksibilitas pengambilan data yang ditawarkan oleh platform UAV (Colomina & Molina, 2018).

Peta klasifikasi tutupan lahan dan peta termal yang dihasilkan memungkinkan pemantauan perubahan lingkungan, identifikasi hotspot, serta deteksi kebocoran atau stres tanaman secara cepat dan efisien. Kombinasi *output* ini mendukung aplikasi praktis dalam pertanian presisi, kehutanan, manajemen bencana, dan inspeksi infrastruktur, sehingga pengambilan keputusan dapat dilakukan berbasis data *real-time* yang valid dan akurat.

Keluaran point *cloud* 3D dan model elevasi juga sangat berguna dalam volumetri pertambangan, perencanaan konstruksi, serta pemodelan struktur bangunan atau kanopi hutan, memberikan informasi topografi dan geometrik yang tidak dapat diperoleh dengan metode konvensional. Penggunaan data ini semakin diperkuat dengan pemrosesan fotogrametri atau LiDAR yang memungkinkan visualisasi, analisis statistik, dan simulasi untuk perencanaan strategis dan penelitian ilmiah.

#### **4. Keunggulan Penggunaan *Drone* untuk *Remote Sensing***

Keunggulan penggunaan *drone* untuk *remote sensing* terletak pada kemampuan menghasilkan data dengan resolusi spasial dan temporal yang tinggi, memungkinkan pemetaan detail area kecil hingga menengah secara cepat dan akurat. Selain itu, fleksibilitas platform UAV dalam membawa berbagai jenis sensor seperti optik, multispektral, termal, dan LiDAR memungkinkan pengumpulan data multidimensional untuk aplikasi yang beragam, mulai dari pertanian presisi hingga pemantauan lingkungan dan infrastruktur (Zhang & Kovacs, 2018).

*Drone* juga menawarkan efisiensi biaya dan waktu dibanding metode konvensional, karena mampu mengakses area yang sulit dijangkau, melakukan misi berulang dalam interval waktu yang singkat, dan menyediakan data *near-real-time* untuk pengambilan keputusan yang cepat. Keunggulan ini membuat *drone* menjadi alat strategis dalam penginderaan jauh modern, mendukung perencanaan, monitoring, dan evaluasi berbasis data yang lebih efektif dan akurat.

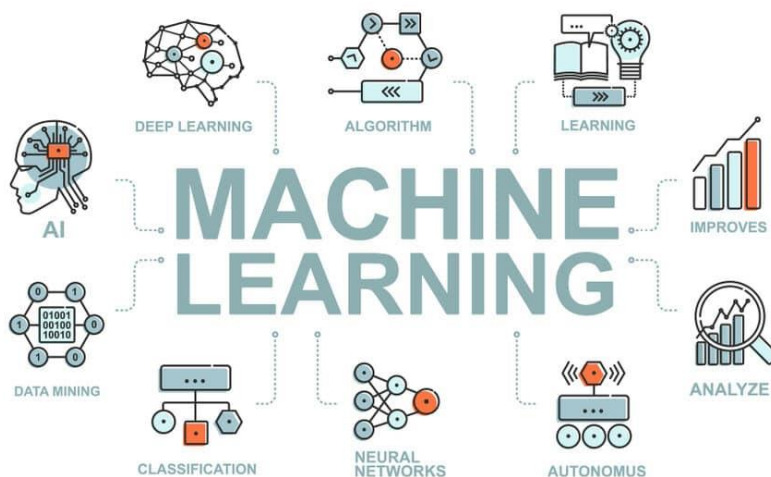
## E. Machine Learning dan Computer Vision

*Machine Learning* (ML) adalah cabang dari kecerdasan buatan yang memungkinkan sistem belajar dari data dan meningkatkan kinerjanya secara otomatis tanpa diprogram secara eksplisit. *Computer Vision* (CV) adalah bidang yang menggunakan algoritma untuk memungkinkan komputer memahami, menganalisis, dan menafsirkan informasi visual dari gambar atau video. Kombinasi ML dan CV memungkinkan pengolahan data visual secara cerdas untuk berbagai aplikasi seperti pengenalan objek, segmentasi citra, dan deteksi pola.

### 1. Jenis dan Metode *Machine Learning* yang Digunakan dalam *Computer Vision*

Jenis dan metode *machine learning* yang digunakan dalam *computer vision* mencakup *supervised learning*, di mana model dilatih menggunakan dataset berlabel untuk mengenali pola dan klasifikasi objek dalam citra, serta *unsupervised learning* yang memungkinkan sistem menemukan struktur atau kluster dalam data visual tanpa label, seperti segmentasi citra atau deteksi anomali. Pendekatan *deep learning*, khususnya *Convolutional Neural Networks* (CNN), telah menjadi metode dominan karena kemampuannya mengekstraksi fitur kompleks dari citra dan meningkatkan akurasi pengenalan objek secara signifikan (Goodfellow, Bengio, & Courville, 2018).

Gambar 4. *Machine Learning*



Sumber: Codepolitan

*Reinforcement learning* juga diterapkan dalam beberapa aplikasi *computer vision*, terutama untuk sistem yang memerlukan pengambilan keputusan berulang melalui interaksi dengan lingkungan visual, misalnya navigasi robotik atau optimasi jalur pengambilan citra. Kombinasi berbagai metode ini memungkinkan *computer vision* untuk menangani berbagai tugas mulai dari deteksi dan klasifikasi objek, segmentasi citra, hingga analisis video secara otomatis dengan ketelitian yang tinggi dan kemampuan adaptasi terhadap data baru.

## **2. Komponen Utama Sistem *Computer Vision***

Komponen utama sistem *computer vision* mencakup perangkat *input* seperti kamera RGB, kamera kedalaman, atau sensor multispektral yang menangkap data visual dari lingkungan, pra-pemrosesan citra untuk normalisasi, penghapusan *noise*, dan augmentasi data, serta ekstraksi fitur yang menyorot pola, tepi, tekstur, dan bentuk dari citra untuk dianalisis lebih lanjut oleh model pembelajaran mesin. Kombinasi komponen ini dengan algoritma *machine learning* atau *deep learning* memungkinkan sistem untuk melakukan klasifikasi, deteksi, segmentasi, dan pengambilan keputusan otomatis berdasarkan informasi visual yang diperoleh (Szeliski, 2019).

*Post-processing* dan modul analisis mendukung transformasi *output* model menjadi hasil yang dapat diinterpretasikan, seperti peta segmentasi, bounding box objek, atau anotasi video, sehingga aplikasi *computer vision* dapat digunakan dalam berbagai bidang mulai dari keamanan, otomasi industri, hingga pertanian presisi. Integrasi antara sensor, pemrosesan citra, algoritma pembelajaran, dan interpretasi hasil inilah yang menjadi fondasi kemampuan *computer vision* dalam memahami dan menafsirkan dunia visual secara efisien dan akurat.

## **3. Aplikasi *Machine Learning* dan *Computer Vision***

Aplikasi *machine learning* dan *computer vision* mencakup pengenalan wajah dan identitas, deteksi serta klasifikasi objek dalam citra dan video, segmentasi citra, serta analisis perilaku manusia atau objek secara otomatis, yang sangat berguna dalam bidang keamanan, otomasi industri, dan monitoring lingkungan. Teknologi ini juga diterapkan dalam pertanian presisi untuk mendeteksi kesehatan tanaman, dalam medis untuk mendiagnosis penyakit melalui citra radiologi, dan dalam otomotif untuk kendaraan otonom, sehingga memungkinkan

pengambilan keputusan berbasis data secara cepat dan akurat (Li *et al.*, 2020).

*Computer vision* dengan *machine learning* digunakan dalam peningkatan kualitas citra, *super-resolution*, pemodelan 3D, *augmented reality* (AR), dan *virtual reality* (VR), memberikan pengalaman visual yang lebih realistis dan interaktif sekaligus mendukung analisis spasial dan temporal data visual dalam skala besar. Integrasi berbagai metode ini memungkinkan sistem untuk belajar dari data baru, menyesuaikan model, dan meningkatkan ketelitian prediksi, sehingga memperluas jangkauan aplikasi praktis di berbagai sektor industri dan penelitian ilmiah.

#### 4. Keunggulan Menggunakan *Machine Learning* dalam *Computer Vision*

Keunggulan menggunakan *machine learning* dalam *computer vision* terletak pada kemampuannya untuk mengotomatisasi analisis citra dan video dalam skala besar, sehingga mengurangi ketergantungan pada intervensi manusia sekaligus meningkatkan efisiensi dan kecepatan pengolahan data visual. Selain itu, algoritma *deep learning* mampu mengenali pola kompleks dan fitur tersembunyi yang sulit dideteksi oleh metode tradisional, memungkinkan akurasi tinggi dalam tugas seperti klasifikasi objek, segmentasi citra, dan deteksi anomali (Shin *et al.*, 2019).

*Machine learning* juga memberikan fleksibilitas dan kemampuan adaptasi, karena model dapat terus belajar dari data baru dan menyesuaikan prediksi tanpa perlu pemrograman ulang, sehingga meningkatkan performa sistem secara berkelanjutan. Keunggulan ini menjadikan *computer vision* berbasis *machine learning* sangat efektif dalam aplikasi yang membutuhkan pengolahan data *real-time*, presisi tinggi, dan kemampuan generalisasi terhadap kondisi lingkungan yang berbeda, mulai dari medis hingga otomasi industri dan kendaraan otonom.

## F. Isu Teknis dan Etika Data Primer Digital

Data primer digital adalah data yang dikumpulkan secara langsung dari sumber atau responden melalui metode digital seperti survei *online*, sensor, aplikasi *mobile*, atau platform interaktif. Dalam

pengumpulan dan penggunaan data ini muncul berbagai isu teknis dan etika yang harus diperhatikan untuk memastikan kualitas, keamanan, dan integritas data serta perlindungan hak privasi responden.

## 1. Isu Teknis dalam Data Primer Digital

Isu teknis dalam pengumpulan data primer digital mencakup berbagai tantangan yang dapat memengaruhi kualitas, keandalan, dan akurasi informasi yang dikumpulkan, termasuk kesalahan *input*, data hilang (*missing values*), dan duplikasi yang dapat menyebabkan bias atau interpretasi yang salah dalam analisis. Selain itu, keterbatasan perangkat keras, kegagalan aplikasi, atau gangguan jaringan juga dapat menghambat proses pengumpulan data secara *real-time*, sehingga dibutuhkan sistem yang handal dan redundansi untuk memastikan kelangsungan operasional. Keamanan data juga menjadi aspek krusial karena data digital rentan terhadap serangan siber, hacking, atau malware, sehingga penerapan protokol enkripsi, firewall, dan backup rutin sangat penting untuk menjaga integritas dan kerahasiaan informasi (Kitchin, 2018).

Validitas dan integritas data merupakan isu teknis penting karena data yang dikumpulkan melalui perangkat digital harus akurat dan dapat dipercaya untuk tujuan analisis ilmiah atau pengambilan keputusan strategis. Kesalahan sensor, kesalahan *input* responden, atau kegagalan sistem otomatis dapat menimbulkan ketidakakuratan, sehingga proses kalibrasi, validasi lapangan, dan pembersihan data menjadi langkah penting untuk memastikan kualitas dataset. Integrasi sistem pengumpulan data yang memadai dengan monitoring kontinu memungkinkan peneliti mendeteksi dan memperbaiki kesalahan secara cepat sebelum data digunakan untuk analisis lebih lanjut.

Isu teknis lain yang signifikan adalah penyimpanan dan manajemen data dalam skala besar, karena data primer digital seringkali memiliki volume yang tinggi dan memerlukan kapasitas penyimpanan serta kemampuan pemrosesan yang mumpuni. Infrastruktur *cloud* atau server lokal harus dirancang dengan redundansi dan skalabilitas agar mampu menampung dan mengolah data secara efisien, sambil tetap menjaga keamanan dan aksesibilitas informasi bagi pengguna yang sah. Pemanfaatan standar metadata, backup rutin, dan kontrol akses yang ketat menjadi praktik terbaik untuk mengatasi isu teknis yang muncul dalam pengumpulan data primer digital.

## 2. Isu Etika dalam Data Primer Digital

Isu etika dalam data primer digital berkaitan dengan perlindungan privasi dan hak individu, di mana pengumpulan, penyimpanan, dan penggunaan data harus dilakukan dengan persetujuan eksplisit dari responden serta transparansi penuh mengenai tujuan dan pihak yang memiliki akses terhadap data tersebut. Penggunaan data tanpa persetujuan, atau untuk tujuan yang merugikan seperti diskriminasi, manipulasi, atau pelanggaran hak individu, dapat menimbulkan konsekuensi hukum dan merusak kepercayaan publik terhadap penelitian atau organisasi. Oleh karena itu, penerapan prinsip etika seperti *informed consent*, anonimisasi data, dan kebijakan akses terbatas menjadi sangat penting untuk menjaga integritas dan tanggung jawab sosial dalam pengolahan data digital (Floridi *et al.*, 2018).

Isu etika muncul dalam konteks keadilan dan transparansi algoritma yang digunakan untuk menganalisis data primer digital, karena algoritma yang bias atau tidak diaudit dapat menghasilkan kesimpulan yang diskriminatif atau menyesatkan. Peneliti dan pengelola data harus memastikan bahwa proses pengolahan data dapat dipertanggungjawabkan, dapat diaudit, dan meminimalkan risiko bias sistematis agar hasil penelitian tetap sah dan adil. Pendekatan ini juga mencakup kepatuhan terhadap regulasi perlindungan data, seperti GDPR atau peraturan lokal lainnya, yang menetapkan standar legal dan etis untuk pengumpulan, penyimpanan, dan penggunaan data individu.

Isu etika juga berkaitan dengan keamanan data dan risiko kebocoran informasi pribadi yang dapat membahayakan responden, sehingga organisasi harus menerapkan protokol keamanan yang kuat, enkripsi, dan prosedur backup untuk melindungi data sensitif. Pemantauan berkelanjutan terhadap akses dan penggunaan data digital membantu mencegah penyalahgunaan serta memastikan tanggung jawab etis terhadap semua pihak yang terlibat. Dengan memadukan prinsip etika, regulasi hukum, dan praktik keamanan teknis, pengelolaan data primer digital dapat dijalankan secara bertanggung jawab, aman, dan dapat dipercaya.

## G. Pengambilan dan Akuisisi Data Primer Digital Menggunakan Python

Pengambilan dan akuisisi data primer digital merupakan proses mengumpulkan data asli yang belum pernah ada sebelumnya dari berbagai sumber digital menggunakan Python sebagai alat otomatisasi. Proses ini penting dalam penelitian, analisis bisnis, maupun sains data karena memastikan data yang dikumpulkan sesuai kebutuhan penelitian dan memiliki kualitas yang tinggi. Python dipilih karena fleksibilitasnya, ketersediaan pustaka (*library*) yang lengkap, serta kemudahan integrasi dengan berbagai format data dan sumber digital.

### 1. Membaca Data Sensor dan Log

Membaca data sensor dan log adalah proses penting dalam sistem digital modern yang melibatkan pengambilan, pengolahan serta interpretasi data yang dihasilkan oleh perangkat sensor dan catatan operasi (*log*) untuk mendapatkan informasi yang bermakna; data sensor merekam respons perangkat terhadap stimulus fisik dari lingkungan dan ketika digabung dengan log operasi sistem, memberikan gambaran komprehensif tentang performa serta sejarah kejadian sistem tersebut, yang sangat berguna untuk pemantauan dan pengambilan keputusan. Menurut Salmon (2025), data logging adalah proses mengumpulkan dan menganalisis data dari sensor untuk melakukan analisis data, dimana data yang dicatat secara terus-menerus mencakup pengukuran seperti suhu, tekanan atau arus serta metadata waktu yang memungkinkan analisis tren dan pola seiring waktu untuk perbaikan proses atau diagnosa masalah. Dalam konteks aplikasi teknis dan ilmiah, pembacaan yang efektif terhadap data sensor dan log memerlukan penanganan format digital yang tepat, validasi nilai, serta penyimpanan yang terstruktur sehingga dapat digunakan untuk analisis lanjutan, visualisasi, dan sistem kontrol otomatis yang responsif terhadap perubahan kondisi nyata

### 2. Akuisisi Data *Streaming* Sederhana

Akuisisi data *streaming* sederhana adalah proses pengumpulan data secara *real-time* dari sumber yang terus-menerus menghasilkan informasi, seperti sensor, perangkat IoT, atau aliran media digital, dengan tujuan untuk memperoleh data yang langsung dapat dianalisis atau disimpan untuk pemrosesan lebih lanjut, sehingga memungkinkan

pemantauan dan pengambilan keputusan yang cepat. Data *streaming* ini sering dilakukan menggunakan metode yang ringan dan efisien, seperti buffer sederhana atau pustaka Python untuk membaca data secara terus-menerus, sehingga sistem dapat menangani aliran data yang masuk secara konstan tanpa kehilangan informasi penting. Menurut Chen dan Zhang (2020), *real-time data acquisition frameworks provide scalable and low-latency solutions for capturing continuous data streams from multiple sources, ensuring timely analysis and response*, yang menekankan pentingnya kecepatan dan skalabilitas dalam pengelolaan aliran data untuk aplikasi industri maupun penelitian.

Pada praktiknya, akuisisi data *streaming* sederhana memerlukan perhatian khusus terhadap sinkronisasi waktu, manajemen memori, serta validasi data agar informasi yang diperoleh tetap akurat dan konsisten, terutama ketika data berasal dari berbagai sumber yang berbeda. Teknik pemrosesan sederhana seperti pengambilan batch kecil atau penggunaan antrian (*queue*) sering digunakan untuk memastikan aliran data dapat diproses dengan lancar, sementara sistem penyimpanan sementara (*buffer*) menjaga data tidak hilang saat terjadi lonjakan *input*. Dengan pendekatan ini, peneliti atau praktisi dapat membangun sistem yang responsif dan handal dalam menangani data *streaming* secara *real-time*, sambil tetap menjaga integritas dan kualitas data untuk analisis lanjutan.

### 3. Contoh Script Python untuk Data IoT & Log

Script ini:

- a. Mensimulasikan sensor suhu & kelembaban
- b. Menyimpan data ke file log (iot\_log.csv)
- c. Mencatat log aktivitas sistem (system.log)

```
import random
import time
import csv
import logging
from datetime import datetime
# Setup logging sistem
logging.basicConfig(
    filename='system.log',
    level=logging.INFO,
    format='%(asctime)s - %(levelname)s - %(message)s')
# Nama file data IoT
```

```

iot_file = 'iot_log.csv'
# Membuat header CSV jika file belum ada
try:
    with open(iot_file, 'x', newline='') as file:
        writer = csv.writer(file)
        writer.writerow(['timestamp', 'temperature', 'humidity'])
except FileExistsError:
    pass
def generate_sensor_data():
    temperature = round(random.uniform(20.0, 35.0), 2)
    humidity = round(random.uniform(40.0, 80.0), 2)
    return temperature, humidity

def save_to_csv(temp, hum):
    with open(iot_file, 'a', newline='') as file:
        writer = csv.writer(file)
        writer.writerow([datetime.now(), temp, hum])

try:
    for _ in range(10): # Simulasi 10 data
        temp, hum = generate_sensor_data()
        save_to_csv(temp, hum)

        logging.info(f'Data tersimpan: Temp={temp}°C,
Humidity={hum}%")
        print(f'Data dikirim: {temp}°C | {hum}%")

        time.sleep(2)

except Exception as e:
    logging.error(f'Terjadi error: {e}")

```

# BAB VI

## DATA SEKUNDER

---

Data sekunder dalam metode pengumpulan data untuk sains data merujuk pada data yang telah tersedia sebelumnya dan dikumpulkan oleh pihak lain untuk tujuan tertentu, seperti data dari lembaga resmi, platform digital, atau arsip organisasi. Dalam praktik sains data, data sekunder berperan penting karena efisien dari segi waktu dan biaya serta memungkinkan analisis pola, tren, dan pengambilan keputusan berbasis data dalam skala besar.

### A. Pengertian dan Karakteristik Data Sekunder

#### 1. Pengertian Data Sekunder

Data sekunder merupakan data yang telah tersedia sebelumnya dan dikumpulkan oleh individu, lembaga, atau organisasi lain untuk tujuan tertentu di luar kebutuhan penelitian yang sedang dilakukan, namun masih relevan untuk dianalisis kembali dalam konteks baru. Dalam bidang sains data dan penelitian empiris, data sekunder sering dimanfaatkan karena dapat memberikan gambaran historis, tren jangka panjang, serta konteks makro yang mendukung proses analisis tanpa harus melalui tahapan pengumpulan data langsung yang memerlukan sumber daya besar.

Secara konseptual, data sekunder dipahami sebagai sumber data yang berasal dari dokumentasi, publikasi, basis data digital, atau arsip resmi yang telah melalui proses pengumpulan sebelumnya, sehingga penggunaannya menuntut pemahaman terhadap konteks asal data tersebut. Menurut Saunders, Lewis, dan Thornhill (2019), data sekunder adalah data yang telah dikumpulkan untuk tujuan selain penelitian saat ini dan dapat digunakan kembali untuk menjawab pertanyaan penelitian baru selama relevan dan dapat dipertanggungjawabkan secara metodologis.

## 2. Karakteristik Data Sekunder

Karakteristik data sekunder dapat dilihat dari beberapa aspek utama yang membedakannya dari data primer, yaitu sebagai berikut:

### a. Berasal dari Sumber Tidak Langsung

Karakteristik data sekunder yang berasal dari sumber tidak langsung berarti bahwa data tersebut tidak dikumpulkan sendiri oleh peneliti, melainkan diperoleh dari dokumen, laporan, atau basis data yang telah disusun sebelumnya oleh pihak lain untuk kepentingan tertentu. Hal ini sejalan dengan pandangan Sekaran dan Bougie (2020) yang menyatakan bahwa data sekunder merupakan data yang telah tersedia karena dikumpulkan oleh individu atau organisasi lain untuk tujuan berbeda, sehingga penggunaannya memerlukan penilaian kritis terhadap relevansi dan konteks asal data.

### b. Telah Melalui Proses Pengolahan Awal

Karakteristik data sekunder yang telah melalui proses pengolahan awal menunjukkan bahwa data tersebut umumnya sudah mengalami tahapan seperti pembersihan, pengelompokan, pengkodean, agregasi, atau peringkasan sebelum dipublikasikan dan digunakan kembali oleh peneliti. Kondisi ini berarti peneliti tidak berhadapan dengan data mentah sepenuhnya, karena sebagaimana dijelaskan oleh Hair, Black, Babin, dan Anderson (2019), data sekunder biasanya telah diproses dan disusun dalam format tertentu untuk tujuan analisis sebelumnya sehingga penggunaannya memerlukan pemahaman terhadap prosedur pengolahan yang telah dilakukan agar interpretasi hasil analisis tetap valid.

### c. Efisien dari Segi Waktu dan Biaya

Pemanfaatan data sekunder dinilai efisien dari segi waktu dan biaya karena peneliti tidak perlu melakukan perencanaan survei, pengumpulan data lapangan, maupun pengolahan awal yang umumnya membutuhkan sumber daya besar. Efisiensi ini memungkinkan peneliti untuk lebih fokus pada tahap analisis dan interpretasi data, karena sebagaimana dijelaskan oleh Creswell dan Creswell (2018), penggunaan data yang telah tersedia dapat mengurangi biaya penelitian secara signifikan serta mempercepat proses penelitian tanpa mengorbankan kedalaman analisis apabila data tersebut relevan dengan tujuan studi.

d. Cakupan Data Relatif Luas

Cakupan data sekunder relatif luas karena data tersebut umumnya dikumpulkan oleh lembaga besar seperti pemerintah, organisasi internasional, atau perusahaan skala nasional dan global, sehingga mencakup wilayah geografis yang luas, periode waktu yang panjang, serta jumlah unit analisis yang besar. Kondisi ini memungkinkan peneliti melakukan analisis komparatif, longitudinal, dan makro, karena sebagaimana dijelaskan oleh Bryman dan Bell (2019), data sekunder sering memiliki jangkauan yang lebih luas dibandingkan data primer sehingga sangat berguna untuk memahami pola dan tren pada skala besar yang sulit dicapai melalui pengumpulan data langsung.

e. Kesesuaian Terbatas dengan Tujuan Penelitian

Kesesuaian data sekunder dengan tujuan penelitian sering kali terbatas karena data tersebut dikumpulkan untuk kebutuhan dan konteks yang berbeda, sehingga variabel, definisi operasional, atau tingkat detailnya tidak selalu sejalan dengan fokus penelitian yang sedang dilakukan. Hal ini menuntut peneliti untuk melakukan evaluasi kritis terhadap relevansi data, karena menurut Schutt (2018), penggunaan data sekunder berpotensi menimbulkan keterbatasan analitis apabila tujuan awal pengumpulan data tidak sepenuhnya selaras dengan pertanyaan penelitian yang ingin dijawab.

## **B. Kelebihan dan Keterbatasan Data Sekunder**

Data sekunder merupakan salah satu sumber data yang banyak dimanfaatkan dalam penelitian dan sains data karena ketersediaannya yang luas serta kemudahan aksesnya. Namun, di balik berbagai keunggulan yang ditawarkan, data sekunder juga memiliki sejumlah keterbatasan yang perlu dipahami agar penggunaannya tetap relevan dan menghasilkan analisis yang valid.

### **1. Kelebihan Data Sekunder**

Data sekunder memiliki sejumlah keunggulan yang menjadikannya banyak digunakan dalam penelitian dan sains data, terutama pada studi berskala besar dan analisis lanjutan. Kelebihan tersebut antara lain:

- a. Efisiensi Waktu dan Biaya  
Efisiensi waktu dan biaya merupakan salah satu kelebihan utama data sekunder karena peneliti dapat langsung memanfaatkan data yang telah tersedia tanpa harus melalui tahapan perencanaan, pengumpulan data lapangan, serta pengolahan awal yang umumnya membutuhkan sumber daya besar. Keunggulan ini memungkinkan penelitian dilakukan lebih cepat dan ekonomis, karena menurut Trochim, Donnelly, dan Arora (2019), penggunaan data yang sudah ada dapat secara signifikan mengurangi biaya penelitian dan mempercepat proses analisis selama data tersebut relevan dengan tujuan studi yang dilakukan.
- b. Cakupan Data yang Luas  
Cakupan data yang luas merupakan kelebihan penting data sekunder karena data tersebut umumnya dikumpulkan oleh lembaga besar seperti pemerintah, organisasi internasional, atau institusi penelitian yang memiliki kemampuan menjangkau populasi besar, wilayah geografis luas, dan periode waktu yang panjang. Keunggulan ini memungkinkan analisis komparatif dan longitudinal dalam skala besar, karena menurut Knafllic (2020), data yang berasal dari sumber terorganisasi dan berskala luas memberikan peluang yang lebih kuat untuk mengidentifikasi pola, tren, dan hubungan antarvariabel yang sulit diperoleh melalui data primer terbatas.
- c. Ketersediaan dalam Berbagai Format  
Ketersediaan data sekunder dalam berbagai format merupakan kelebihan penting karena data dapat diakses dalam bentuk tabel statistik, laporan teks, arsip dokumen, basis data digital, hingga data multimedia yang memudahkan peneliti menyesuaikan metode analisis dengan jenis data yang digunakan. Keberagaman format ini mendukung fleksibilitas analisis lintas metode, karena menurut Heaton (2020), data sekunder sering tersedia dalam beragam bentuk yang memungkinkan peneliti melakukan analisis kuantitatif maupun kualitatif tanpa harus membatasi diri pada satu jenis representasi data saja.
- d. Mendukung Analisis Historis dan Tren  
Data sekunder mendukung analisis historis dan tren karena banyak data dikumpulkan secara berkala dan terdokumentasi dalam rentang waktu yang panjang, sehingga memungkinkan

peneliti menelusuri perubahan, perkembangan, serta pola dinamika suatu fenomena dari waktu ke waktu. Keunggulan ini sangat penting dalam analisis longitudinal dan prediktif, karena menurut Wooldridge (2020), data yang tersusun berdasarkan urutan waktu memberikan dasar yang kuat untuk mengamati tren, mengevaluasi perubahan struktural, dan memahami hubungan sebab akibat dalam jangka panjang.

## **2. Keterbatasan Data Sekunder**

Di samping kelebihan, data sekunder juga memiliki keterbatasan yang perlu diperhatikan agar hasil penelitian tetap valid dan dapat dipertanggungjawabkan. Keterbatasan tersebut meliputi:

### **a. Kesesuaian Terbatas dengan Tujuan Penelitian**

Kesesuaian data sekunder dengan tujuan penelitian sering kali terbatas karena data tersebut dikumpulkan untuk kepentingan, konteks, dan kerangka analisis yang berbeda, sehingga variabel, indikator, maupun tingkat kedalamannya tidak selalu selaras dengan kebutuhan penelitian yang sedang dilakukan. Keterbatasan ini menuntut kehati-hatian dalam penggunaan data, karena menurut Johnston (2017) yang masih banyak dirujuk dalam metodologi penelitian mutakhir, data sekunder dapat membatasi kemampuan peneliti dalam menjawab pertanyaan penelitian secara spesifik apabila tujuan awal pengumpulan data tidak sepenuhnya sesuai, meskipun data tersebut kaya dan mudah diakses.

### **b. Ketergantungan pada Kualitas Sumber**

Ketergantungan pada kualitas sumber menjadi keterbatasan utama data sekunder karena akurasi, kelengkapan, dan keandalan data sepenuhnya ditentukan oleh metode, instrumen, serta standar yang digunakan oleh pihak pengumpul data sebelumnya. Keterbatasan ini berpotensi memengaruhi validitas hasil analisis, karena menurut Bryman dan Bell (2021), penggunaan data sekunder menuntut evaluasi kritis terhadap kualitas sumber data mengingat peneliti tidak memiliki kontrol langsung atas proses pengumpulan dan pengolahan data yang dapat mengandung bias atau kesalahan.

- c. **Kurangnya Kontrol terhadap Proses Pengumpulan Data**  
Kurangnya kontrol terhadap proses pengumpulan data menjadi keterbatasan penting data sekunder karena peneliti tidak terlibat secara langsung dalam penentuan desain penelitian, teknik sampling, instrumen pengukuran, maupun prosedur pengumpulan data yang digunakan oleh pihak pengumpul sebelumnya. Kondisi ini dapat memengaruhi ketepatan analisis dan interpretasi hasil, karena menurut Babbie (2020), penggunaan data sekunder membatasi kemampuan peneliti untuk memastikan kesesuaian metodologis antara cara data dikumpulkan dan tujuan penelitian yang sedang dilakukan.
- d. **Kemungkinan Data Tidak Mutakhir**  
Kemungkinan data sekunder tidak mutakhir menjadi keterbatasan karena data tersebut sering kali dipublikasikan setelah jeda waktu tertentu, sehingga kondisi atau fenomena yang direpresentasikan tidak lagi mencerminkan situasi terkini. Keterbatasan ini dapat memengaruhi relevansi hasil analisis, karena menurut Saunders, Lewis, dan Thornhill (2019), data sekunder berisiko kehilangan aktualitasnya apabila tidak diperbarui secara berkala, terutama ketika digunakan untuk menganalisis dinamika yang cepat berubah.

## **C. Validitas dan Relevansi Data Sekunder**

### **1. Validitas Data Sekunder**

Validitas data sekunder merujuk pada tingkat ketepatan dan keandalan data dalam merepresentasikan fenomena yang diteliti sesuai dengan konteks penggunaannya. Meskipun data sekunder tidak dikumpulkan langsung oleh peneliti, validitasnya tetap dapat dijaga melalui evaluasi kritis terhadap sumber, metode pengumpulan, serta proses pengolahan data yang telah dilakukan sebelumnya. Beberapa aspek penting yang berkaitan dengan validitas data sekunder meliputi:

- a. **Kredibilitas Sumber Data**

Kredibilitas sumber data merupakan aspek fundamental dalam validitas data sekunder karena kualitas data sangat ditentukan oleh reputasi, kompetensi, dan integritas lembaga atau individu yang mengumpulkan serta mempublikasikan data tersebut. Sumber yang kredibel umumnya memiliki standar

metodologis yang jelas, prosedur pengumpulan data yang terdokumentasi, serta mekanisme pengendalian mutu yang dapat meningkatkan kepercayaan terhadap akurasi dan konsistensi data.

Pada konteks penelitian dan sains data, penggunaan data sekunder dari sumber yang kurang kredibel berisiko menghasilkan kesimpulan yang bias atau menyesatkan, meskipun data tersebut mudah diakses dan kaya secara kuantitas. Oleh karena itu, sebagaimana dijelaskan oleh Hair, Page, dan Brunsveld (2021), peneliti wajib menilai reputasi dan keandalan sumber data sekunder karena validitas temuan sangat bergantung pada sejauh mana data tersebut dikumpulkan dan disajikan oleh sumber yang dapat dipercaya.

b. Kejelasan Metode Pengumpulan Data

Kejelasan metode pengumpulan data merupakan komponen penting dalam validitas data sekunder karena memungkinkan peneliti memahami bagaimana data diperoleh, instrumen apa yang digunakan, serta prosedur apa yang diterapkan selama proses pengumpulan. Informasi metodologis yang jelas membantu peneliti menilai potensi bias, kesalahan pengukuran, serta tingkat konsistensi data sehingga hasil analisis yang dihasilkan dapat dipertanggungjawabkan secara ilmiah.

Pada penelitian berbasis data sekunder, ketidakjelasan metode pengumpulan data dapat menurunkan tingkat kepercayaan terhadap data meskipun data tersebut berasal dari sumber yang dikenal luas. Oleh karena itu, menurut Creswell dan Creswell (2018), transparansi metode pengumpulan data sangat diperlukan agar peneliti dapat mengevaluasi kesesuaian dan validitas data dalam menjawab pertanyaan penelitian yang dirumuskan.

c. Konsistensi dan Akurasi Data

Konsistensi dan akurasi data merupakan aspek penting dalam validitas data sekunder karena keduanya menunjukkan sejauh mana data disajikan secara stabil, bebas dari kesalahan pencatatan, serta memiliki kesesuaian nilai antarvariabel dan antarperiode pengamatan. Data yang konsisten dan akurat memungkinkan peneliti memperoleh hasil analisis yang dapat dipercaya, karena ketidaksesuaian atau kesalahan kecil sekalipun

dapat memengaruhi interpretasi dan kesimpulan penelitian secara signifikan.

Pada penggunaan data sekunder, konsistensi dan akurasi harus dievaluasi secara cermat karena peneliti tidak terlibat langsung dalam proses pengumpulan data awal. Oleh karena itu, sebagaimana dijelaskan oleh Hair, Black, Babin, dan Anderson (2019), kualitas analisis sangat bergantung pada tingkat akurasi dan konsistensi data yang digunakan, sehingga pemeriksaan terhadap kesalahan, duplikasi, dan ketidakwajaran data menjadi langkah penting untuk menjaga validitas penelitian.

## **2. Relevansi Data Sekunder**

Relevansi data sekunder berkaitan dengan tingkat kesesuaian data terhadap tujuan, pertanyaan penelitian, dan kerangka analisis yang digunakan. Data yang valid belum tentu relevan apabila variabel, waktu pengambilan, atau konteks pengumpulan tidak selaras dengan fokus penelitian. Oleh karena itu, penilaian relevansi menjadi tahap krusial sebelum data sekunder digunakan dalam analisis. Poin-poin utama dalam menilai relevansi data sekunder antara lain:

### **a. Kesesuaian Variabel dengan Tujuan Penelitian**

Kesesuaian variabel dengan tujuan penelitian merupakan aspek utama dalam relevansi data sekunder karena variabel yang tersedia harus mampu merepresentasikan konsep, fenomena, atau hubungan yang ingin dianalisis oleh peneliti. Apabila variabel yang ada tidak sejalan dengan rumusan masalah dan tujuan penelitian, maka data sekunder berpotensi menghasilkan analisis yang kurang tepat meskipun jumlah dan kualitas datanya memadai.

Pada konteks ini, peneliti perlu menelaah definisi operasional, skala pengukuran, serta konteks penggunaan variabel agar interpretasi hasil analisis tidak menyimpang dari tujuan penelitian yang ditetapkan. Hal ini sejalan dengan pendapat Neuman (2019) yang menegaskan bahwa relevansi data sekunder sangat bergantung pada tingkat kesesuaian variabel yang tersedia dengan pertanyaan penelitian, karena perbedaan tujuan awal pengumpulan data dapat membatasi kemampuan data tersebut dalam menjelaskan fenomena yang diteliti.

b. Kesesuaian Konteks Waktu dan Wilayah

Kesesuaian konteks waktu dan wilayah merupakan aspek krusial dalam relevansi data sekunder karena data harus merefleksikan periode dan lokasi yang selaras dengan fenomena yang sedang diteliti agar hasil analisis tidak menyimpang dari realitas yang ingin dipahami. Perbedaan waktu pengumpulan data dapat menyebabkan perubahan kondisi sosial, ekonomi, atau lingkungan, sehingga penggunaan data yang tidak sesuai konteks temporal berisiko menurunkan ketepatan interpretasi penelitian.

Kesesuaian wilayah juga menentukan relevansi data sekunder karena karakteristik suatu daerah dapat berbeda secara signifikan antara satu lokasi dan lokasi lainnya. Oleh karena itu, sebagaimana dijelaskan oleh Bryman (2018), peneliti perlu memastikan bahwa data sekunder yang digunakan memiliki kecocokan konteks waktu dan geografis dengan tujuan penelitian, karena ketidaksesuaian konteks tersebut dapat membatasi validitas dan makna temuan penelitian.

c. Kelengkapan Data

Kelengkapan data merupakan unsur penting dalam relevansi data sekunder karena data yang digunakan harus mencakup seluruh variabel, indikator, dan unit analisis yang dibutuhkan untuk menjawab tujuan penelitian secara utuh. Data yang tidak lengkap dapat menyebabkan kesenjangan informasi, membatasi ruang analisis, serta berpotensi menghasilkan kesimpulan yang parsial atau bias.

Pada konteks penelitian berbasis data sekunder, kelengkapan data juga berkaitan dengan ketersediaan metadata, dokumentasi, dan detail pendukung yang memungkinkan peneliti memahami struktur serta makna data secara menyeluruh. Oleh karena itu, sebagaimana dijelaskan oleh Creswell dan Creswell (2018), data sekunder yang relevan harus cukup lengkap untuk mendukung analisis yang mendalam, karena kekurangan informasi penting dapat mengurangi kemampuan data dalam menjelaskan fenomena yang diteliti secara komprehensif.

## D. Peran Data Sekunder dalam Sains Data

Data sekunder memiliki peran strategis dalam sains data karena menjadi sumber utama yang menyediakan volume data besar, beragam, dan historis yang diperlukan untuk proses analisis, pemodelan, dan pengambilan keputusan berbasis data. Dalam praktik sains data, data sekunder memungkinkan analis dan data *scientist* untuk melihat pola, tren, serta hubungan antarvariabel tanpa harus selalu melakukan pengumpulan data primer yang memerlukan waktu dan biaya besar. Peran data sekunder dalam sains data dapat dijelaskan melalui beberapa fungsi utama berikut:

### 1. Sebagai Sumber Utama Eksplorasi Data

Data sekunder berperan sebagai sumber utama eksplorasi data karena menyediakan kumpulan data awal yang kaya dan beragam untuk memahami struktur, pola, anomali, serta karakteristik dasar data sebelum dilakukan analisis lanjutan. Pada tahap eksplorasi, data sekunder memungkinkan data *scientist* melakukan ringkasan statistik, visualisasi, dan identifikasi hubungan awal antarvariabel tanpa harus menunggu proses pengumpulan data primer yang memakan waktu.

Pada praktik sains data, eksplorasi berbasis data sekunder membantu perumusan hipotesis analitis, pemilihan metode pemodelan yang tepat, serta penentuan kebutuhan data tambahan jika diperlukan. Hal ini sejalan dengan pandangan Han, Kamber, dan Pei (2022) yang menyatakan bahwa dataset yang telah tersedia sangat penting dalam tahap awal analisis karena memungkinkan eksplorasi data secara efisien untuk menemukan pola dan pengetahuan awal yang menjadi dasar proses data mining dan analitik lanjutan.

### 2. Mendukung Pembangunan dan Pelatihan Model Analitik

Data sekunder mendukung pembangunan dan pelatihan model analitik karena menyediakan volume data yang besar, beragam, dan siap pakai yang dibutuhkan untuk melatih algoritma statistik dan pembelajaran mesin secara efektif. Ketersediaan data tersebut memungkinkan proses *feature engineering*, pembagian data latih dan uji, serta evaluasi performa model dilakukan secara lebih cepat dan sistematis. Dalam sains data, kualitas dan kuantitas data sekunder sangat berpengaruh terhadap kemampuan model dalam menangkap pola dan

melakukan generalisasi terhadap data baru. Hal ini sejalan dengan pendapat Géron (2022) yang menekankan bahwa model pembelajaran mesin membutuhkan dataset yang cukup besar dan representatif agar proses pelatihan menghasilkan model analitik yang akurat, stabil, dan dapat diandalkan.

### **3. Memungkinkan Analisis Historis Dan Prediktif**

Data sekunder memungkinkan analisis historis karena banyak data dikumpulkan dan disimpan secara berkala dalam rentang waktu yang panjang, sehingga peneliti dapat menelusuri perubahan, pola, dan dinamika suatu fenomena dari masa ke masa secara sistematis. Dengan data historis tersebut, sains data dapat digunakan untuk memahami perilaku masa lalu sebagai dasar dalam merumuskan model analisis yang lebih akurat dan kontekstual. Selain itu, ketersediaan data historis yang memadai juga menjadi fondasi utama bagi analisis prediktif, karena pola masa lalu digunakan untuk memperkirakan kecenderungan dan kondisi di masa depan. Hal ini sejalan dengan pandangan Hyndman dan Athanasopoulos (2021) yang menegaskan bahwa data deret waktu historis merupakan elemen kunci dalam membangun model peramalan yang mampu menghasilkan prediksi yang andal dan berbasis bukti empiris.

### **4. Menjadi Dasar Pengambilan Keputusan Strategis**

Data sekunder menjadi dasar pengambilan keputusan strategis karena menyediakan informasi empiris yang komprehensif mengenai kondisi masa lalu dan saat ini, sehingga organisasi dapat memahami pola kinerja, risiko, serta peluang secara lebih objektif. Dengan memanfaatkan data yang telah terakumulasi dari berbagai sumber, pengambil keputusan dapat merumuskan strategi berbasis bukti tanpa bergantung semata pada intuisi atau asumsi. Dalam konteks sains data, analisis data sekunder memungkinkan penyusunan rekomendasi strategis yang lebih akurat dan terukur, baik dalam bidang bisnis, kebijakan publik, maupun perencanaan organisasi. Hal ini sejalan dengan pandangan Provost dan Fawcett (2019) yang menegaskan bahwa pemanfaatan data historis dan terstruktur melalui analitik data memberikan dasar yang kuat bagi pengambilan keputusan strategis yang rasional dan berorientasi pada nilai.



# BAB VII

## METODE PENGUMPULAN DATA SEKUNDER KONVENSIONAL

---

Metode pengumpulan data sekunder konvensional merupakan pendekatan penelitian yang memanfaatkan data yang telah tersedia sebelumnya dan diperoleh dari berbagai sumber non-digital seperti buku, arsip cetak, laporan resmi, dokumen institusi, serta publikasi statistik. Metode ini berperan penting dalam mendukung kajian ilmiah karena memungkinkan peneliti memperoleh gambaran konseptual dan empiris secara efisien, sistematis, serta berkelanjutan tanpa melakukan pengumpulan data langsung di lapangan.

### A. Studi Dokumen dan Arsip

Studi Dokumen dan Arsip merupakan metode pengumpulan data sekunder yang dilakukan dengan cara menelaah, mengkaji, dan menganalisis berbagai dokumen serta arsip yang telah tersedia sebelumnya. Dokumen dan arsip tersebut dapat berupa dokumen tertulis, cetak, maupun visual yang dihasilkan oleh individu, organisasi, atau lembaga resmi, sehingga metode ini sangat berguna untuk memahami peristiwa, kebijakan, proses sosial, dan perkembangan historis tanpa keterlibatan langsung peneliti dalam proses pengumpulan data primer.

#### 1. Pengertian Studi Dokumen dan Arsip

Studi dokumen dan arsip merupakan metode penelitian yang berfokus pada pengumpulan serta analisis sistematis terhadap berbagai dokumen dan arsip yang telah ada sebelumnya, baik dalam bentuk tertulis, visual, maupun digital, yang dihasilkan oleh individu, organisasi, atau lembaga tertentu dalam konteks sosial dan historis tertentu. Metode ini digunakan untuk memperoleh pemahaman mendalam mengenai kebijakan, proses, peristiwa, atau fenomena sosial

melalui penafsiran isi dokumen yang dilakukan secara kritis dan terstruktur, sebagaimana dijelaskan bahwa analisis dokumen adalah proses sistematis untuk meninjau atau mengevaluasi dokumen guna memperoleh makna, pemahaman, dan pengetahuan empiris yang relevan dengan tujuan penelitian (Creswell & Poth, 2018).

Pada praktik penelitian, studi dokumen dan arsip tidak hanya berfungsi sebagai sumber data pelengkap, tetapi juga dapat menjadi sumber data utama ketika peneliti ingin menelusuri perkembangan kebijakan, perubahan institusional, atau dinamika sosial dalam rentang waktu tertentu tanpa melakukan pengumpulan data langsung. Keandalan metode ini sangat ditentukan oleh kemampuan peneliti dalam menilai autentisitas, kredibilitas, dan relevansi dokumen, serta dalam mengaitkan temuan dokumen dengan kerangka teoritis dan konteks penelitian secara menyeluruh agar hasil analisis memiliki validitas dan kedalaman ilmiah yang memadai.

## **2. Jenis Dokumen dan Arsip**

Dokumen dan arsip yang digunakan dalam penelitian dapat berasal dari berbagai bentuk dan sumber, antara lain:

### **a. Dokumen Resmi**

Dokumen resmi merupakan sumber data tertulis yang dihasilkan dan dikeluarkan secara sah oleh lembaga pemerintah, organisasi formal, atau institusi berwenang sebagai bagian dari fungsi administratif, hukum, dan kebijakan, sehingga isinya mencerminkan keputusan, aturan, serta aktivitas kelembagaan yang bersifat formal dan terstruktur. Dalam konteks penelitian, dokumen resmi memiliki nilai ilmiah yang tinggi karena disusun melalui prosedur baku, dapat dipertanggungjawabkan secara hukum dan institusional, serta dapat digunakan untuk menelusuri kebijakan, proses pengambilan keputusan, dan dinamika organisasi secara objektif, di mana dokumen semacam ini dipandang sebagai sumber data yang stabil, sistematis, dan kaya konteks untuk analisis kualitatif (Bowen, 2018).

### **b. Dokumen Nonresmi**

Dokumen nonresmi adalah catatan atau arsip yang tidak diterbitkan secara formal oleh lembaga resmi, seperti surat pribadi, memo internal, catatan harian, atau laporan informal, yang merekam pengalaman, pandangan, dan interaksi sosial

individu atau kelompok dalam konteks tertentu. Dokumen jenis ini penting dalam penelitian karena memberikan wawasan kontekstual dan subjektif yang tidak tersedia dalam dokumen resmi, memungkinkan analisis mendalam terhadap motivasi, persepsi, dan praktik sosial, di mana dokumen nonresmi dianggap sebagai sumber data yang kaya makna dan fleksibel untuk analisis kualitatif (Rapley, 2019).

c. Arsip Statistik dan Administrasi

Arsip statistik dan administrasi merupakan kumpulan data terstruktur yang dihasilkan oleh instansi pemerintah, lembaga penelitian, atau organisasi formal melalui pencatatan rutin, sensus, survei, dan laporan keuangan atau administrasi, yang mencerminkan informasi kuantitatif dan kualitatif mengenai kegiatan, populasi, maupun kinerja institusi. Arsip ini sangat penting dalam penelitian karena menyediakan data historis yang akurat dan sistematis, memungkinkan analisis tren, evaluasi kebijakan, serta perbandingan antarperiode atau antarunit, di mana arsip statistik dan administrasi dianggap sebagai sumber data primer sekunder yang dapat diverifikasi dan diandalkan untuk keperluan analisis ilmiah (Punch, 2018).

d. Dokumen Visual dan Audio

Dokumen visual dan audio mencakup materi seperti foto, video, peta, rekaman suara, dan film dokumenter yang merekam peristiwa, interaksi, atau praktik sosial dalam bentuk visual maupun audio, sehingga memberikan representasi yang kaya akan konteks, nuansa, dan dinamika yang sulit ditangkap melalui dokumen tertulis saja. Dalam penelitian, dokumen jenis ini memungkinkan analisis mendalam terhadap simbol, ekspresi, dan komunikasi nonverbal, serta mendukung interpretasi kontekstual yang lebih hidup, karena materi visual dan audio dianggap sebagai sumber data yang kaya makna dan mampu memperkuat validitas temuan kualitatif (Pink, 2018).

### 3. Prosedur Pelaksanaan Studi Dokumen dan Arsip

Pelaksanaan studi dokumen dan arsip dimulai dengan identifikasi kebutuhan data yang jelas, termasuk jenis dokumen yang relevan dengan tujuan penelitian, sumber dokumen, serta kriteria seleksi untuk memastikan data yang dikumpulkan bersifat valid, kredibel, dan

representatif. Tahapan ini diikuti oleh pengumpulan dokumen secara sistematis, pencatatan metadata, dan evaluasi awal terhadap keaslian serta relevansi dokumen, sehingga peneliti dapat menyiapkan bahan yang siap dianalisis secara mendalam (Bowen, 2018).

Tahap berikutnya adalah analisis isi dokumen yang dilakukan secara sistematis menggunakan teknik coding, kategorisasi, dan interpretasi tematik, sehingga informasi yang diperoleh dapat dikaitkan dengan pertanyaan penelitian dan kerangka teori yang digunakan. Analisis ini memungkinkan peneliti mengekstrak pola, hubungan, dan makna dari dokumen, sekaligus melakukan triangulasi data untuk meningkatkan validitas hasil penelitian.

Tahap akhir dalam prosedur studi dokumen dan arsip meliputi penyusunan temuan, penafsiran hasil analisis dalam konteks penelitian, dan penyajian data secara koheren dalam laporan penelitian. Proses ini memastikan bahwa seluruh dokumen dianalisis secara kritis, hasilnya dapat dipertanggungjawabkan, serta memberikan kontribusi yang signifikan terhadap pemahaman fenomena yang diteliti.

## **B. Literature Review (Kajian Pustaka)**

*Literature Review* (Kajian Pustaka) merupakan salah satu tahap penting dalam penelitian yang berfungsi untuk menelaah, mengevaluasi, dan mensintesis penelitian terdahulu, teori, dan literatur terkait topik penelitian sehingga memberikan dasar ilmiah yang kuat untuk merumuskan masalah, kerangka teori, dan metodologi penelitian. Kajian pustaka membantu peneliti memahami perkembangan ilmu, menemukan kesenjangan penelitian, serta menghindari duplikasi studi, sekaligus membangun landasan konseptual yang jelas dan sistematis.

### **1. Tujuan Literature Review**

Tujuan *literature review* adalah untuk memberikan dasar ilmiah yang kuat bagi penelitian dengan menelaah, mengevaluasi, dan mensintesis temuan penelitian terdahulu, teori, dan literatur terkait sehingga peneliti dapat memahami konteks, tren, dan perkembangan terbaru dalam bidang studi yang diteliti. Dengan demikian, kajian pustaka membantu mengidentifikasi kesenjangan penelitian, memperkuat kerangka teori, dan membimbing perumusan pertanyaan penelitian atau hipotesis secara sistematis, di mana literature review

dianggap sebagai alat penting untuk membangun landasan konseptual dan metodologis penelitian yang kredibel (Hart, 2018).

## **2. Jenis Literatur yang Dikaji**

Jenis literatur yang dikaji dalam sebuah literature review mencakup berbagai sumber akademik dan ilmiah yang relevan dengan topik penelitian, seperti buku teks, artikel jurnal, laporan penelitian, tesis, disertasi, dokumen resmi, serta publikasi institusi, yang memberikan kerangka teori, data empiris, dan pemahaman konseptual yang mendalam. Pemilihan jenis literatur ini dilakukan secara kritis untuk memastikan bahwa sumber yang digunakan memiliki kredibilitas, relevansi, dan kualitas ilmiah yang tinggi sehingga dapat mendukung analisis dan sintesis temuan penelitian secara menyeluruh (Ridley, 2018).

Literatur nonformal atau *grey literature* juga dapat dikaji, termasuk laporan teknis, artikel konferensi, white paper, dan materi digital yang belum diterbitkan secara resmi, karena sumber ini sering menyediakan informasi kontekstual, data terbaru, atau perspektif yang tidak tersedia dalam publikasi akademik konvensional. Dengan memasukkan literatur semacam ini, peneliti dapat memperoleh wawasan yang lebih luas, melacak tren terkini, dan memperkuat validitas serta kedalaman analisis penelitian.

## **3. Tahapan Pelaksanaan *Literature Review***

Tahapan pelaksanaan *literature review* dimulai dengan identifikasi topik dan perumusan pertanyaan penelitian yang jelas, diikuti dengan penentuan kata kunci dan strategi pencarian literatur yang sistematis di berbagai basis data akademik, jurnal ilmiah, buku, dan sumber relevan lainnya. Proses ini memastikan bahwa literatur yang dikumpulkan sesuai dengan fokus penelitian dan mampu memberikan landasan teoritis serta empiris yang solid, sehingga kajian pustaka dapat dilakukan secara efektif dan terarah (Booth, Sutton, & Papaioannou, 2019).

Tahap berikutnya adalah evaluasi dan seleksi literatur, di mana setiap sumber dianalisis dari segi kredibilitas, relevansi, metodologi, dan kualitas ilmiah, sehingga hanya literatur yang memenuhi standar akademik yang digunakan untuk sintesis dan analisis. Evaluasi ini membantu peneliti menyingkirkan sumber yang bias, kurang valid, atau

tidak relevan, sehingga hasil kajian pustaka tetap objektif dan dapat dipertanggungjawabkan.

Tahap terakhir adalah sintesis dan penulisan kajian pustaka, yang melibatkan pengelompokan literatur berdasarkan tema, konsep, atau metodologi, diikuti dengan analisis kritis terhadap hubungan antartemuan penelitian. Penulisan ini tidak hanya merangkum literatur, tetapi juga mengaitkan temuan dengan kerangka teori, mengidentifikasi kesenjangan penelitian, dan menegaskan kontribusi penelitian baru terhadap bidang studi yang relevan (Booth, Sutton, & Papaioannou, 2019).

### C. *Database* Statistik Resmi

*Database* Statistik Resmi adalah kumpulan data yang disusun, dikelola, dan diterbitkan oleh lembaga pemerintah, badan resmi, atau institusi publik yang bertugas melakukan pencatatan dan pemantauan informasi kuantitatif mengenai berbagai aspek sosial, ekonomi, demografi, dan lingkungan. *Database* ini bersifat terstruktur, dapat diakses publik, dan biasanya diperbarui secara periodik sehingga memungkinkan penelitian, perencanaan kebijakan, dan pengambilan keputusan berbasis data yang akurat.

#### 1. Karakteristik *Database* Statistik Resmi

Karakteristik *database* statistik resmi mencakup keandalan, keteraturan, dan keterpercayaan data yang disusun berdasarkan prosedur pengumpulan, pengolahan, dan verifikasi yang baku oleh lembaga pemerintah atau institusi resmi, sehingga hasilnya dapat digunakan sebagai sumber informasi yang sah untuk penelitian, perencanaan, dan pengambilan keputusan. *Database* ini juga bersifat sistematis, tersusun secara terstruktur dalam bentuk tabel, grafik, dan basis data elektronik, serta diperbarui secara periodik agar mencerminkan kondisi terkini, memungkinkan analisis tren dan evaluasi kebijakan secara akurat (United Nations, 2019).

*Database* statistik resmi memiliki karakteristik aksesibilitas yang tinggi, karena sebagian besar tersedia untuk publik dan dapat digunakan oleh peneliti, akademisi, maupun masyarakat umum, sehingga mendukung transparansi dan akuntabilitas institusi penyedia data. Sifat ini menjadikan *database* statistik resmi sebagai sumber data yang tidak

hanya kredibel, tetapi juga mudah diintegrasikan ke dalam berbagai studi empiris dan analisis kebijakan berbasis bukti.

## **2. Jenis Data dalam *Database* Statistik Resmi**

- a. Data demografi: Informasi mengenai jumlah penduduk, usia, jenis kelamin, kependudukan, migrasi, dan keluarga.
- b. Data ekonomi: Statistik terkait pertumbuhan ekonomi, inflasi, pengangguran, produksi, perdagangan, dan pendapatan.
- c. Data sosial: Data pendidikan, kesehatan, kriminalitas, ketenagakerjaan, dan kesejahteraan masyarakat.
- d. Data lingkungan: Informasi mengenai kualitas udara, air, energi, sumber daya alam, dan perubahan lingkungan.

## **3. Manfaat *Database* Statistik Resmi**

Manfaat *database* statistik resmi terletak pada kemampuannya menyediakan data yang akurat, sahih, dan terstruktur sehingga dapat digunakan sebagai dasar penelitian ilmiah, analisis sosial-ekonomi, serta perencanaan dan evaluasi kebijakan publik secara sistematis. Data yang disediakan memungkinkan peneliti dan pembuat kebijakan untuk mengidentifikasi tren, membandingkan indikator antarwilayah atau periode waktu, serta mengambil keputusan berbasis bukti yang dapat dipertanggungjawabkan secara akademik dan administratif (OECD, 2019).

*Database* statistik resmi mendukung transparansi dan akuntabilitas lembaga publik karena akses publik terhadap data memungkinkan masyarakat, akademisi, dan pihak terkait lainnya memantau kinerja institusi dan hasil kebijakan secara objektif. Sifat data yang lengkap, konsisten, dan diperbarui secara berkala menjadikannya sumber informasi yang andal untuk analisis jangka panjang, perencanaan strategis, serta evaluasi program dan kebijakan publik.

## **D. Data dari Organisasi dan Institusi**

Data dari Organisasi dan Institusi merupakan jenis data sekunder yang diperoleh dari lembaga formal, baik pemerintah maupun non-pemerintah, yang dikumpulkan, diolah, dan disimpan untuk kepentingan administrasi, pelaporan, penelitian, atau evaluasi program. Data ini sering digunakan dalam penelitian karena bersifat resmi, terstruktur, dan mencerminkan aktivitas, kebijakan, atau kondisi tertentu yang dikelola

oleh organisasi atau institusi, sehingga menjadi sumber informasi yang kredibel untuk analisis ilmiah maupun perencanaan strategis.

### **1. Sumber Data dari Organisasi dan Institusi**

Sumber data dari organisasi dan institusi mencakup informasi yang dikumpulkan dan dikelola oleh lembaga formal, baik pemerintah maupun non-pemerintah, yang menyediakan data terkait aktivitas, kebijakan, demografi, ekonomi, sosial, maupun program-program tertentu, sehingga menjadi sumber informasi yang kredibel dan terstruktur untuk penelitian maupun perencanaan kebijakan. Data ini meliputi laporan resmi, *database* internal, publikasi institusi, dan statistik organisasi yang telah diverifikasi, sehingga memungkinkan peneliti untuk menganalisis tren, evaluasi program, dan pengambilan keputusan berbasis bukti secara sistematis (Fink, 2019).

Sumber data dari organisasi dan institusi memberikan perspektif yang lebih kontekstual dan spesifik karena mencerminkan kondisi nyata atau praktik yang berlangsung dalam lingkup organisasi tersebut, sekaligus memfasilitasi triangulasi data dengan sumber lain untuk meningkatkan validitas temuan penelitian. Aksesibilitas sebagian data untuk publik juga mendukung transparansi dan akuntabilitas, sehingga data ini menjadi fondasi penting dalam analisis ilmiah maupun evaluasi kebijakan yang berbasis fakta.

### **2. Karakteristik Data dari Organisasi dan Institusi**

Karakteristik data dari organisasi dan institusi mencakup keandalan, keteraturan, dan keterverifikasian karena data dikumpulkan melalui prosedur resmi dan standar yang baku, sehingga informasi yang diperoleh dapat dipertanggungjawabkan baik secara akademik maupun administratif. Data ini biasanya tersimpan dalam bentuk laporan, basis data, atau arsip yang sistematis, memungkinkan analisis kuantitatif maupun kualitatif yang konsisten dengan tujuan penelitian atau evaluasi kebijakan (Babbie, 2018).

Data dari organisasi dan institusi bersifat kontekstual dan spesifik karena merefleksikan kegiatan, kebijakan, atau kondisi nyata dalam lingkup lembaga penyedia data, sehingga memberikan gambaran yang akurat mengenai fenomena yang diteliti. Aksesibilitas data untuk publik atau pihak tertentu juga menjadikan sumber ini penting dalam mendukung transparansi, akuntabilitas, dan validitas analisis penelitian serta pengambilan keputusan berbasis bukti.

### 3. Manfaat Data dari Organisasi dan Institusi

Manfaat data dari organisasi dan institusi terletak pada kemampuannya menyediakan informasi yang resmi, kredibel, dan terstruktur sehingga dapat digunakan sebagai dasar penelitian empiris, evaluasi program, perencanaan kebijakan, serta pengambilan keputusan berbasis bukti secara sistematis. Data ini memungkinkan peneliti dan pembuat kebijakan untuk menganalisis tren, membandingkan indikator antarunit atau periode waktu, serta menilai efektivitas kegiatan atau kebijakan yang telah diterapkan, sehingga hasil analisis menjadi valid dan dapat dipertanggungjawabkan secara akademik maupun praktis (Johnston, 2017).

Data dari organisasi dan institusi membantu menghemat waktu dan sumber daya penelitian karena data primer telah tersedia, serta memperkuat kredibilitas temuan penelitian melalui dukungan bukti yang terdokumentasi secara resmi. Aksesibilitas sebagian data untuk publik juga meningkatkan transparansi dan akuntabilitas lembaga penyedia data, sehingga informasi ini dapat dimanfaatkan oleh berbagai pihak termasuk peneliti, pemerintah, dan masyarakat luas untuk tujuan analisis, evaluasi, maupun pengembangan kebijakan.

## E. Internet dan Portal Online

### 1. Internet

Internet adalah jaringan global yang menghubungkan jutaan komputer, server, dan perangkat lainnya di seluruh dunia, memungkinkan pertukaran data dan komunikasi secara *real-time*. Internet berfungsi sebagai infrastruktur utama bagi berbagai layanan digital, seperti *website*, email, media sosial, aplikasi *cloud*, dan layanan berbasis IoT. Selain itu, internet juga menjadi media penting untuk penyebaran informasi, edukasi, hiburan, dan transaksi bisnis. Fungsi utama internet meliputi:

#### a. Komunikasi

Internet berperan sebagai alat utama komunikasi global yang memungkinkan individu, kelompok, dan organisasi untuk saling bertukar informasi dan berinteraksi tanpa batasan geografis maupun waktu, sehingga mempercepat proses penyampaian pesan dan memperluas jaringan sosial dalam skala global. Sebagaimana dijelaskan, “*The Internet revolutionized*

*communication and social interaction, connecting the world like never before...*” yang menunjukkan bahwa internet telah secara fundamental mengubah cara manusia berkomunikasi dan berinteraksi satu sama lain sejak munculnya jaringan digital ini, termasuk melalui email, pesan instan, dan media sosial yang melampaui hambatan tradisional (Ahuja, 2023).

Pada konteks kehidupan modern, fungsi komunikasi internet tidak hanya terbatas pada penyampaian pesan teks, tetapi juga mencakup berbagai bentuk interaksi multimedia seperti panggilan suara, *video conference*, dan kolaborasi digital yang mendukung hubungan personal serta profesional. Dengan demikian, internet menjadi sarana komunikasi yang fleksibel, efisien, dan kaya fitur yang mendukung pertukaran informasi *real-time* di seluruh dunia, sekaligus menyatukan beragam komunitas digital dalam satu jaringan global yang dinamis.

b. Akses Informasi

Internet memungkinkan akses informasi secara cepat, luas, dan mudah bagi siapa saja yang tersambung ke jaringan global, sehingga individu dan komunitas dapat mencari, menemukan, dan memanfaatkan data dan fakta dari berbagai sumber tanpa terikat batasan fisik maupun waktu. Dalam era digital seperti sekarang, akses ini menjadi dasar bagi pendidikan, pekerjaan, penelitian, layanan publik, dan partisipasi sosial karena internet menyediakan sumber informasi yang kaya dan beragam. Menurut Petković (2025), *“The Internet today forms the foundation for access to education, healthcare, employment, political participation, and information,”* yang mencerminkan peran fundamental internet dalam menyediakan akses informasi penting untuk kehidupan modern.

Kemampuan untuk mengakses informasi melalui internet tidak otomatis sama bagi semua orang karena akses yang efektif bergantung pada faktor konektivitas, perangkat, literasi digital, dan kualitas jaringan, sehingga disparitas akses tetap menjadi tantangan nyata dalam memanfaatkan potensi penuh internet. Karena itu, selain menyediakan infrastruktur, perlu adanya upaya meningkatkan kemampuan pengguna dalam mencari dan mengevaluasi informasi yang tersedia secara daring agar manfaat

akses informasi melalui internet dapat dirasakan secara optimal dan bertanggung jawab oleh masyarakat luas.

c. Transaksi Digital

Internet berperan sebagai media utama yang memungkinkan terjadinya transaksi digital, di mana individu dan organisasi dapat membeli dan menjual barang atau jasa secara *online* tanpa perlu bertemu secara fisik, karena jaringan global ini menyediakan infrastruktur yang menghubungkan pembeli, penjual, dan sistem pembayaran digital dengan cepat dan efisien. Hal ini terlihat dari bagaimana volume transaksi melalui e-commerce dan metode pembayaran berbasis internet semakin meningkat secara signifikan, dengan sistem seperti dompet digital, *mobile banking*, dan gateway pembayaran yang mempercepat proses transaksi serta mengurangi hambatan geografis dan waktu. Menurut hasil penelitian Amit Kumar Singh & Gupta (2025), digital payments and transactions “*have transformed online buying and selling by improving convenience, security, and global reach, making digital transactions a crucial enabler of e-commerce growth.*” (Singh & Gupta, 2025)

Transaksi digital melalui internet tidak hanya mempermudah proses jual-beli, tetapi juga mendukung inklusi ekonomi yang lebih luas dengan memberikan akses kepada pelaku usaha kecil hingga pasar global, meningkatkan efisiensi operasional, dan membuka peluang partisipasi dalam ekonomi digital secara *real-time*. Dengan dukungan teknologi internet yang terus berkembang, transaksi digital kini menjadi fondasi penting dalam kegiatan ekonomi modern yang mendorong pertumbuhan sektor bisnis daring serta menciptakan pengalaman pengguna yang lebih responsif dan aman.

## 2. Portal Online

Portal *Online* adalah platform digital atau situs web yang menyediakan akses terpusat ke berbagai informasi, layanan, atau sumber daya tertentu. Portal *online* berfungsi sebagai “gerbang” yang mempermudah pengguna untuk mendapatkan konten dan layanan dari satu titik akses, sehingga pengalaman interaksi menjadi lebih efisien dan terstruktur. Portal ini sering digunakan oleh perusahaan, institusi

pendidikan, lembaga pemerintah, dan organisasi publik maupun privat. Fungsi utama portal *online* meliputi:

a. Pusat Informasi

Portal *online* berfungsi sebagai pusat informasi yang menyediakan akses terstruktur ke berbagai konten, termasuk berita, artikel, data statistik, dan dokumen resmi, sehingga pengguna dapat memperoleh informasi yang relevan secara cepat dan efisien tanpa harus mencari dari berbagai sumber terpisah. Dengan desain yang terorganisir dan sistem navigasi yang intuitif, portal *online* memungkinkan pengelolaan informasi secara terpusat, mendukung pengambilan keputusan, penelitian, dan aktivitas profesional maupun pribadi. Menurut Chen dan Zhang (2022), "*Online portals serve as centralized hubs for information dissemination, enhancing accessibility and efficiency for users across sectors,*" yang menekankan peran portal *online* sebagai media pengelolaan dan penyebaran informasi yang efisien. (Chen & Zhang, 2022)

Portal *online* juga meningkatkan kemampuan pengguna untuk menilai, membandingkan, dan memanfaatkan data yang tersedia secara terintegrasi, sehingga mempercepat proses pengambilan keputusan dan mendukung kolaborasi antarindividu atau organisasi. Dengan demikian, portal *online* tidak hanya berfungsi sebagai sumber informasi statis, tetapi juga sebagai platform dinamis yang menghubungkan pengguna dengan berbagai layanan dan konten relevan secara *real-time*, meningkatkan efektivitas komunikasi dan manajemen informasi.

b. Layanan Interaktif

Portal *online* menyediakan layanan interaktif yang memungkinkan pengguna untuk melakukan berbagai aktivitas secara langsung melalui platform, seperti pendaftaran, pengajuan dokumen, pembayaran, pengaduan, maupun konsultasi layanan, sehingga interaksi antara pengguna dan penyedia layanan menjadi lebih cepat, efisien, dan terdokumentasi dengan baik. Fungsi interaktif ini memanfaatkan teknologi digital untuk meningkatkan keterlibatan pengguna, memberikan umpan balik *real-time*, dan memungkinkan penyesuaian layanan sesuai kebutuhan individu, sehingga portal *online* berperan tidak hanya sebagai sumber informasi, tetapi juga sebagai platform

partisipatif. Menurut Al-Haderi dan Ismail (2021), “*Interactive services on online portals significantly enhance user engagement and streamline administrative processes, making digital interaction more effective and responsive,*” yang menekankan bahwa fitur interaktif pada portal *online* meningkatkan efektivitas komunikasi dan layanan digital. (Al-Haderi & Ismail, 2021)

Layanan interaktif juga mendukung transparansi dan akuntabilitas karena setiap transaksi atau pengajuan dapat dilacak dan dicatat secara sistematis, sehingga mengurangi risiko kesalahan manual dan meningkatkan kepercayaan pengguna terhadap platform. Dengan demikian, portal *online* yang dilengkapi layanan interaktif tidak hanya meningkatkan efisiensi operasional organisasi, tetapi juga memperkuat pengalaman pengguna dengan menyediakan akses yang responsif, personal, dan aman untuk berbagai layanan digital.

c. Personalisasi Konten

Portal *online* menyajikan personalisasi konten sebagai salah satu fungsi utama yang menyesuaikan tampilan informasi dan rekomendasi secara spesifik berdasarkan perilaku, preferensi, dan interaksi pengguna, sehingga setiap pengunjung mendapatkan pengalaman yang lebih relevan dengan kebutuhan pribadinya daripada sekadar tampilan konten umum. Personalisasi ini memanfaatkan data pengguna untuk mengoptimalkan penyajian konten yang mungkin paling menarik atau berguna bagi individu tertentu, yang pada gilirannya meningkatkan keterlibatan pengguna, kepuasan, dan efektivitas interaksi digital di dalam portal. Menurut penelitian terbaru dalam konteks digital, personalisasi konten berarti “*adaptation of information according to the user’s individual needs, preferences, or behavior,*” yang menunjukkan bahwa fungsi ini bukan sekadar pilihan estetis, tetapi merupakan strategi untuk membuat konten lebih bermakna dan relevan bagi setiap pengguna.

Implementasi personalisasi konten dalam portal *online* juga membantu mengurangi informasi yang tidak relevan dan mempermudah pengguna dalam menemukan materi penting dengan lebih cepat, sehingga menghemat waktu dan mengurangi beban kognitif saat menavigasi banyaknya pilihan konten yang

tersedia. Dengan demikian, portal yang menerapkan personalisasi konten dapat menciptakan pengalaman pengguna yang lebih adaptif dan responsif terhadap kebutuhan individu, sekaligus memperkuat keterikatan pengunjung terhadap platform melalui konten yang terasa lebih “disesuaikan.”

## **F. Media Massa dan Publikasi Populer**

Media massa dan publikasi populer merupakan saluran komunikasi yang berfungsi untuk menyebarkan informasi, hiburan, opini, dan edukasi kepada masyarakat luas. Media ini meliputi berbagai bentuk, mulai dari media cetak seperti koran, majalah, dan tabloid; media penyiaran seperti radio dan televisi; hingga media digital dan *online* seperti portal berita, blog, serta platform media sosial populer. Fungsi utamanya adalah menjangkau audiens yang besar dan heterogen dengan penyampaian informasi yang ringkas, mudah dipahami, dan menarik perhatian.

### **1. Karakteristik Media Massa dan Publikasi Populer**

Media massa dan publikasi populer memiliki karakteristik utama berupa jangkauan luas yang memungkinkan informasi disebarkan secara cepat kepada audiens heterogen, serta kemampuan untuk menyajikan konten dengan bahasa yang mudah dipahami dan visual menarik sehingga pesan dapat diterima secara efektif oleh masyarakat (McQuail, 2019). Selain itu, media ini memiliki siklus produksi yang cepat dan dinamis, terutama pada platform digital, yang memungkinkan update informasi secara *real-time* sehingga relevansi berita tetap terjaga.

Media massa dan publikasi populer juga menonjolkan orientasi audiens dan nilai komersial, di mana konten disesuaikan dengan minat, kebutuhan, dan preferensi masyarakat untuk meningkatkan engagement dan menarik iklan, yang kemudian memengaruhi pemilihan topik dan penyajian berita. Karakteristik ini membuat media populer menjadi alat penting untuk membentuk opini publik dan menentukan agenda sosial, politik, maupun budaya, namun tetap rentan terhadap bias editorial dan sensasionalisme.

Karakteristik lain yang menonjol adalah keberagaman format dan jenis media, mulai dari cetak, elektronik, hingga digital interaktif, yang menyediakan fleksibilitas bagi peneliti maupun masyarakat untuk

mengakses informasi sesuai kebutuhannya. Keberagaman ini memungkinkan pengumpulan data sekunder dalam bentuk teks, audio, dan visual sehingga analisis konten dapat dilakukan dengan metode kuantitatif maupun kualitatif untuk memahami tren, opini, dan pola komunikasi publik secara lebih menyeluruh.

## **2. Jenis Media Massa dan Publikasi Populer**

Media massa dan publikasi populer mencakup beragam jenis saluran komunikasi yang digunakan untuk menyebarkan informasi dan opini kepada audiens luas, seperti media cetak (surat kabar dan majalah), penyiaran (radio dan televisi), serta media digital termasuk portal berita *online* dan blog dengan jangkauan besar; jenis-jenis ini berfungsi secara bersamaan untuk memenuhi kebutuhan informasi, hiburan, dan interaksi sosial yang terus berubah di tengah perkembangan teknologi komunikasi massa (Deuze, 2020). Dalam konteks digital, portal berita *online* dan platform media sosial juga mengambil peran signifikan sebagai media massa karena kapasitasnya untuk mempercepat peredaran informasi dan memungkinkan partisipasi audiens dalam produksi serta distribusi konten, sehingga batas antara produser dan konsumen informasi semakin kabur dan dinamis.

Tipe media cetak seperti surat kabar dan majalah tetap eksis meskipun menghadapi tekanan digitalisasi karena kemampuan menyediakan liputan mendalam dan versi arsip yang dapat dirujuk kembali, sedangkan media penyiaran radio dan televisi terus memanfaatkan keunggulan audio-visual untuk menarik perhatian audiens melalui siaran langsung, talkshow, dan program dokumenter yang memiliki nada populer dan relevan bagi kehidupan sehari-hari. Di era modern, blog populer dan kanal mikroblogging memberikan ruang bagi individu atau kelompok kecil untuk menyebarluaskan konten yang sering kali mencerminkan perspektif personal atau komunitas tertentu, memperluas variasi genre publikasi populer di luar format tradisional sekaligus menantang dominasi media arus utama dalam membentuk wacana publik.

Publikasi populer seperti majalah gaya hidup, majalah sains populer, serta kanal video berbasis internet menghadirkan konten dengan pendekatan visual kuat dan narasi yang mudah diakses oleh masyarakat luas, sehingga jenis media ini tidak hanya berperan sebagai sumber informasi tetapi juga sebagai alat refleksi budaya dan nilai sosial yang

sedang tren. Variasi jenis media massa dan publikasi populer tersebut menunjukkan bagaimana masyarakat modern mengonsumsi informasi melalui kombinasi format cetak, audio, visual, dan interaktif, yang kesemuanya berkontribusi dalam pembentukan opini publik dan dinamika komunikasi massa kontemporer.

### **3. Peran dan Fungsi**

Media massa dan publikasi populer berperan penting dalam menyampaikan informasi secara luas kepada masyarakat, tidak hanya sebagai saluran berita tetapi juga sebagai alat pendidikan dan pembentukan opini publik karena dapat menyampaikan rangkaian peristiwa aktual serta interpretasi makna sosial kepada audiens yang heterogen (McQuail, 2019). Peran tersebut mencakup fungsi informatif dan edukatif yang membantu masyarakat memahami dinamika sosial, politik, ekonomi, dan budaya secara lebih menyeluruh, sekaligus menjaga keterhubungan antara fenomena lokal dan global dalam konteks kehidupan sehari-hari.

Di samping fungsi informatif, media massa dan publikasi populer juga berfungsi sebagai arena diskusi publik di mana beragam pandangan dan wacana dikemukakan sehingga memperluas ruang keterlibatan masyarakat dalam kehidupan berbangsa dan bernegara, termasuk dalam proses pengambilan keputusan yang bersifat kolektif. Fungsi ini memungkinkan media menjadi mediator sosial yang menghubungkan antara pembuat kebijakan dan publik umum, sehingga isu-isu strategis dapat dipahami dan direspons dengan lebih cepat oleh berbagai lapisan masyarakat.

Media massa dan publikasi populer berperan dalam hiburan dan penguatan budaya, menyediakan konten teks, audio, dan visual yang tidak hanya informatif tetapi juga menarik secara estetis sehingga membantu mengurangi kejenuhan sekaligus memperkaya pengalaman budaya masyarakat. Peran dan fungsi media tersebut mendukung terwujudnya masyarakat informasi yang dinamis dan partisipatif, di mana akses terhadap sumber daya komunikasi memungkinkan setiap individu untuk menjadi bagian dari jaringan penyebaran wawasan dan nilai sosial yang relevan di era digital saat ini.

## G. Ekstraksi dan Integrasi Data Sekunder Menggunakan Python

Ekstraksi dan integrasi data sekunder menggunakan Python adalah proses mengambil data yang sudah tersedia dari berbagai sumber dan menggabungkannya ke dalam satu format yang siap dianalisis. Python sangat populer untuk tugas ini karena memiliki berbagai pustaka (*library*) yang mendukung pengambilan, pembersihan, dan integrasi data secara efisien dari berbagai format, seperti CSV, JSON, Excel, *database*, dan API *online*.

### 1. Membaca File Statistik Resmi

Membaca file statistik resmi merupakan langkah penting dalam pengumpulan data sekunder karena data ini berasal dari lembaga terpercaya seperti badan statistik nasional, kementerian, atau organisasi internasional yang menyediakan informasi yang tervalidasi dan terdokumentasi dengan baik, sehingga memungkinkan analisis yang akurat dan dapat dipertanggungjawabkan (Saunders *et al.*, 2019). Proses ini biasanya melibatkan identifikasi sumber data, pemilihan format file yang sesuai seperti CSV, Excel, JSON, atau PDF, serta pemanfaatan perangkat lunak atau bahasa pemrograman seperti Python untuk ekstraksi dan pembacaan data secara efisien. Selain itu, tahap awal ini juga menekankan pentingnya memahami metadata, definisi variabel, serta periode pengumpulan data untuk menghindari kesalahan interpretasi dalam analisis selanjutnya.

Langkah selanjutnya adalah pembersihan dan normalisasi, yang mencakup penghapusan duplikasi, penanganan *missing values*, pengaturan format tanggal, dan konsistensi nama kolom agar dataset siap untuk analisis lebih lanjut. Tahap ini penting karena file statistik resmi sering kali berisi informasi dalam berbagai format yang memerlukan harmonisasi agar dapat digabungkan dengan dataset lain, sehingga integrasi data dapat dilakukan tanpa kehilangan makna atau kualitas informasi. Dengan demikian, pembacaan file statistik resmi tidak hanya sekadar mengambil data, tetapi juga memastikan bahwa data tersebut siap untuk dianalisis secara sistematis dan digunakan dalam pengambilan keputusan berbasis bukti.

Membaca file statistik resmi memungkinkan peneliti atau analis untuk melakukan validasi silang terhadap sumber lain dan melihat tren

historis maupun pola data secara lebih komprehensif, yang mendukung penelitian kuantitatif maupun kualitatif. Proses ini juga memudahkan pemanfaatan data untuk visualisasi, laporan, dan dashboard interaktif yang membantu penyampaian informasi secara jelas dan mudah dipahami oleh pemangku kepentingan. Dengan pendekatan yang tepat, membaca file statistik resmi menjadi fondasi yang kuat untuk memperoleh *insight* yang relevan, akurat, dan berbasis bukti dalam berbagai konteks penelitian maupun praktik profesional.

## 2. Normalisasi dan Pembersihan Data

Normalisasi dan pembersihan data adalah tahap krusial dalam proses pengolahan data yang bertujuan untuk memastikan kualitas, konsistensi, dan keterandalan dataset sebelum dianalisis, terutama ketika data berasal dari berbagai sumber sekunder yang memiliki format, struktur, dan skala yang berbeda-beda (Kotu & Deshpande, 2019). Proses pembersihan mencakup identifikasi dan penanganan *missing values*, penghapusan duplikasi, koreksi kesalahan penulisan, serta pengaturan tipe data agar sesuai dengan kebutuhan analisis, sedangkan normalisasi menstandarisasi skala dan format data sehingga berbagai variabel dapat dibandingkan dan digabungkan secara valid. Dengan demikian, normalisasi dan pembersihan data tidak hanya meningkatkan akurasi hasil analisis tetapi juga meminimalkan bias dan kesalahan interpretasi yang bisa muncul akibat inkonsistensi atau kualitas data yang buruk.

Normalisasi dan pembersihan data memungkinkan integrasi dataset dari berbagai sumber menjadi satu kerangka data yang kohesif, sehingga memudahkan analisis lanjutan baik secara kuantitatif maupun kualitatif. Proses ini juga meliputi standarisasi format tanggal, satuan ukuran, dan kategori variabel, serta transformasi nilai ekstrem atau *outlier* agar tidak memengaruhi kesimpulan analisis secara tidak proporsional. Pendekatan sistematis ini membantu peneliti atau analis untuk memastikan bahwa data yang digunakan mencerminkan informasi yang sebenarnya dan siap untuk diproses menggunakan algoritma statistik atau metode *machine learning*.

Normalisasi dan pembersihan data juga penting untuk meningkatkan efisiensi pemrosesan dan meminimalkan risiko error pada tahap analisis, terutama ketika dataset besar atau kompleks digunakan dalam proyek penelitian atau pengambilan keputusan bisnis. Proses ini

dapat dibantu oleh *tools* atau pustaka pemrograman seperti Python (pandas, numpy) yang menyediakan fungsi khusus untuk deteksi anomali, pengisian *missing values*, dan standarisasi variabel secara otomatis. Dengan implementasi yang tepat, normalisasi dan pembersihan data menjadi fondasi penting untuk menghasilkan *insight* yang valid, dapat dipercaya, dan dapat dipertanggungjawabkan.

### 3. Contoh Script Python Data Sekunder

#### Contoh Mengolah Data Sekunder dari File CSV

Misalnya kita memiliki file `data_penjualan.csv`.

Contoh isi CSV:

```
tanggal,produk,jumlah,harga
2025-01-01,A,10,5000
2025-01-02,B,5,7000
2025-01-03,A,8,5000
```

Script Python:

```
import pandas as pd
# Membaca data sekunder dari file CSV
data = pd.read_csv('data_penjualan.csv')
# Menampilkan 5 data pertama
print("Data Awal:")
print(data.head())
# Menambahkan kolom total penjualan
data['total'] = data['jumlah'] * data['harga']
# Menghitung total penjualan per produk
summary = data.groupby('produk')['total'].sum()
print("\nTotal Penjualan per Produk:")
print(summary)
# Menyimpan hasil analisis ke file baru
summary.to_csv('ringkasan_penjualan.csv')
```



# BAB VIII

## METODE PENGUMPULAN DATA SEKUNDER BERBASIS TEKNOLOGI

---

Metode pengumpulan data sekunder berbasis teknologi memanfaatkan perangkat digital dan internet untuk mengakses informasi yang telah tersedia. Pendekatan ini memungkinkan pengumpulan data lebih cepat, efisien, dan akurat dibandingkan metode konvensional.

### A. Konsep Otomatisasi Pengumpulan Data

Otomatisasi pengumpulan data adalah proses sistematis yang menggunakan teknologi untuk menangkap, memindahkan, dan menyimpan data dari berbagai sumber secara otomatis, tanpa campur tangan manusia secara langsung. Konsep ini berfokus pada peningkatan efisiensi, kecepatan, dan akurasi dalam pengelolaan data, sekaligus memastikan ketersediaan data secara konsisten untuk analisis atau pengambilan keputusan. Dengan otomatisasi, organisasi dapat mengurangi kesalahan manual, mempercepat proses pengumpulan, serta memungkinkan pemrosesan data dalam jumlah besar dengan lebih efektif.

#### 1. Prinsip Dasar

Prinsip dasar otomatisasi pengumpulan data menekankan integrasi antara teknologi, sumber data, dan proses yang terstruktur sehingga pengambilan informasi dapat berlangsung secara konsisten dan minim kesalahan manusia, termasuk penggunaan trigger otomatis, jadwal terprogram, serta mekanisme validasi untuk memastikan data yang dikumpulkan akurat dan siap diproses lebih lanjut (Chen *et al.*, 2020). Pendekatan ini memungkinkan organisasi untuk mengoptimalkan aliran data dari berbagai platform digital, seperti API, sensor IoT, dan

log sistem, sehingga setiap unit informasi dapat diakses secara *real-time* tanpa perlu intervensi manual yang intensif.

Prinsip otomasi juga mengedepankan ketahanan dan skalabilitas sistem, di mana pipeline pengumpulan harus mampu menangani perubahan volume data, kegagalan koneksi, dan variasi format sumber data, sambil tetap menjaga integritas serta keamanan informasi yang dikumpulkan (Chen *et al.*, 2020). Dengan penerapan prinsip ini, perusahaan atau institusi dapat mempercepat proses analisis, mendukung pengambilan keputusan berbasis data, dan memastikan bahwa semua data yang diterima dapat ditelusuri asal-usulnya melalui mekanisme logging dan metadata yang memadai.

## 2. Komponen Utama Sistem Otomatisasi

### a. Pengumpul Data (*Collector*)

Pengumpul data (*collector*) dalam sistem otomatisasi berfungsi sebagai komponen inti yang bertugas menangkap informasi dari berbagai sumber secara langsung, baik melalui API, sensor, log aplikasi, maupun *scraping web*, sehingga data mentah dapat diteruskan ke proses transformasi dan penyimpanan secara efisien tanpa intervensi manusia (Singh *et al.*, 2021). Selain itu, *collector* dirancang untuk mendukung pemrosesan skala besar dan *real-time*, termasuk pengelolaan retry otomatis, filter data, dan pengaturan prioritas aliran informasi agar setiap data yang dikumpulkan tetap akurat, lengkap, dan sesuai kebutuhan analitik.

### b. Penyalur Data

Penyalur data dalam sistem otomatisasi berperan sebagai mediator yang mentransfer informasi dari pengumpul data ke modul transformasi dan penyimpanan, menggunakan mekanisme seperti *message queue*, *event bus*, atau *webhook* untuk memastikan aliran data berlangsung secara efisien, terstruktur, dan andal, sekaligus mengurangi risiko kehilangan atau duplikasi data (Zhang *et al.*, 2019). Selain itu, penyalur data memungkinkan sistem untuk menangani volume data yang besar dan bersifat *real-time* dengan skalabilitas tinggi, mendukung proses pipeline otomatis yang responsif dan konsisten terhadap berbagai jenis sumber dan format data.

### c. Transformasi Data (ETL/ELT)

Transformasi data (ETL/ELT) dalam sistem otomatisasi berfungsi untuk membersihkan, menormalkan, dan mengubah data mentah menjadi format yang konsisten dan siap dianalisis, di mana proses ini mencakup ekstraksi dari sumber, transformasi sesuai aturan bisnis, dan pemuatan ke repositori tujuan seperti data warehouse atau data lake (Patel & Mehta, 2022). Proses transformasi ini dirancang agar mendukung pipeline berskala besar dan *real-time*, memastikan integritas, akurasi, serta konsistensi data, sekaligus memungkinkan agregasi, enrichment, dan validasi otomatis untuk meningkatkan kualitas analitik.

d. Penyimpanan

Penyimpanan dalam sistem otomatisasi berperan sebagai repositori akhir yang menampung data yang telah dikumpulkan dan ditransformasikan, baik di data warehouse, data lake, atau basis data terstruktur maupun semi-terstruktur, sehingga memungkinkan akses cepat, integritas data terjaga, dan dukungan untuk analitik skala besar (Kim & Lee, 2021). Sistem penyimpanan ini juga dirancang agar skalabel, aman, dan andal, termasuk kemampuan untuk menangani volume data yang terus meningkat, menyediakan mekanisme backup, enkripsi, dan kontrol akses yang memadai untuk menjamin keamanan serta kepatuhan terhadap regulasi.

### 3. Metode Pengumpulan Data Otomatis

Metode pengumpulan data otomatis mencakup berbagai pendekatan yang dirancang untuk menangkap informasi dari sumber digital secara efisien dan berkelanjutan, termasuk polling terjadwal, pengumpulan berbasis event, batch processing, dan *streaming* data, sehingga setiap data dapat dikumpulkan secara tepat waktu dan siap digunakan untuk analisis tanpa intervensi manual. Pendekatan ini memanfaatkan integrasi teknologi seperti API, webhook, sensor IoT, dan agen pengumpul untuk memastikan data yang masuk akurat, relevan, dan konsisten dengan kebutuhan proses bisnis atau penelitian (Almeida *et al.*, 2019).

Metode otomatis menekankan fleksibilitas dalam menangani variasi sumber dan format data, termasuk kemampuan untuk menyesuaikan pipeline dengan volume data yang berubah-ubah dan menangani kegagalan pengambilan melalui retry otomatis atau

mekanisme backoff, sehingga aliran data tetap stabil dan dapat diandalkan. Sistem ini juga mendukung integrasi dengan monitoring dan logging untuk memastikan observabilitas penuh, sehingga setiap kegagalan atau anomali dapat segera terdeteksi dan ditangani, serta metadata yang dihasilkan memungkinkan audit dan kepatuhan regulasi.

Implementasi metode pengumpulan data otomatis memungkinkan pengelolaan data skala besar secara efisien, mempercepat proses transformasi dan analitik, serta mendukung pengambilan keputusan berbasis data secara *real-time* tanpa menambah beban tenaga manusia secara signifikan. Dengan pendekatan ini, organisasi dapat memastikan kualitas, keamanan, dan ketersediaan data tetap terjaga secara konsisten.

#### **4. Langkah Alur Kerja Otomatisasi**

Langkah alur kerja otomatisasi mencakup serangkaian tahapan yang terstruktur mulai dari deteksi sumber data dan autentikasi, pengambilan informasi secara otomatis, validasi dan pembersihan, hingga transformasi menjadi format yang konsisten dan siap disimpan, sehingga setiap data dapat diproses tanpa campur tangan manual secara berulang (Hernandez *et al.*, 2020). Dalam tahap ini, mekanisme pemantauan dan notifikasi diterapkan untuk memastikan pipeline berjalan lancar, mendeteksi kegagalan atau anomali, serta menyediakan informasi diagnostik yang diperlukan agar proses tetap andal dan efisien.

Data dimuat ke penyimpanan yang sesuai, baik data warehouse, data lake, atau basis data khusus, di mana sistem memastikan integritas, keamanan, dan ketersediaan data, sekaligus mendukung akses untuk analisis dan pengambilan keputusan berbasis data secara *real-time*. Seluruh proses alur kerja ini juga mencakup pengelolaan retensi data, backup, dan kontrol akses untuk menjamin kepatuhan terhadap regulasi dan menjaga kualitas informasi secara berkelanjutan.

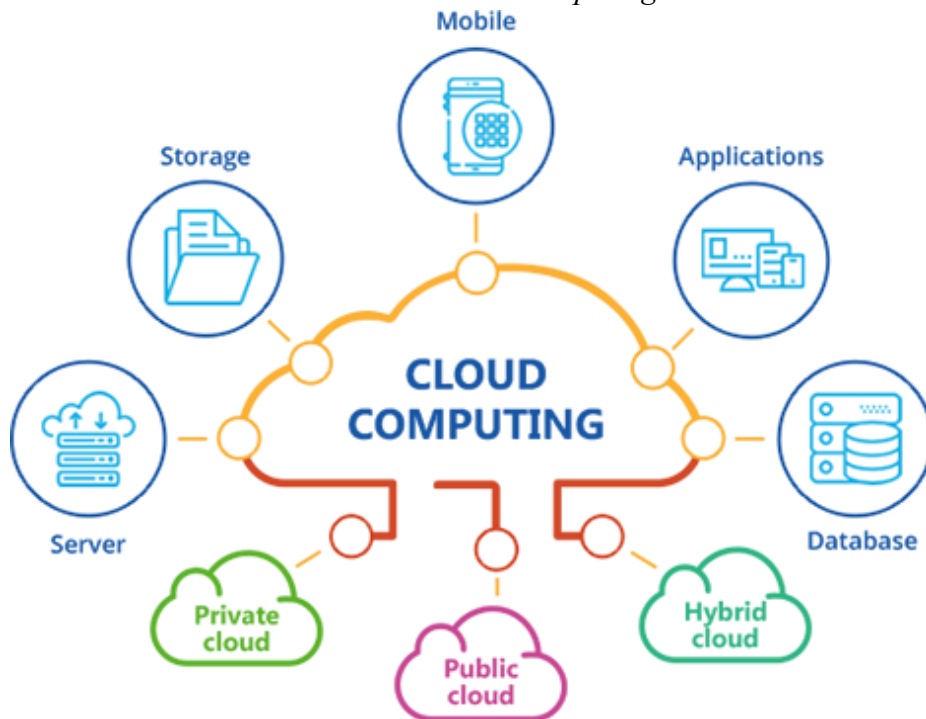
Langkah-langkah otomatisasi dirancang agar skalabel dan fleksibel, memungkinkan organisasi menyesuaikan pipeline dengan volume dan variasi data yang terus berubah, sambil memanfaatkan logging, metadata, dan mekanisme audit untuk meningkatkan observabilitas dan transparansi seluruh proses pengelolaan data (Hernandez *et al.*, 2020). Dengan penerapan alur kerja ini, sistem tidak hanya mempercepat pengumpulan dan pemrosesan data, tetapi juga mengurangi risiko kesalahan manual, memaksimalkan efisiensi

operasional, dan memastikan data selalu siap digunakan untuk analitik lanjutan.

## B. Infrastruktur Digital dan *Cloud Data*

Infrastruktur Digital dan *Cloud Data* merupakan fondasi teknologi yang memungkinkan organisasi mengelola, menyimpan, memproses, dan mendistribusikan data secara efisien melalui jaringan berbasis internet maupun sistem internal. Infrastruktur ini mencakup kombinasi perangkat keras, perangkat lunak, jaringan, pusat data, serta layanan komputasi awan (*cloud computing*) yang terintegrasi untuk mendukung operasional, analisis data, dan transformasi digital secara menyeluruh.

Gambar 5. *Cloud Computing*



Sumber: Btech

Secara umum, infrastruktur digital berperan sebagai kerangka kerja utama dalam pengelolaan sistem informasi modern, sedangkan *cloud data* berfungsi sebagai media penyimpanan dan pemrosesan data berbasis internet yang fleksibel dan skalabel.

### 1. Komponen Infrastruktur Digital

Infrastruktur digital terdiri atas beberapa elemen utama yang saling terhubung:

a. Perangkat Keras (*Hardware*)

Perangkat keras (*hardware*) dalam infrastruktur digital merupakan komponen fisik yang menjadi fondasi operasional sistem informasi, meliputi server, perangkat penyimpanan (*storage*), komputer, perangkat jaringan seperti router dan switch, serta pusat data (*data center*) yang bekerja secara terintegrasi untuk mendukung pemrosesan, transmisi, dan penyimpanan data secara stabil dan berkelanjutan. Dalam konteks komputasi modern, *hardware* berperan penting dalam mendukung virtualisasi, komputasi awan, serta pengolahan *big data* karena kapasitas dan kinerjanya menentukan kecepatan, skalabilitas, dan keandalan layanan digital, sejalan dengan pandangan Mell dan Grance (2018) yang menyatakan bahwa infrastruktur komputasi menyediakan sumber daya pemrosesan, penyimpanan, dan jaringan sebagai fondasi layanan berbasis *cloud*.

b. Perangkat Lunak (*Software*)

Perangkat lunak (*software*) dalam infrastruktur digital merupakan komponen non-fisik yang mengatur, mengendalikan, dan mengintegrasikan seluruh proses komputasi melalui sistem operasi, aplikasi bisnis, basis data, *middleware*, serta platform analitik yang memungkinkan pengolahan, penyimpanan, dan distribusi data berjalan secara terstruktur dan efisien dalam suatu ekosistem teknologi. Peran strategis software semakin krusial dalam lingkungan berbasis *cloud* dan transformasi digital karena ia menjembatani interaksi antara pengguna dan perangkat keras serta memungkinkan otomatisasi dan orkestrasi layanan, sejalan dengan Pressman dan Maxim (2020) yang menyatakan bahwa perangkat lunak merupakan produk logis yang menginstruksikan komputer untuk menjalankan fungsi tertentu dalam suatu sistem terintegrasi.

c. Jaringan dan Konektivitas

Jaringan dan konektivitas dalam infrastruktur digital merupakan sistem yang menghubungkan berbagai perangkat, server, pusat data, dan layanan *cloud* melalui media transmisi seperti fiber

optik, kabel tembaga, maupun jaringan nirkabel sehingga memungkinkan pertukaran data secara cepat, stabil, dan aman dalam skala lokal maupun global. Perannya sangat krusial dalam menjamin ketersediaan layanan digital, mendukung komunikasi *real-time*, serta memastikan interoperabilitas sistem, sejalan dengan penjelasan Kurose dan Ross (2021) bahwa jaringan komputer adalah infrastruktur yang memungkinkan perangkat komputasi saling bertukar data melalui protokol komunikasi yang terstandarisasi.

d. Sistem Keamanan Digital

Sistem keamanan digital dalam infrastruktur digital merupakan mekanisme terpadu yang dirancang untuk melindungi data, jaringan, perangkat keras, dan perangkat lunak dari ancaman siber melalui penerapan enkripsi, firewall, intrusion detection system, autentikasi multi-faktor, serta manajemen kontrol akses yang terstandarisasi guna menjaga kerahasiaan, integritas, dan ketersediaan informasi. Pentingnya sistem ini semakin meningkat seiring transformasi digital dan adopsi *cloud computing*, karena sebagaimana dijelaskan oleh Whitman dan Mattord (2022) keamanan informasi adalah upaya perlindungan terhadap informasi dan elemen-elemen pentingnya untuk menjamin keberlangsungan operasional organisasi dan meminimalkan risiko kerugian akibat ancaman siber.

## 2. Konsep dan Karakteristik *Cloud Data*

Konsep *cloud data* merujuk pada sistem penyimpanan dan pengelolaan data yang memanfaatkan infrastruktur komputasi awan sehingga data tidak lagi tersimpan secara lokal pada perangkat fisik pengguna, melainkan pada server terdistribusi yang dapat diakses melalui jaringan internet dengan tingkat fleksibilitas dan skalabilitas yang tinggi. Dalam praktiknya, *cloud data* memungkinkan organisasi melakukan pemrosesan, integrasi, dan analisis data secara dinamis sesuai kebutuhan kapasitas, sehingga mendukung efisiensi operasional dan transformasi digital berbasis data.

Karakteristik utama *cloud data* mencakup layanan mandiri sesuai permintaan (*on-demand self-service*), akses jaringan luas (*broad network access*), elastisitas sumber daya, serta pengelolaan terpusat berbasis virtualisasi yang memungkinkan alokasi sumber daya secara efisien dan

terukur. Hal ini sejalan dengan Mell dan Grance (2018) yang menjelaskan bahwa komputasi awan memiliki karakteristik esensial berupa kemampuan menyediakan sumber daya komputasi yang dapat dikonfigurasi secara cepat dan elastis dengan interaksi manajemen yang minimal dari penyedia layanan.

### 3. Model Layanan *Cloud*

*Cloud computing* memiliki tiga model layanan utama:

#### a. *Infrastructure as a Service* (IaaS)

*Infrastructure as a Service* (IaaS) dalam model layanan *cloud* merupakan bentuk layanan komputasi yang menyediakan sumber daya infrastruktur dasar seperti server virtual, penyimpanan, jaringan, dan sistem operasi melalui internet sehingga organisasi tidak perlu membangun atau memelihara pusat data fisik sendiri. Dalam model ini pengguna memiliki kendali atas sistem operasi dan aplikasi yang dijalankan di atas infrastruktur tersebut, sedangkan penyedia layanan bertanggung jawab atas pengelolaan perangkat keras dan virtualisasi, sebagaimana dijelaskan oleh Erl, Puttini, dan Mahmood (2020) bahwa IaaS menyediakan infrastruktur TI yang dapat diprovisikan secara elastis sebagai layanan berbasis *cloud* untuk mendukung kebutuhan komputasi pengguna.

#### b. *Platform as a Service* (PaaS)

*Platform as a Service* (PaaS) dalam model layanan *cloud* merupakan layanan yang menyediakan lingkungan pengembangan dan deployment aplikasi secara lengkap melalui infrastruktur berbasis *cloud*, sehingga pengembang dapat membangun, menguji, dan mengelola aplikasi tanpa harus mengatur perangkat keras, sistem operasi, maupun konfigurasi jaringan secara langsung. Dalam model ini penyedia layanan mengelola infrastruktur dan platform seperti *middleware*, *database*, serta *tools* pengembangan, sedangkan pengguna berfokus pada pengembangan aplikasi, sebagaimana dijelaskan oleh Mell dan Grance (2018) bahwa PaaS memungkinkan konsumen untuk menyebarkan aplikasi yang dibuat menggunakan bahasa pemrograman dan alat yang didukung oleh penyedia *cloud* tanpa mengelola infrastruktur *cloud* yang mendasarinya.

c. *Software as a Service* (SaaS)

*Software as a Service* (SaaS) dalam model layanan *cloud* merupakan bentuk layanan di mana aplikasi perangkat lunak disediakan dan dijalankan sepenuhnya oleh penyedia *cloud* sehingga pengguna dapat mengaksesnya melalui internet tanpa perlu melakukan instalasi, pemeliharaan, maupun pengelolaan infrastruktur dan platform pendukung. Dalam model ini seluruh pengelolaan aplikasi, pembaruan sistem, keamanan, serta penyimpanan data berada pada tanggung jawab penyedia layanan, sementara pengguna hanya memanfaatkan fungsi aplikasi sesuai kebutuhan, sebagaimana dijelaskan oleh Stair dan Reynolds (2021) bahwa SaaS adalah model distribusi perangkat lunak berbasis *cloud* di mana aplikasi dihosting oleh vendor dan diakses oleh pelanggan melalui jaringan.

#### 4. Manfaat Infrastruktur Digital dan *Cloud Data*

Manfaat infrastruktur digital dan *cloud data* terlihat pada peningkatan efisiensi operasional, fleksibilitas pengelolaan sumber daya TI, serta kemampuan organisasi dalam mengakses, menyimpan, dan menganalisis data secara *real-time* tanpa harus melakukan investasi besar pada perangkat keras dan pusat data fisik. Selain itu, penerapan *cloud* memungkinkan skalabilitas elastis, kolaborasi lintas lokasi, serta peningkatan ketahanan sistem melalui mekanisme backup dan disaster recovery terintegrasi, sebagaimana dijelaskan oleh Laudon dan Laudon (2022) bahwa komputasi awan memberikan organisasi kemampuan untuk mengurangi biaya TI sekaligus meningkatkan kelincahan dan skalabilitas operasional melalui penyediaan sumber daya komputasi sesuai permintaan.

### C. Tantangan Teknis dan Etika Data Sekunder Digital

Tantangan Teknis dan Etika Data Sekunder Digital muncul seiring meningkatnya pemanfaatan data yang diperoleh dari sumber digital seperti basis data daring, media sosial, portal pemerintah, repositori ilmiah, dan platform *big data*, di mana data tersebut tidak dikumpulkan secara langsung oleh peneliti tetapi digunakan kembali untuk tujuan analisis tertentu. Meskipun data sekunder digital menawarkan efisiensi dan cakupan luas, penggunaannya menghadirkan

berbagai persoalan teknis dan etis yang perlu dikelola secara cermat agar hasil analisis tetap valid, akurat, serta tidak melanggar prinsip perlindungan data.

## 1. Tantangan Teknis Data Sekunder Digital

### a. Kualitas dan Konsistensi Data

Kualitas dan konsistensi data merupakan tantangan utama dalam pemanfaatan data sekunder digital karena data yang diperoleh dari berbagai platform daring, basis data terbuka, maupun sistem informasi organisasi sering kali mengandung nilai yang hilang, duplikasi, inkonsistensi format, serta perbedaan standar pencatatan yang dapat memengaruhi akurasi analisis. Permasalahan ini menjadi krusial karena data sekunder tidak dikumpulkan secara langsung oleh peneliti sehingga proses verifikasi, validasi, dan pembersihan data (*data cleaning*) harus dilakukan secara sistematis untuk memastikan keandalan hasil penelitian atau pengambilan keputusan.

Konsistensi data antar sumber juga menuntut adanya standarisasi struktur, metadata, serta definisi variabel agar integrasi dataset tidak menimbulkan distorsi informasi atau bias interpretasi dalam proses analitik. Sebagaimana dijelaskan oleh Batini dan Scannapieco (2020), kualitas data berkaitan dengan sejauh mana data memenuhi persyaratan akurasi, kelengkapan, konsistensi, dan ketepatan waktu sehingga dapat digunakan secara efektif untuk tujuan tertentu.

### b. Interoperabilitas dan Integrasi Data

Interoperabilitas dan integrasi data menjadi tantangan teknis penting dalam pemanfaatan data sekunder digital karena data yang bersumber dari berbagai sistem, platform, dan organisasi umumnya memiliki format, struktur, standar metadata, serta protokol yang berbeda sehingga menyulitkan proses penggabungan dan pertukaran informasi secara langsung. Ketidaksesuaian skema basis data, perbedaan definisi variabel, serta variasi kualitas dokumentasi teknis dapat menimbulkan inkonsistensi dan kesalahan interpretasi apabila tidak dilakukan proses transformasi, harmonisasi, dan standarisasi data secara sistematis.

Pada konteks ini, integrasi data memerlukan pendekatan arsitektur dan tata kelola yang mampu menjamin kompatibilitas

antar sistem serta menjaga makna semantik informasi agar tetap konsisten setelah digabungkan. Sebagaimana dijelaskan oleh Lankhorst (2021), interoperabilitas mengacu pada kemampuan sistem yang berbeda untuk bertukar dan menggunakan informasi secara efektif melalui kesepahaman teknis dan semantik yang terstandarisasi.

c. Validitas dan Relevansi Kontekstual

Validitas dan relevansi kontekstual merupakan tantangan teknis krusial dalam penggunaan data sekunder digital karena data tersebut pada awalnya dikumpulkan untuk tujuan, populasi, dan kondisi tertentu yang belum tentu selaras dengan kebutuhan analisis baru, sehingga berpotensi menimbulkan bias apabila konteks asal, metode pengumpulan, serta batasan variabel tidak dipahami secara menyeluruh. Ketidaksesuaian antara tujuan awal pengumpulan data dan tujuan pemanfaatan kembali dapat memengaruhi ketepatan inferensi, terutama ketika terjadi perbedaan waktu, lokasi, karakteristik responden, atau definisi operasional variabel yang digunakan.

Peneliti perlu melakukan evaluasi kritis terhadap sumber, periode, metodologi, dan cakupan data untuk memastikan bahwa informasi yang dianalisis benar-benar merepresentasikan fenomena yang diteliti dalam konteks yang relevan. Sebagaimana dijelaskan oleh Saunders, Lewis, dan Thornhill (2019), penggunaan data sekunder menuntut penilaian yang cermat terhadap kesesuaian tujuan awal pengumpulan data dengan tujuan penelitian saat ini agar validitas temuan tetap terjaga.

d. Keamanan dan Kerentanan Sistem

Keamanan dan kerentanan sistem merupakan tantangan teknis yang signifikan dalam pemanfaatan data sekunder digital karena proses pengambilan, penyimpanan, dan distribusi data melalui jaringan internet, API, maupun platform *cloud* membuka peluang terjadinya akses tidak sah, kebocoran informasi, serangan malware, serta manipulasi data apabila tidak dilindungi oleh mekanisme pengamanan yang memadai. Kompleksitas arsitektur digital yang melibatkan berbagai sistem terintegrasi juga meningkatkan risiko celah keamanan (*vulnerabilities*) akibat kesalahan konfigurasi, kelemahan perangkat lunak, atau

kurangnya pembaruan sistem yang dapat dimanfaatkan oleh pihak tidak bertanggung jawab.

Penerapan kontrol keamanan seperti enkripsi data, autentikasi multi-faktor, manajemen akses berbasis peran, serta pemantauan sistem secara berkelanjutan menjadi langkah penting untuk menjaga integritas dan kerahasiaan data sekunder digital. Sebagaimana dikemukakan oleh Whitman dan Mattord (2022), keamanan informasi bertujuan melindungi informasi dan sistem pendukungnya dari berbagai ancaman guna memastikan keberlangsungan operasional dan meminimalkan dampak risiko siber.

## 2. Tantangan Etika Data Sekunder Digital

### a. Privasi dan Perlindungan Data Pribadi

Privasi dan perlindungan data pribadi merupakan isu etika utama dalam pemanfaatan data sekunder digital karena data yang diperoleh dari platform daring, media sosial, maupun basis data institusi sering kali memuat informasi yang dapat mengidentifikasi individu secara langsung maupun tidak langsung, sehingga penggunaannya berpotensi melanggar hak privasi apabila tidak dikelola secara hati-hati. Risiko tersebut semakin meningkat ketika data dikombinasikan dari berbagai sumber (*data linkage*) yang dapat menghasilkan profil individu secara lebih rinci tanpa persetujuan eksplisit dari subjek data.

Penerapan prinsip minimisasi data, anonimisasi, pembatasan tujuan penggunaan, serta kepatuhan terhadap regulasi perlindungan data menjadi langkah penting untuk menjaga integritas etis dalam penelitian maupun analisis berbasis data sekunder digital. Sebagaimana dijelaskan oleh Véliz (2020), perlindungan data pribadi bertujuan membatasi akses dan penggunaan informasi individu agar tidak disalahgunakan serta tetap menghormati hak otonomi dan privasi setiap orang.

### b. Persetujuan dan Izin Penggunaan (*Consent*)

Persetujuan dan izin penggunaan (*consent*) merupakan aspek etika fundamental dalam pemanfaatan data sekunder digital karena data yang tersedia secara daring atau dalam basis data institusi tidak selalu berarti bebas digunakan untuk semua tujuan, terutama apabila data tersebut berkaitan dengan individu yang

tidak secara eksplisit memberikan persetujuan untuk analisis lanjutan. Tantangan muncul ketika data dikumpulkan untuk tujuan awal tertentu namun kemudian dimanfaatkan kembali untuk kepentingan penelitian, komersial, atau analitik yang berbeda, sehingga berpotensi menimbulkan pelanggaran hak subjek data apabila tidak disertai mekanisme persetujuan yang sah dan transparan.

Prinsip *informed consent*, kejelasan tujuan penggunaan, serta pemberian informasi yang memadai kepada subjek data menjadi landasan penting dalam menjaga legitimasi etis penggunaan data sekunder digital. Sebagaimana dijelaskan oleh Saunders, Lewis, dan Thornhill (2019), persetujuan yang diinformasikan mengharuskan partisipan memahami tujuan dan implikasi penggunaan data sebelum memberikan izin, sehingga integritas dan tanggung jawab penelitian tetap terjaga.

c. Transparansi dan Akuntabilitas

Transparansi dan akuntabilitas dalam penggunaan data sekunder digital menuntut keterbukaan mengenai sumber data, metode pengumpulan, proses pembersihan, teknik analisis, serta batasan interpretasi agar pemanfaatan data dapat dipahami dan dievaluasi secara kritis oleh pihak lain. Tanpa transparansi, risiko manipulasi data, bias analitik, dan penyalahgunaan informasi meningkat karena pemangku kepentingan tidak memiliki akses yang memadai untuk menilai validitas maupun implikasi etis dari hasil yang dihasilkan.

Akuntabilitas juga mengharuskan individu atau organisasi yang mengelola data bertanggung jawab atas dampak sosial, hukum, dan etis dari keputusan berbasis data, termasuk kesiapan untuk diaudit dan dievaluasi sesuai standar tata kelola yang berlaku. Sebagaimana dinyatakan oleh Floridi dan Cowls (2019), prinsip tata kelola data yang etis menekankan transparansi dan pertanggungjawaban sebagai fondasi dalam memastikan sistem digital beroperasi secara adil dan dapat dipercaya.

d. Bias Algoritmik dan Representasi Data

Bias algoritmik dan representasi data merupakan tantangan etika penting dalam pemanfaatan data sekunder digital karena dataset yang digunakan untuk pelatihan model atau analisis sering kali tidak merepresentasikan populasi secara seimbang,

sehingga dapat menghasilkan keputusan yang diskriminatif atau tidak adil terhadap kelompok tertentu. Ketidakseimbangan distribusi data, keterbatasan cakupan demografis, serta bias historis yang tertanam dalam sumber data digital dapat memperkuat ketimpangan sosial ketika diproses melalui sistem algoritmik tanpa evaluasi kritis terhadap konteks dan struktur datanya.

Permasalahan ini semakin kompleks ketika algoritma dianggap netral padahal secara implisit merefleksikan pola dan preferensi yang terkandung dalam data pelatihan yang digunakan. Sebagaimana dijelaskan oleh O'Neil (2018), model berbasis data yang tidak diawasi secara etis dapat memperbesar ketidakadilan karena mengandalkan pola statistik yang mereproduksi bias sosial yang sudah ada.

# BAB IX

## WEB SCRAPING DAN WEB CRAWLING

---

*Web scraping* dan *web crawling* merupakan dua teknik penting dalam proses pengumpulan data digital yang memanfaatkan sumber daya informasi di internet secara sistematis dan terstruktur. *Web scraping* berfokus pada proses ekstraksi data spesifik dari halaman web untuk diolah lebih lanjut, sedangkan *web crawling* bertujuan menelusuri dan mengindeks berbagai halaman situs secara otomatis guna memetakan struktur serta ketersediaan informasi secara luas.

### A. Pengertian Web Scraping dan Web Crawling

*Web scraping* dan *web crawling* merupakan dua teknik utama dalam proses akuisisi data berbasis web yang sering digunakan dalam bidang data science, riset pasar, analisis bisnis, hingga pengembangan mesin pencari. Keduanya sama-sama memanfaatkan program atau bot otomatis, namun memiliki tujuan, mekanisme kerja, dan cakupan proses yang berbeda.

#### 1. Pengertian *Web Scraping*

*Web scraping* adalah teknik ekstraksi data dari halaman web secara otomatis dengan tujuan mengambil informasi tertentu yang dibutuhkan pengguna. Proses ini biasanya dilakukan dengan membaca struktur HTML suatu halaman, kemudian mengidentifikasi elemen-elemen spesifik seperti teks, tabel, gambar, tautan, atau metadata untuk dikumpulkan dan disimpan dalam format terstruktur seperti CSV, JSON, atau *database*. Secara rinci, *web scraping* memiliki karakteristik sebagai berikut:

- a. Berorientasi pada Data Spesifik

Berorientasi pada data spesifik merupakan karakteristik utama *web scraping* karena proses ini dirancang untuk mengekstraksi elemen informasi tertentu dari struktur halaman web, seperti harga produk, rating pengguna, judul berita, atau data tabular, tanpa harus mengambil seluruh isi situs secara menyeluruh. Fokus yang terarah ini memungkinkan sistem untuk mengidentifikasi pola pada struktur HTML atau DOM menggunakan selector tertentu sehingga data yang diperoleh bersifat terstruktur, relevan, dan siap digunakan untuk analisis lanjutan seperti data mining, *machine learning*, maupun business intelligence.

Pendekatan yang berorientasi pada data spesifik juga menekankan efisiensi dalam proses akuisisi data karena hanya bagian yang dibutuhkan yang diambil dan disimpan, sehingga mengurangi beban penyimpanan serta meningkatkan kecepatan pemrosesan data dalam skala besar. Hal ini sejalan dengan definisi yang menyatakan bahwa *web scraping* merupakan teknik otomatis untuk mengekstraksi informasi terstruktur dari situs web guna digunakan kembali dalam konteks lain (Mitchell, 2018), yang menegaskan bahwa orientasi utamanya bukan pada penelusuran halaman, melainkan pada pengambilan data yang telah ditargetkan secara jelas.

b. Mengandalkan Struktur HTML/DOM

Mengandalkan struktur HTML/DOM merupakan karakteristik fundamental dalam *web scraping* karena proses ekstraksi data dilakukan dengan membaca dan menavigasi elemen-elemen penyusun halaman web yang direpresentasikan dalam bentuk tag HTML dan hierarki *Document Object Model* (DOM). Melalui pemanfaatan selector seperti tag name, class, id, XPath, atau CSS selector, scraper dapat mengidentifikasi lokasi data secara presisi di dalam struktur dokumen sehingga informasi yang diambil tetap konsisten meskipun berada dalam halaman dengan konten yang kompleks dan berlapis.

Ketergantungan pada struktur HTML/DOM juga menuntut pemahaman terhadap bagaimana browser membangun representasi pohon dokumen dari kode sumber halaman, sebab perubahan kecil pada struktur atau atribut elemen dapat memengaruhi keberhasilan proses ekstraksi secara signifikan.

Hal ini sejalan dengan penjelasan Mitchell (2018) yang menyatakan bahwa *web scraping* bekerja dengan membaca struktur HTML halaman dan memanfaatkan pola elemen untuk mengambil data tertentu secara otomatis, sehingga efektivitas *scraping* sangat ditentukan oleh kemampuan dalam menganalisis dan menavigasi DOM secara akurat.

c. Digunakan untuk Analisis Lanjutan

Karakteristik *web scraping* yang digunakan untuk analisis lanjutan menegaskan bahwa proses ekstraksi data dari web bertujuan menghasilkan dataset terstruktur yang dapat diolah kembali melalui teknik statistik, analitik komputasional, maupun kecerdasan buatan untuk menemukan pola dan tren tertentu. Data yang diperoleh dari berbagai sumber daring seperti e-commerce, media sosial, atau portal berita dapat digabungkan, dibersihkan, dan dianalisis guna mendukung peramalan, segmentasi pasar, analisis sentimen, serta pengambilan keputusan berbasis data secara lebih akurat dan terukur.

Pemanfaatan *web scraping* untuk analisis lanjutan juga memperlihatkan perannya dalam ekosistem *big data*, di mana volume dan variasi informasi dari web menjadi sumber penting bagi organisasi untuk memperoleh keunggulan kompetitif melalui *insight* yang relevan dan aktual. Hal ini sejalan dengan pernyataan Krotov dan Silva (2018) yang menjelaskan bahwa *web scraping* memungkinkan organisasi mengumpulkan data eksternal berbasis web secara otomatis untuk dianalisis dalam mendukung kebutuhan bisnis dan penelitian, sehingga teknik ini menjadi bagian integral dari strategi data *analytics* modern.

d. Dapat Bersifat Statis Maupun Dinamis

Karakteristik *web scraping* yang dapat bersifat statis maupun dinamis menunjukkan bahwa teknik ini mampu mengekstraksi data baik dari halaman web yang kontennya langsung tersedia dalam kode HTML awal (statis) maupun dari halaman yang memuat data melalui eksekusi JavaScript atau permintaan asinkron setelah halaman dimuat (dinamis). Pada halaman statis, scraper cukup membaca respons HTML dari server, sedangkan pada halaman dinamis dibutuhkan pendekatan tambahan seperti headless browser atau otomatisasi rendering agar struktur DOM

yang telah diperbarui oleh skrip dapat diakses dan diekstraksi secara akurat.

Perbedaan ini menuntut pemahaman teknis terhadap mekanisme komunikasi *client-server* serta bagaimana browser memproses dan memperbarui *Document Object Model* setelah menerima data tambahan dari API atau sumber lain di belakang layar. Sebagaimana dijelaskan oleh Russell (2018) bahwa pengambilan data web modern sering kali melibatkan interaksi dengan konten yang dihasilkan secara dinamis melalui JavaScript dan layanan web, sehingga strategi *scraping* perlu menyesuaikan diri dengan arsitektur aplikasi web yang terus berkembang.

## 2. Pengertian *Web Crawling*

*Web crawling* adalah proses penelusuran dan pengindeksan halaman-halaman web secara otomatis dengan mengikuti tautan (*hyperlink*) dari satu halaman ke halaman lainnya. Teknik ini biasanya digunakan untuk memetakan struktur situs atau mengumpulkan daftar URL yang tersedia dalam suatu domain atau di seluruh internet. Secara rinci, *web crawling* memiliki karakteristik sebagai berikut:

### a. Berorientasi pada Penelusuran Halaman

Berorientasi pada penelusuran halaman merupakan karakteristik utama *web crawling* karena proses ini dirancang untuk menemukan, mengunjungi, dan mengindeks halaman-halaman web dengan mengikuti struktur tautan (*hyperlink*) yang saling terhubung dalam suatu situs maupun antar situs. Crawler bekerja secara sistematis dengan memulai dari sekumpulan URL awal (seed URLs), kemudian mengekstrak tautan pada setiap halaman yang dikunjungi untuk memperluas cakupan penelusuran sehingga terbentuk peta struktur web yang komprehensif dan terorganisasi.

Orientasi pada penelusuran ini menekankan cakupan dan kontinuitas proses eksplorasi dibandingkan pengambilan data spesifik, karena tujuan utamanya adalah membangun indeks atau repositori halaman yang dapat digunakan untuk sistem pencarian, analisis jaringan, maupun pemantauan konten secara berkala. Hal ini sejalan dengan penjelasan Olston dan Najork (2018) yang menyatakan bahwa *web crawling* merupakan proses otomatis untuk menemukan dan mengambil halaman web dengan

mengikuti tautan guna membangun koleksi dokumen yang luas dan terstruktur untuk kebutuhan pengindeksan dan pencarian informasi.

b. Menggunakan Bot atau Crawler (Spider)

Karakteristik *web crawling* yang menggunakan bot atau crawler (spider) menunjukkan bahwa proses penelusuran halaman web dilakukan oleh program otomatis yang dirancang untuk mengakses, membaca, dan mengikuti tautan secara berulang tanpa intervensi manusia. Bot ini bekerja dengan mekanisme terjadwal dan terstruktur, dimulai dari kumpulan URL awal, kemudian mengunduh konten halaman, mengekstraksi hyperlink, serta menambahkan tautan baru ke dalam antrian penelusuran sehingga proses eksplorasi dapat berlangsung secara berkelanjutan dan sistematis dalam skala besar.

Penggunaan crawler atau spider juga memungkinkan efisiensi tinggi dalam pengindeksan jutaan hingga miliaran halaman web karena bot mampu beroperasi dengan algoritma tertentu yang mengatur prioritas kunjungan, frekuensi akses, serta kepatuhan terhadap protokol seperti robots.txt untuk menjaga etika dan stabilitas sistem. Sebagaimana dijelaskan oleh Olston dan Najork (2018) bahwa web crawler adalah program otomatis yang secara sistematis menelusuri web dengan mengambil halaman dan mengikuti tautan untuk membangun koleksi dokumen berskala besar, yang menegaskan bahwa peran bot menjadi fondasi utama dalam arsitektur mesin pencari modern.

c. Digunakan dalam Mesin Pencari

Karakteristik *web crawling* yang digunakan dalam mesin pencari menunjukkan bahwa proses penelusuran halaman web merupakan tahap fundamental dalam membangun indeks pencarian yang memungkinkan pengguna menemukan informasi secara cepat dan relevan. Crawler mesin pencari bekerja dengan mengunjungi halaman-halaman web, mengunduh kontennya, serta menyimpan representasi dokumen dan struktur tautannya ke dalam sistem indeks yang kemudian diproses oleh algoritma peringkat untuk menentukan urutan hasil pencarian.

Peran crawling dalam mesin pencari juga mencakup pembaruan berkala terhadap konten yang telah diindeks agar perubahan informasi di web dapat segera tercermin dalam hasil pencarian, sehingga akurasi dan relevansi tetap terjaga dalam lingkungan web yang dinamis. Sebagaimana dijelaskan oleh Manning, Raghavan, dan Schütze (2021) bahwa mesin pencari membangun indeks dokumen web melalui proses crawling yang mengumpulkan dan memproses halaman sebelum dilakukan tahap pengindeksan dan pemeringkatan, yang menegaskan bahwa crawling merupakan fondasi operasional sistem pencarian modern.

d. Bersifat Luas dan Sistematis

Karakteristik *web crawling* yang bersifat luas dan sistematis menunjukkan bahwa proses penelusuran dilakukan dengan cakupan yang besar, mencakup banyak domain dan jutaan halaman, serta mengikuti strategi terstruktur dalam menentukan urutan kunjungan dan prioritas URL. Crawler dirancang menggunakan algoritma tertentu seperti breadth-first search atau pendekatan berbasis prioritas untuk memastikan eksplorasi berjalan terorganisasi, menghindari duplikasi kunjungan, serta mengelola sumber daya jaringan dan penyimpanan secara efisien dalam skala besar.

Sifat luas dan sistematis ini memungkinkan mesin pencari maupun sistem pengindeksan lainnya membangun repositori dokumen web yang komprehensif dan terus diperbarui sesuai perubahan konten di internet. Hal ini sejalan dengan penjelasan Olston dan Najork (2018) yang menyatakan bahwa *web crawling* dilakukan secara terstruktur dan berskala besar untuk mengumpulkan halaman web melalui strategi penjadwalan dan pemilihan URL yang efisien, sehingga mampu mendukung pembangunan indeks informasi yang andal dan menyeluruh.

## B. Perbedaan Web Scraping dan Web Crawling

*Web scraping* dan *web crawling* merupakan dua teknik pengumpulan data berbasis web yang sering digunakan secara bersamaan, namun memiliki tujuan, mekanisme kerja, serta cakupan proses yang berbeda secara mendasar. Perbedaan keduanya dapat

dipahami dari aspek orientasi, fungsi, ruang lingkup, hingga hasil akhir yang dihasilkan dari proses tersebut. Berikut penjelasan rinci mengenai perbedaan *web scraping* dan *web crawling*:

## 1. Perbedaan Berdasarkan Tujuan Utama

Perbedaan berdasarkan tujuan utama antara *web crawling* dan *web scraping* terletak pada orientasi akhir dari proses otomatis yang dijalankan terhadap sumber daya web, di mana *web crawling* bertujuan untuk menemukan, mengumpulkan, dan memperbarui kumpulan halaman secara luas guna membangun indeks atau basis dokumen, sedangkan *web scraping* berfokus pada pengambilan data tertentu yang telah ditargetkan dari halaman-halaman tersebut untuk kebutuhan analisis spesifik. *Web crawling* menekankan pada eksplorasi menyeluruh terhadap struktur tautan dan ketersediaan halaman dalam suatu domain atau di seluruh web, sementara *web scraping* menekankan pada selektivitas dan presisi dalam mengekstraksi elemen data seperti teks, angka, tabel, atau metadata yang relevan dengan tujuan tertentu. Dengan demikian, perbedaan tujuan ini menunjukkan bahwa *crawling* berorientasi pada cakupan dan pemetaan informasi, sedangkan *scraping* berorientasi pada ekstraksi konten yang siap diolah lebih lanjut.

Pada konteks sistem informasi modern, *web crawling* menjadi fondasi bagi mesin pencari dan sistem pengindeksan karena berfungsi membangun koleksi dokumen yang dapat diproses dan diberi peringkat, sedangkan *web scraping* lebih banyak dimanfaatkan dalam analisis bisnis, penelitian, dan pengembangan aplikasi berbasis data untuk memperoleh *insight* dari informasi web yang terstruktur. Perbedaan orientasi tujuan tersebut juga berdampak pada desain arsitektur sistem, algoritma, serta strategi pengelolaan data yang digunakan dalam masing-masing teknik. Hal ini sejalan dengan pernyataan Manning, Raghavan, dan Schütze (2021) yang menjelaskan bahwa mesin pencari bergantung pada proses *crawling* untuk mengumpulkan dokumen web sebelum dilakukan pengindeksan dan pemeringkatan, yang menegaskan bahwa tujuan *crawling* berbeda secara mendasar dari ekstraksi data spesifik seperti pada *web scraping*.

## 2. Perbedaan Berdasarkan Orientasi Proses

Perbedaan berdasarkan orientasi proses antara *web crawling* dan *web scraping* terletak pada bagaimana masing-masing teknik

menjalankan tahapan operasionalnya dalam mengakses dan memanfaatkan sumber daya web, di mana *web crawling* berorientasi pada proses navigasi dan eksplorasi tautan secara berkelanjutan untuk memperluas cakupan halaman yang dikunjungi, sedangkan *web scraping* berorientasi pada proses identifikasi, seleksi, dan ekstraksi elemen data tertentu dari struktur halaman yang telah diakses. Dalam *web crawling*, fokus prosesnya adalah mengelola antrean URL, menentukan prioritas kunjungan, serta menjaga efisiensi dan cakupan eksplorasi secara sistematis, sementara dalam *web scraping* prosesnya lebih terpusat pada analisis struktur HTML atau DOM guna memastikan data yang diambil sesuai dengan kriteria yang telah ditentukan. Dengan demikian, orientasi proses *crawling* menekankan perluasan jaringan halaman, sedangkan *scraping* menekankan pendalaman isi halaman.

Secara teknis, perbedaan orientasi ini memengaruhi desain algoritma, penggunaan sumber daya, serta metode pemrosesan data yang diterapkan pada masing-masing teknik dalam sistem berbasis web. *Web crawling* dirancang untuk bekerja secara iteratif dan luas dengan strategi traversal tertentu, sedangkan *web scraping* dirancang untuk bekerja secara selektif dan presisi terhadap elemen konten tertentu yang menjadi target ekstraksi. Hal ini sejalan dengan penjelasan Russell (2018) yang menyatakan bahwa proses pengambilan data dari web dapat berfokus pada eksplorasi sumber daya secara luas maupun pada ekstraksi konten spesifik dari halaman tertentu, yang menunjukkan adanya perbedaan mendasar dalam orientasi proses antara *crawling* dan *scraping*.

### **3. Perbedaan Berdasarkan Ruang Lingkup**

Perbedaan berdasarkan ruang lingkup antara *web crawling* dan *web scraping* terlihat dari cakupan area kerja masing-masing teknik dalam mengakses dan memproses sumber daya web, di mana *web crawling* memiliki ruang lingkup yang luas karena dirancang untuk melihat banyak halaman bahkan lintas domain melalui mekanisme penelusuran tautan yang berkelanjutan, sedangkan *web scraping* memiliki ruang lingkup yang lebih sempit dan terfokus karena hanya menargetkan halaman atau bagian tertentu sesuai kebutuhan data yang telah ditentukan. *Web crawling* dapat mencakup ribuan hingga miliaran halaman dalam satu sistem terdistribusi untuk membangun indeks yang komprehensif, sementara *web scraping* biasanya dibatasi pada subset halaman yang relevan dengan topik, kata kunci, atau sumber tertentu

sehingga prosesnya lebih terkontrol dan spesifik. Dengan demikian, perbedaan ruang lingkup ini menunjukkan bahwa crawling beroperasi pada level makro dalam struktur web, sedangkan *scraping* bekerja pada level mikro dalam konten halaman.

Pada praktiknya, ruang lingkup crawling yang luas menuntut pengelolaan sumber daya jaringan, penyimpanan, serta strategi penjadwalan yang kompleks agar proses tetap efisien dan tidak membebani server target, sedangkan *scraping* lebih menitikberatkan pada ketepatan pemilihan elemen data daripada perluasan jumlah halaman yang dikunjungi. Perbedaan ini juga memengaruhi desain arsitektur sistem, di mana crawling sering menggunakan pendekatan terdistribusi untuk menangani skala besar, sementara *scraping* dapat dijalankan pada skala lebih kecil namun dengan analisis struktur halaman yang lebih mendalam. Hal ini sejalan dengan penjelasan Olston dan Najork (2018) yang menyatakan bahwa *web crawling* dilakukan dalam skala besar untuk mengumpulkan koleksi halaman web yang luas melalui strategi sistematis, yang menegaskan perbedaan ruang lingkungannya dibandingkan praktik ekstraksi data terarah seperti pada *web scraping*.

#### **4. Perbedaan Berdasarkan Mekanisme Kerja**

Perbedaan berdasarkan mekanisme kerja antara *web crawling* dan *web scraping* terletak pada tahapan operasional serta logika sistem yang digunakan dalam mengakses dan memproses informasi dari web, di mana *web crawling* bekerja dengan mekanisme penelusuran otomatis yang mengelola antrean URL, mengunduh halaman, serta mengekstraksi tautan baru untuk memperluas jangkauan eksplorasi secara berulang dan terstruktur. Sebaliknya, *web scraping* bekerja dengan mekanisme parsing dokumen yang berfokus pada analisis struktur HTML atau DOM dari halaman tertentu untuk mengidentifikasi dan mengekstraksi elemen data spesifik menggunakan selector atau pola tertentu yang telah ditentukan sebelumnya. Dengan demikian, crawling mengutamakan proses navigasi dan penemuan halaman secara sistematis, sedangkan *scraping* mengutamakan proses ekstraksi konten secara selektif dan presisi.

Secara teknis, crawler biasanya dilengkapi dengan modul pengatur jadwal (*scheduler*), manajemen antrean (*queue manager*), serta sistem deduplikasi URL untuk memastikan efisiensi dan menghindari

pengulangan kunjungan, sementara scraper lebih menitikberatkan pada teknik parsing, validasi struktur data, serta transformasi hasil ekstraksi ke dalam format terstruktur seperti JSON atau basis data. Perbedaan mekanisme ini mencerminkan bahwa crawling beroperasi pada tingkat jaringan dan struktur tautan, sedangkan *scraping* beroperasi pada tingkat konten dan elemen dokumen halaman. Hal ini sejalan dengan penjelasan Mitchell (2018) yang menyatakan bahwa *web scraping* berfokus pada teknik penguraian (parsing) dokumen HTML untuk mengambil data tertentu dari halaman web, yang secara implisit membedakan mekanismenya dari proses crawling yang berorientasi pada penelusuran tautan.

## 5. Perbedaan Berdasarkan Hasil Akhir

Perbedaan berdasarkan hasil akhir antara *web crawling* dan *web scraping* dapat dilihat dari bentuk *output* yang dihasilkan setelah proses otomatisasi selesai dijalankan, di mana *web crawling* menghasilkan kumpulan URL, dokumen web, serta metadata seperti struktur tautan dan informasi indeks yang digunakan untuk membangun repositori pencarian berskala besar. Sebaliknya, *web scraping* menghasilkan data terstruktur yang telah diekstraksi dari elemen tertentu dalam halaman web, seperti teks, angka, tabel, atau atribut spesifik, yang kemudian disimpan dalam format siap olah seperti CSV, JSON, atau basis data relasional untuk keperluan analisis lanjutan. Dengan demikian, hasil akhir crawling bersifat dokumentatif dan berorientasi pada pengindeksan, sedangkan hasil akhir *scraping* bersifat analitis dan berorientasi pada pemanfaatan data secara langsung.

Secara konseptual, crawling berkontribusi pada pembangunan infrastruktur informasi yang memungkinkan sistem pencarian dan pemetaan web berfungsi secara efektif, sementara *scraping* berkontribusi pada penyediaan dataset yang dapat dimanfaatkan untuk kebutuhan riset, intelijen bisnis, dan pengembangan model prediktif. Perbedaan hasil akhir ini menunjukkan bahwa meskipun keduanya sama-sama mengakses sumber daya web, produk data yang dihasilkan memiliki tujuan penggunaan yang berbeda dalam ekosistem pengolahan informasi digital. Hal ini sejalan dengan penjelasan Manning, Raghavan, dan Schütze (2021) yang menyatakan bahwa proses crawling dalam mesin pencari bertujuan mengumpulkan dokumen untuk dibangun menjadi indeks pencarian, yang menegaskan bahwa *output* crawling

berbeda secara fundamental dari ekstraksi data terstruktur seperti pada *web scraping*.

## C. Arsitektur dan Alur Kerja

Arsitektur dan alur kerja dalam sistem pengumpulan data berbasis web, baik untuk *web crawling* maupun *web scraping*, merupakan kerangka teknis yang mengatur bagaimana data diakses, diproses, dan disimpan secara sistematis. Pemahaman terhadap arsitektur dan alur kerja penting karena menentukan efisiensi, skalabilitas, dan akurasi dalam memperoleh data dari berbagai sumber *online*.

### 1. Arsitektur

Arsitektur dalam sistem *web crawling* dan *web scraping* merupakan rancangan struktural yang menggambarkan bagaimana komponen-komponen perangkat lunak saling terhubung dan bekerja sama untuk mengakses, memproses, serta menyimpan data dari web secara otomatis. Arsitektur yang baik bersifat modular, skalabel, dan mampu menangani volume data besar serta dinamika perubahan halaman web, sehingga sistem dapat beroperasi secara efisien, stabil, dan berkelanjutan. Berikut penjelasan rinci mengenai komponen utama dalam arsitektur tersebut:

#### a. Komponen Pengambil Data (*Fetching Layer*)

Komponen Pengambil Data (*Fetching Layer*) merupakan bagian fundamental dalam arsitektur *web crawling* dan *web scraping* karena bertanggung jawab melakukan komunikasi langsung dengan server web melalui protokol HTTP/HTTPS untuk mengirim permintaan (*request*) dan menerima respons berupa dokumen HTML, JSON, atau format data lainnya yang akan diproses lebih lanjut. Lapisan ini harus mampu mengelola konfigurasi teknis seperti pengaturan header, user-agent, manajemen koneksi, penanganan redirect, kontrol kecepatan akses, serta mekanisme retry ketika terjadi kegagalan koneksi agar proses pengambilan data berjalan efisien, stabil, dan tetap mematuhi kebijakan akses situs, sebagaimana dijelaskan oleh Mitchell (2018) bahwa tahap awal dalam *web scraping* melibatkan pengambilan halaman web melalui permintaan HTTP sebelum dilakukan proses parsing dan ekstraksi data.

b. *Crawler Engine*

*Crawler Engine* merupakan komponen inti dalam arsitektur *web crawling* yang berfungsi mengoordinasikan proses penelusuran halaman dengan mengelola daftar URL awal, mengekstraksi tautan dari setiap halaman yang dikunjungi, serta mengatur logika kunjungan ulang untuk memastikan cakupan eksplorasi yang luas dan efisien. Komponen ini bekerja dengan algoritma traversal tertentu, mekanisme deduplikasi URL, serta kebijakan penjadwalan akses agar proses pengumpulan dokumen web berlangsung sistematis dan terkontrol, sebagaimana dijelaskan oleh Olston dan Najork (2018) bahwa crawler adalah sistem otomatis yang mengunduh halaman web dan mengikuti tautan untuk membangun koleksi dokumen dalam skala besar guna keperluan pengindeksan.

c. *Scheduler dan Queue Manager*

*Scheduler dan Queue Manager* merupakan komponen penting dalam arsitektur *web crawling* karena bertanggung jawab mengatur antrian URL yang akan dikunjungi, menentukan prioritas akses berdasarkan kebijakan tertentu, serta mengendalikan frekuensi permintaan agar proses penelusuran berjalan efisien dan tidak membebani server target. Komponen ini memastikan distribusi tugas *crawling* berlangsung sistematis melalui mekanisme penjadwalan, penghindaran duplikasi, serta pengaturan ulang kunjungan halaman untuk pembaruan konten, sebagaimana dijelaskan oleh Manning, Raghavan, dan Schütze (2021) bahwa sistem *crawling* memerlukan pengelolaan antrian dan strategi penjadwalan untuk mengatur dokumen mana yang harus diambil dan kapan proses tersebut dilakukan dalam skala besar.

d. *Parser dan Extractor*

*Parser dan Extractor* merupakan komponen utama dalam arsitektur *web scraping* yang berfungsi menganalisis struktur dokumen HTML atau DOM setelah halaman berhasil diambil oleh fetching layer, kemudian mengidentifikasi dan mengekstraksi elemen data tertentu berdasarkan pola atau selector yang telah ditentukan seperti tag, class, id, XPath, atau CSS selector. Komponen ini memastikan bahwa data mentah dari halaman web dapat diubah menjadi informasi terstruktur yang

siap diproses lebih lanjut, sebagaimana dijelaskan oleh Russell (2018) bahwa proses ekstraksi data web melibatkan penguraian struktur halaman untuk mengambil konten spesifik yang relevan dengan kebutuhan analisis.

e. *Data Processing Layer*

*Data Processing Layer* merupakan komponen utama dalam arsitektur *web crawling* dan *web scraping* yang bertugas mengolah data mentah hasil ekstraksi agar menjadi terstruktur, bersih, dan siap digunakan untuk analisis lanjutan melalui proses seperti pembersihan (*cleaning*), transformasi, normalisasi, deduplikasi, serta validasi format data. Lapisan ini berperan penting dalam menjamin kualitas dan konsistensi dataset sebelum disimpan atau diintegrasikan ke sistem analitik, sebagaimana dijelaskan oleh Rahm dan Do (2018) bahwa proses pembersihan dan transformasi data merupakan tahap krusial dalam memastikan kualitas data agar dapat digunakan secara efektif dalam sistem informasi dan analitik.

f. *Storage System*

*Storage System* merupakan komponen utama dalam arsitektur *web crawling* dan *web scraping* yang berfungsi menyimpan data hasil ekstraksi dan pemrosesan secara terstruktur maupun tidak terstruktur ke dalam media penyimpanan seperti *database* relasional, NoSQL, data warehouse, maupun distributed file system agar dapat diakses kembali untuk analisis, integrasi, maupun visualisasi lebih lanjut. Peran sistem penyimpanan ini sangat krusial dalam menjamin skalabilitas, konsistensi, ketersediaan, serta keandalan data dalam lingkungan *big data*, sebagaimana dinyatakan oleh Kleppmann (2018) bahwa sistem penyimpanan data modern harus dirancang untuk mendukung reliabilitas, skalabilitas, dan konsistensi dalam pengelolaan data berskala besar pada sistem terdistribusi.

## 2. Alur Kerja

Alur kerja dalam sistem *web crawling* dan *web scraping* merupakan rangkaian tahapan operasional yang menggambarkan bagaimana data diambil dari web, diproses, hingga disimpan dalam bentuk yang dapat dimanfaatkan lebih lanjut. Alur kerja yang terstruktur memastikan proses berjalan sistematis, efisien, serta mampu menangani

perubahan konten web secara dinamis tanpa mengorbankan kualitas data yang dihasilkan. Berikut penjelasan rinci mengenai tahapan alur kerja tersebut:

a. Penentuan Tujuan dan Sumber Data

Penentuan tujuan dan sumber data merupakan tahapan awal dalam alur kerja *web crawling* dan *web scraping* yang berfungsi menetapkan fokus pengumpulan data serta memilih situs atau halaman web yang relevan dengan kebutuhan analisis, sehingga proses selanjutnya dapat berjalan efisien dan tepat sasaran. Tahap ini mencakup identifikasi jenis data yang dibutuhkan, penentuan domain atau URL target, serta pengecekan kepatuhan terhadap kebijakan akses situs, sebagaimana dijelaskan oleh Russell (2018) bahwa penentuan sumber dan tujuan pengumpulan data merupakan langkah krusial dalam memastikan relevansi dan kualitas dataset yang diperoleh dari web sebelum dilakukan ekstraksi lebih lanjut.

b. Inisialisasi URL (Seed URLs)

Inisialisasi URL (seed URLs) merupakan tahapan awal dalam alur kerja *web crawling* yang berfungsi sebagai titik awal penelusuran sistem untuk mengakses dan mengidentifikasi halaman-halaman web yang akan diproses lebih lanjut, di mana daftar URL awal ini dimasukkan ke dalam antrian (*queue*) agar crawler dapat mulai melakukan pengambilan konten serta menemukan tautan-tautan baru secara sistematis. Tahap ini sangat menentukan cakupan dan arah penelusuran data karena kualitas dan relevansi seed URLs akan memengaruhi struktur graf penelusuran yang terbentuk, sebagaimana dijelaskan oleh Olston dan Najork (2018) bahwa pemilihan URL awal dalam sistem crawling berperan penting dalam menentukan efektivitas eksplorasi web dan distribusi sumber daya perayapan pada sistem berskala besar.

c. Proses *Crawling* (Penelusuran Halaman)

Proses *crawling* (penelusuran halaman) merupakan tahapan inti dalam alur kerja *web crawling* yang melibatkan pengambilan halaman web secara otomatis melalui permintaan HTTP, diikuti dengan identifikasi dan ekstraksi tautan (*hyperlink*) yang terdapat pada halaman tersebut untuk dimasukkan kembali ke dalam antrian sehingga sistem dapat menelusuri struktur web secara

berkelanjutan dan terstruktur. Tahapan ini harus dikelola dengan mekanisme pengaturan frekuensi akses, penghindaran duplikasi URL, serta pemenuhan kebijakan robots.txt agar efisien dan etis, sebagaimana dijelaskan oleh Baeza-Yates dan Ribeiro-Neto (2019) bahwa crawling merupakan proses sistematis untuk mengunduh halaman web dan mengikuti tautan yang ditemukan guna membangun koleksi dokumen yang dapat diproses lebih lanjut dalam sistem temu kembali informasi.

d. Parsing dan Identifikasi Struktur

Parsing dan identifikasi struktur merupakan tahapan alur kerja setelah proses crawling yang bertujuan untuk membahas dokumen HTML atau XML ke dalam representasi terstruktur seperti *Document Object Model* (DOM) sehingga elemen-elemen penting seperti judul, paragraf, tabel, metadata, maupun atribut tertentu dapat dikenali dan dipisahkan sesuai kebutuhan ekstraksi data. Tahap ini sangat menentukan akurasi hasil karena pemahaman terhadap struktur hierarkis dan tag markup memungkinkan sistem mengidentifikasi pola konten secara presisi, sebagaimana dijelaskan oleh Kleppmann (2018) bahwa data yang diperoleh dari berbagai sumber perlu diuraikan dan ditransformasikan ke dalam struktur yang dapat diproses secara sistematis agar dapat dianalisis dan disimpan secara andal dalam sistem berbasis data.

e. Proses *Scraping* (Ekstraksi Data)

Proses *scraping* (ekstraksi data) merupakan tahapan lanjutan dalam alur kerja setelah parsing, yang berfungsi mengambil elemen-elemen spesifik dari struktur dokumen web seperti teks, angka, atribut, maupun metadata berdasarkan pola tertentu menggunakan teknik selektor, ekspresi reguler, atau query DOM sehingga data yang sebelumnya tidak terstruktur dapat diubah menjadi format terorganisir dan siap dianalisis. Tahap ini menuntut ketepatan dalam menentukan aturan ekstraksi agar relevansi dan akurasi data terjaga, sebagaimana dijelaskan oleh Russell (2018) bahwa proses ekstraksi data dari web melibatkan identifikasi pola konten yang konsisten untuk mengubah informasi semi-terstruktur menjadi dataset terstruktur yang dapat diproses lebih lanjut dalam analisis data.

f. Pembersihan dan Normalisasi Data

Pembersihan dan normalisasi data merupakan tahapan penting dalam alur kerja setelah proses ekstraksi yang bertujuan menghilangkan data duplikat, memperbaiki kesalahan format, menangani nilai yang hilang, serta menyeragamkan struktur dan representasi data agar konsisten serta siap digunakan dalam analisis lebih lanjut. Tahap ini memastikan kualitas dan integritas dataset sehingga hasil analitik menjadi lebih akurat dan dapat diandalkan, sebagaimana dijelaskan oleh Rahm dan Do (2018) bahwa proses data cleaning berfokus pada deteksi dan koreksi kesalahan serta inkonsistensi dalam data untuk meningkatkan kualitas dan kegunaan data dalam sistem informasi.

## D. Tools dan Teknologi Web Scraping

*Tools* dan teknologi *web scraping* merupakan seperangkat perangkat lunak, pustaka (*library*), *framework*, serta infrastruktur pendukung yang digunakan untuk mengekstraksi data dari situs web secara otomatis, terstruktur, dan efisien. Pemilihan *tools* yang tepat sangat bergantung pada kompleksitas situs, jenis data yang diambil (statis atau dinamis), skala proyek, serta kebutuhan integrasi dengan sistem analitik atau penyimpanan data. Secara rinci, berikut adalah tools dan teknologi utama dalam *web scraping* beserta deskripsinya:

### 1. Bahasa Pemrograman

Bahasa pemrograman adalah sebuah formal language yang dirancang untuk menulis instruksi-instruksi yang dapat dipahami dan dijalankan oleh komputer dalam rangka mengendalikan perilaku perangkat lunak atau sistem yang dikembangkan, sehingga memungkinkan programmer memetakan logika dan algoritma menjadi rangkaian perintah konkret yang tereksekusi oleh mesin. Menurut definisi dalam literatur teknik komputer, “*A programming language is an engineered language for expressing computer programs*” yang menunjukkan bahwa bahasa pemrograman dirancang sebagai sistem komunikasi antara manusia dan komputer melalui struktur sintaks dan semantik yang terstandarisasi untuk tujuan pembuatan perangkat lunak yang efektif dan efisien (Wikipedia, 2026).

Bahasa pemrograman memiliki aturan baku berupa sintaks dan semantik yang membedakannya dari bahasa alami, di mana sintaks mengatur bentuk penulisan kode dan semantik menentukan makna dari setiap konstruksi yang ditulis, sehingga program yang dihasilkan konsisten dan dapat dieksekusi sesuai dengan tujuan perintah. Beragam bahasa pemrograman dikembangkan dengan paradigma dan tujuan yang berbeda, mulai dari bahasa tingkat rendah yang dekat dengan hardware hingga bahasa tingkat tinggi yang lebih mudah dipahami manusia, dengan setiap jenis bahasa membawa kekuatan tersendiri dalam konteks penggunaan tertentu.

## **2. *Library dan Framework Web Scraping***

*Library* dan *framework web scraping* merupakan komponen perangkat lunak yang dirancang untuk mempermudah proses pengambilan, penguraian, dan pengolahan data dari halaman web secara otomatis dengan menyediakan fungsi-fungsi siap pakai yang mengurangi kebutuhan penulisan kode dari awal. Dalam praktiknya, penggunaan *library* seperti BeautifulSoup atau Requests dalam Python memungkinkan pengembang melakukan parsing HTML dan pengiriman HTTP request secara efisien, sementara *framework* seperti Scrapy menyediakan arsitektur terstruktur untuk manajemen crawling, pipeline data, dan pengendalian proses *scraping* skala besar, sebagaimana dijelaskan oleh Mitchell (2018) bahwa *web scraping* modern memanfaatkan kombinasi request handling dan HTML parsing untuk mengekstrak data secara sistematis dari situs web yang kompleks.

Secara teknis, *library web scraping* berfungsi sebagai kumpulan modul yang dapat dipanggil sesuai kebutuhan tertentu seperti membaca struktur DOM, menavigasi elemen berdasarkan tag atau atribut, serta mengekstrak teks dan metadata dengan presisi tinggi tanpa harus membangun sistem secara menyeluruh. Sebaliknya, *framework* menyediakan kerangka kerja terpadu yang mencakup pengaturan alur crawling, pengelolaan antrian URL, pengaturan *concurrency*, *middleware*, serta integrasi dengan sistem penyimpanan data sehingga lebih cocok untuk proyek berskala besar dan berkelanjutan.

## **3. *HTTP Client dan Request Handling***

HTTP *client* dan *request handling* merupakan komponen fundamental dalam arsitektur komunikasi web yang berfungsi untuk

mengirimkan permintaan (*request*) dari sisi klien ke server serta menerima dan mengelola respons (*response*) yang dikembalikan oleh server berdasarkan protokol HTTP atau HTTPS. Dalam konteks pengambilan data digital, HTTP client seperti Requests pada Python atau Axios pada JavaScript memungkinkan pengembang mengatur metode permintaan seperti GET, POST, PUT, dan DELETE, sekaligus mengelola header, parameter kueri, autentikasi, dan enkripsi sehingga interaksi dengan server berlangsung secara terstruktur dan sesuai standar protokol jaringan modern.

Secara teknis, request handling mencakup proses pembentukan paket permintaan yang berisi URL tujuan, metode HTTP, header tambahan seperti *User-Agent* atau *Authorization*, serta payload data apabila diperlukan, kemudian memproses respons yang diterima dalam bentuk kode status, header respons, dan body konten seperti HTML atau JSON. Fielding dan Reschke (2022) dalam spesifikasi HTTP Semantics RFC 9110 menjelaskan bahwa HTTP merupakan protokol aplikasi berbasis pesan yang mendefinisikan metode, kode status, dan representasi sumber daya untuk memungkinkan komunikasi terstandarisasi antara klien dan server, sehingga mekanisme request dan response menjadi inti dari pertukaran informasi di web modern.

Implementasi HTTP client yang efektif juga melibatkan pengelolaan error handling seperti timeout, koneksi terputus, atau respons dengan kode status 4xx dan 5xx, serta penerapan strategi seperti retry mechanism dan rate limiting untuk menjaga stabilitas sistem dan mencegah beban berlebih pada server tujuan. Dengan pengelolaan request handling yang tepat, proses komunikasi data menjadi lebih andal, aman, dan efisien, terutama dalam sistem otomatisasi seperti *web scraping*, integrasi API, dan layanan berbasis microservices yang sangat bergantung pada pertukaran data melalui protokol HTTP.

#### **4. Parsing dan Ekstraksi Data**

Parsing dan ekstraksi data merupakan tahapan krusial dalam proses pengolahan informasi digital yang bertujuan untuk mengubah data mentah, seperti dokumen HTML, XML, atau JSON, menjadi struktur yang lebih terorganisasi sehingga dapat dipahami, dianalisis, dan disimpan secara sistematis. Dalam konteks komputasi web, parsing dilakukan dengan membaca struktur sintaks suatu dokumen berdasarkan aturan tertentu untuk mengidentifikasi elemen, atribut, dan hierarki data,

kemudian ekstraksi data memfokuskan pada pengambilan informasi spesifik yang relevan sesuai kebutuhan pengguna atau sistem.

Secara teknis, proses parsing memanfaatkan parser yang bekerja dengan membangun representasi pohon (*tree structure*) seperti *Document Object Model* (DOM) sehingga setiap elemen dalam dokumen dapat diakses melalui jalur tertentu menggunakan XPath atau CSS selector, sementara teknik ekstraksi memanfaatkan pola tertentu termasuk ekspresi reguler untuk menyaring teks, angka, atau metadata dari struktur tersebut. Sebagaimana dijelaskan oleh Grune dan Jacobs (2019) bahwa parsing adalah proses analisis struktur sintaktis dari rangkaian simbol sesuai dengan aturan tata bahasa formal untuk menentukan struktur hierarkisnya, yang menegaskan bahwa parsing menjadi dasar utama dalam memahami dan mengekstrak informasi dari dokumen terstruktur (*Parsing Techniques: A Practical Guide*, 2019).

## 5. *Headless Browser dan Rendering Engine*

*Headless browser* dan *rendering engine* merupakan komponen penting dalam ekosistem aplikasi web modern yang memungkinkan proses pemuatan, interpretasi, dan penampilan konten web tanpa antarmuka grafis pengguna, sehingga sangat berguna untuk otomatisasi, pengujian, dan pengambilan data dari situs berbasis JavaScript dinamis. *Headless browser* adalah peramban yang berjalan tanpa tampilan visual tetapi tetap menjalankan seluruh fungsi seperti pemrosesan HTML, CSS, dan JavaScript, sedangkan *rendering engine* bertugas menerjemahkan kode sumber halaman web menjadi representasi visual atau struktur internal yang dapat diproses oleh sistem.

Secara teknis, *rendering engine* seperti Blink (digunakan pada Chromium) atau WebKit bekerja dengan memarsing dokumen HTML untuk membangun *Document Object Model* (DOM), memproses CSS untuk membentuk *CSS Object Model* (CSSOM), lalu menggabungkannya menjadi *render tree* yang selanjutnya diproses melalui tahap layout dan painting untuk menghasilkan tampilan halaman. Google Developers (2020) menjelaskan bahwa mesin *rendering* pada Chrome memproses HTML dan CSS untuk membangun struktur DOM dan *render tree* sebelum melakukan layout dan paint guna menampilkan halaman secara utuh, yang menunjukkan bahwa proses *rendering* merupakan tahapan sistematis dalam memvisualisasikan konten web berbasis standar terbuka.

## E. Penerapan dalam Sains Data

Penerapan *web scraping* dan *web crawling* dalam sains data berperan sebagai fondasi dalam proses akuisisi data dari sumber terbuka di internet untuk mendukung analisis, pemodelan, dan pengambilan keputusan berbasis data. *Web crawling* digunakan untuk menelusuri dan mengindeks halaman-halaman web secara sistematis guna menemukan sumber data yang relevan, sedangkan *web scraping* berfokus pada ekstraksi informasi spesifik dari halaman tersebut agar dapat diubah menjadi dataset terstruktur yang siap diproses lebih lanjut. Berikut adalah penerapan *web scraping* dan *web crawling* dalam sains data secara rinci:

### 1. Pengumpulan Data Skala Besar (*Big Data Acquisition*)

Pengumpulan Data Skala Besar (*Big Data Acquisition*) merupakan proses sistematis untuk memperoleh data dalam volume, variasi, dan kecepatan tinggi dari berbagai sumber heterogen seperti platform digital, sensor, media sosial, transaksi daring, serta repositori terbuka yang tersebar di internet. Dalam konteks sains data dan komputasi modern, proses ini tidak hanya berfokus pada kuantitas data, tetapi juga pada kemampuan infrastruktur dalam menangani karakteristik *big data* yang mencakup *volume*, *velocity*, dan *variety* sebagaimana dijelaskan oleh Marr (2018) bahwa *big data* merujuk pada kumpulan data dengan ukuran dan kompleksitas yang melampaui kemampuan perangkat lunak tradisional untuk mengelola dan memprosesnya secara efektif.

Secara teknis, pengumpulan data skala besar melibatkan penggunaan teknologi terdistribusi seperti sistem penyimpanan berbasis *cloud*, data lake, serta mekanisme *streaming* data yang memungkinkan akuisisi data secara *real-time* maupun batch processing dalam jumlah sangat besar. Proses ini sering kali didukung oleh teknik otomatisasi seperti *web crawling*, integrasi API, message broker (misalnya Kafka), serta pipeline ETL (*Extract, Transform, Load*) yang memastikan data dapat ditransformasikan ke dalam format terstruktur dan siap dianalisis tanpa kehilangan integritas maupun konsistensi.

## 2. Analisis Sentimen dan Opini Publik

Analisis sentimen dan opini publik merupakan metode dalam sains data yang bertujuan untuk mengidentifikasi, mengekstraksi, dan mengklasifikasikan sikap, emosi, serta persepsi masyarakat terhadap suatu topik, produk, kebijakan, atau peristiwa berdasarkan data teks yang diperoleh dari media sosial, ulasan daring, forum diskusi, maupun portal berita digital. Teknik ini memanfaatkan pendekatan pemrosesan bahasa alami (*Natural Language Processing/NLP*), pembelajaran mesin, dan representasi vektor teks untuk mengubah data tidak terstruktur menjadi informasi terukur yang dapat dikategorikan ke dalam polaritas seperti positif, negatif, atau netral guna mendukung evaluasi berbasis data secara sistematis.

Secara metodologis, analisis sentimen melibatkan tahapan prapemrosesan teks seperti tokenisasi, normalisasi, penghapusan stopwords, dan stemming sebelum dilakukan ekstraksi fitur menggunakan metode seperti TF-IDF, word embedding, atau model berbasis transformer yang mampu menangkap konteks semantik secara lebih mendalam. Liu (2020) menjelaskan bahwa sentiment analysis adalah bidang studi yang menganalisis opini, sentimen, evaluasi, sikap, dan emosi orang terhadap entitas seperti produk, layanan, organisasi, individu, isu, peristiwa, dan topik, yang menunjukkan bahwa pendekatan ini tidak hanya berfokus pada klasifikasi polaritas tetapi juga pada pemahaman struktur opini secara komprehensif dalam data teks berskala besar.

## 3. Monitoring Harga dan Analisis Kompetitor

Monitoring harga dan analisis kompetitor merupakan strategi berbasis data yang digunakan perusahaan untuk mengamati perubahan harga produk, pola promosi, ketersediaan stok, serta strategi pasar pesaing secara sistematis melalui pemanfaatan teknologi digital dan pengumpulan data daring. Dalam lingkungan bisnis yang sangat kompetitif dan dinamis, proses ini memungkinkan organisasi mengidentifikasi tren harga, membandingkan posisi pasar, serta merespons perubahan strategi pesaing secara cepat dan terukur guna mempertahankan daya saing.

Secara teknis, monitoring harga dilakukan dengan mengumpulkan data dari berbagai platform e-commerce, marketplace, dan situs resmi kompetitor menggunakan teknik otomatisasi seperti *web*

*scraping* dan integrasi API, kemudian data tersebut dianalisis untuk mendeteksi fluktuasi harga, diskon musiman, atau strategi bundling produk. Menurut Kotler dan Keller (2022), harga merupakan satu-satunya elemen bauran pemasaran yang menghasilkan pendapatan dan dapat disesuaikan dengan cepat untuk merespons dinamika pasar, sehingga pemantauan harga pesaing menjadi bagian penting dalam strategi penetapan harga yang efektif dan kompetitif (Marketing Management, 16th ed., 2022).

Analisis kompetitor tidak hanya berfokus pada aspek harga, tetapi juga mencakup evaluasi posisi merek, variasi produk, ulasan pelanggan, serta strategi komunikasi pemasaran yang diterapkan oleh pesaing dalam ekosistem digital. Dengan memanfaatkan analitik data dan visualisasi dashboard, perusahaan dapat memperoleh wawasan komprehensif mengenai kekuatan dan kelemahan pesaing sehingga keputusan strategis seperti penyesuaian harga dinamis, diferensiasi produk, atau optimalisasi promosi dapat dilakukan secara berbasis bukti dan berkelanjutan.

#### **4. Riset Akademik dan Ilmiah**

Riset akademik dan ilmiah merupakan proses sistematis yang bertujuan untuk menghasilkan pengetahuan baru, memverifikasi teori yang telah ada, serta memberikan kontribusi empiris maupun konseptual terhadap perkembangan ilmu pengetahuan melalui pendekatan metodologis yang terstruktur dan dapat dipertanggungjawabkan. Dalam konteks perkembangan teknologi digital, kegiatan penelitian semakin didukung oleh ketersediaan data daring, repositori ilmiah terbuka, serta perangkat analitik komputasional yang memungkinkan peneliti mengakses, mengelola, dan menganalisis data dalam skala besar dengan tingkat presisi yang tinggi.

Secara metodologis, riset ilmiah melibatkan tahapan perumusan masalah, penyusunan kerangka teori, pengumpulan dan analisis data, hingga penarikan kesimpulan yang didasarkan pada bukti empiris yang valid dan reliabel agar temuan penelitian dapat diuji ulang serta direplikasi oleh peneliti lain. Creswell dan Creswell (2018) menyatakan bahwa penelitian ilmiah adalah suatu proses langkah demi langkah yang digunakan untuk mengumpulkan dan menganalisis informasi guna meningkatkan pemahaman terhadap suatu topik atau isu tertentu, yang menegaskan bahwa penelitian bukan sekadar aktivitas pengumpulan data

melainkan proses sistematis yang mengikuti prosedur ilmiah yang ketat (*Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 2018).

## 5. Prediksi dan Pemodelan *Machine Learning*

Prediksi dan pemodelan *machine learning* merupakan proses analisis data yang menggunakan algoritma matematis dan statistik untuk membangun model yang mampu mengenali pola dalam dataset historis, kemudian memproyeksikan hasil atau perilaku di masa depan secara otomatis. Dalam konteks sains data, metode ini digunakan untuk mengubah data mentah menjadi wawasan prediktif yang mendukung pengambilan keputusan, mulai dari rekomendasi produk, prediksi permintaan pasar, hingga peramalan risiko operasional, sehingga memaksimalkan nilai informasi yang terkandung dalam data.

Secara teknis, pemodelan *machine learning* melibatkan tahapan pra-pemrosesan data seperti normalisasi, penghapusan nilai hilang, dan transformasi fitur, diikuti dengan pemilihan algoritma yang sesuai seperti regresi, pohon keputusan, random forest, atau jaringan saraf tiruan, serta proses validasi dan evaluasi model untuk memastikan akurasi, generalisasi, dan ketahanan terhadap overfitting. James *et al.* (2019) menyatakan bahwa *machine learning* adalah sekumpulan metode statistik yang memungkinkan komputer belajar dari data untuk membuat prediksi atau keputusan tanpa diprogram secara eksplisit, yang menunjukkan bahwa inti dari pemodelan *machine learning* adalah kemampuan sistem untuk menyesuaikan diri dengan pola data yang kompleks dan dinamis (An Introduction to Statistical Learning, 2019).

## F. Isu Legal dan Kepatuhan

Isu legal dan kepatuhan dalam konteks *web scraping*, *web crawling*, dan pengolahan data digital merujuk pada aturan hukum, regulasi, dan standar etika yang harus dipatuhi oleh organisasi atau individu ketika mengakses, mengumpulkan, menyimpan, dan menggunakan data dari internet maupun sumber digital lainnya. Kepatuhan terhadap regulasi ini sangat penting untuk mencegah risiko hukum, melindungi privasi individu, dan menjaga integritas data, terutama karena data digital sering kali bersifat sensitif, dilindungi hak

cipta, atau tunduk pada peraturan perlindungan data pribadi. Berikut adalah rincian isu legal dan kepatuhan yang relevan:

## **1. Hak Cipta dan Kekayaan Intelektual (*Intellectual Property Rights*)**

Hak cipta dan kekayaan intelektual (*Intellectual Property Rights*) merupakan mekanisme hukum yang memberikan perlindungan terhadap karya orisinal seperti tulisan, gambar, musik, perangkat lunak, dan konten digital lainnya, sehingga pencipta atau pemilikinya memiliki hak eksklusif untuk memperbanyak, mendistribusikan, atau memodifikasi karya tersebut. Dalam konteks digital dan *web scraping*, pemahaman yang jelas tentang hak cipta sangat penting karena pengambilan konten tanpa izin dapat menimbulkan risiko hukum yang signifikan, termasuk tuntutan ganti rugi atau pemblokiran akses.

Secara praktis, hak cipta mencakup perlindungan atas karya yang bersifat tetap dan kreatif, sementara kekayaan intelektual lebih luas termasuk paten, merek dagang, dan rahasia dagang yang mengatur inovasi, identitas merek, dan proses bisnis unik. Menurut WIPO (2021), *Intellectual Property Rights provide creators with the legal tools to protect their works, incentivize innovation, and facilitate economic growth, indicating that proper management and compliance with IP laws are essential in digital content creation and usage* (World Intellectual Property Organization, 2021).

## **2. Perlindungan Data Pribadi dan Privasi**

Perlindungan data pribadi dan privasi adalah upaya hukum, teknis, dan organisasi untuk menjaga keamanan informasi identitas individu yang dikumpulkan, disimpan, dan diproses oleh entitas digital, sehingga mencegah akses tidak sah, penyalahgunaan, atau kebocoran data yang dapat merugikan subjek data. Dalam era digital, praktik ini sangat penting karena data pribadi sering menjadi target bagi berbagai pihak yang ingin memanfaatkan informasi untuk tujuan komersial, kriminal, atau politik tanpa izin dari pemilik data.

Secara operasional, perlindungan data pribadi mencakup penerapan kebijakan privasi, mekanisme enkripsi, kontrol akses berbasis peran, audit keamanan, dan prosedur penghapusan atau anonimasi data untuk memastikan bahwa setiap pengolahan data tetap mematuhi regulasi lokal maupun internasional. Solove dan Hartzog (2020)

menekankan bahwa *privacy protection is not only a legal obligation but a fundamental component of trust between individuals and organizations, highlighting the need for comprehensive privacy frameworks in the digital environment* (Solove & Hartzog, 2020).

### **3. Ketentuan Layanan Situs Web (*Terms of Service / ToS*)**

Ketentuan Layanan Situs Web (*Terms of Service/ToS*) adalah perjanjian hukum yang mengatur hak, kewajiban, dan batasan penggunaan situs web antara pemilik situs dan pengunjung atau pengguna, termasuk ketentuan mengenai konten, privasi, tanggung jawab hukum, dan cara akses data yang sah. Dalam konteks *web scraping* dan pengolahan data digital, memahami dan mematuhi ToS menjadi sangat penting karena pelanggaran terhadap ketentuan ini dapat menimbulkan konsekuensi hukum, pemblokiran akses, atau tuntutan ganti rugi dari pemilik situs.

Secara operasional, ToS biasanya mencakup klausul mengenai larangan pengambilan data secara otomatis, batasan penggunaan API, pembatasan jumlah request, hak cipta konten, dan prosedur penyelesaian sengketa, sehingga setiap aktivitas digital yang dilakukan harus selaras dengan ketentuan tersebut. According to Balkin (2018), *terms of service agreements function as private regulatory frameworks that shape user behavior and define permissible activities on digital platforms, emphasizing that compliance with ToS is both a legal and ethical responsibility in the digital ecosystem* (Balkin, 2018).

### **4. Etika dan Kode Praktik Profesional**

Etika dan kode praktik profesional dalam konteks sains data dan teknologi informasi adalah seperangkat prinsip dan pedoman yang menuntun profesional dalam melakukan aktivitas pengumpulan, pengolahan, dan analisis data secara bertanggung jawab, transparan, dan menghormati hak pihak lain. Prinsip ini mencakup integritas, kejujuran, kerahasiaan, dan tanggung jawab sosial, sehingga setiap keputusan berbasis data tidak merugikan individu atau masyarakat serta tetap mematuhi standar hukum dan profesional.

Secara praktis, kode etik profesional menuntun pengembang, analis data, dan peneliti untuk menghindari praktik manipulasi data, pelanggaran privasi, diskriminasi algoritmik, serta penggunaan data untuk tujuan yang tidak sah atau merugikan pihak lain. According to

Jobin, Ienca, and Vayena (2019), *ethical frameworks and professional codes in data science serve to guide responsible innovation, ensure fairness, and maintain public trust, highlighting that adherence to ethical standards is crucial for sustainable and socially responsible technological development* (Jobin, Ienca, & Vayena, 2019).

## G. Implementasi *Web Scraping* Menggunakan Python

Implementasi *web scraping* menggunakan Python merupakan proses teknis untuk mengambil, membaca, dan mengekstraksi data dari halaman web secara otomatis dengan memanfaatkan berbagai library yang dirancang khusus untuk komunikasi HTTP, parsing dokumen, dan otomatisasi browser. Python menjadi pilihan utama dalam praktik *scraping* karena sintaksnya sederhana, ekosistem pustaka yang lengkap, serta kemampuannya untuk diintegrasikan dengan sistem analitik, *database*, dan *machine learning*.

### 1. Struktur HTML dan DOM

Struktur HTML merupakan kerangka dasar dari sebuah halaman web yang dibangun menggunakan elemen-elemen bertingkat seperti `<html>`, `<head>`, `<body>`, serta berbagai tag semantik lainnya yang membentuk hierarki dokumen secara terorganisasi dan terstandarisasi sesuai spesifikasi web. Setiap elemen HTML memiliki atribut dan konten yang tersusun dalam pola parent-child sehingga membentuk struktur pohon yang memungkinkan browser memahami bagaimana konten harus ditampilkan dan diinterpretasikan.

*Document Object Model* (DOM) adalah representasi terstruktur dari dokumen HTML atau XML dalam bentuk objek yang dapat diakses dan dimanipulasi secara dinamis melalui bahasa pemrograman seperti JavaScript. World Wide Web Consortium (W3C) (2021) menjelaskan bahwa *the DOM defines a platform- and language-neutral interface that allows programs and scripts to dynamically access and update the content, structure, and style of documents*, yang menegaskan bahwa DOM berfungsi sebagai jembatan antara struktur dokumen dan interaksi pemrograman dalam lingkungan web modern.

## 2. *Scraping* Data Tabel dan Teks

*Scraping* data tabel dan teks merupakan teknik ekstraksi informasi dari elemen terstruktur maupun semi-terstruktur pada halaman web dengan memanfaatkan pola HTML seperti tag `<table>`, `<tr>`, `<td>`, serta elemen paragraf, heading, dan div yang mengandung konten tekstual. Dalam praktiknya, proses ini dilakukan dengan mengirim permintaan HTTP ke halaman target, kemudian memarsing struktur HTML untuk mengidentifikasi baris dan kolom tabel atau blok teks tertentu yang relevan sebelum mengubahnya menjadi format terstruktur seperti CSV, JSON, atau DataFrame untuk analisis lebih lanjut.

Secara teknis, *scraping* tabel membutuhkan pemahaman terhadap hierarki DOM karena setiap tabel tersusun atas baris dan sel yang saling terkait, sehingga proses ekstraksi harus mampu membaca setiap elemen secara iteratif dan mempertahankan relasi antar kolom agar tidak terjadi pergeseran data. Mitchell (2018) menjelaskan bahwa *web scraping involves programmatically collecting information from web pages and extracting structured data from HTML content, which often includes parsing tables and textual elements for analysis*, yang menegaskan bahwa pengambilan data terstruktur dari tabel dan teks merupakan praktik umum dalam proses *scraping* modern.

## 3. Contoh Script Python *Web Scraping*

Berikut adalah beberapa contoh Script Python *Web Scraping* menggunakan library yang umum dipakai seperti requests, BeautifulSoup, dan pandas.

⚠️ Pastikan *scraping* dilakukan pada *website* yang mengizinkan *scraping* (cek robots.txt & Terms of Service).

### 1 Contoh *Web Scraping* Sederhana (Ambil Judul Artikel)

Misalnya kita ingin mengambil semua judul artikel dari sebuah halaman *website*.

```
import requests
from bs4 import BeautifulSoup

# URL target
url = "https://example.com"
```

```

# Mengirim request ke website
response = requests.get(url)

# Parsing HTML
soup = BeautifulSoup(response.text, 'html.parser')

# Ambil semua tag h2 (misalnya judul artikel)
titles = soup.find_all('h2')

print("Daftar Judul:")
for i, title in enumerate(titles, 1):
    print(f'{i}. {title.text.strip()}')

```

## **2** Contoh *Scraping Data Tabel Website*

```

import requests
from bs4 import BeautifulSoup
import pandas as pd

url = "https://example.com/tabel"

response = requests.get(url)
soup = BeautifulSoup(response.text, 'html.parser')

# Ambil tabel pertama
table = soup.find('table')

rows = table.find_all('tr')

data = []
for row in rows:
    cols = row.find_all(['td', 'th'])
    cols = [col.text.strip() for col in cols]
    data.append(cols)

# Simpan ke DataFrame
df = pd.DataFrame(data)

```

```
print(df)

# Simpan ke CSV
df.to_csv('hasil_scraping.csv', index=False)
```

### **3** Contoh *Scraping Website Berita (Judul + Link)*

```
import requests
from bs4 import BeautifulSoup

url = "https://example-news.com"
response = requests.get(url)

soup = BeautifulSoup(response.text, 'html.parser')

articles = soup.find_all('a')

print("Daftar Artikel:")
for article in articles[:10]:
    title = article.text.strip()
    link = article.get('href')

    if title and link:
        print(f'Judul: {title}')
        print(f'Link : {link}')
        print("-" * 40)
```

### **4** Contoh *Web Scraping dengan User-Agent (Menghindari Blocking)*

```
import requests
from bs4 import BeautifulSoup

url = "https://example.com"

headers = {
    "User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64)"
}
```

```
response = requests.get(url, headers=headers)
soup = BeautifulSoup(response.text, 'html.parser')

print(soup.title.text)
```

### **5** Contoh Menggunakan Selenium (Untuk *Website* Dinamis / JavaScript)

Jika *website* menggunakan JavaScript (tidak bisa diambil dengan requests saja):

```
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.chrome.service import Service
from selenium.webdriver.chrome.options import Options
import time

options = Options()
options.add_argument("--headless")

driver = webdriver.Chrome(options=options)

driver.get("https://example.com")
time.sleep(3)

elements = driver.find_elements(By.TAG_NAME, "h2")

for el in elements:
    print(el.text)

driver.quit()
```

# BAB X

## APPLICATION PROGRAMMING INTERFACE (API) SEBAGAI SUMBER DATA

---

---

*Application Programming Interface* (API) memungkinkan pertukaran data secara langsung antara sistem atau aplikasi, sehingga mempermudah akses informasi secara *real-time*. API menjadi sumber data penting karena menyediakan data terstruktur yang siap digunakan untuk analisis dan pengembangan aplikasi.

### A. Konsep Dasar API

Application Programming Interface (API) adalah sekumpulan aturan, protokol, dan alat yang memungkinkan aplikasi atau sistem perangkat lunak yang berbeda untuk saling berkomunikasi dan bertukar data. API berfungsi sebagai jembatan antara klien (aplikasi yang meminta data) dan server (sumber data atau layanan), sehingga proses pertukaran informasi dapat dilakukan secara terstruktur, efisien, dan otomatis. Dengan API, pengembang dapat memanfaatkan fungsi atau data dari aplikasi lain tanpa harus mengetahui detail implementasi internalnya, cukup dengan memahami cara memanggil API sesuai kontrak yang telah ditentukan.

#### 1. Komponen Utama API

##### a. Endpoint / URL

API endpoint atau endpoint adalah lokasi digital spesifik berupa URL yang menjadi tempat API menerima permintaan dan mengirimkan respons untuk mengakses atau memanipulasi data pada server, sehingga endpoint berfungsi sebagai titik akhir komunikasi antara aplikasi klien dan layanan yang disediakan oleh API. Menurut Kinza Yasar (2024), “*API endpoints are the specific digital location where requests or API calls for information are sent by one program to retrieve the digital*

*resource that exists there,*” yang menegaskan bahwa endpoint merupakan URL unik yang menunjukkan sumber daya yang dituju dalam pertukaran data melalui API.

Pada praktiknya, URL pada endpoint terdiri dari base URL yang menunjukkan server utama dan path yang merinci sumber daya atau fungsionalitas tertentu, sehingga setiap request dikirimkan ke alamat tersebut dengan metode HTTP untuk mendapatkan atau memodifikasi data sesuai kebutuhan. Desain endpoint yang tepat harus jelas, intuitif, dan terdokumentasi dengan baik agar memudahkan pengembang dalam menentukan lokasi akses data yang diinginkan serta memastikan komunikasi antara sistem berjalan efektif dan akurat.

b. Metode HTTP

Metode HTTP dalam API adalah instruksi standar yang digunakan untuk menentukan jenis operasi yang ingin dilakukan pada sumber daya tertentu di server, seperti GET untuk mengambil data, POST untuk membuat data baru, PUT atau PATCH untuk memperbarui data, dan DELETE untuk menghapus data, sehingga setiap metode membawa makna semantik yang berbeda dan memandu server dalam merespons permintaan dengan cara yang sesuai. Menurut Fielding (2018), *“HTTP methods define a uniform interface to manipulate web resources, where each method indicates a specific action to be applied to a given resource,”* yang menekankan bahwa metode HTTP adalah fondasi interaksi klien-server yang terstandarisasi dalam arsitektur web dan API.

Pemilihan metode HTTP yang tepat sangat penting untuk menjaga integritas data, keamanan, dan konsistensi sistem, karena kesalahan dalam penggunaan metode dapat menyebabkan efek yang tidak diinginkan seperti duplikasi data, kehilangan data, atau kebocoran informasi sensitif. Desain API yang baik selalu menyertakan dokumentasi jelas mengenai metode HTTP yang digunakan pada setiap endpoint, sehingga pengembang dapat memahami cara berinteraksi dengan API secara akurat dan mengoptimalkan alur komunikasi antara aplikasi klien dan server.

c. Header

Header dalam API adalah bagian dari permintaan atau respons yang membawa informasi meta penting seperti tipe konten, autentikasi, bahasa, dan instruksi caching, sehingga memungkinkan server dan klien memahami konteks, format data, serta aturan keamanan dari setiap komunikasi. Menurut Richardson dan Amundsen (2019), “*HTTP headers provide metadata about the request or response, enabling clients and servers to exchange information beyond the main message body, such as authentication credentials, content types, and caching policies,*” yang menekankan peran header sebagai elemen krusial dalam memastikan komunikasi API berjalan aman dan terstruktur.

Header API juga memungkinkan pengendalian lebih lanjut atas perilaku permintaan dan respons, misalnya melalui header *Authorization* untuk autentikasi token, *Content-Type* untuk menentukan format payload, serta *Accept* untuk memberi tahu server format yang diharapkan oleh klien, sehingga setiap interaksi dapat diatur dengan fleksibel tanpa mengubah isi data utama. Desain dan dokumentasi header yang jelas sangat penting dalam pengembangan API karena membantu pengembang memahami persyaratan komunikasi dan mengurangi risiko kesalahan implementasi yang dapat memengaruhi performa atau keamanan sistem.

d. *Query & Path Parameters*

*Query* dan *path parameters* adalah mekanisme dalam API yang digunakan untuk menyaring, memodifikasi, atau mengidentifikasi sumber daya tertentu melalui URL, di mana *path parameters* menandai sumber daya spesifik dalam struktur endpoint sedangkan *query parameters* memberikan opsi tambahan seperti filter, sort, atau pagination, sehingga kombinasi keduanya memungkinkan pengambilan data yang lebih tepat dan fleksibel. Menurut Richardson dan Amundsen (2019), “*Path and query parameters in RESTful APIs provide a structured way to specify the resource being accessed and to pass additional information to filter, sort, or paginate the response data,*” yang menekankan pentingnya parameter ini untuk memberikan kontrol granular terhadap data yang dikonsumsi oleh klien.

Penggunaan query dan path parameters yang konsisten dan terdokumentasi dengan baik juga meningkatkan kemudahan integrasi dan pemeliharaan API, karena pengembang dapat memahami dengan jelas cara membangun URL yang tepat untuk berbagai skenario permintaan. Selain itu, parameter ini mendukung skalabilitas sistem dengan memungkinkan pengambilan data yang selektif dan mengurangi beban server, sehingga operasi API menjadi lebih efisien tanpa perlu memproses seluruh dataset secara menyeluruh.

e. *Body / Payload*

*Body* atau *payload* dalam API adalah bagian dari permintaan atau respons yang membawa data utama yang dikirim atau diterima antara klien dan server, biasanya dalam format JSON atau XML, sehingga memungkinkan pertukaran informasi yang kompleks seperti objek, daftar, atau struktur bersarang secara terstruktur. Menurut Massé (2019), “*The body of an HTTP request or response carries the actual data being transmitted between client and server, enabling the exchange of structured information such as JSON or XML objects,*” yang menegaskan bahwa *payload* adalah elemen krusial untuk menyampaikan konten yang relevan dalam komunikasi API.

*Desain payload* yang baik sangat penting untuk memastikan kompatibilitas antara klien dan server, karena format yang jelas, konsisten, dan terdokumentasi memungkinkan pengembang untuk memahami struktur data yang diperlukan serta mengurangi risiko kesalahan interpretasi data. Selain itu, *payload* yang efisien dapat meningkatkan performa sistem dengan meminimalkan ukuran data yang dikirim, mendukung integritas informasi, dan mempermudah validasi serta pengolahan data di sisi penerima.

## 2. Tipe dan Gaya API

Tipe dan gaya API mencakup berbagai pendekatan dalam desain dan arsitektur yang menentukan cara aplikasi berkomunikasi dan bertukar data, termasuk RESTful API yang memanfaatkan prinsip sumber daya dan metode HTTP, GraphQL yang memungkinkan klien menentukan struktur data yang dibutuhkan dalam satu permintaan, serta SOAP yang berbasis XML dan menyediakan fitur keamanan dan transaksi bawaan untuk lingkungan enterprise. Menurut Pautasso *et al.*

(2018), “*API styles such as REST, GraphQL, and SOAP offer distinct approaches to designing interfaces that facilitate communication and data exchange between distributed systems, each with trade-offs in flexibility, performance, and complexity,*” yang menekankan bahwa pemilihan gaya API harus disesuaikan dengan kebutuhan fungsional dan teknis sistem.

RESTful API banyak digunakan karena kesederhanaannya, kemampuan skalabilitas tinggi, dan kesesuaian dengan protokol HTTP standar, sehingga pengembang dapat membangun aplikasi yang modular dan mudah diintegrasikan dengan layanan lain. GraphQL menawarkan fleksibilitas yang lebih besar dibanding REST karena klien dapat mengambil hanya data yang dibutuhkan, mengurangi overfetching dan underfetching, namun membutuhkan setup yang lebih kompleks di sisi server dan dokumentasi query yang tepat. Sementara itu, SOAP tetap digunakan di beberapa sistem legacy dan enterprise karena menyediakan standar protokol yang kuat untuk transaksi kompleks, keamanan tingkat tinggi, dan interoperabilitas antarplatform. Memahami perbedaan tipe dan gaya API membantu pengembang menentukan strategi integrasi dan desain API yang optimal agar komunikasi antar sistem berjalan efisien, aman, dan dapat diandalkan.

### **3. Keamanan dan Otorisasi**

Keamanan dan otorisasi dalam API sangat penting untuk melindungi data dan memastikan hanya pengguna atau aplikasi yang berwenang yang dapat mengakses sumber daya tertentu, dengan mekanisme seperti API key, OAuth 2.0, dan JSON Web Tokens (JWT) yang memberikan kontrol akses berbasis peran dan autentikasi yang aman. Menurut Hardt (2018), “*OAuth 2.0 provides a framework for delegated authorization, allowing third-party applications to obtain limited access to user resources without exposing credentials, thereby enhancing security and control in API interactions,*” yang menekankan pentingnya protokol otorisasi modern untuk menjaga integritas dan kerahasiaan data.

Implementasi enkripsi TLS/HTTPS pada API menjamin keamanan data saat ditransmisikan antara klien dan server, sementara praktik seperti validasi *input*, sanitasi data, dan kontrol CORS/CSRF membantu mencegah serangan injeksi, akses lintas domain, dan manipulasi data. Desain API yang aman tidak hanya fokus pada

otentikasi, tetapi juga mengatur hak akses seminimal mungkin (*principle of least privilege*), sehingga mengurangi risiko kebocoran data atau penyalahgunaan oleh pihak yang tidak berwenang.

Monitoring, logging, dan audit trail juga merupakan bagian penting dari keamanan API, karena memungkinkan deteksi dini aktivitas mencurigakan, analisis insiden, dan pemeliharaan kepatuhan terhadap standar keamanan dan regulasi data. Dengan kombinasi mekanisme otorisasi, enkripsi, validasi, dan pemantauan, API dapat beroperasi secara aman dan andal, sekaligus memberikan fleksibilitas bagi pengembang untuk membangun aplikasi yang terintegrasi namun tetap terlindungi.

#### **4. Kinerja, Skalabilitas, dan Keandalan**

Kinerja, skalabilitas, dan keandalan API merupakan aspek penting yang menentukan kemampuan sistem untuk menangani beban permintaan yang tinggi, merespons dengan cepat, dan tetap konsisten dalam menyediakan layanan meskipun terjadi lonjakan trafik atau gangguan sistem. Menurut Richardson dan Smith (2020), *“High-performance APIs require careful design considerations including efficient request handling, caching strategies, and load balancing to ensure that services remain responsive and scalable under varying workloads,”* yang menekankan bahwa perencanaan arsitektur dan optimasi mekanisme distribusi beban sangat penting untuk menjaga kualitas layanan API.

Penerapan teknik seperti rate limiting membatasi jumlah permintaan per klien untuk mencegah penyalahgunaan dan menjaga stabilitas server, sementara caching dapat mengurangi latensi dan beban pemrosesan dengan menyimpan respons yang sering diminta, sehingga meningkatkan kecepatan akses data tanpa membebani sumber daya backend. Selain itu, mekanisme pagination dan batch processing memungkinkan pengambilan data dalam jumlah besar secara efisien, sehingga API tetap responsif meskipun menangani dataset yang besar atau kompleks.

Keandalan API juga didukung oleh konsep idempotency, di mana operasi yang sama menghasilkan efek yang konsisten meskipun dipanggil berulang kali, serta sistem monitoring dan logging yang aktif untuk mendeteksi masalah atau kegagalan layanan secara dini. Dengan menggabungkan strategi optimasi performa, pengaturan beban, dan

pemantauan yang efektif, API dapat beroperasi secara andal dan scalable, memastikan pengalaman pengguna tetap optimal dan risiko downtime diminimalkan.

## **B. Jenis-Jenis API**

API dapat diklasifikasikan berdasarkan akses, tujuan, dan arsitektur yang digunakan, sehingga pengembang dapat memilih jenis API yang paling sesuai dengan kebutuhan integrasi, keamanan, dan interoperabilitas sistem. Berikut adalah jenis-jenis API beserta penjelasannya:

### **1. Open API (Public API)**

Open API atau public API adalah antarmuka pemrograman aplikasi yang tersedia secara publik di internet dan memungkinkan pengembang dari berbagai pihak untuk mengakses, mengambil, dan memanfaatkan data atau layanan yang disediakan oleh suatu aplikasi atau sistem tanpa perlu izin khusus, sehingga memfasilitasi integrasi, kolaborasi, dan inovasi antaraplikasi secara luas. Menurut TechTarget (2021), *“An open API, also called public API, is an application programming interface made publicly available to software developers,”* yang menegaskan bahwa Open API dirancang untuk dikonsumsi oleh siapa saja di luar organisasi penyedia layanan guna memperluas jangkauan fungsi dan ekosistem layanan digital.

Pada praktiknya, Open API sering menggunakan standar komunikasi seperti REST melalui HTTP dan format data seperti JSON agar dapat dengan mudah diintegrasikan oleh pengembang dari berbagai lingkungan teknologi, sehingga aplikasi baru dapat dibangun dengan memanfaatkan fungsionalitas atau sumber daya yang sudah ada tanpa harus membuat semuanya dari awal. Desain dan dokumentasi Open API yang jelas sangat penting untuk memastikan interoperabilitas, keamanan akses, serta pengalaman integrasi yang konsisten bagi pengembang eksternal dan internal, sekaligus mendorong pertumbuhan ekosistem digital yang lebih luas dan dinamis.

### **2. Partner API**

Partner API adalah antarmuka pemrograman aplikasi yang disediakan secara terbatas dan hanya dapat diakses oleh mitra bisnis tertentu, dengan mekanisme autentikasi dan kontrol akses yang ketat

sehingga menjamin keamanan data dan memungkinkan kolaborasi antarorganisasi secara aman. Menurut Dahanayake dan Sol (2019), *“Partner APIs provide controlled access to selected business partners, enabling secure and managed integration while protecting sensitive organizational data,”* yang menegaskan bahwa Partner API dirancang untuk mendukung kerjasama bisnis tanpa mengorbankan keamanan informasi kritis.

Implementasi Partner API biasanya melibatkan penggunaan API key, token OAuth, atau mekanisme autentikasi khusus yang membatasi akses hanya kepada pihak yang memiliki izin, sehingga memungkinkan integrasi layanan secara efisien dengan menjaga privasi dan integritas data. Desain dan dokumentasi Partner API yang jelas sangat penting untuk memastikan mitra bisnis dapat memanfaatkan layanan secara optimal, sekaligus meminimalkan risiko kesalahan penggunaan atau penyalahgunaan akses yang dapat merugikan kedua belah pihak.

### **3. Internal API (Private API)**

Internal API atau Private API adalah antarmuka pemrograman aplikasi yang dirancang khusus untuk digunakan secara internal oleh organisasi, sehingga hanya sistem, layanan, atau pengembang internal yang memiliki akses untuk berinteraksi dengan sumber daya atau fungsionalitas tertentu, sehingga mendukung modularisasi dan integrasi aplikasi internal tanpa mengekspos data ke pihak luar. Menurut Richardson dan Smith (2020), *“Private APIs are designed for internal organizational use, enabling secure and efficient interaction between internal systems while maintaining strict access control and operational boundaries,”* yang menekankan bahwa Private API berfokus pada keamanan, efisiensi, dan manajemen interaksi antar sistem internal.

Penggunaan Internal API memungkinkan organisasi untuk mengembangkan layanan secara terpisah namun tetap terintegrasi, mempercepat proses pengembangan, dan mempermudah pemeliharaan sistem karena setiap modul dapat diuji, diperbarui, atau diganti secara independen. Selain itu, dokumentasi dan standar penggunaan Private API yang baik sangat penting untuk memastikan konsistensi dalam pengembangan internal, mengurangi risiko kesalahan integrasi, dan menjaga integritas serta kerahasiaan data organisasi.

#### 4. Composite API

Composite API adalah antarmuka pemrograman aplikasi yang memungkinkan penggabungan beberapa permintaan ke berbagai layanan atau sumber daya menjadi satu panggilan tunggal, sehingga meminimalkan jumlah request yang dikirim oleh klien dan meningkatkan efisiensi pengambilan data dari sistem yang kompleks. Menurut Pautasso *et al.* (2019), “*Composite APIs enable clients to bundle multiple service calls into a single request, thereby reducing network latency and simplifying the orchestration of data from multiple sources,*” yang menekankan bahwa penggunaan Composite API meningkatkan kinerja dan efisiensi interaksi antara aplikasi dengan berbagai layanan backend.

Implementasi Composite API sangat berguna dalam aplikasi yang memerlukan agregasi data dari beberapa endpoint atau sistem secara bersamaan, karena klien tidak perlu melakukan banyak panggilan terpisah yang dapat meningkatkan latensi dan beban jaringan. Desain dan dokumentasi Composite API yang baik juga memastikan integritas data, konsistensi respons, dan memudahkan pengembang dalam memanfaatkan fungsionalitas kompleks tanpa harus memahami detail setiap layanan yang digabungkan.

#### 5. REST API

REST API adalah jenis antarmuka pemrograman aplikasi yang menggunakan prinsip *Representational State Transfer* (REST) dan protokol HTTP untuk mengakses dan memanipulasi sumber daya melalui URL, sehingga memungkinkan pengembang membangun aplikasi yang ringan, skalabel, dan mudah diintegrasikan dengan layanan lain menggunakan metode HTTP standar seperti GET, POST, PUT, PATCH, dan DELETE. Menurut Richardson dan Amundsen (2019), “*RESTful APIs provide a simple, stateless, and resource-oriented approach to designing web services that can be easily consumed by clients across diverse platforms,*” yang menegaskan bahwa REST API menekankan kesederhanaan, independensi klien-server, dan interoperabilitas lintas platform.

Implementasi REST API sangat cocok untuk aplikasi web dan *mobile* karena fleksibilitas dalam pemanggilan sumber daya, kemampuan caching, dan kompatibilitas dengan berbagai format data seperti JSON dan XML. Selain itu, desain REST API yang konsisten dan

terdokumentasi dengan baik membantu pengembang memahami struktur endpoint dan metode yang tersedia, sehingga integrasi menjadi lebih cepat, aman, dan dapat diandalkan dalam skala besar maupun sistem terdistribusi.

## 6. SOAP API

SOAP API adalah antarmuka pemrograman aplikasi yang menggunakan protokol *Simple Object Access Protocol* berbasis XML untuk pertukaran data antar sistem, dengan standar keamanan, transaksi, dan validasi pesan yang ketat sehingga cocok untuk aplikasi enterprise dan sistem legacy yang memerlukan keandalan tinggi. Menurut Alonso *et al.* (2019), “*SOAP provides a protocol-based approach for exchanging structured information in distributed environments, offering built-in support for security, transactions, and reliable messaging,*” yang menegaskan bahwa SOAP API menekankan interoperabilitas, keamanan, dan konsistensi dalam komunikasi antar sistem.

Implementasi SOAP API umumnya menggunakan *Web Services Description Language* (WSDL) untuk mendefinisikan layanan dan kontrak yang dapat dikonsumsi oleh klien, sehingga pengembang dapat memahami operasi yang tersedia dan struktur pesan yang harus dikirimkan. Meskipun lebih kompleks dibanding REST, SOAP tetap digunakan dalam konteks yang membutuhkan standar protokol tinggi, integrasi lintas platform yang ketat, serta jaminan pengiriman pesan yang andal dan terverifikasi.

## 7. GraphQL API

GraphQL API adalah jenis antarmuka pemrograman aplikasi yang menggunakan GraphQL, sebuah bahasa kueri dan runtime open-source yang dirancang untuk API sehingga klien dapat mengirim satu permintaan kueri ke server dan menerima hanya data yang dibutuhkan sesuai struktur yang ditentukan, tanpa perlu berulang kali memanggil berbagai endpoint seperti pada API tradisional. Menurut.org (2025), “*GraphQL is an open-source query language for APIs and a server-side runtime that provides a strongly-typed schema to define relationships between data, making APIs more flexible and predictable,*” yang menegaskan bahwa GraphQL API menawarkan fleksibilitas dan efisiensi pengambilan dan manipulasi data dengan memanfaatkan skema yang terstruktur dan satu titik panggilan kueri.

Implementasi GraphQL API memungkinkan pengembang backend untuk menyusun schema yang mengatur tipe dan hubungan data yang tersedia dan memberikan klien kemampuan untuk menentukan dengan tepat bidang data apa yang diperlukan dalam respons, sehingga mengurangi over-fetching atau under-fetching data. Pendekatan ini tidak hanya meningkatkan efisiensi jaringan dan respons aplikasi tetapi juga menyederhanakan pengelolaan API dalam lingkungan kompleks dengan permintaan data yang bervariasi, karena klien tidak perlu memproses data yang tidak relevan atau membuat banyak permintaan untuk memenuhi kebutuhan informasi.

## 8. gRPC API

gRPC API adalah sebuah framework open-source yang dikembangkan untuk menyediakan komunikasi high-performance antara klien dan server dalam sistem terdistribusi dengan menggunakan model Remote Procedure Call (RPC) yang memungkinkan satu aplikasi memanggil fungsi di aplikasi lain seolah-olah fungsi itu berada di sistem lokal. Menurut Techtarget (2022), “*gRPC is a high-performance remote procedure call framework released by Google that accommodates client-server interactions needed for cross-platform development and microservice architecture design,*” yang menegaskan bahwa gRPC API dirancang untuk efisiensi, interoperabilitas lintas bahasa, serta performa tinggi dalam pertukaran data antar layanan yang berbeda.

Implementasi gRPC API biasanya memanfaatkan HTTP/2 sebagai protokol transportasi dan Protocol Buffers sebagai bahasa definisi antarmuka (*Interface Definition Language*), sehingga pesan yang dikirim bersifat terkompresi, kuat secara tipe, dan efisien dalam pemrosesan. Pendekatan ini mengurangi latensi dan meningkatkan throughput sambil mendukung fitur-fitur lanjutan seperti *streaming* dua arah, penanganan kesalahan terstandarisasi, serta kemampuan load balancing dan tracing yang sangat berguna dalam arsitektur mikroservis berskala besar.

### C. Mekanisme Request dan Response

Mekanisme request dan response adalah inti dari komunikasi API, di mana klien mengirim permintaan (*request*) ke server untuk mengakses, memodifikasi, atau memproses data, dan server kemudian

merespons dengan data atau status yang sesuai (*response*). Proses ini mengikuti protokol komunikasi tertentu (biasanya HTTP/HTTPS) dan dapat dibagi menjadi beberapa komponen penting:

### 1. *Request* (Permintaan)

Permintaan (*request*) dalam API adalah pesan terstruktur yang dikirimkan oleh aplikasi klien ke server API untuk meminta tindakan tertentu, seperti mengambil, membuat, memperbarui, atau menghapus data, dan biasanya dibangun atas komponen-komponen seperti metode HTTP, endpoint (URL), header, parameter, dan body sesuai kebutuhan operasi yang diinginkan. Sebagai bagian dari mekanisme komunikasi API, API request mencerminkan instruksi eksplisit dari klien kepada server yang menjelaskan apa yang diminta, di mana permintaan ditujukan, serta bagaimana server harus memprosesnya berdasarkan informasi yang disertakan.

Menurut Uptimia (2024), “*An API request is a message sent from a client application to an API endpoint asking for data or services, and it typically includes the HTTP method, URL, headers, and any required parameters or body content,*” yang menegaskan bahwa komponen-komponen tersebut saling berkolaborasi untuk membentuk sebuah permintaan yang valid dan dapat dipahami oleh server sehingga fungsi atau data yang diminta dapat dipenuhi dengan benar oleh API.

Struktur permintaan API dimulai ketika klien menentukan metode HTTP yang sesuai (misalnya GET, POST, PUT, DELETE) dan endpoint yang ingin diakses, kemudian menambahkan header seperti autentikasi ataupun tipe konten yang diharapkan, serta parameter atau body jika operasi memerlukan data tambahan. Keberhasilan permintaan sangat bergantung pada keakuratan format pesan, kelengkapan komponen, serta kepatuhan terhadap aturan yang ditetapkan oleh API penyedia layanan, sehingga server dapat memproses permintaan tersebut dan menghasilkan respons yang tepat.

### 2. *Response* (Respons)

Respons (*response*) dalam API adalah balasan terstruktur yang dikirimkan oleh server kepada klien setelah menerima dan memproses permintaan (*request*), berisi informasi tentang hasil permintaan tersebut serta data yang diminta atau penjelasan kesalahan jika permintaan tidak dapat dipenuhi. Komponen utama dalam response mencakup status code

yang merupakan kode tiga digit untuk menunjukkan hasil eksekusi (misalnya sukses, kesalahan klien, atau kesalahan server), header yang berisi metadata seperti tipe konten dan instruksi caching, serta body yang membawa data aktual atau pesan yang relevan sesuai format yang disepakati seperti JSON atau XML.

Menurut GetStream (2025), “*An API response is what an API provider (the server) returns to you (the client) after handling your API request,*” yang menegaskan bahwa response adalah komponen kunci dalam komunikasi API yang menunjukkan apakah permintaan berhasil diproses dan menyediakan hasil atau informasi yang diperlukan aplikasi klien. Dalam praktiknya, respons API memungkinkan klien memahami konteks hasil permintaan secara otomatis melalui kode status, memanfaatkan metadata dari header untuk menyesuaikan penanganan data, serta mengakses body yang berisi data yang diminta atau penjelasan tentang kesalahan yang terjadi. Desain respons yang konsisten dan terdokumentasi dengan baik sangat penting agar klien dapat menafsirkan hasil dengan benar, mengelola berbagai skenario seperti pengambilan data, pembuatan sumber daya, atau penanganan error, serta membangun interaksi antar aplikasi yang efektif dan dapat diandalkan.

### **3. Alur Kerja Umum *Request-Response***

Alur kerja umum *request-response* dalam API adalah rangkaian komunikasi standar antara klien dan server di mana klien memulai pertukaran dengan mengirim permintaan (*request*) yang berisi metode, endpoint, header, dan data yang diperlukan, kemudian server menerima permintaan tersebut, memrosesnya (melakukan validasi, logika bisnis, akses *database*, dan sebagainya), serta menghasilkan respons (*response*) yang sesuai dengan hasil pemrosesan itu. Setelah server selesai menyiapkan jawaban berupa status code, header, dan body data, respons tersebut dikirim kembali ke klien melalui protokol komunikasi yang disepakati seperti HTTP/HTTPS sehingga klien dapat menggunakan atau menampilkan hasilnya sesuai kebutuhan aplikasi.

Menurut sumber referensi umum dalam arsitektur request-response, “*the Request-Response cycle in HTTP represents the flow of communication between a client and a server, starting when the client sends an HTTP request and ending when the server returns an HTTP response containing a status code, headers, and often a response*

*body*,” yang menegaskan bahwa urutan inilah yang menjadi mekanisme dasar pertukaran data dan instruksi dalam API modern.

Proses ini biasanya dimulai dengan klien membangun dan mengirim permintaan ke endpoint tertentu sesuai metode yang ditetapkan, perjalanan pesan tersebut melalui jaringan hingga mencapai server, server menjalankan perintah atau pengambilan data yang diminta, lalu server membungkus hasilnya dalam sebuah respons dan mengirimkannya kembali ke klien. Klien kemudian menerima respons tersebut, memeriksa status code untuk menentukan apakah permintaan berhasil, lalu memproses body atau data yang diterima untuk ditampilkan atau digunakan dalam logika aplikasi lebih lanjut.

#### **D. API sebagai Sumber Data Sekunder**

API (*Application Programming Interface*) telah menjadi salah satu sumber data sekunder yang penting bagi analisis, penelitian, dan pengembangan aplikasi modern, karena memungkinkan pengambilan data yang telah tersedia di server atau sistem lain tanpa harus mengumpulkan data secara langsung dari sumber primer. Data sekunder melalui API bersifat terstruktur dan dapat diakses secara otomatis, sehingga mempermudah proses integrasi, transformasi, dan analisis data secara efisien. Berikut penjelasan rinci mengenai API sebagai sumber data sekunder:

##### **1. Akses Terstruktur dan Otomatis**

Akses terstruktur dan otomatis melalui API memungkinkan pengembang atau peneliti untuk mengambil data sekunder dari sistem atau layanan eksternal secara konsisten dan efisien, karena API menyediakan format data standar seperti JSON, XML, atau CSV yang mudah diolah oleh aplikasi dan sistem analitik. Mekanisme ini memungkinkan otomatisasi pengambilan data menggunakan skrip atau aplikasi, sehingga data dapat diperbarui secara *real-time* atau periodik tanpa intervensi manual, meningkatkan produktivitas dan mengurangi risiko kesalahan manusia. Dengan struktur yang konsisten dan akses yang dapat diotomatisasi, pengguna dapat menggabungkan data dari berbagai sumber secara lancar dan melakukan analisis lintas sistem dengan lebih cepat dan akurat.

Menurut Toma *et al.* (2020), “*APIs provide structured access to external data sources, enabling automated data retrieval, integration, and processing in a consistent and efficient manner,*” yang menekankan bahwa API mendukung pengambilan data secara terstruktur dan otomatis sehingga mempermudah integrasi dan analisis data dari berbagai sumber. Selain itu, akses terstruktur dan otomatis memungkinkan organisasi untuk membangun data pipelines yang terotomatisasi, di mana data dari API dapat langsung masuk ke data warehouse, sistem analitik, atau aplikasi visualisasi untuk diproses lebih lanjut. Pendekatan ini sangat penting untuk memastikan data selalu up-to-date dan siap digunakan, mendukung pengambilan keputusan yang cepat dan berbasis bukti, serta memungkinkan skalabilitas pengolahan data ketika volume atau kompleksitas data meningkat secara signifikan.

## **2. Beragam Jenis Data**

Beragam jenis data yang tersedia melalui API sebagai sumber data sekunder mencakup informasi publik, data organisasi, maupun data yang dihasilkan oleh pengguna atau sensor digital, sehingga memberikan fleksibilitas dalam pemanfaatan untuk berbagai keperluan penelitian, analisis, maupun pengembangan aplikasi. Data ini bisa berupa statistik, teks, gambar, multimedia, atau format lain yang telah distandarisasi, memungkinkan pengembang untuk mengolah dan mengintegrasikannya secara efisien ke dalam sistem atau aplikasinya. Keberagaman jenis data ini juga memfasilitasi kombinasi dari berbagai sumber untuk menghasilkan analisis yang lebih komprehensif dan mendalam, mendukung pengambilan keputusan berbasis data yang lebih akurat.

Menurut Chou *et al.* (2021), “*APIs enable access to a wide variety of data types, including structured, semi-structured, and unstructured data, allowing organizations and researchers to integrate and analyze diverse information sources efficiently,*” yang menegaskan bahwa API menyediakan akses ke berbagai tipe data sehingga memungkinkan pemrosesan dan analisis lintas sumber dengan lebih mudah dan konsisten.

Keberagaman data melalui API memungkinkan organisasi untuk menyesuaikan penggunaan data dengan kebutuhan spesifik, seperti melakukan segmentasi pelanggan, pemantauan kondisi pasar, atau analisis tren sosial, tanpa harus mengumpulkan data primer secara langsung. Pendekatan ini tidak hanya menghemat waktu dan biaya, tetapi

juga meningkatkan kualitas data yang digunakan, karena sebagian besar data sekunder melalui API telah melalui proses validasi, normalisasi, atau pembersihan oleh penyedia layanan, sehingga lebih andal untuk analisis lanjut.

### **3. Efisiensi Waktu dan Biaya**

Pemanfaatan API sebagai sumber data sekunder dapat secara signifikan mengurangi waktu dan biaya yang dibutuhkan untuk pengumpulan, integrasi, dan pemrosesan data karena API memungkinkan otomatisasi, akses *real-time*, dan pemanfaatan fungsi atau data yang sudah ada tanpa harus membangun sistem dari nol. Dengan API, organisasi dapat mengotomatisasi alur kerja seperti sinkronisasi data, pembaruan informasi, dan integrasi antar sistem, yang tidak hanya mempercepat proses tetapi juga mengurangi tenaga manual dan risiko kesalahan manusia yang memakan waktu dan biaya tambahan.

Menurut tinjauan sistematis oleh Khanyi dkk. (2024), adopsi jaringan data dan API “*significantly improve operational efficiency by increasing process speed and reducing operational costs,*” yang menegaskan bahwa penggunaan API tidak hanya mempercepat proses bisnis tetapi juga memberikan penghematan biaya operasional melalui efisiensi integrasi data dan peningkatan produktivitas kerja.

Keuntungan efisiensi waktu dan biaya ini sangat penting terutama dalam lingkungan bisnis atau penelitian yang dinamis, karena memungkinkan tim untuk lebih fokus pada analisis data, inovasi produk, dan pengambilan keputusan strategis daripada terjebak dalam tugas manual atau pengembangan ulang fungsi yang sudah tersedia. Selain itu, penggunaan API yang konsisten dan terstandarisasi membantu organisasi menurunkan biaya pemeliharaan sistem dan mempercepat waktu go-to-market untuk aplikasi baru atau fitur analitik yang bergantung pada sumber data eksternal.

### **4. Kontrol dan Keamanan Akses Data**

Kontrol dan keamanan akses data dalam konteks API adalah mekanisme penting yang menjamin bahwa hanya pengguna atau aplikasi yang berwenang yang dapat mengakses sumber daya dan informasi sensitif, sekaligus mencegah akses tidak sah, pencurian data, atau penyalahgunaan sistem melalui titik akhir API. Praktik ini mencakup autentikasi identity (misalnya API key, token OAuth 2.0, atau JWT)

untuk memverifikasi identitas peminta serta otorisasi yang memberikan hak berbeda berdasarkan peran atau izin, sehingga setiap permintaan dievaluasi terhadap kebijakan akses yang ketat sebelum data dapat dikirimkan. Mekanisme ini semakin relevan karena API sering menjadi jembatan penghubung berbagai layanan, basis data, dan sistem *cloud*, sehingga kegagalan dalam menetapkan kontrol akses yang kuat dapat membuka celah keamanan yang dieksploitasi oleh peretas.

Menurut IBM (2025), “*API security is a set of practices and procedures that protect application programming interfaces (APIs) and the data they transmit from misuse, malicious bot attacks and other cybersecurity threats,*” yang menegaskan bahwa kontrol dan keamanan akses adalah bagian utama dari menjaga integritas, kerahasiaan, dan ketersediaan data yang diakses melalui API serta memastikan bahwa hanya entitas yang sah yang dapat berinteraksi dengan sumber daya digital.

Implementasi kontrol dan keamanan akses data yang efektif tidak hanya mencegah kebocoran informasi sensitif seperti data pengguna atau finansial, tetapi juga mendukung kepatuhan terhadap regulasi privasi dan perlindungan data seperti GDPR atau PCI-DSS, serta membantu mempertahankan kepercayaan pengguna dan mitra bisnis. Desain API yang aman biasanya melibatkan enkripsi komunikasi (TLS/HTTPS), validasi *input*, enkripsi data dalam transit, serta monitoring dan logging aktivitas untuk deteksi dini anomali atau serangan, sehingga lingkungan data tetap terlindungi secara menyeluruh.

## **5. Kemudahan Integrasi dan Analisis Lanjutan**

Kemudahan integrasi dan analisis lanjutan melalui API memungkinkan organisasi atau peneliti untuk menggabungkan data dari berbagai sumber secara cepat dan konsisten, sehingga mempermudah pembuatan data pipeline yang mengalirkan informasi secara otomatis ke sistem analitik, dashboard, atau aplikasi visualisasi. Dengan struktur data yang standar dan akses yang terotomatisasi, API mendukung pemrosesan data lintas platform, memungkinkan pengolahan, transformasi, dan analisis lanjutan untuk menghasilkan wawasan yang mendalam, prediksi berbasis data, serta pengambilan keputusan yang lebih tepat. Kemampuan integrasi ini juga mempercepat inovasi karena data dari berbagai layanan dapat digabungkan tanpa perlu membangun sistem

pengumpulan data primer dari awal, sehingga menghemat waktu dan sumber daya.

Menurut Zhang *et al.* (2022), “APIs facilitate seamless integration of heterogeneous data sources and enable advanced analytics by providing structured, machine-readable data that can be processed and combined efficiently,” yang menegaskan bahwa API tidak hanya mempermudah integrasi data, tetapi juga mendukung analisis lanjutan dengan menyediakan data yang siap untuk diproses secara otomatis.

Kemudahan integrasi melalui API memungkinkan pengembangan sistem *real-time analytics* atau *predictive analytics*, di mana data yang diterima dapat langsung digunakan untuk memantau tren, performa bisnis, atau pola perilaku pengguna. Pendekatan ini meningkatkan fleksibilitas operasional karena organisasi dapat menyesuaikan analisis dengan kebutuhan dinamis, menggabungkan data internal dan eksternal, serta menghasilkan *insight* yang relevan tanpa hambatan teknis dalam pemrosesan data.

## E. Isu Keamanan dan Etika

Isu keamanan dalam penggunaan API sebagai sumber data sekunder menjadi perhatian utama karena API bertindak sebagai pintu gerbang antara sistem klien dan server penyedia data, sehingga rentan terhadap serangan siber seperti penyadapan data, injeksi kode, serangan DDoS, dan eksploitasi celah autentikasi. Untuk menjaga keamanan, diperlukan mekanisme autentikasi dan otorisasi yang kuat, misalnya melalui API key, OAuth 2.0, atau token JWT, sehingga hanya pihak yang sah yang dapat mengakses data. Selain itu, enkripsi komunikasi menggunakan protokol TLS/HTTPS menjadi penting untuk melindungi data selama transmisi agar informasi sensitif tidak disadap oleh pihak tidak bertanggung jawab.

Mekanisme kontrol lain yang krusial adalah pembatasan jumlah permintaan (*rate limiting*) dan *throttling*, yang berfungsi untuk mencegah penyalahgunaan API dan menjaga kinerja server tetap optimal. Praktik logging dan monitoring aktivitas API juga perlu diterapkan secara rutin, karena hal ini memungkinkan deteksi dini terhadap upaya akses tidak sah atau aktivitas mencurigakan yang dapat menimbulkan risiko terhadap integritas dan ketersediaan data. Dengan

langkah-langkah ini, organisasi dapat mengurangi kemungkinan terjadinya pelanggaran keamanan yang berdampak pada kerugian finansial maupun reputasi.

Dari perspektif etika, penggunaan API harus memperhatikan privasi pengguna, kepatuhan terhadap regulasi, dan transparansi penggunaan data. Data sekunder yang diperoleh melalui API sering kali berisi informasi sensitif, termasuk identitas, lokasi, dan perilaku pengguna, sehingga wajib mematuhi regulasi perlindungan data seperti GDPR, CCPA, atau undang-undang lokal lainnya. Etika penggunaan data juga menuntut agar data tidak disalahgunakan, misalnya dengan menjual informasi tanpa izin, menyebarkan data yang bersifat diskriminatif, atau memanipulasi informasi untuk keuntungan pribadi atau korporasi.

Tanggung jawab API terbagi antara penyedia dan pengguna API. Penyedia API bertugas menyediakan dokumentasi yang jelas, endpoint yang aman, dan mekanisme kontrol akses yang efektif, sementara pengguna API memiliki kewajiban untuk menggunakan data secara sah, mencatat sumber data, dan memastikan bahwa analisis atau aplikasi yang dikembangkan tidak merugikan individu atau kelompok tertentu. Pendekatan ini membangun ekosistem yang aman dan dapat dipercaya, sehingga data sekunder dapat dimanfaatkan secara optimal untuk penelitian, bisnis, atau pengembangan teknologi.

Secara ringkas, beberapa poin penting yang perlu diperhatikan terkait isu keamanan dan etika API meliputi: autentikasi dan otorisasi yang ketat, enkripsi data dalam transit, kepatuhan terhadap regulasi dan privasi pengguna, transparansi dan tanggung jawab penyedia serta pengguna, serta monitoring dan logging aktivitas API. Dengan penerapan prinsip-prinsip ini secara konsisten, API dapat digunakan sebagai sumber data sekunder yang aman, terpercaya, dan berkelanjutan, mendukung pengambilan keputusan berbasis data tanpa melanggar keamanan dan etika digital.

## **F. Pengambilan Data API Menggunakan Python**

Pengambilan data dari API menggunakan Python adalah proses memperoleh data sekunder dari berbagai sumber eksternal melalui permintaan HTTP, yang kemudian dapat diproses, dianalisis, dan diintegrasikan ke dalam sistem atau aplikasi. Python menjadi salah satu

bahasa pemrograman paling populer untuk tujuan ini karena memiliki pustaka (*library*) yang lengkap, mudah digunakan, dan mendukung berbagai format data seperti JSON, XML, atau CSV. Penggunaan Python memungkinkan otomatisasi proses pengambilan data, sehingga pengguna dapat mengambil data secara berkala atau *real-time* tanpa intervensi manual.

## 1. HTTP Request dan JSON

HTTP Request adalah mekanisme komunikasi antara klien dan server dalam protokol HTTP di mana klien mengirimkan permintaan berisi metode (seperti GET, POST, PUT, DELETE), endpoint (URL), header, dan opsi data lain kepada server untuk meminta atau mengirimkan data, sehingga server dapat memproses permintaan tersebut dan memberikan respons yang sesuai. HTTP Request berperan penting dalam interaksi API karena ia menentukan apa yang diminta, di mana sumber daya berada, serta bagaimana server harus menanggapi permintaan tersebut berdasarkan semua informasi yang disertakan dalam permintaan. Format dan struktur HTTP Request harus mengikuti standar protokol agar server dapat memprosesnya dengan benar, terutama ketika mengirimkan atau menerima data yang kompleks seperti JSON.

JSON (*JavaScript Object Notation*) adalah format pertukaran data berbasis teks yang ringan dan mudah dibaca oleh manusia maupun mesin, digunakan secara luas dalam API modern untuk merepresentasikan objek dan struktur data dalam permintaan maupun respons karena sifatnya yang sederhana dan interoperable antar bahasa pemrograman. JSON memungkinkan pengiriman objek bersarang dan daftar data dalam format *key-value* yang konsisten sehingga klien dan server dapat dengan cepat mem-parsing (mengurai) serta memproses data tanpa ambiguitas, yang sangat penting ketika berhadapan dengan sistem heterogen dan lintas platform.

Menurut Sakimura *et al.* (2019), “*JSON has become the de facto standard for structured data exchange in web services and APIs due to its simplicity, readability, and compatibility with many programming environments,*” yang menegaskan bahwa JSON telah menjadi standar *de facto* dalam pertukaran data terstruktur pada layanan web dan API modern karena kemudahan penerapan dan dukungan luas di berbagai teknologi.

## 2. Autentikasi API

Autentikasi API adalah proses verifikasi identitas pengguna atau aplikasi yang mencoba mengakses API untuk memastikan bahwa hanya entitas yang sah yang dapat memanfaatkan layanan atau data yang disediakan, sehingga mencegah akses tidak sah dan potensi penyalahgunaan. Proses ini biasanya melibatkan penggunaan API key, token OAuth 2.0, atau JSON Web Token (JWT) yang disertakan dalam header HTTP atau parameter permintaan, sehingga server dapat memvalidasi identitas peminta sebelum memberikan respons data. Implementasi autentikasi yang kuat menjadi kunci untuk menjaga integritas, kerahasiaan, dan keamanan sistem, terutama saat API mengakses data sensitif atau terhubung dengan layanan kritikal.

Pada praktiknya, autentikasi API tidak hanya melindungi data, tetapi juga memungkinkan kontrol granular terhadap hak akses pengguna, misalnya memberikan izin baca (*read-only*), tulis (*write*), atau admin untuk endpoint tertentu. Python dan bahasa pemrograman modern lainnya menyediakan pustaka yang memudahkan pengelolaan autentikasi, seperti menambahkan header otorisasi, menyimpan token secara aman, dan memperbarui token secara otomatis sesuai kebijakan API. Dengan mekanisme ini, organisasi dapat memastikan bahwa data yang diperoleh melalui API hanya digunakan oleh pihak yang berhak dan aktivitas akses dapat dilacak untuk audit atau analisis keamanan.

Menurut Chou *et al.* (2021), “*API authentication mechanisms such as OAuth and API keys are critical for ensuring secure and controlled access to data services, preventing unauthorized access and potential misuse of sensitive information,*” yang menegaskan bahwa autentikasi API menjadi komponen penting untuk menjaga keamanan dan kontrol akses terhadap data dan layanan digital.

## 3. Contoh Script Python Pengambilan Data API

Berikut adalah beberapa Contoh Script Python Pengambilan Data API (API Data Retrieval) menggunakan berbagai metode yang umum dipakai dalam penelitian, aplikasi, maupun analisis data.

### **[1] Contoh Pengambilan Data API (GET Request Sederhana)**

Mengambil data dari API publik (JSONPlaceholder).

```
import requests
```

```

import pandas as pd

# URL API
url = "https://jsonplaceholder.typicode.com/posts"

# Kirim request GET
response = requests.get(url)

# Cek status
if response.status_code == 200:
    data = response.json()

    # Ubah ke DataFrame
    df = pd.DataFrame(data)

    print("5 Data Pertama:")
    print(df.head())
else:
    print("Gagal mengambil data:", response.status_code)

```

## **2** Contoh API dengan Parameter (Query String)

Misalnya mengambil data berdasarkan parameter tertentu.

```
import requests
```

```
url = "https://jsonplaceholder.typicode.com/comments"
```

```
params = {
    "postId": 1
}
```

```
response = requests.get(url, params=params)
```

```
if response.status_code == 200:
```

```
    data = response.json()
```

```
    for item in data:
```

```
        print("Nama:", item['name'])
```

```
print("Email:", item['email'])
print("-" * 30)
```

### **3** Contoh API Menggunakan API Key (*Authentication*)

Beberapa API memerlukan API key.

```
import requests
```

```
url = "https://api.openweathermap.org/data/2.5/weather"
```

```
params = {
    "q": "Jakarta",
    "appid": "API_KEY_ANDA",
    "units": "metric"
}
```

```
response = requests.get(url, params=params)
```

```
if response.status_code == 200:
    data = response.json()
    print("Suhu:", data['main']['temp'], "°C")
    print("Cuaca:", data['weather'][0]['description'])
else:
    print("Error:", response.status_code)
```

### **4** Contoh POST Request (Mengirim Data ke API)

```
import requests
```

```
url = "https://jsonplaceholder.typicode.com/posts"
```

```
payload = {
    "title": "Data API",
    "body": "Contoh pengiriman data ke API",
    "userId": 1
}
```

```
response = requests.post(url, json=payload)
```

```
print("Status:", response.status_code)
print("Response JSON:", response.json())
```

### **5** Contoh Menyimpan Data API ke CSV

```
import requests
import pandas as pd

url = "https://jsonplaceholder.typicode.com/users"
response = requests.get(url)

if response.status_code == 200:
    data = response.json()
    df = pd.DataFrame(data)

    df.to_csv("data_users_api.csv", index=False)
    print("Data berhasil disimpan ke CSV")
```

# BAB XI

## INTEGRASI METODE PENGUMPULAN DATA DALAM PROYEK SAINS DATA

---

Integrasi metode pengumpulan data dalam proyek sains data memungkinkan penggunaan berbagai sumber informasi untuk menghasilkan analisis yang lebih akurat. Pendekatan ini meningkatkan efisiensi dan kualitas keputusan yang diambil dari data.

### A. Pemilihan Metode Berdasarkan Tujuan Analisis

Pemilihan metode pengumpulan data berdasarkan tujuan analisis merupakan langkah penting dalam proyek sains data untuk memastikan data yang dikumpulkan relevan, akurat, dan mampu mendukung *insight* yang diinginkan. Proses ini menekankan kesesuaian antara jenis data, teknik pengumpulan, dan tujuan analisis, sehingga hasil akhirnya dapat dipercaya dan berguna dalam pengambilan keputusan. Beberapa aspek penting dalam pemilihan metode ini antara lain:

#### 1. Menentukan Tujuan Analisis Secara Jelas

Menentukan tujuan analisis secara jelas merupakan langkah krusial dalam proyek sains data karena tujuan yang spesifik membantu peneliti atau analis dalam merancang strategi pengumpulan dan pemrosesan data yang tepat, sehingga mengurangi risiko bias dan meningkatkan relevansi hasil analisis (Kumar & Singh, 2020). Dengan tujuan yang terdefinisi dengan baik, organisasi dapat memfokuskan sumber daya secara efisien untuk memperoleh *insight* yang langsung mendukung pengambilan keputusan strategis.

Tujuan analisis yang jelas juga memungkinkan penentuan metode pengumpulan data yang sesuai, apakah data yang dibutuhkan bersifat kuantitatif untuk pengukuran objektif atau kualitatif untuk pemahaman mendalam terhadap fenomena tertentu, sehingga setiap langkah pengumpulan data dapat disesuaikan dengan jenis informasi

yang diinginkan. Hal ini penting untuk menjaga konsistensi dan validitas data yang dikumpulkan serta meminimalkan kesalahan interpretasi pada tahap analisis.

Tujuan yang terstruktur dengan baik memudahkan evaluasi hasil analisis, karena indikator keberhasilan dan metrik yang digunakan dapat diukur sesuai konteks awal yang telah ditetapkan. Dengan demikian, seluruh proses pengumpulan dan analisis data tidak hanya sistematis tetapi juga mampu memberikan *output* yang bermanfaat bagi pengembangan strategi dan perencanaan lebih lanjut dalam organisasi.

## **2. Menyesuaikan Dengan Jenis Data**

Menyesuaikan dengan jenis data merupakan aspek penting dalam proses sains data karena setiap jenis data memiliki karakteristik, format, dan kebutuhan analisis yang berbeda sehingga penggunaan metode yang tepat akan memengaruhi validitas dan relevansi hasil yang diperoleh. Data numerik (kuantitatif) dapat dianalisis dengan teknik statistik dan model matematis untuk menemukan pola, tren, dan hubungan antar variabel, sedangkan data naratif atau deskriptif (kualitatif) membutuhkan pendekatan interpretatif untuk memahami konteks dan makna di balik fenomena yang diamati (Sowmiya, 2025).

Pada praktiknya, pemilihan teknik analisis yang sesuai dengan jenis data membantu analisis menjadi lebih efisien dan akurat, karena metode yang tidak cocok dapat menghasilkan kesimpulan yang bias atau tidak bermakna; misalnya, menerapkan uji statistik parametris pada data non-numerik akan menghasilkan *output* yang tidak valid. Dengan demikian, pemahaman yang baik terhadap jenis data yang dimiliki dan penyesuaian metode analisis yang digunakan memungkinkan proses pengolahan data yang sistematis, interpretasi yang tepat, serta *insight* yang lebih kuat untuk mendukung keputusan atau rekomendasi berbasis bukti.

## **3. Mempertimbangkan Sumber Data**

Mempertimbangkan sumber data merupakan langkah penting dalam sains data karena kualitas, keandalan, dan relevansi data sangat bergantung pada asalnya, sehingga pemilihan sumber yang tepat memastikan hasil analisis dapat dipercaya dan mendukung pengambilan keputusan yang efektif (Chen *et al.*, 2019). Data dapat diperoleh dari sumber internal seperti basis data perusahaan atau sistem operasional,

maupun dari sumber eksternal seperti publikasi, portal *online*, atau lembaga riset, dan setiap sumber memiliki karakteristik serta batasan yang perlu dievaluasi sebelum digunakan.

Evaluasi sumber data meliputi aspek kredibilitas, akurasi, dan keterkinian informasi, karena data yang usang atau tidak valid dapat menimbulkan bias dan kesalahan dalam interpretasi hasil analisis. Dengan mempertimbangkan sumber secara seksama, analis dapat memilih metode pengumpulan dan pembersihan data yang sesuai, sekaligus mengurangi risiko ketidakakuratan yang dapat memengaruhi keputusan strategis.

#### **4. Evaluasi Biaya dan Waktu**

Evaluasi biaya dan waktu dalam suatu proyek merupakan langkah penting untuk memastikan bahwa anggaran yang dialokasikan dan durasi yang direncanakan dapat terpenuhi secara efektif, karena ketidaksesuaian antara biaya dan waktu sering menjadi indikator adanya risiko, inefisiensi, atau kurangnya kontrol yang dapat menyebabkan kegagalan proyek. Dengan melakukan evaluasi secara sistematis, manajer proyek dapat mengidentifikasi deviasi dari rencana awal, menilai dampaknya terhadap kinerja keseluruhan proyek, dan mengambil tindakan korektif sebelum masalah menjadi lebih besar (Suksmono & Sari, 2024).

Proses evaluasi ini tidak hanya mencakup pengukuran pengeluaran aktual dan waktu yang telah digunakan, tetapi juga membandingkannya dengan anggaran dan jadwal yang telah ditetapkan sehingga memungkinkan perencanaan ulang atau penyesuaian strategi bila diperlukan. Pendekatan ini sangat penting dalam menjaga efisiensi alokasi sumber daya, meminimalkan pembengkakan biaya, dan mencegah keterlambatan penyelesaian yang berdampak negatif pada hasil akhir proyek serta kepuasan pemangku kepentingan.

#### **5. Menjamin Validitas dan Reliabilitas**

Menjamin validitas dan reliabilitas dalam proses pengumpulan dan analisis data merupakan aspek penting untuk memastikan bahwa hasil yang diperoleh benar-benar mencerminkan realitas serta dapat diandalkan secara konsisten, sehingga keputusan yang diambil berdasarkan data tersebut menjadi sah dan tidak menyesatkan (Karnia, 2024). Validitas memastikan bahwa alat ukur atau metode yang

digunakan benar-benar mengukur apa yang dimaksud untuk diukur, sedangkan reliabilitas menjamin bahwa pengukuran tersebut menghasilkan hasil yang konsisten bila diulang dalam kondisi yang serupa.

Jika validitas dan reliabilitas data tidak dijaga dengan baik, risiko terjadinya bias, kesalahan interpretasi, dan kesimpulan yang keliru akan meningkat drastis, sehingga integritas keseluruhan proyek sains data menjadi lemah dan *outputnya* kurang dapat dipercaya. Proses pengujian validitas dan reliabilitas mencakup berbagai teknik evaluasi, seperti pengujian kriteria pengukuran dan pengujian konsistensi internal, yang membantu peneliti atau analis dalam mengidentifikasi kelemahan pada instrumen dan memperbaikinya sebelum data digunakan dalam tahap analisis yang lebih lanjut.

## **B. Integrasi Data Primer dan Sekunder**

Integrasi data primer dan sekunder merupakan proses sistematis dalam menggabungkan data yang dikumpulkan secara langsung oleh peneliti dengan data yang telah tersedia dari sumber lain guna menghasilkan analisis yang lebih komprehensif, akurat, dan kontekstual. Data primer biasanya diperoleh melalui survei, wawancara, observasi, eksperimen, atau kuesioner yang dirancang khusus sesuai kebutuhan penelitian, sedangkan data sekunder berasal dari laporan pemerintah, publikasi ilmiah, basis data institusi, laporan industri, arsip perusahaan, maupun sumber digital lainnya. Integrasi keduanya menjadi penting karena masing-masing memiliki keunggulan dan keterbatasan; data primer cenderung spesifik dan kontekstual, tetapi terbatas dalam cakupan, sementara data sekunder umumnya lebih luas dan historis, namun tidak selalu sepenuhnya sesuai dengan kebutuhan penelitian.

Secara konseptual, integrasi data primer dan sekunder bertujuan untuk memperkuat validitas hasil penelitian melalui pendekatan triangulasi, memperluas cakupan analisis, serta meningkatkan efisiensi waktu dan biaya. Dengan menggabungkan kedua jenis data, peneliti dapat membandingkan temuan lapangan dengan data makro atau tren jangka panjang sehingga interpretasi menjadi lebih mendalam. Selain itu, integrasi ini mendukung pengembangan model analitik yang lebih robust karena variabel yang digunakan berasal dari sumber yang

beragam dan saling melengkapi. Berikut penjelasan rinci mengenai tahapan dan aspek penting dalam integrasi data primer dan sekunder:

### **1. Identifikasi Tujuan dan Kebutuhan Data**

Identifikasi tujuan dan kebutuhan data merupakan tahap fundamental dalam setiap penelitian maupun proyek sains data karena menjadi dasar dalam menentukan arah analisis, jenis data yang dikumpulkan, serta metode yang digunakan untuk mengolahnya. Pada tahap ini, peneliti atau analis harus merumuskan pertanyaan penelitian secara spesifik, membahas variabel yang relevan, serta memastikan bahwa data yang dikumpulkan benar-benar selaras dengan tujuan strategis yang ingin dicapai agar tidak terjadi pemborosan sumber daya dan distorsi analisis. Selain itu, proses ini membantu membatasi ruang lingkup penelitian sehingga fokus pada informasi yang esensial, mengurangi *noise* data, serta meningkatkan validitas dan reliabilitas hasil yang diperoleh melalui perencanaan kebutuhan data yang terstruktur dan berbasis masalah.

Pada konteks manajemen dan analitik modern, penentuan kebutuhan data harus mempertimbangkan kesesuaian antara tujuan bisnis atau penelitian dengan indikator kinerja yang terukur, sehingga setiap variabel yang dipilih memiliki kontribusi langsung terhadap pengambilan keputusan berbasis bukti. Kebutuhan data juga harus dianalisis dari sisi ketersediaan, kualitas, relevansi waktu, serta aspek etika dan legalitas penggunaannya agar proses pengumpulan dan pengolahan data tidak menimbulkan risiko metodologis maupun hukum di kemudian hari. Sebagaimana ditegaskan oleh Provost dan Fawcett (2018), *“formulating the problem precisely and determining what data are needed is often more critical than the choice of algorithm,”* yang menunjukkan bahwa keberhasilan analisis sangat ditentukan oleh kejelasan tujuan serta ketepatan dalam mengidentifikasi kebutuhan data sejak awal perencanaan.

### **2. Evaluasi Kualitas Data**

Evaluasi kualitas data merupakan proses krusial dalam penelitian dan proyek sains data karena menentukan sejauh mana data yang digunakan layak dijadikan dasar analisis dan pengambilan keputusan. Proses ini mencakup penilaian terhadap dimensi akurasi, kelengkapan, konsistensi, validitas, ketepatan waktu, dan relevansi, sehingga data yang tidak memenuhi standar dapat segera diperbaiki atau dieliminasi

sebelum memasuki tahap analitik yang lebih kompleks. Tanpa evaluasi yang sistematis, data yang bias, tidak lengkap, atau tidak mutakhir berpotensi menghasilkan kesimpulan yang keliru serta merugikan organisasi atau peneliti dalam jangka panjang.

Pada praktiknya, evaluasi kualitas data dilakukan melalui teknik pembersihan data, pemeriksaan anomali, validasi silang dengan sumber lain, serta pengujian reliabilitas instrumen pengumpulan data, terutama jika melibatkan data primer. Selain aspek teknis, evaluasi juga mempertimbangkan kesesuaian data dengan tujuan penelitian agar tidak terjadi ketidaksinkronan antara variabel yang tersedia dan kebutuhan analisis yang telah dirumuskan sebelumnya. Sebagaimana dikemukakan oleh Kelleher dan Tierney (2018), “*data quality is a critical determinant of the success of any data analytics project,*” yang menegaskan bahwa keberhasilan analisis sangat bergantung pada mutu data yang digunakan sejak tahap awal pengolahan.

### **3. Harmonisasi dan Standarisasi Variabel**

Harmonisasi dan standarisasi variabel merupakan proses penyesuaian dan penyelarasan definisi, format, skala pengukuran, serta kategori data agar variabel yang berasal dari berbagai sumber dapat dibandingkan dan digabungkan secara konsisten dalam satu kerangka analisis. Proses ini menjadi sangat penting terutama dalam integrasi data primer dan sekunder, karena perbedaan metode pengumpulan, perbedaan periode waktu, maupun variasi definisi operasional sering kali menimbulkan inkonsistensi yang dapat mengganggu validitas hasil penelitian. Melalui harmonisasi, peneliti menetapkan kesamaan makna antarvariabel, sedangkan standarisasi memastikan bahwa unit, kode, dan struktur data mengikuti aturan yang seragam sehingga analisis dapat dilakukan tanpa distorsi metodologis.

Pada praktiknya, harmonisasi dapat mencakup konversi satuan ukuran, penyatuan kategori klasifikasi, penyesuaian rentang waktu, serta rekoding variabel agar sesuai dengan kerangka konseptual penelitian, sementara standarisasi dapat melibatkan penggunaan pedoman metadata, data dictionary, dan standar internasional tertentu untuk menjamin interoperabilitas data. Proses ini juga memerlukan dokumentasi yang transparan agar setiap perubahan terhadap variabel dapat ditelusuri kembali dan direplikasi oleh peneliti lain, sehingga menjaga akuntabilitas ilmiah serta konsistensi analitik. Sebagaimana dinyatakan

oleh Knafllic (2019), “*standardizing definitions and measures is essential to ensure consistency and comparability in data analysis,*” yang menegaskan bahwa keseragaman definisi dan ukuran merupakan fondasi utama dalam menghasilkan analisis yang dapat dipercaya.

#### **4. Pembersihan dan Transformasi Data**

Pembersihan dan transformasi data merupakan tahapan esensial dalam siklus pengolahan data yang bertujuan untuk memastikan bahwa dataset berada dalam kondisi yang akurat, konsisten, dan siap dianalisis sesuai dengan tujuan penelitian atau kebutuhan bisnis. Pembersihan data mencakup identifikasi dan penanganan nilai hilang, duplikasi, inkonsistensi format, kesalahan entri, serta *outlier* yang dapat mengganggu distribusi dan memengaruhi hasil analisis secara signifikan apabila tidak ditangani secara sistematis. Sementara itu, transformasi data melibatkan proses perubahan struktur atau representasi data seperti normalisasi, standarisasi skala, agregasi, encoding variabel kategorikal, serta pembuatan variabel turunan agar data lebih sesuai dengan model analitik yang akan digunakan.

Pada praktiknya, proses ini tidak hanya bersifat teknis, tetapi juga konseptual karena memerlukan pemahaman terhadap konteks data dan tujuan analisis sehingga setiap keputusan pembersihan maupun transformasi tidak menghilangkan makna substantif yang terkandung di dalamnya. Kesalahan dalam tahap ini dapat menghasilkan bias, distorsi interpretasi, atau bahkan kesimpulan yang menyesatkan, sehingga dokumentasi setiap langkah menjadi bagian penting dari akuntabilitas dan reproduktibilitas penelitian. Sebagaimana ditegaskan oleh Kelleher dan Tierney (2018), “*a large proportion of effort in data science projects is spent on cleaning and preparing data,*” yang menunjukkan bahwa keberhasilan analisis sangat ditentukan oleh kualitas proses persiapan data sebelum pemodelan dilakukan.

#### **5. Proses Penggabungan Data**

Proses penggabungan data merupakan tahapan penting dalam integrasi berbagai sumber informasi yang bertujuan untuk menyatukan dataset berbeda ke dalam satu struktur analisis yang koheren dan saling terhubung. Penggabungan ini umumnya dilakukan melalui teknik seperti *merging*, *joining*, *concatenation*, atau *record linkage* berdasarkan variabel kunci yang sama, seperti identitas individu, kode wilayah, atau

periode waktu, sehingga setiap entitas dapat dipetakan secara akurat antar sumber. Ketepatan dalam menentukan kunci penggabungan dan metode yang digunakan sangat menentukan kualitas hasil integrasi, karena kesalahan dalam tahap ini dapat menyebabkan duplikasi, kehilangan data, atau pencocokan yang tidak valid yang berdampak pada kesimpulan akhir.

Pada praktik sains data modern, proses penggabungan tidak hanya bersifat teknis, tetapi juga membutuhkan pemahaman mendalam terhadap struktur, konteks, dan keterbatasan masing-masing dataset agar integrasi yang dilakukan tetap mempertahankan makna substantif data tersebut. Evaluasi pasca-penggabungan melalui pemeriksaan distribusi, jumlah observasi, dan konsistensi variabel menjadi langkah penting untuk memastikan bahwa integrasi tidak menimbulkan distorsi analitis maupun bias sistematis. Sebagaimana dinyatakan oleh Provost dan Fawcett (2018), *“combining data from multiple sources can provide a richer view of the problem than any single dataset alone,”* yang menegaskan bahwa penggabungan data yang tepat mampu meningkatkan kedalaman dan kualitas pemahaman terhadap suatu permasalahan.

## 6. Validasi dan Uji Konsistensi

Validasi dan uji konsistensi merupakan tahap krusial dalam pengolahan dan integrasi data yang bertujuan untuk memastikan bahwa data yang telah dikumpulkan, dibersihkan, dan digabungkan benar-benar akurat, logis, serta sesuai dengan tujuan analisis yang telah ditetapkan. Validasi dilakukan dengan memeriksa kesesuaian data terhadap aturan, definisi operasional, maupun sumber pembanding yang kredibel, sedangkan uji konsistensi berfokus pada pemeriksaan keteraturan hubungan antarvariabel, keselarasan nilai dalam rentang waktu tertentu, serta stabilitas distribusi data setelah proses transformasi atau penggabungan. Tanpa validasi dan uji konsistensi yang sistematis, risiko terjadinya bias, kesalahan interpretasi, maupun kesimpulan yang menyesatkan akan meningkat karena data yang digunakan belum terjamin keandalannya secara metodologis.

Pada praktiknya, validasi dapat dilakukan melalui teknik pemeriksaan statistik deskriptif, pengecekan *outlier*, *cross-validation* dengan dataset lain, serta pengujian reliabilitas instrumen apabila data berasal dari survei atau eksperimen, sementara uji konsistensi mencakup

pemeriksaan integritas referensial dan logika relasional antarvariabel. Proses ini juga melibatkan dokumentasi setiap temuan anomali dan tindakan korektif yang dilakukan agar seluruh tahapan analisis dapat direplikasi dan diaudit secara transparan. Sebagaimana ditegaskan oleh Kelleher dan Tierney (2018), “*ensuring that data are consistent and correctly represent the real-world entities they describe is fundamental to reliable analytics*,” yang menunjukkan bahwa konsistensi dan representasi yang tepat merupakan fondasi utama dalam menghasilkan analisis yang dapat dipercaya.

### C. Pengumpulan Data Multisumber

Pengumpulan data multisumber merupakan pendekatan sistematis dalam memperoleh data dari berbagai sumber yang berbeda untuk menghasilkan pemahaman yang lebih komprehensif, akurat, dan kontekstual terhadap suatu fenomena atau permasalahan. Dalam proyek penelitian maupun sains data, pendekatan ini sangat penting karena satu sumber data saja sering kali tidak mampu memberikan gambaran yang utuh, baik dari sisi kedalaman informasi maupun cakupan populasi. Dengan memanfaatkan berbagai sumber, organisasi atau peneliti dapat meningkatkan validitas temuan, mengurangi bias, serta memperkaya perspektif analisis melalui kombinasi data yang saling melengkapi.

Secara konseptual, pengumpulan data multisumber mengacu pada integrasi data yang berasal dari sumber primer dan sekunder, sumber internal dan eksternal, serta data terstruktur maupun tidak terstruktur. Data primer dapat diperoleh melalui survei, wawancara, observasi lapangan, eksperimen, atau kuesioner yang dirancang khusus sesuai kebutuhan penelitian. Sementara itu, data sekunder dapat berasal dari laporan pemerintah, publikasi ilmiah, *database* perusahaan, laporan industri, media sosial, sensor digital, maupun platform daring lainnya. Kombinasi ini memungkinkan peneliti untuk menghubungkan informasi mikro (misalnya persepsi individu) dengan data makro (misalnya statistik populasi atau tren industri).

#### 1. Identifikasi Tujuan dan Ruang Lingkup

Identifikasi tujuan dan ruang lingkup merupakan tahap awal yang sangat menentukan dalam perencanaan penelitian maupun proyek sains data karena menjadi dasar dalam menetapkan arah analisis, batasan

kajian, serta strategi pengumpulan dan pengolahan data. Tujuan yang dirumuskan secara jelas membantu peneliti atau analis menentukan pertanyaan utama yang ingin dijawab, indikator kinerja yang relevan, serta hasil yang diharapkan sehingga seluruh proses berjalan secara terarah dan efisien. Ruang lingkup yang terdefinisi dengan baik juga berfungsi membatasi fokus kajian agar tidak melebar secara berlebihan, sehingga sumber daya, waktu, dan tenaga dapat dialokasikan secara optimal untuk aspek yang paling strategis dan berdampak signifikan terhadap hasil akhir.

Pada praktiknya, identifikasi tujuan dan ruang lingkup mencakup penentuan populasi atau objek penelitian, batasan temporal dan geografis, variabel yang akan dianalisis, serta asumsi yang digunakan sebagai dasar analisis, sehingga setiap keputusan metodologis memiliki landasan yang rasional dan terukur. Proses ini juga membantu mengurangi ambiguitas dalam interpretasi hasil karena sejak awal telah ditetapkan parameter yang jelas mengenai apa yang termasuk dan tidak termasuk dalam penelitian atau proyek tersebut. Sebagaimana ditegaskan oleh Provost dan Fawcett (2018), *“formulating the problem precisely is the most important step in the data science process,”* yang menunjukkan bahwa kejelasan dalam merumuskan tujuan dan batasan masalah menjadi fondasi utama bagi keberhasilan analisis dan pengambilan keputusan berbasis data.

## **2. Klasifikasi Jenis Sumber Data**

Klasifikasi jenis sumber data merupakan langkah penting dalam penelitian dan proyek sains data karena membantu peneliti memahami karakteristik, asal, serta potensi pemanfaatan setiap data yang digunakan dalam analisis. Secara umum, sumber data dapat dikategorikan menjadi data primer yang diperoleh langsung melalui survei, wawancara, observasi, atau eksperimen, dan data sekunder yang berasal dari laporan resmi, publikasi ilmiah, arsip institusi, maupun basis data yang telah tersedia sebelumnya.

Sumber data juga dapat dibedakan berdasarkan asalnya menjadi sumber internal yang dihasilkan dari aktivitas organisasi seperti data transaksi dan laporan operasional, serta sumber eksternal yang diperoleh dari lingkungan luar seperti data pasar, statistik pemerintah, dan data industri. Pengelompokan ini memudahkan proses perencanaan pengumpulan dan integrasi data karena setiap kategori memiliki tingkat

kontrol, aksesibilitas, serta reliabilitas yang berbeda sehingga memerlukan strategi pengelolaan yang sesuai.

### **3. Metode Pengumpulan Data**

Metode pengumpulan data merupakan prosedur sistematis yang digunakan untuk memperoleh informasi yang relevan, akurat, dan sesuai dengan tujuan penelitian atau kebutuhan analisis dalam suatu proyek. Pemilihan metode yang tepat sangat menentukan kualitas data yang dihasilkan karena setiap teknik memiliki karakteristik, kelebihan, dan keterbatasan yang berbeda tergantung pada konteks, jenis variabel, serta populasi yang diteliti.

Secara umum, metode pengumpulan data dapat dibedakan menjadi metode primer seperti survei, wawancara, observasi, dan eksperimen yang memungkinkan peneliti memperoleh data langsung dari sumber pertama, serta metode sekunder yang memanfaatkan dokumen, arsip, laporan resmi, dan basis data yang telah tersedia sebelumnya. Dalam perkembangan teknologi digital, metode pengumpulan juga mencakup teknik berbasis teknologi seperti *web scraping*, penggunaan *Application Programming Interface* (API), sensor *Internet of Things* (IoT), dan pengumpulan data media sosial yang memungkinkan akses terhadap data dalam jumlah besar dan waktu nyata.

### **4. Evaluasi Kredibilitas dan Kualitas Sumber**

Evaluasi kredibilitas dan kualitas sumber merupakan langkah penting dalam memastikan bahwa data yang digunakan dalam penelitian atau proyek sains data berasal dari sumber yang dapat dipercaya, relevan, dan memiliki standar metodologis yang jelas. Proses ini mencakup penilaian terhadap reputasi institusi atau penulis, transparansi metode pengumpulan data, kelengkapan metadata, konsistensi pelaporan, serta kesesuaian konteks dan periode waktu dengan tujuan analisis yang telah ditetapkan.

Evaluasi kualitas sumber juga mempertimbangkan aspek akurasi, objektivitas, validitas, dan reliabilitas informasi agar data yang digunakan tidak mengandung bias sistematis atau kesalahan yang dapat memengaruhi hasil analisis secara signifikan. Sebagaimana ditegaskan oleh Creswell dan Creswell (2018), “*the quality of research depends on the validity and reliability of the information collected,*” yang menunjukkan bahwa kredibilitas sumber menjadi fondasi utama dalam

menghasilkan temuan yang dapat dipertanggungjawabkan secara ilmiah dan profesional.

## D. Pipeline Pengumpulan Data Berbasis Python

Pipeline pengumpulan data berbasis Python merupakan rangkaian proses terstruktur dan terotomatisasi yang digunakan untuk memperoleh, memproses, menyimpan, serta memvalidasi data dari berbagai sumber secara sistematis menggunakan bahasa pemrograman Python. Pipeline ini dirancang agar proses pengumpulan data dapat berjalan secara berulang (*reproducible*), efisien, dan terdokumentasi dengan baik, sehingga meminimalkan kesalahan manual serta meningkatkan kualitas data yang siap dianalisis. Dalam konteks proyek sains data modern, pipeline tidak hanya berfungsi sebagai alat teknis, tetapi juga sebagai kerangka kerja manajemen data yang memastikan integritas, konsistensi, dan skalabilitas sistem. Secara umum, pipeline pengumpulan data berbasis Python terdiri dari beberapa tahapan utama berikut:

### 1. Perencanaan dan Spesifikasi Kebutuhan Data

Perencanaan dan spesifikasi kebutuhan data merupakan tahap fundamental yang menentukan arah, kedalaman, serta relevansi suatu penelitian atau proyek sains data karena pada fase ini ditetapkan secara sistematis jenis data apa yang dibutuhkan, untuk tujuan apa data tersebut dikumpulkan, dan bagaimana data tersebut akan digunakan dalam proses analisis. Perencanaan yang matang membantu memastikan bahwa data yang dihimpun benar-benar selaras dengan rumusan masalah, indikator kinerja, serta hasil yang ingin dicapai, sehingga meminimalkan pengumpulan data yang tidak relevan atau berlebihan yang dapat menghambat efisiensi dan meningkatkan kompleksitas pengolahan.

Spesifikasi kebutuhan data juga mencakup penentuan format data, sumber yang akan digunakan, periode waktu yang dicakup, tingkat granularitas informasi, serta standar kualitas yang harus dipenuhi agar data dapat dianalisis secara valid dan reliabel. Selain itu, tahap ini membantu mengidentifikasi potensi kendala seperti keterbatasan akses, isu privasi, serta risiko bias, sehingga strategi mitigasi dapat dirancang sejak awal dan proyek berjalan secara terarah serta terkendali.

## 2. Akuisisi Data (*Data Ingestion*)

Akuisisi data atau data ingestion merupakan proses awal dalam pipeline data yang berfungsi untuk mengumpulkan dan memasukkan data dari berbagai sumber ke dalam sistem penyimpanan atau lingkungan analitik agar dapat diproses lebih lanjut. Proses ini mencakup pengambilan data dari sumber terstruktur maupun tidak terstruktur seperti *database* relasional, file CSV atau JSON, API, sensor IoT, media sosial, hingga sistem *streaming real-time*, dengan tetap memperhatikan aspek keamanan, autentikasi, serta integritas transfer data.

Pada praktiknya, akuisisi data dapat dilakukan secara batch maupun *real-time* tergantung pada kebutuhan analisis dan karakteristik sistem yang digunakan, sehingga organisasi harus menentukan arsitektur ingestion yang sesuai dengan volume, kecepatan, dan variasi data yang dikelola. Ketepatan dalam merancang mekanisme ingestion sangat penting karena kesalahan pada tahap ini dapat menyebabkan kehilangan data, duplikasi, inkonsistensi format, atau gangguan pada proses analitik berikutnya.

## 3. Validasi Awal dan Logging

Validasi awal dan logging merupakan tahapan penting dalam pipeline pengelolaan data yang bertujuan untuk memastikan bahwa data yang baru diakuisisi memenuhi standar struktur, format, dan kualitas sebelum diproses lebih lanjut. Validasi awal dilakukan dengan memeriksa kesesuaian tipe data, kelengkapan atribut, konsistensi nilai, serta mendeteksi anomali atau kesalahan dasar yang dapat mengganggu proses transformasi dan analisis berikutnya.

Proses logging berfungsi sebagai mekanisme pencatatan sistematis terhadap setiap aktivitas dalam pipeline, termasuk waktu eksekusi, jumlah data yang diproses, kesalahan yang terjadi, serta status keberhasilan atau kegagalan suatu proses. Dengan adanya logging yang terstruktur, tim pengelola data dapat melakukan pelacakan (*traceability*), audit, serta debugging secara lebih efisien apabila terjadi gangguan atau inkonsistensi dalam sistem.

## 4. Pembersihan Data (*Data Cleaning*)

Pembersihan data (*data cleaning*) merupakan proses sistematis untuk mengidentifikasi, memperbaiki, atau menghapus data yang tidak akurat, tidak lengkap, duplikat, inkonsisten, maupun mengandung

anomali sehingga dataset menjadi layak digunakan dalam analisis dan pengambilan keputusan. Proses ini mencakup berbagai teknik seperti penanganan *missing values*, standarisasi format, normalisasi variabel, penghapusan *outlier* yang tidak relevan, serta rekonsiliasi inkonsistensi antar-sumber data guna meningkatkan reliabilitas dan validitas hasil analitik.

Pada praktiknya, pembersihan data tidak hanya berfokus pada aspek teknis, tetapi juga pada pemahaman konteks bisnis dan tujuan analisis agar setiap tindakan korektif tidak menghilangkan informasi penting yang justru bernilai strategis. Pendekatan ini menjadi krusial terutama dalam lingkungan *big data* dan sistem otomatis, karena kesalahan kecil pada tahap awal dapat menghasilkan bias sistemik yang berdampak besar pada model prediktif maupun laporan manajerial.

## 5. Transformasi Data

Transformasi data merupakan proses mengubah struktur, format, atau nilai data mentah menjadi bentuk yang lebih terorganisasi, konsisten, dan sesuai dengan kebutuhan analisis maupun sistem target seperti data warehouse atau model pembelajaran mesin. Proses ini mencakup berbagai teknik seperti agregasi, normalisasi, encoding variabel kategorikal, scaling numerik, integrasi lintas sumber, serta konversi skema agar data dapat digunakan secara optimal dalam proses komputasi dan pengambilan keputusan.

Pada lingkungan manajemen data modern, transformasi tidak hanya berfungsi sebagai tahap teknis dalam skema ETL (*Extract, Transform, Load*), tetapi juga sebagai mekanisme untuk memastikan keselarasan makna data (*semantic consistency*) di seluruh sistem organisasi. Transformasi yang dirancang dengan baik mampu mengurangi redundansi, meningkatkan interoperabilitas antar-platform, serta mempercepat proses analitik karena data telah disiapkan dalam format yang sesuai dengan kebutuhan pengguna akhir.

## 6. Integrasi dan Penyimpanan Data

Integrasi dan penyimpanan data merupakan tahapan strategis dalam manajemen data modern yang bertujuan untuk menggabungkan berbagai sumber data heterogen ke dalam satu sistem terpadu sehingga informasi dapat diakses secara konsisten dan terstruktur. Proses integrasi mencakup harmonisasi skema, penyelarasan format, resolusi konflik

data, serta penggabungan entitas agar tercipta satu sumber kebenaran (*single source of truth*) yang mendukung analisis lintas fungsi dalam organisasi.

Tahap penyimpanan dilakukan dengan memilih arsitektur yang sesuai seperti data warehouse, data lake, atau data lakehouse yang mampu menampung volume besar data terstruktur maupun tidak terstruktur dengan performa dan skalabilitas yang optimal. Penyimpanan yang dirancang dengan baik harus mempertimbangkan aspek keamanan, kontrol akses, efisiensi query, retensi data, serta kepatuhan terhadap regulasi agar sistem tetap andal dan berkelanjutan dalam jangka panjang.

Sebagaimana ditegaskan oleh Kimball dan Ross (2018), “*the data warehouse is the foundation for business intelligence*,” yang menunjukkan bahwa integrasi dan penyimpanan data yang terstruktur menjadi dasar utama dalam menghasilkan informasi strategis bagi pengambilan keputusan organisasi. Dengan demikian, keberhasilan analitik dan pelaporan manajerial sangat bergantung pada kemampuan organisasi dalam mengintegrasikan serta menyimpan data secara terstandar, aman, dan terkelola dengan baik.

## **E. Studi Kasus Proyek *End-to-End***

### **PT NUSANTARA MART**

PT Nusantara Mart memiliki lebih dari 150 cabang di berbagai kota dan menghadapi masalah overstock pada beberapa produk serta stockout pada produk lain, yang menyebabkan kerugian finansial dan turunya kepuasan pelanggan. Selama ini data penjualan, data persediaan gudang, data promosi, dan data pelanggan tersimpan di sistem yang terpisah sehingga manajemen kesulitan memperoleh gambaran menyeluruh untuk pengambilan keputusan strategis. Proyek *end-to-end* dimulai dari tahap identifikasi kebutuhan bisnis, yaitu meningkatkan akurasi prediksi permintaan produk hingga 20% dalam satu tahun dan menurunkan tingkat kehabisan stok sebesar 15%. Tim proyek terdiri dari data *engineer*, data *analyst*, data *scientist*, dan perwakilan divisi operasional yang bekerja secara kolaboratif untuk memastikan solusi teknis selaras dengan kebutuhan bisnis.

Pada tahap pengumpulan data, tim mengintegrasikan berbagai sumber seperti *database* transaksi kasir (POS), sistem manajemen gudang, data kampanye promosi, serta data eksternal seperti tren

musiman dan hari libur nasional. Data kemudian melalui proses validasi awal untuk memastikan konsistensi format tanggal, kesesuaian kode produk, serta kelengkapan atribut penting seperti lokasi toko dan kategori barang. Tahap berikutnya adalah pembersihan dan transformasi data, termasuk penanganan *missing values*, penghapusan duplikasi transaksi, normalisasi harga, serta agregasi data harian menjadi mingguan untuk kebutuhan model prediksi. Setelah itu, data dimuat ke dalam data warehouse berbasis *cloud* yang memungkinkan akses terpusat dan mendukung analitik skala besar.

Pada tahap analitik, tim data *scientist* membangun model *machine learning* untuk memprediksi permintaan produk berdasarkan histori penjualan, pola musiman, dan efek promosi. Model tersebut diuji menggunakan data historis (*backtesting*) untuk memastikan akurasi dan stabilitas sebelum diterapkan ke sistem produksi. Tahap implementasi dilakukan dengan mengintegrasikan hasil prediksi ke dalam dashboard manajemen berbasis business intelligence sehingga manajer cabang dapat melihat rekomendasi jumlah pemesanan stok secara real time. Sistem juga dilengkapi logging dan monitoring untuk mendeteksi error pada pipeline serta memastikan performa model tetap optimal dari waktu ke waktu.

Hasil proyek menunjukkan peningkatan akurasi prediksi sebesar 23% dan penurunan tingkat stockout sebesar 18% dalam enam bulan pertama implementasi. Studi kasus ini menunjukkan bahwa proyek *end-to-end* tidak hanya mencakup aspek teknis seperti pengolahan dan analisis data, tetapi juga melibatkan perencanaan strategis, kolaborasi lintas fungsi, manajemen perubahan, serta evaluasi berkelanjutan untuk memastikan solusi memberikan nilai bisnis yang nyata.

# BAB XII

## PENUTUP

---

Penutup mengenai metode pengumpulan data untuk sains data menegaskan bahwa kualitas analisis dan ketepatan model sangat ditentukan oleh ketelitian dalam merancang strategi pengumpulan data sejak tahap awal penelitian. Berbagai metode seperti observasi, survei, eksperimen, *web scraping*, serta pemanfaatan basis data sekunder harus dipilih secara selektif agar sesuai dengan tujuan analisis dan karakteristik permasalahan yang dikaji. Selain itu, aspek validitas, reliabilitas, etika, serta keamanan data menjadi faktor krusial yang tidak dapat diabaikan dalam proses pengumpulan data. Dengan perencanaan yang sistematis dan pengelolaan data yang terstruktur, metode pengumpulan data yang tepat akan menghasilkan fondasi yang kuat bagi pengembangan *insight*, pengambilan keputusan, serta inovasi berbasis sains data.

### A. Ringkasan Metode Pengumpulan Data

Ringkasan metode pengumpulan data dalam konteks sains data menekankan bahwa proses ini merupakan fondasi utama dalam keseluruhan siklus analitik, mulai dari perumusan masalah hingga pengambilan keputusan berbasis model. Metode pengumpulan data tidak hanya berfungsi sebagai sarana memperoleh informasi, tetapi juga menentukan kualitas, representativitas, dan integritas dataset yang akan dianalisis. Oleh karena itu, pemilihan metode harus mempertimbangkan tujuan penelitian, jenis data yang dibutuhkan (kuantitatif atau kualitatif), serta konteks operasional tempat data tersebut dihasilkan. Ketidaktepatan dalam tahap ini berpotensi menimbulkan bias, inkonsistensi, dan kesalahan interpretasi pada tahap analisis lanjutan.

Secara umum, metode pengumpulan data dapat diklasifikasikan menjadi data primer dan data sekunder. Data primer diperoleh secara langsung melalui survei, wawancara, observasi, eksperimen, maupun sensor digital dan perangkat IoT, sehingga memungkinkan peneliti atau

analisis mengontrol variabel dan memastikan kesesuaian dengan kebutuhan analitik. Sementara itu, data sekunder bersumber dari basis data internal organisasi, laporan resmi, publikasi pemerintah, repositori daring, hingga platform digital yang menyediakan data dalam skala besar. Dalam praktik sains data modern, integrasi antara kedua jenis data ini sering dilakukan untuk memperkaya perspektif analisis dan meningkatkan akurasi model prediktif.

Perkembangan teknologi informasi turut memperluas teknik pengumpulan data melalui pemanfaatan *Application Programming Interface* (API), *web scraping*, *data streaming*, serta sistem pencatatan transaksi otomatis. Metode ini memungkinkan pengumpulan data dalam volume besar (*big data*) dengan kecepatan tinggi dan variasi format yang beragam, seperti teks, gambar, audio, dan log aktivitas pengguna. Namun demikian, kompleksitas tersebut menuntut adanya mekanisme validasi, pembersihan (*data cleaning*), serta standarisasi format agar data yang dikumpulkan siap diproses lebih lanjut dalam tahap eksplorasi dan pemodelan. Tanpa pengelolaan yang tepat, kelimpahan data justru dapat menurunkan kualitas *insight* yang dihasilkan.

Ringkasan metode pengumpulan data juga mencakup pertimbangan etika, legalitas, dan keamanan informasi. Perlindungan terhadap privasi individu, kepatuhan terhadap regulasi perlindungan data, serta transparansi dalam penggunaan data menjadi prinsip yang harus dijunjung tinggi dalam praktik sains data. Dengan pendekatan yang sistematis, terukur, dan bertanggung jawab, metode pengumpulan data tidak hanya menghasilkan dataset yang andal, tetapi juga memastikan bahwa proses analitik berjalan secara berkelanjutan dan memberikan nilai strategis bagi organisasi maupun masyarakat.

## **B. Tren Masa Depan Pengumpulan Data**

Tren masa depan pengumpulan data dalam sains data berkembang seiring kemajuan teknologi digital, peningkatan kapasitas komputasi, serta kebutuhan organisasi terhadap analisis yang semakin cepat dan presisi. Perubahan ini tidak hanya memengaruhi cara data dikumpulkan, tetapi juga bagaimana data diintegrasikan, divalidasi, diamankan, dan dimanfaatkan secara strategis. Berikut penjelasan rinci mengenai tren-tren utama tersebut:

## 1. Otomatisasi dan *Real-time Data Collection*

Otomatisasi dan *real-time data collection* merupakan pendekatan modern dalam pengumpulan data yang memanfaatkan teknologi digital untuk merekam, mentransmisikan, dan memproses data secara langsung tanpa intervensi manual yang signifikan, sehingga memungkinkan organisasi memperoleh informasi secara cepat, akurat, dan berkelanjutan dalam berbagai konteks operasional seperti industri, keuangan, kesehatan, dan pemerintahan. Perkembangan *Internet of Things* (IoT), sensor cerdas, sistem enterprise berbasis *cloud*, serta integrasi *Application Programming Interface* (API) telah mendorong terciptanya ekosistem data yang mampu mengalir secara simultan dari berbagai sumber ke pusat analitik, sehingga proses monitoring, evaluasi kinerja, hingga deteksi anomali dapat dilakukan dalam hitungan detik. Menurut Kitchin (2018), “*real-time analytics refers to the processing and analysis of data as it is generated, enabling immediate insights and responsive decision-making*,” yang menegaskan bahwa kecepatan pemrosesan data menjadi elemen krusial dalam mendukung pengambilan keputusan yang adaptif di era digital.

Pada praktiknya, otomatisasi pengumpulan data mengurangi potensi kesalahan manusia (*human error*), meningkatkan konsistensi pencatatan, serta memperluas kapasitas pengumpulan data dalam skala besar yang sebelumnya sulit dilakukan melalui metode konvensional seperti survei manual atau *input* data administratif. Sistem berbasis *machine learning* kini bahkan mampu melakukan penyaringan otomatis, validasi awal, serta klasifikasi data sebelum disimpan ke dalam data warehouse atau data lake, sehingga kualitas data dapat dijaga sejak tahap awal siklus analitik. Integrasi *edge computing* juga memungkinkan sebagian proses komputasi dilakukan di dekat sumber data, yang tidak hanya mempercepat respons sistem tetapi juga mengurangi beban jaringan dan meningkatkan efisiensi operasional.

## 2. Integrasi *Big Data* dan *Multi-Source Data*

Integrasi *Big data* dan *multi-source data* merupakan pendekatan strategis dalam sains data yang menggabungkan berbagai sumber data heterogen, baik terstruktur maupun tidak terstruktur, ke dalam satu sistem terpadu guna menghasilkan analisis yang lebih komprehensif dan

bernilai tinggi. Dalam praktiknya, data dapat berasal dari sistem internal perusahaan seperti ERP dan CRM, media sosial, perangkat IoT, transaksi digital, data geospasial, hingga basis data publik, yang kemudian dikonsolidasikan melalui teknologi seperti data warehouse modern, data lake, serta integrasi berbasis API. Menurut Mayer-Schönberger dan Cukier (2018), “*Big data refers to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value,*” yang menegaskan bahwa nilai strategis *big data* terletak pada kemampuannya mengolah volume dan variasi data dalam skala besar untuk menghasilkan wawasan baru.

Secara teknis, integrasi multi-source data menuntut adanya proses ekstraksi, transformasi, dan pemuatan (ETL/ELT) yang efektif agar data dari berbagai format dan struktur dapat distandarisasi sebelum dianalisis lebih lanjut. Tantangan utama dalam integrasi ini meliputi perbedaan skema data, inkonsistensi kualitas, duplikasi informasi, serta kebutuhan sinkronisasi waktu (*time alignment*) agar analisis lintas sumber tetap relevan dan akurat. Oleh karena itu, organisasi memerlukan arsitektur data yang fleksibel dan skalabel, termasuk penggunaan metadata management serta data catalog untuk memastikan transparansi dan keterlacakan sumber data.

### **3. Pemanfaatan *Artificial Intelligence* (AI) dalam Pengumpulan Data**

Pemanfaatan *Artificial Intelligence* (AI) dalam pengumpulan data merepresentasikan transformasi signifikan dalam cara organisasi memperoleh, menyaring, dan memvalidasi informasi secara otomatis dan adaptif, terutama ketika berhadapan dengan volume data yang besar serta variasi format yang kompleks seperti teks, gambar, audio, dan log aktivitas digital. Teknologi seperti *machine learning*, *natural language processing* (NLP), dan computer vision memungkinkan sistem untuk secara cerdas mengidentifikasi pola, mendeteksi anomali, serta mengekstraksi informasi relevan sejak tahap awal akuisisi data tanpa ketergantungan penuh pada intervensi manusia. Menurut Russell dan Norvig (2021), “*Artificial intelligence is the study of agents that receive percepts from the environment and perform actions,*” yang menunjukkan bahwa sistem AI dirancang untuk merespons *input* lingkungan secara otonom, termasuk dalam konteks pengumpulan dan pemrosesan data.

Pada praktik operasional, AI digunakan untuk mengotomatisasi survei digital melalui chatbot cerdas, melakukan *scraping* data berbasis algoritma adaptif, serta mengklasifikasikan data secara *real-time* sebelum disimpan dalam sistem penyimpanan terpusat seperti data lake atau *cloud infrastructure*. Sistem berbasis AI juga mampu melakukan data validation dan *data cleaning* secara otomatis dengan mendeteksi inkonsistensi, duplikasi, maupun nilai ekstrem yang berpotensi mengganggu kualitas analisis lanjutan. Dengan demikian, proses pengumpulan data tidak lagi sekadar tahap pasif pengambilan informasi, tetapi menjadi proses dinamis yang terus belajar dan meningkatkan akurasi melalui pembaruan model secara berkelanjutan.

#### **4. Data Privacy, Security, dan Ethical Data Governance**

*Data privacy, security, dan ethical data governance* merupakan elemen fundamental dalam pengelolaan data modern yang memastikan bahwa proses pengumpulan, penyimpanan, pemrosesan, dan distribusi data dilakukan secara bertanggung jawab serta sesuai dengan prinsip hukum dan etika yang berlaku. *Data privacy* berfokus pada perlindungan hak individu atas informasi pribadinya, termasuk bagaimana data tersebut dikumpulkan, digunakan, dan dibagikan, sementara data security menitikberatkan pada perlindungan teknis melalui enkripsi, kontrol akses, autentikasi berlapis, dan sistem deteksi ancaman siber. Menurut Floridi dan Cows (2019), “*Data governance should ensure that data practices respect fundamental rights and promote the flourishing of individuals and society,*” yang menegaskan bahwa tata kelola data tidak hanya bersifat teknis, tetapi juga harus berlandaskan prinsip moral dan perlindungan hak asasi.

Pada praktik organisasi, penerapan *privacy by design* menjadi pendekatan strategis yang mengintegrasikan perlindungan privasi sejak tahap perancangan sistem dan bukan sebagai tambahan setelah sistem berjalan, sehingga risiko kebocoran dan penyalahgunaan data dapat diminimalkan secara sistematis. Penguatan keamanan data dilakukan melalui kebijakan manajemen risiko, audit keamanan berkala, penerapan standar internasional seperti ISO/IEC 27001, serta penggunaan teknologi seperti blockchain dan *zero-trust architecture* untuk meningkatkan integritas serta ketahanan sistem terhadap serangan siber. Selain itu, kepatuhan terhadap regulasi perlindungan data menuntut adanya

transparansi, akuntabilitas, serta dokumentasi yang jelas mengenai alur pemrosesan data dalam organisasi.

## 5. *Edge Computing* dan Desentralisasi Data

*Edge computing* dan desentralisasi data merupakan pendekatan arsitektur komputasi yang memindahkan proses pemrosesan dan analisis data dari pusat data terpusat (*centralized cloud*) ke lokasi yang lebih dekat dengan sumber data, seperti perangkat IoT, sensor industri, atau node jaringan lokal, sehingga mampu mengurangi latensi dan meningkatkan efisiensi sistem secara keseluruhan. Pendekatan ini menjadi semakin relevan dalam lingkungan yang menghasilkan data dalam jumlah besar dan berkecepatan tinggi, seperti *smart city*, kendaraan otonom, manufaktur cerdas, serta layanan kesehatan berbasis perangkat wearable, di mana respons waktu nyata sangat krusial. Menurut Shi *et al.* (2018), “*Edge computing refers to enabling technologies allowing computation to be performed at the edge of the network, on downstream data on behalf of cloud services and upstream data on behalf of IoT services,*” yang menegaskan bahwa *edge computing* berfungsi sebagai perpanjangan kemampuan komputasi *cloud* dengan mendekati proses analitik ke sumber data.

Desentralisasi data dalam konteks ini berarti distribusi penyimpanan dan pemrosesan data ke berbagai node atau lokasi yang saling terhubung, sehingga tidak seluruh data harus dikirim ke pusat untuk diproses, yang pada akhirnya mengurangi beban bandwidth dan risiko *bottleneck* jaringan. Model ini memungkinkan sebagian data diproses secara lokal untuk kebutuhan respons cepat, sementara data yang telah difilter atau diringkas dapat dikirim ke *cloud* untuk analisis lanjutan, pelaporan strategis, atau pelatihan model *machine learning* dalam skala besar. Dengan demikian, arsitektur terdistribusi ini menciptakan keseimbangan antara efisiensi operasional di tingkat lokal dan kapasitas komputasi skala besar di tingkat pusat.

## 6. *Data-as-a-Service (DaaS)* dan Ekonomi Data

*Data-as-a-Service (DaaS)* merupakan model layanan berbasis *cloud* yang memungkinkan organisasi mengakses, mengelola, dan memanfaatkan data melalui mekanisme berlangganan tanpa harus membangun infrastruktur penyimpanan dan pengolahan data secara

mandiri, sehingga menciptakan fleksibilitas dan efisiensi dalam pengelolaan aset informasi. Dalam skema ini, penyedia layanan bertanggung jawab atas integrasi, pembersihan, standarisasi, serta distribusi data kepada pengguna melalui platform digital yang aman dan terukur, sehingga data dapat digunakan secara langsung untuk kebutuhan analitik, pelaporan, maupun pengembangan model prediktif. Menurut Vial (2019), “*Digital technologies enable new business models that transform how value is created, delivered, and captured,*” yang menunjukkan bahwa model seperti DaaS menjadi bagian dari transformasi digital yang mengubah cara organisasi menciptakan dan memonetisasi nilai dari data.

Konsep ekonomi data (*data economy*) merujuk pada sistem ekonomi di mana data diperlakukan sebagai aset strategis yang memiliki nilai komersial dan dapat diperdagangkan, dianalisis, serta dikombinasikan untuk menghasilkan produk dan layanan baru. Dalam konteks ini, data tidak lagi sekadar hasil sampingan aktivitas operasional, melainkan sumber daya utama yang mendorong inovasi, efisiensi, serta diferensiasi kompetitif melalui pengembangan *insight* berbasis analitik dan kecerdasan buatan. Perusahaan dapat memperoleh pendapatan tambahan dengan menjual dataset terkurasi, menyediakan akses API, atau membangun ekosistem kolaboratif lintas industri yang memungkinkan pertukaran data secara aman dan terstandar.

### C. Implikasi bagi Mahasiswa dan Praktisi Sains Data

Implikasi perkembangan metode pengumpulan data, otomatisasi, *big data*, AI, dan tata kelola data memberikan dampak signifikan bagi mahasiswa maupun praktisi sains data. Perubahan ini tidak hanya menuntut penguasaan teknis, tetapi juga kemampuan adaptif, etis, dan strategis dalam menghadapi ekosistem data yang semakin kompleks. Berikut penjelasan rinci mengenai implikasi tersebut:

#### 1. Penguatan Kompetensi Teknis dan Analitis

Penguatan kompetensi teknis dan analitis menjadi kebutuhan fundamental dalam pengembangan kapasitas mahasiswa dan praktisi sains data karena kompleksitas data yang terus meningkat menuntut penguasaan keterampilan yang tidak hanya bersifat operasional tetapi juga konseptual dan strategis. Kompetensi teknis mencakup kemampuan

pemrograman, pengelolaan basis data, pemrosesan *big data*, serta penggunaan perangkat analitik dan *machine learning*, sedangkan kompetensi analitis menuntut kemampuan berpikir kritis, pemodelan statistik, interpretasi hasil, dan penerjemahan temuan menjadi rekomendasi yang aplikatif. Menurut Russell dan Norvig (2021), “*Artificial intelligence is the study of agents that receive percepts from the environment and perform actions,*” yang mengindikasikan bahwa pemahaman mendalam terhadap sistem cerdas dan proses analitik menjadi bagian integral dari kompetensi teknis modern dalam ekosistem data.

Pada praktiknya, penguatan kompetensi teknis tidak hanya berarti menguasai alat atau bahasa pemrograman tertentu, tetapi juga memahami arsitektur sistem data seperti *cloud computing*, data warehouse, data lake, serta integrasi API yang mendukung pengumpulan dan pengolahan data secara *real-time*. Sementara itu, kompetensi analitis menuntut kemampuan mengevaluasi kualitas data, mendeteksi bias, melakukan validasi model, serta menginterpretasikan *output* algoritma secara objektif dan kontekstual agar hasil analisis tidak disalahartikan. Kombinasi antara penguasaan teknis dan ketajaman analitis memungkinkan individu untuk menghasilkan *insight* yang akurat sekaligus relevan dengan kebutuhan organisasi atau permasalahan yang dihadapi.

## **2. Kesiapan Menghadapi Otomatisasi dan AI**

Kesiapan menghadapi otomatisasi dan *Artificial Intelligence* (AI) menjadi aspek krusial dalam era transformasi digital karena perkembangan teknologi ini tidak hanya mengubah proses kerja, tetapi juga mendefinisikan ulang peran manusia dalam sistem produksi dan pengambilan keputusan berbasis data. Otomatisasi memungkinkan tugas-tugas rutin dan repetitif dilakukan oleh sistem cerdas dengan tingkat akurasi dan kecepatan yang tinggi, sehingga individu dituntut untuk meningkatkan kapasitas pada aspek kognitif tingkat tinggi seperti analisis strategis, evaluasi etis, serta pemecahan masalah kompleks. Menurut Brynjolfsson dan McAfee (2018), “*The key to winning the race is not to compete against machines but to compete with machines,*” yang menegaskan bahwa kesiapan menghadapi AI terletak pada kemampuan manusia untuk berkolaborasi dengan teknologi, bukan sekadar menggantikannya.

Pada konteks profesional dan akademik, kesiapan tersebut mencakup pemahaman mendalam mengenai cara kerja algoritma, logika *machine learning*, serta keterbatasan sistem otomatis agar pengguna tidak menerima hasil analitik secara pasif tanpa proses verifikasi kritis. Individu juga perlu mengembangkan literasi data dan literasi AI yang mencakup kemampuan membaca *output* model, mengenali potensi bias, serta memahami implikasi sosial dari keputusan yang dihasilkan oleh sistem otomatis. Dengan demikian, kesiapan menghadapi otomatisasi bukan hanya soal keterampilan teknis, tetapi juga kesiapan mental dan intelektual untuk beradaptasi dengan dinamika kerja yang semakin terdigitalisasi.

### **3. Kesadaran Etika, Privasi, dan Tata Kelola Data**

Kesadaran etika, privasi, dan tata kelola data merupakan fondasi penting dalam praktik sains data modern karena pemanfaatan data dalam skala besar membawa konsekuensi sosial, hukum, dan moral yang signifikan terhadap individu maupun masyarakat. Etika data tidak hanya berkaitan dengan kepatuhan terhadap regulasi, tetapi juga mencakup tanggung jawab profesional dalam memastikan bahwa proses pengumpulan, analisis, dan distribusi data tidak merugikan hak asasi, tidak menimbulkan diskriminasi, serta tidak disalahgunakan untuk kepentingan yang melanggar norma. Menurut Floridi dan Cowls (2019), “*Data governance should ensure that data practices respect fundamental rights and promote the flourishing of individuals and society,*” yang menegaskan bahwa tata kelola data harus berorientasi pada perlindungan hak dasar serta kesejahteraan sosial.

Pada praktik organisasi, kesadaran privasi diwujudkan melalui penerapan prinsip transparansi, persetujuan yang diinformasikan (*informed consent*), pembatasan tujuan penggunaan data, serta minimisasi pengumpulan data yang tidak relevan dengan kebutuhan analitik. Tata kelola data yang efektif juga memerlukan struktur kebijakan yang jelas, termasuk pembagian tanggung jawab, mekanisme audit internal, serta sistem pengendalian akses yang ketat untuk mencegah kebocoran atau penyalahgunaan informasi. Selain itu, pengembangan sistem berbasis *privacy by design* memastikan bahwa perlindungan privasi telah terintegrasi sejak tahap perancangan teknologi dan bukan sekadar tambahan setelah sistem beroperasi.

#### 4. Kemampuan Interdisipliner dan Kolaboratif

Kemampuan interdisipliner dan kolaboratif menjadi kompetensi kunci dalam sains data karena permasalahan yang dianalisis umumnya bersifat kompleks dan melibatkan berbagai dimensi seperti teknologi, bisnis, sosial, hukum, dan kebijakan publik. Seorang praktisi sains data tidak hanya dituntut memahami teknik analitik dan pemrograman, tetapi juga mampu mengintegrasikan perspektif dari bidang lain agar hasil analisis relevan dengan konteks dan kebutuhan pemangku kepentingan. Menurut Waller dan Fawcett (2018), “*Data science requires the integration of domain knowledge, statistics, and computer science to extract meaningful insights,*” yang menegaskan bahwa kekuatan sains data terletak pada perpaduan lintas disiplin untuk menghasilkan pemahaman yang komprehensif.

Pada praktik profesional, kemampuan interdisipliner tercermin dari kapasitas untuk menerjemahkan masalah bisnis atau sosial ke dalam kerangka analitik yang terstruktur, sekaligus mengomunikasikan hasil model secara jelas kepada pihak non-teknis seperti manajer, pembuat kebijakan, atau tim operasional. Kolaborasi lintas fungsi menjadi penting karena proyek sains data sering melibatkan tim yang terdiri atas ahli IT, analis bisnis, pakar hukum, hingga manajemen risiko yang harus bekerja secara sinergis dalam satu siklus pengembangan solusi berbasis data. Tanpa kemampuan komunikasi dan koordinasi yang efektif, potensi teknis yang dimiliki seorang *data scientist* dapat terhambat oleh kesalahpahaman tujuan, perbedaan terminologi, maupun ketidaksinkronan strategi organisasi.

#### 5. Adaptasi terhadap Teknologi Baru

Adaptasi terhadap teknologi baru merupakan kompetensi strategis yang harus dimiliki oleh mahasiswa dan praktisi sains data karena perkembangan teknologi digital seperti *cloud computing*, *artificial intelligence*, *big data analytics*, dan *edge computing* berlangsung secara cepat dan berkelanjutan. Perubahan ini tidak hanya memengaruhi alat dan platform yang digunakan, tetapi juga mengubah proses kerja, model bisnis, serta ekspektasi organisasi terhadap kecepatan dan akurasi pengambilan keputusan berbasis data. Menurut Vial (2019), “*Digital transformation can be understood as a process that aims to improve an entity by triggering significant changes to its properties through combinations of information, computing,*

*communication, and connectivity technologies,*” yang menegaskan bahwa adaptasi teknologi merupakan bagian dari transformasi menyeluruh dalam organisasi.

Pada konteks profesional, kemampuan beradaptasi mencakup kesiapan untuk mempelajari sistem baru, memahami arsitektur teknologi terkini, serta mengevaluasi kelebihan dan keterbatasan inovasi sebelum diimplementasikan dalam lingkungan kerja. Mahasiswa dan praktisi perlu mengembangkan pola *pikir growth mindset* yang mendorong pembelajaran berkelanjutan, sertifikasi kompetensi, serta eksplorasi teknologi baru agar tidak tertinggal oleh dinamika industri yang sangat kompetitif. Adaptasi ini juga memerlukan kemampuan analitis untuk menilai dampak teknologi terhadap efisiensi operasional, keamanan data, serta keberlanjutan strategi organisasi.

## **6. Peningkatan Daya Saing dan Peluang Karier**

Peningkatan daya saing dan peluang karier dalam bidang sains data merupakan konsekuensi langsung dari transformasi digital yang menjadikan data sebagai aset strategis dalam berbagai sektor industri, mulai dari keuangan, kesehatan, manufaktur, hingga pemerintahan. Permintaan terhadap tenaga profesional yang mampu mengolah, menganalisis, dan menginterpretasikan data secara akurat terus meningkat karena organisasi membutuhkan dasar pengambilan keputusan yang berbasis bukti dan *real-time analytics*. Menurut Brynjolfsson dan McAfee (2018), “*The key to winning the race is not to compete against machines but to compete with machines,*” yang menunjukkan bahwa individu yang mampu memanfaatkan teknologi cerdas secara efektif akan memiliki keunggulan kompetitif dalam pasar kerja modern.

Pada konteks pengembangan karier, kompetensi di bidang sains data membuka peluang kerja yang luas seperti *data analyst, data scientist, machine learning engineer, business intelligence specialist*, hingga *data engineer* yang berperan dalam membangun arsitektur sistem data skala besar. Selain peluang kerja formal, perkembangan ekonomi digital juga memungkinkan munculnya peluang kewirausahaan berbasis teknologi, seperti pengembangan startup analitik, penyedia layanan *Data-as-a-Service*, maupun konsultan transformasi digital. Hal ini menunjukkan bahwa kemampuan analitik dan teknis tidak hanya

meningkatkan peluang kerja, tetapi juga memperluas opsi karier yang fleksibel dan inovatif.

# DAFTAR PUSTAKA

---

- Ahuja, P. (2023). The impact of the Internet on communication and social interaction. LinkedIn.
- Ajayi, V. O. (2023). A Review on Primary Sources of Data and Secondary Sources of Data. *European Journal of Education and Pedagogy*.
- AL Beam, *et al.* (2020). Challenges to the Reproducibility of *Machine Learning* and Scientific Research. PMC – National Center for Biotechnology Information.
- Al-Fedaghi, S. (2019). A framework for structured logging in computer systems. *International Journal of Computer Applications*, 182(5), 1–10.
- Al-Haderi, S., & Ismail, N. (2021). Enhancing user engagement through interactive services on *online* portals. *Journal of Digital Innovation and Technology*.
- Almeida, P., Sousa, R., & Carvalho, J. (2019). Automated Data Collection Techniques for Large-Scale Digital Systems. *International Journal of Data Science and Analytics*, 7(3), 145–162.
- Alonso, G., Casati, F., Kuno, H., & Machiraju, V. (2019). *Web Services: Concepts, Architectures and Applications*. Springer.
- Ávila, E., & Sulé, A. (2025). The Role of the Information and Documentation Professional in *Big data* Management for Scientific Research. *JLIS.it*, 16(3), 43–57.
- Babbie, E. (2018). *The Practice of Social Research* (14th ed.). Boston, MA: Cengage *Learning*.
- Balkin, J. M. (2018). The Three Laws of Robotics in the Age of Digital Platforms. *Yale Journal of Law & Technology*, 20(1), 1–35.
- Barriball, K. L., & While, A. (2019). Collecting data using a semi-structured interview: a discussion paper. *Journal of Advanced Nursing*, 70(6), 1155–1164.
- Batini, C., & Scannapieco, M. (2020). *Data and Information Quality: Dimensions, Principles and Techniques*. Springer.

- Booth, A., Sutton, A., & Papaioannou, D. (2019). *Systematic Approaches to a Successful Literature Review* (3rd ed.). London: SAGE Publications.
- Borgman, C. L. (2018). *Big data, little data, no data: Scholarship in the networked world*. MIT Press.
- Bowen, G. A. (2018). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27–40.
- Bryman, A. (2018). *Social Research Methods* (5th ed.). Oxford University Press.
- Brynjolfsson, E., & McAfee, A. (2018). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W. W. Norton & Company.
- Cândido, J., Aniche, M., & van Deursen, A. (2021). Log-based software monitoring: a systematic mapping study.
- CASRAI. (2025). *Reproducibility Standards*.
- Chen, H., Chiang, R. H. L., & Storey, V. C. (2019). Business Intelligence and *Analytics*: From *Big data* to Big Impact. *MIS Quarterly*, 36(4), 1165–1188.
- Chen, L., & Zhang, H. (2022). The role of *online* portals as centralized information hubs: Accessibility and efficiency in digital environments. *Journal of Information Systems and Technology*.
- Chen, M. (2019). *Big data: A Survey*. *Mobile Networks and Applications*, 23(2), 1–39.
- Chen, M., Mao, S., & Liu, Y. (2019). *Big data: A survey*. *Mobile Networks and Applications*, 23(2), 171–209.
- Chen, Y., & Zhang, X. (2020). *Real-time* Data Acquisition Frameworks for Continuous Data Streams. *Journal of Information Processing Systems*, 16(3), 567–582.
- Chodak, G., Suchacka, G., & Rovetta, S. (2020). HTTP-level e-commerce data based on server access logs for an *online* store. *Computer Networks*.
- Chou, T., Lin, Y., & Chen, P. (2021). Harnessing APIs for Diverse Data Types in Modern *Analytics*. *Information Systems Journal*, 31(4), 567–589.
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research Methods in Education* (8th ed.). Routledge.

- Collis, J., & Hussey, R. (2018). *Business Research: A Practical Guide for Undergraduate and Postgraduate Students* (4th ed.). London: Palgrave.
- Colomina, I., & Molina, P. (2018). Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 92, 79–97.
- Condon, P., Simpson, J., & Emanuel, M. E. (2022). Research data integrity: A cornerstone of rigorous and reproducible research. *IASSIST Quarterly*, 46(3).
- Creswell, J. W. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (5th ed.). Thousand Oaks, CA: SAGE Publications.
- Creswell, J. W., & Guetterman, T. C. (2025). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research* (7th ed.). Pearson.
- Dahanayake, A., & Sol, H. (2019). APIs and Business Partner Integration: Security and Management Perspectives. *Journal of Systems and Software*, 155, 110-123.
- Deuze, M. (2020). *Media Life*. Cambridge: Polity Press.
- DiCicco-Bloom, B., & Crabtree, B. F. (2019). The qualitative research interview. *Medical Education*, 53(4), 351–358.
- Erl, T., Puttini, R., & Mahmood, Z. (2020). *Cloud Computing: Concepts, Technology & Architecture*. Pearson Education.
- Fielding, R. T. (2018). *Architectural Styles and the Design of Network-based Software Architectures*. University of California, Irvine.
- Fielding, R., & Reschke, J. (2022). HTTP Semantics (RFC 9110). Internet *Engineering* Task Force (IETF).
- Fink, A. (2019). *Conducting Research Literature Reviews: From the Internet to Paper* (5th ed.). Thousand Oaks, CA: SAGE Publications.
- Flick, U. (2018). *An Introduction to Qualitative Research* (6th ed.). SAGE Publications.
- Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1).
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707.

- Fowler, F. J. (2019). *Survey Research Methods* (6th ed.). Thousand Oaks, CA: SAGE Publications.
- Gandomi, A., & Haider, M. (2019). Beyond the hype: *Big data* concepts, methods, and *analytics*. *International Journal of Information Management*, 45, 112–123.
- George, T. (2025, January 22). Unstructured Interview | Definition, Guide & Examples. Scribbr.
- Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (3rd ed.). O'Reilly Media.
- Gillis, A. S., & Yasar, K. (2023). Edge device. TechTarget.
- Gillis, A. S., & Yasar, K. (2024). *Cloud platforms in IoT*. TechTarget.
- Gillis, A., & Yasar, K. (2025). What is IoT (Internet of Things)? TechTarget.
- Goodfellow, I., Bengio, Y., & Courville, A. (2018). *Deep Learning*. MIT Press.
- Google Developers. (2020). How browsers work. Google Chrome Developers Documentation.
- GraphQL.org (2025). What is GraphQL?
- Gravetter, F. J., & Forzano, L. A. B. (2018). *Research Methods for the Behavioral Sciences* (6th ed.). Cengage *Learning*.
- Grove, S. K., Burns, N., & Gray, J. (2019). *The practice of nursing research: Appraisal, synthesis, and generation of evidence* (8th ed.). Elsevier.
- Grune, D., & Jacobs, C. J. H. (2019). *Parsing Techniques: A Practical Guide* (2nd ed.). Springer.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (8th ed.). Cengage *Learning*.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (8th ed.). Cengage.
- Hair, J. F., Page, M., & Brunsveld, N. (2021). *Essentials of Business Research Methods* (5th ed.). Routledge.
- Han, J., Kamber, M., & Pei, J. (2022). *Data Mining: Concepts and Techniques* (4th ed.). Morgan Kaufmann.
- Hardt, D. (2018). The OAuth 2.0 Authorization Framework. RFC 6749. IETF.
- Hart, C. (2018). *Doing a Literature Review: Releasing the Research Imagination* (2nd ed.). London: SAGE Publications.

- Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2018). The rise of “*big data*” on *cloud computing*: Review and open research issues. *Information Systems*, 47, 98–115.
- Heale, R., & Twycross, A. (2018). Validity and reliability in quantitative studies. *Evidence-Based Nursing*, 21(3), 66–67.
- Heaton, J. (2020). *Secondary Analysis of Qualitative Data: An Overview*. SAGE Publications.
- Hernandez, M., Lopez, R., & Torres, J. (2020). Workflow Automation for Efficient Data Pipeline Management. *Journal of Information Systems and Technology*, 15(2), 102–118.
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts.
- IBM Think (2025). What Is API Security? IBM.
- ISO/IEC. (2019). ISO/IEC 25012:2019 Systems and software *engineering* — Systems and software Quality Requirements and Evaluation (SQuaRE) — Data quality model. International Organization for Standardization.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2019). *An Introduction to Statistical Learning: with Applications in R* (2nd ed.). Springer.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399.
- Johnston, M. P. (2017). Secondary Data Analysis: A Method of which the Time Has Come. *Qualitative and Quantitative Methods in Libraries*, 3, 619–626.
- Karnia, R. (2024). *Importance of Reliability and Validity in Research*. Science Publishing Group.
- Katal, A., Wazid, M., & Goudar, R. H. (2018). *Big data: Issues, Challenges, Tools and Good Practices*. In 2018 IEEE International Conference on *Big data (Big data)*, 16–24. IEEE.
- Kelleher, J. D., & Tierney, B. (2018). *Data Science*. Cambridge, MA: MIT Press.
- Khanyi, M. B., Xaba, S. N., Mlotshwa, N. A., & Thango, B. (2024). The Role of Data Networks and APIs in Enhancing Operational Efficiency in SMEs: A Systematic Review. ResearchGate.

- Kim, H., & Lee, S. (2021). Scalable and Secure Data Storage Solutions for Automated Data Pipelines. *Journal of Data Management Systems*, 12(4), 221–238.
- Kimball, R., & Ross, M. (2018). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (3rd ed.). Hoboken, NJ: John Wiley & Sons.
- Kimberlin, C. L., & Winterstein, A. G. (2018). Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy*, 65(23), 2276–2284.
- Kitchin, R. (2018). *The data revolution: Big data, open data, data infrastructures and their consequences* (2nd ed.). SAGE Publications.
- Knaflic, C. N. (2019). *Storytelling with Data: Let's Practice!*. Hoboken, NJ: Wiley.
- Knaflic, S. N. (2020). *Storytelling with Data: Let's Practice!* Wiley.
- Kotler, P., & Keller, K. L. (2022). *Marketing Management* (16th ed.). Pearson Education Limited.
- Kotu, V., & Deshpande, B. (2019). *Data Science: Concepts and Practice* (2nd ed.). Burlington: Morgan Kaufmann.
- Koukaras, P. & Tjortjis, C. (2025). Data Preprocessing and Feature *Engineering* for Data Mining: Techniques, Tools, and Best Practices. *MDPI AI*, 6, 257.
- Krosnick, J. A., & Presser, S. (2018). Question and Questionnaire Design. In D. L. Vannette & J. A. Krosnick (Eds.), *The Palgrave Handbook of Survey Research*. Palgrave Macmillan.
- Krotov, V., & Silva, L. (2018). Legality and ethics of *web scraping*. *Proceedings of the Americas Conference on Information Systems (AMCIS)*.
- Krueger, R. A., & Casey, M. A. (2019). *Focus Groups: A Practical Guide for Applied Research* (5th ed.). SAGE Publications.
- Kumar, R., & Singh, A. (2020). *Data Analytics for Strategic Decision Making: A Practical Approach*. Springer.
- Kumar, V. (2024). Primary Data Research. In *International Marketing Research* (pp. 183–210). Springer Nature.
- Kuner, C., Bygrave, L. A., & Docksey, C. (2020). *The EU General Data Protection Regulation (GDPR): A Commentary*. Oxford University Press.

- Kurose, J. F., & Ross, K. W. (2021). *Computer Networking: A Top-Down Approach* (8th ed.). Pearson.
- Kvale, S., & Brinkmann, S. (2018). *InterViews: Learning the Craft of Qualitative Research Interviewing* (3rd ed.). SAGE Publications.
- Lankhorst, M. (2021). *Enterprise Architecture at Work: Modelling, Communication and Analysis* (4th ed.). Springer.
- Laudon, K. C., & Laudon, J. P. (2020). *Management information systems: Managing the digital firm* (16th ed.). Pearson Education.
- Li, X., Chen, W., Zhang, Y., & Yang, Y. (2020). Deep *learning* for computer vision: A brief review. *Neurocomputing*, 396, 414–429.
- Liu, B. (2020). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions* (2nd ed.). Cambridge University Press.
- Loola Bokonda, P., Ouazzani-Touhami, K., & Souissi, N. (2020). A Practical Analysis of *Mobile* Data Collection Apps. *International Journal of Interactive Mobile Technologies (iJIM)*.
- Malhotra, N. K., Nunan, D., & Birks, D. F. (2019). *Marketing Research: An Applied Approach* (5th ed.). Harlow: Pearson Education.
- Manning, C. D., Raghavan, P., & Schütze, H. (2021). *Introduction to Information Retrieval* (Updated ed.). Cambridge University Press.
- Marr, B. (2018). *Big data in Practice: How 45 Successful Companies Used Big data Analytics to Deliver Extraordinary Results*. John Wiley & Sons.
- Massé, M. (2019). *REST API Design Rulebook*. O'Reilly Media.
- Mayer-Schönberger, V., & Cukier, K. (2018). *Big data: A Revolution That Will Transform How We Live, Work, and Think*. Boston: Houghton Mifflin Harcourt.
- McKinney, W. (2018). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython* (2nd ed.). Sebastopol, CA: O'Reilly Media.
- McQuail, D. (2019). *McQuail's Mass Communication Theory* (7th ed.). London: SAGE Publications.
- Mell, P., & Grance, T. (2018). *The NIST Definition of Cloud Computing*. National Institute of Standards and Technology (NIST Special Publication 800-145).

- Mitchell, R. (2018). *Web Scraping with Python: Collecting More Data from the Modern Web* (2nd ed.). O'Reilly Media.
- Montgomery, D. C. (2019). *Design and Analysis of Experiments* (10th ed.). John Wiley & Sons.
- Moravec, V. (2025). Algorithmic personalization: adaptation of information according to the user's individual needs, preferences, or behavior. *Nature Human Behaviour*.
- Morgan, D. L. (2018). *Focus Groups as Qualitative Research* (2nd ed.). SAGE Publications.
- Mutka, A., Hämäläinen, R., & Virtanen, T. (2020). *Mobile data collection systems in field research: Enhancing accuracy and efficiency*. ScienceDirect.
- Nasution, A. H. (2024). Ilmu Data. Universitas Islam Riau Repository.
- National Academies of Sciences, *Engineering*, and Medicine. (2019). *Reproducibility and replicability in science*. Washington, DC: The National Academies Press.
- Neuman, W. L. (2019). *Social Research Methods: Qualitative and Quantitative Approaches* (8th ed.). Pearson Education.
- Nex, F., & Remondino, F. (2018). UAV for 3D mapping applications: A review. *Applied Geomatics*, 10(1), 1–15.
- O'Neil, C. (2018). *Weapons of Math Destruction: How Big data Increases Inequality and Threatens Democracy*. Crown.
- OECD. (2019). *The Value of Official Statistics in Evidence-Based Policy Making*. Paris: OECD Publishing.
- Olston, C., & Najork, M. (2018). *Web crawling*. *Foundations and Trends in Information Retrieval*, 4(3), 175–246.
- Opdenakker, R. (2019). Advantages and disadvantages of four interview techniques in qualitative research. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 20(3), Art. 11.
- Patel, S., & Mehta, R. (2022). Modern ETL and ELT Processes for Automated Data Systems. *Journal of Big data Analytics*, 9(1), 58–74.
- Patterson, B. (2024). Event tracking gives businesses a competitive advantage by enabling a deeper understanding of customer movements across marketing channels (RudderStack).

- Pautasso, C., Zimmermann, O., & Leymann, F. (2018). RESTful Web Services vs. Big Web Services: Making the Right Architectural Decision. *ACM Computing Surveys*, 50(6), 1-34.
- Petković, A. (2025). The Internet today forms the foundation for access to education, healthcare, employment, political participation, and information. *Science International Journal*.
- Pink, S. (2018). *Doing Visual Ethnography* (4th ed.). London: SAGE Publications.
- Polit, D. F., & Beck, C. T. (2018). *Nursing research: Generating and assessing evidence for nursing practice* (10th ed.). Wolters Kluwer.
- Pressman, R. S., & Maxim, B. R. (2020). *Software Engineering: A Practitioner's Approach* (9th ed.). McGraw-Hill Education.
- Provost, F., & Fawcett, T. (2018). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking* (Revised Edition). Sebastopol, CA: O'Reilly Media.
- Punch, K. F. (2018). *Introduction to Social Research: Quantitative and Qualitative Approaches* (3rd ed.). London: SAGE Publications.
- Pushkarna, M., Zaldivar, A., & Kjartansson, O. (2022). Data Cards: Purposeful and Transparent Dataset Documentation for Responsible AI. arXiv.
- Qiu, F. (2025). A Review on Integrating IoT, IIoT, and Industry 4.0: A Pathway .... IET Research.
- Qu, S. Q., & Dumay, J. (2018). The qualitative research interview.
- Qualtrics. (2023). Primary Research vs Secondary Research – What's the Difference? Qualtrics.
- Rahm, E., & Do, H. H. (2018). Data cleaning: Problems and current approaches. *IEEE Data Engineering Bulletin*, 23(4), 3–13.
- Rajaraman, A., & Ullman, J. D. (2020). *Mining of Massive Datasets* (3rd ed.). Cambridge University Press.
- Rapley, T. (2019). *Doing Conversation, Discourse and Document Analysis*. London: SAGE Publications.
- Reis, J., & Housley, M. (2022). *Fundamentals of Data Engineering: Plan and Build Robust Data Systems*. Sebastopol, CA: O'Reilly Media.
- Richardson, L., & Amundsen, M. (2019). *RESTful Web APIs*. O'Reilly Media.

- Ridley, D. (2018). *The Literature Review: A Step-by-Step Guide for Students* (2nd ed.). London: SAGE Publications.
- Russell, M. A. (2018). *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Instagram, GitHub, and More* (3rd ed.). O'Reilly Media.
- Salmon, N.H.A. (2025). The Design of a Data Logger with WEB-based Storage. *Jurnal IKATEMI ELC*.
- Saunders, M., Lewis, P., & Thornhill, A. (2019). *Research Methods for Business Students* (8th ed.). Pearson.
- Schutt, R. K. (2018). *Investigating the Social World: The Process and Practice of Research* (8th ed.). SAGE Publications.
- Seifert, A. (2018). *Mobile* data collection with smartphones offers unique and innovative opportunities. National Center for Biotechnology Information (PMC).
- Sekaran, U., & Bougie, R. (2020). *Research Methods for Business: A Skill-Building Approach* (8th ed.). Wiley.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2018). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (2nd ed.). Wadsworth Cengage *Learning*.
- Sharma, A. (2019). Data-Driven Decision Making: A Primer for IT Professionals. *Journal of Decision Systems*, 28(3), 145–160.
- Sheldon, R. (2022). Definition of sensor. TechTarget.
- Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2018). Edge *computing*: Vision and challenges. *IEEE Internet of Things Journal*, 3(5), 637–646.
- Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., ... Summers, R. M. (2019). Deep *learning* in medical image analysis: A review. *Medical Image Analysis*, 42, 60–88.
- Shmueli, G., Bruce, P. C., Gedeck, P., & Patel, N. R. (2020). *Data Mining for Business Analytics: Concepts, Techniques, and Applications in Python* (3rd ed.). Wiley.
- Silberschatz, A., Korth, H. F., & Sudarshan, S. (2019). *Database system concepts* (7th ed.). McGraw-Hill Education.
- Singh, A. K., & Gupta, V. K. (2025). Transforming e-commerce: the critical role of digital transactions. *International Journal of Innovation Studies*.
- Solove, D. J. (2021). *Understanding Privacy*. Harvard University Press.
- Sowmiya, R. (2025). Types of Data in Research. LinkedIn.

- Stair, R., & Reynolds, G. (2021). *Principles of Information Systems* (14th ed.). Cengage *Learning*.
- Sugiyono. (2019). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Suksmono, & Sari. (2024). Cost and Time Analysis Using the Earned Value Analysis Method. *Eduvest - Journal of Universal Studies*, 4(9).
- Suresh, K. P., & Chandrashekara, S. (2018). Sample size estimation and power analysis for clinical research studies. *Journal of Human Reproductive Sciences*, 11(1), 41–46.
- Suresh, K. P., & Chandrashekara, S. (2018). Sample size estimation and power analysis for clinical research studies. *Journal of Human Reproductive Sciences*, 11(1), 41–46.
- Szeliski, R. (2019). *Computer Vision: Algorithms and Applications* (2nd ed.). Springer.
- Taherdoost, H. (2021). *Data Collection Methods and Tools for Research: A Step-by-Step Guide to Choose Data Collection Technique for Academic and Business Research Projects*. *International Journal of Academic Research in Management*.
- TechTarget (2021). What is open API (public API)?
- Toma, M., Popescu, D., & Stoica, R. (2020). Leveraging APIs for Structured and Automated Data Access. *Journal of Data and Information Quality*, 12(3), 1–18.
- Trochim, W. M. K., Donnelly, J. P., & Arora, K. (2019). *Research Methods: The Essential Knowledge Base* (2nd ed.). Cengage *Learning*.
- U Sulung. (2024). Primary Data Offers Information Most Relevant and Accurate. *Journal International Journal of Education Research (IICLS)*.
- United Nations. (2019). *Fundamentals of Official Statistics*. New York: United Nations.
- Vassio, L., Drago, I., Mellia, M., Ben Houidi, Z., & Lamali, M. (2018). You, the Web and Your Device: Longitudinal Characterization of Browsing Habits. *arXiv*.
- Véliz, C. (2020). *Privacy Is Power: Why and How You Should Take Back Control of Your Data*. Bantam Press.

- Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, 28(2), 118–144.
- Waller, M. A., & Fawcett, S. E. (2018). Data science, predictive *analytics*, and *big data*: A revolution that will transform supply chain design and management. *Journal of Business Logistics*, 39(1), 5–17.
- Whitman, M. E., & Mattord, H. J. (2022). *Principles of Information Security* (7th ed.). Cengage *Learning*.
- Wikipedia. (2026). Programming language. Diakses dari versi terbaru artikel Wikipedia tentang programming language.
- Wooldridge, J. M. (2020). *Introductory Econometrics: A Modern Approach* (7th ed.). MIT Press.
- World Intellectual Property Organization (WIPO). (2021). *Understanding Intellectual Property*. WIPO Publication.
- World Wide Web Consortium (W3C). (2021). Document Object Model (DOM) Standard. W3C Recommendation.
- Yasar, K. (2024). What is an API Endpoint? TechTarget.
- Zhang, L., Chen, Y., & Kumar, S. (2022). API-Driven Data Integration and Advanced *Analytics*. ResearchGate.
- Zhou, D. Z., Fokaefs, M. (2024). AI Assistants for Incident Lifecycle in a Microservice Environment: A Systematic Literature Review. arXiv.
- Zhou, Z., Zhang, X., Xie, L., & Qian, Y. (2019). *Big data Security and Privacy Protection: Advances and Challenges*. *IEEE Access*, 7, 123-136.
- Zikmund, W. G., Babin, B. J., Carr, J. C., & Griffin, M. (2019). *Business Research Methods* (13th ed.). Cengage *Learning*.

# GLOSARIUM

---

<b>API</b>	Antarmuka pemrograman aplikasi yang memungkinkan sistem atau aplikasi berbeda untuk saling berkomunikasi dan bertukar data secara terstandarisasi.
<b>Bigdata</b>	Kumpulan data dalam volume sangat besar, beragam format, dan kecepatan tinggi yang memerlukan teknologi komputasi khusus untuk penyimpanan, pemrosesan, dan analisisnya.
<b>Data</b>	Sekumpulan fakta, angka, simbol, teks, gambar, atau rekaman digital yang dikumpulkan dari berbagai sumber dan masih bersifat mentah sehingga memerlukan proses pengolahan agar dapat menghasilkan informasi yang bermakna dan mendukung pengambilan keputusan.
<b>Eksperimen</b>	Metode penelitian yang dilakukan dengan memanipulasi satu atau lebih variabel independen dalam kondisi terkontrol untuk menguji hubungan sebab-akibat terhadap variabel dependen.
<b>Instrumen</b>	Alat ukur atau perangkat yang dirancang secara sistematis untuk mengumpulkan data sesuai tujuan penelitian agar menghasilkan informasi yang valid dan reliabel.
<b>Kuesioner</b>	Seperangkat pertanyaan tertulis yang disusun secara sistematis untuk mengumpulkan data dari responden mengenai topik tertentu secara efisien dan terstruktur.
<b>Metadata</b>	Informasi tambahan yang menjelaskan struktur, konteks, sumber, waktu, atau karakteristik lain dari suatu dataset sehingga memudahkan pengelolaan dan interpretasi data.

<b>Observasi</b>	Teknik pengumpulan data melalui proses pengamatan langsung dan sistematis terhadap perilaku, peristiwa, atau fenomena tertentu guna memperoleh informasi faktual sesuai konteks yang terjadi.
<b>Populasi</b>	Keseluruhan individu, objek, atau unit analisis yang memiliki karakteristik tertentu dan menjadi ruang lingkup utama dalam suatu penelitian atau analisis data.
<b>Preprocessing</b>	Tahap awal dalam pengolahan data yang meliputi pembersihan, transformasi, reduksi, dan penyesuaian format data agar siap digunakan dalam proses analisis lanjutan.
<b>Reliabilitas</b>	Tingkat konsistensi dan kestabilan hasil pengukuran suatu instrumen ketika digunakan berulang kali dalam kondisi yang relatif sama.
<b>Responden</b>	Individu atau entitas yang menjadi sumber informasi dalam penelitian dengan memberikan jawaban, tanggapan, atau data sesuai instrumen yang digunakan.
<b>Sampling</b>	Teknik sistematis dalam memilih sebagian anggota dari suatu populasi dengan metode tertentu agar sampel tersebut dapat merepresentasikan karakteristik populasi secara akurat dan efisien.
<b>Scraping</b>	Teknik pengambilan data secara otomatis dari halaman web atau sumber digital lain menggunakan skrip atau perangkat lunak tertentu untuk memperoleh informasi terstruktur.
<b>Sensor</b>	Perangkat elektronik atau digital yang mampu mendeteksi perubahan kondisi fisik atau lingkungan dan mengubahnya menjadi data yang dapat direkam dan dianalisis.
<b>Survei</b>	Metode pengumpulan data yang dilakukan dengan menyebarkan pertanyaan terstruktur kepada sejumlah responden untuk memperoleh informasi kuantitatif maupun kualitatif mengenai sikap, opini, atau karakteristik tertentu.

<b>Validitas</b>	Tingkat ketepatan dan kesesuaian suatu instrumen atau metode dalam mengukur konsep yang seharusnya diukur sesuai dengan tujuan penelitian.
<b>Variabel</b>	Atribut, karakteristik, atau faktor yang dapat diukur dan memiliki variasi nilai antar objek penelitian serta digunakan untuk dianalisis hubungannya dalam studi ilmiah.
<b>Wawancara</b>	Metode pengumpulan data yang melibatkan interaksi langsung antara peneliti dan responden melalui tanya jawab terarah untuk menggali informasi secara mendalam dan kontekstual.



# INDEKS

---

---

---

## A

akademik · 32, 133, 135, 136, 137, 184,  
241

aksesibilitas · 49, 112, 134, 227

*Algoritma* · 10, 63

audit · 13, 38, 44, 45, 47, 106, 107, 152,  
186, 198, 213, 229, 237, 241

---

## B

*behavior* · 103, 141, 187, 252

*big data* · 8, 9, 10, 11, 12, 13, 38, 39, 40,  
51, 154, 157, 165, 175, 182, 230, 234,  
236, 239, 240, 242, 249, 256, 267

*blockchain* · 237

---

## C

*cloud* · 9, 38, 40, 97, 98, 99, 100, 101, 102,  
107, 108, 112, 137, 153, 154, 155, 156,  
157, 159, 182, 209, 232, 235, 237, 238,  
240, 242, 249

---

## D

diferensiasi · 184, 239

digitalisasi · 84, 85, 86, 143

disparitas · 138

distribusi · 143, 154, 157, 159, 162, 174,  
176, 198, 223, 224, 237, 238, 239, 241

---

## E

*e-commerce* · 139, 165, 183, 246, 254

ekonomi · 125, 134, 135, 136, 139, 144,  
239, 243

empiris · 2, 3, 22, 23, 24, 25, 29, 77, 82,  
83, 117, 127, 129, 130, 133, 135, 137,  
184

entitas · 35, 86, 106, 183, 186, 209, 213,  
224, 231, 258

---

## F

finansial · 14, 59, 209, 211, 231

*firewall* · 112, 155

fleksibilitas · 32, 56, 68, 72, 73, 82, 90,  
100, 101, 108, 111, 120, 142, 151, 155,  
157, 197, 198, 201, 202, 207, 210, 239

fluktuasi · 43, 44, 184

fundamental · 15, 21, 41, 47, 53, 85, 96,  
122, 138, 160, 164, 167, 173, 179, 187,  
221, 225, 228, 237, 239, 241

---

## **G**

geografis · 8, 60, 119, 120, 125, 137, 139,  
226

---

## **I**

implikasi · 14, 24, 50, 52, 82, 161, 239,  
241

inflasi · 135

infrastruktur · 11, 12, 40, 97, 98, 107, 108,  
137, 138, 139, 153, 154, 155, 156, 157,  
172, 178, 182, 238

inovatif · 244

integrasi · 2, 5, 8, 12, 13, 38, 39, 40, 85,  
89, 99, 100, 101, 106, 114, 145, 146,  
149, 151, 152, 155, 158, 175, 178, 179,  
180, 182, 184, 196, 197, 199, 200, 202,  
206, 207, 208, 209, 210, 220, 222, 223,  
224, 225, 226, 230, 231, 234, 235, 236,  
239, 240

integritas · 6, 11, 13, 36, 86, 112, 113, 115,  
122, 150, 151, 152, 155, 160, 161, 178,  
182, 185, 187, 194, 196, 197, 200, 201,  
209, 210, 213, 220, 225, 228, 229, 233,  
237

interaktif · 80, 111, 140, 141, 142, 144,  
146

investasi · 59, 157

---

## **K**

kolaborasi · 52, 138, 140, 157, 199, 200,  
232

komparatif · 119, 120

komprehensif · 8, 24, 36, 38, 41, 43, 67,  
70, 73, 74, 79, 80, 99, 101, 114, 125,  
127, 146, 166, 168, 170, 183, 184, 207,  
220, 225, 235, 242, 267

komputasi · 9, 11, 17, 49, 98, 153, 154,  
155, 156, 157, 180, 182, 230, 234, 235,  
238, 257

konkret · 18, 77, 178

konsistensi · 5, 7, 9, 11, 13, 17, 22, 28, 36,  
37, 38, 39, 40, 43, 44, 45, 50, 52, 64,  
71, 72, 82, 83, 86, 87, 89, 95, 106, 123,  
124, 145, 146, 151, 158, 175, 182, 194,  
200, 201, 202, 218, 220, 221, 222, 224,  
227, 228, 229, 232, 235, 258

---

## **L**

*legacy* · 197, 202

---

## **M**

manajerial · 8, 230, 231

manipulasi · 14, 16, 29, 83, 88, 113, 159,  
161, 187, 197, 202

manufaktur · 238, 243

metodologi · 43, 48, 51, 55, 121, 132, 133,  
134, 159

---

## **P**

politik · 142, 144, 186

proyeksi · 2

---

## **R**

rasional · 2, 10, 24, 48, 127, 226

*rates* · 104

*real-time* · 8, 9, 11, 12, 14, 34, 38, 39, 40,  
58, 96, 99, 100, 101, 103, 104, 105,  
108, 111, 112, 114, 115, 137, 138, 139,  
140, 142, 150, 151, 152, 155, 157, 182,  
193, 206, 208, 210, 212, 229, 235, 237,  
240, 243

regulasi · 13, 41, 105, 113, 151, 152, 160,  
185, 186, 198, 209, 211, 231, 234, 237,  
241

relevansi · 1, 5, 13, 21, 27, 32, 33, 51, 53,  
55, 57, 60, 63, 69, 74, 79, 102, 118,  
119, 122, 124, 125, 130, 132, 133, 142,  
159, 168, 176, 177, 217, 218, 221, 228

*Representasi* · 161

*Retrieval* · 213, 251, 252

revolusi · 97

---

## **S**

siber · 13, 103, 112, 155, 160, 210, 237

stabilitas · 43, 44, 167, 180, 198, 224, 232

stakeholder · 13

---

## **T**

transformasi · 2, 6, 11, 12, 42, 47, 51, 85,  
88, 91, 106, 110, 146, 150, 151, 152,  
153, 154, 155, 158, 172, 175, 185, 206,  
209, 223, 224, 229, 230, 232, 236, 239,  
240, 243, 258, 267

transparansi · 47, 48, 49, 51, 65, 113, 123,  
134, 135, 136, 137, 141, 152, 161, 211,  
227, 234, 236, 238, 241



# BIOGRAFI PENULIS

---



**Prof. Dr. Harya Damar Widiputra, S.T.,  
M.Kom., IPM.**

Profesor di bidang Data Mining dan merupakan dosen tetap di Fakultas Teknologi Informasi Perbanas Institute. Aktif dalam pengajaran, penelitian, dan pengembangan kebijakan di bidang data analytics, sistem informasi, dan transformasi digital, dengan pengalaman kolaborasi bersama institusi pemerintah, pendidikan tinggi, dan industri.



# METODE PENGUMPULAN DATA

UNTUK SAINS DATA

Buku referensi “Metode Pengumpulan Data untuk Sains Data” membahas secara komprehensif berbagai pendekatan, teknik, dan strategi dalam memperoleh data yang berkualitas sebagai fondasi utama analisis dan pengambilan keputusan berbasis data. Buku referensi ini membahas konsep dasar pengumpulan data, jenis dan sumber data, desain instrumen, teknik sampling, serta metode pengumpulan data terstruktur dan tidak terstruktur, termasuk survei, wawancara, observasi, eksperimen, web scraping, API, sensor digital, dan media sosial. Selain itu, pembahasan dilengkapi dengan prinsip validitas, reliabilitas, etika penelitian, perlindungan privasi, serta tata kelola data dalam konteks big data dan transformasi digital. Buku referensi ini ditujukan bagi mahasiswa, peneliti, akademisi, dan praktisi yang ingin membahas metodologis yang kuat dalam proses pengumpulan data yang akurat, relevan, dan bertanggung jawab.



 [mediapenerbitindonesia.com](http://mediapenerbitindonesia.com)  
 +6281362150605  
 Penerbit Idn  
 @pt.mediapenerbitidn

