

# PEMODELAN DAN VISUALISASI DATA

Tim Penulis:

Fifit Alfiah | Agung Yuliyanto Nugroho | Ade Setiadi  
Indo Intan | Wiranti Sri Utami | Hamzah Nurrifqi Fakhri F  
Norbertus Tri Suswanto Saptadi | Tugiman | Mohammad Badrul  
Max Teja Ajie Cipta Widiyanto | Erna Astriyani | Patria Adhistian  
Siti Mutmainah | Nasril Sany | Muhamad Yusup  
Inda Rusdia Sofiani | Aqil Farras



# **PEMODELAN DAN VISUALISASI DATA**

**Fifit Alfiah  
Agung Yuliyanto Nugroho  
Ade Setiadi  
Indo Intan  
Wiranti Sri Utami  
Hamzah Nurrifqi Fakhri F  
Norbertus Tri Suswanto Saptadi  
Tugiman  
Mohammad Badrul  
Max Teja Ajie Cipta Widiyanto  
Erna Astriyani  
Patria Adhistian  
Siti Mutmainah  
Nasril Sany  
Muhamad Yusup  
Inda Rusdia Sofiani  
Aqil Farras**

# PEMODELAN DAN VISUALISASI DATA

## **Tim Penulis:**

Fifit Alfiah  
Agung Yuliyanto Nugroho  
Ade Setiadi  
Indo Intan  
Wiranti Sri Utami  
Hamzah Nurriqfi Fakhri F  
Norbertus Tri Suswanto Saptadi  
Tugiman  
Mohammad Badrul  
Max Teja Ajie Cipta Widiyanto  
Erna Astriyani  
Patria Adhistian  
Siti Mutmainah  
Nasril Sany  
Muhamad Yusup  
Inda Rusdia Sofiani  
Aqil Farras

**Editor** : Ajay Supriadi, M.Kom.  
**Tata Letak** : Asep Nugraha, S.Hum.  
**Desain Cover** : Septimike Yourintan Mutiara, S.Gz.  
**Ukuran** : UNESCO 15,5 x 23 cm  
**Halaman** : x, 287  
**ISBN** : 978-634-7021-51-9  
**Terbit Pada** : Juni 2025  
**Anggota IKAPI** : No. 073/BANTEN/2023

## **Hak Cipta 2025 @ Sada Kurnia Pustaka dan Penulis**

*Hak cipta dilindungi undang-undang dilarang memperbanyak karya tulis ini dalam bentuk dan dengan cara apapun tanpa izin tertulis dari penerbit dan penulis.*

## **PENERBIT PT SADA KURNIA PUSTAKA**

Jl. Warung Selikur Km.6 Sukajaya – Carenang, Kab. Serang-Banten  
Email : sadapenerbit@gmail.com  
Website : sadapenerbit.com & repository.sadapenerbit.com  
Telpon/WA : +62 838 1281 8431

# KATA PENGANTAR

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya sehingga buku yang berjudul "*Pemodelan dan Visualisasi Data*" ini dapat diselesaikan dengan baik. Di era yang terus bergerak maju ini, data telah menjadi nadi bagi hampir setiap aspek kehidupan, mulai dari pengambilan keputusan bisnis, pengembangan ilmiah, hingga pemahaman tren sosial. Namun, data mentah, seberapa pun besarnya, seringkali menyerupai tumpukan puzzle yang belum tersusun. Tanpa kemampuan untuk memahami polanya, menggali wawasan tersembunyi, dan mengomunikasikannya secara efektif, potensi besar data akan tetap terkunci. Di sinilah *Pemodelan dan Visualisasi Data* memainkan peran krusial, berfungsi sebagai jembatan yang menghubungkan kumpulan data kompleks dengan pemahaman yang jelas dan aksi yang tepat.

Buku ini hadir sebagai panduan komprehensif bagi siapa saja yang ingin menguasai seni dan sains di balik pemodelan serta visualisasi data. Kami percaya bahwa setiap individu, terlepas dari latar belakang akademis atau profesionalnya, memiliki potensi untuk mengubah data menjadi narasi yang kuat dan keputusan yang cerdas. Kami akan membawa Anda melangkah dari konsep dasar yang fundamental hingga teknik-teknik lanjutan yang transformatif. Anda akan belajar bagaimana memahami struktur intrinsik data, memilih model statistik atau pembelajaran mesin yang paling tepat untuk mengungkap pola dan hubungan yang tersembunyi, dan yang terpenting, bagaimana menyajikan temuan-temuan tersebut dalam bentuk visual yang tidak hanya menarik secara estetika, tetapi juga mudah dipahami, informatif, dan memiliki dampak nyata.

Dalam setiap bab, kami telah berupaya keras untuk menyajikan materi ini dengan cara yang jelas, praktis, dan relevan dengan tantangan dunia nyata. Anda akan menemukan beragam contoh kasus, studi kasus mendalam, dan latihan praktis yang dirancang khusus untuk memperdalam pemahaman konseptual Anda sekaligus mengasah keterampilan teknis Anda. Kami akan membahas berbagai

jenis pemodelan, mulai dari regresi sederhana hingga metode klustering dan klasifikasi yang lebih kompleks, serta memperkenalkan Anda pada berbagai jenis visualisasi, dari grafik dasar hingga dasbor interaktif.

Buku ini ditujukan bagi berbagai kalangan pembaca: analis data yang ingin menyempurnakan portofolio keterampilan mereka, ilmuwan data yang mencari referensi praktis, mahasiswa yang baru memasuki dunia data, atau profesional di berbagai bidang yang berkeinginan untuk memanfaatkan kekuatan data demi keunggulan kompetitif. Kami berharap buku ini tidak hanya menjadi sumber referensi yang berharga di perpustakaan Anda, tetapi juga menjadi inspirasi yang membangkitkan rasa ingin tahu dan semangat inovasi Anda dalam menjelajahi, memahami, dan mengomunikasikan data.

Selamat membaca, dan semoga perjalanan Anda dalam memahami dan menguasai Pemodelan dan Visualisasi Data ini penuh dengan wawasan baru dan keberhasilan. Mari bersama-sama mengubah data menjadi kekuatan yang mendorong perubahan positif!

Penulis

# DAFTAR ISI

<b>KATA PENGANTAR .....</b>	<b>iii</b>
<b>DAFTAR ISI .....</b>	<b>v</b>
<b>BAB 1 KONSEP DASAR PEMODELAN DAN VISUALISASI DATA .....</b>	<b>1</b>
Pendahuluan .....	2
Konsep Dasar Pemodelan Data .....	2
Proses Pemodelan Data .....	3
Konsep Visualisasi Data .....	4
Jenis-Jenis Visualisasi Data .....	5
Tools Visualisasi Data .....	9
Daftar Pustaka .....	11
Profil Penulis .....	13
<b>BAB 2 PERSIAPAN DATA (<i>DATA PREPARATION</i>) .....</b>	<b>14</b>
Pendahuluan .....	15
Pengumpulan Data ( <i>Data Collection</i> ) .....	16
Pembersihan Data ( <i>Data Cleaning</i> ) .....	17
Transformasi Data ( <i>Data Transformation</i> ) .....	18
Integrasi Data ( <i>Data Integration</i> ) .....	19
Reduksi Data ( <i>Data Reduction</i> ) .....	19
Validasi Data ( <i>Data Validation</i> ) .....	20
Daftar Pustaka .....	23
Profil Penulis .....	24
<b>BAB 3 EKSPLORASI DATA (<i>EXPLORATORY DATA ANALYSIS/EDA</i>) .....</b>	<b>25</b>
Pendahuluan .....	26
Tujuan dan Manfaat <i>Exploratory Data Analysis</i> (EDA) .....	27
Jenis Analisis Data Eksplorasi .....	28
Studi Kasus Sederhana .....	33
Daftar Pustaka .....	36
Profil Penulis .....	38
<b>BAB 4 PEMILIHAN DAN REKAYASA FITUR .....</b>	<b>39</b>
Pentingnya Fitur Dalam <i>Machine Learning</i> .....	40
Peran Fitur Terhadap Kinerja Model .....	40
Konsep Dasar Fitur Dalam <i>Machine Learning</i> .....	41

Pendekatan Utama Dalam Pemilihan Fitur .....	42
Evaluasi Hasil Pemilihan Fitur .....	43
Strategi dan <i>Workflow</i> Dalam Pemilihan dan Rekayasa Fitur...45	
Perbandingan Manual vs Otomatisasi (AutoML, <i>FeatureTools</i> ).....	47
Studi Kasus: Contoh Penerapan Nyata .....	48
Analisis dan Diskusi Hasil.....	50
Tantangan dan <i>Best Practices</i> .....	50
Daftar Pustaka.....	60
Profil Penulis.....	63
<b>BAB 5 PEMODELAN LAYANAN VISUALISASI DATA AWAN UNTUK PEMANTAUAN DATA MULTISENSOR .....</b>	<b>64</b>
Pendahuluan .....	65
Konsep Dasar Visualisasi Data Multisensor.....	65
Komponen Utama Arsitektur Berbasis <i>Cloud</i> .....	67
Alur Kerja Sistem .....	68
Keunggulan Arsitektur Berbasis <i>Cloud</i> .....	69
Teknologi Pendukung dan <i>Tools</i> Visualisasi.....	70
Studi Kasus Implementasi Visualisasi Data Multisensor Berbasis <i>Cloud</i> .....	71
Analisis dan Evaluasi Implementasi .....	72
Kesimpulan.....	74
Daftar Pustaka.....	76
Profil Penulis.....	79
<b>BAB 6 PENGOLAHAN DAN VISUALISASI DATA.....</b>	<b>80</b>
Pendahuluan .....	81
Konsep Dasar Pengolahan Data.....	81
Proses Pengolahan Data .....	83
Visualisasi Data.....	89
Kesimpulan.....	96
Daftar Pustaka.....	97
Profil Penulis.....	99
<b>BAB 7 BIG DATA DAN FRAMEWORK PEMROSESAN DATA.....</b>	<b>100</b>
Pendahuluan <i>Big Data</i> .....	101
Sumber dan Jenis Data Besar .....	102
Arsitektur <i>Big Data</i> .....	104


Teknologi Penyimpanan dan Manajemen <i>Big Data</i> .....	106
<i>Framework</i> Pemrosesan Data .....	108
Pemrosesan Data <i>Real-Time</i> dan <i>Streaming</i> .....	110
<i>Big Data</i> dan <i>Machine Learning</i> .....	112
Keamanan dan Etika Dalam Pengelolaan <i>Big Data</i> .....	114
Tren dan Masa Depan <i>Big Data</i> .....	115
Kesimpulan .....	116
Daftar Pustaka.....	119
Profil Penulis .....	122
<b>BAB 8 PRINSIP DASAR VISUALISASI DATA .....</b>	<b>123</b>
Pentingnya Prinsip Dasar Visualisasi Data.....	124
Prinsip Kejelasan ( <i>Clarity</i> ).....	125
Prinsip Kesederhanaan ( <i>Simplicity</i> ) .....	126
Prinsip Konsistensi ( <i>Consistency</i> ) .....	127
Prinsip Konteks ( <i>Context</i> ).....	128
Prinsip Fokus Pada Audiens ( <i>Audience Awareness</i> ).....	129
Prinsip Relevansi ( <i>Relevance</i> ).....	131
Daftar Pustaka .....	136
Profil Penulis.....	138
<b>BAB 9 JENIS-JENIS VISUALISASI DATA .....</b>	<b>139</b>
Pendahuluan .....	140
<i>Line Plot</i> .....	140
<i>Scatter Plot</i> .....	142
<i>Histogram</i> .....	144
Diagram Batang/ <i>Bar Chart</i> .....	147
Diagram Lingkaran .....	150
Daftar Pustaka .....	152
Profil Penulis .....	153
<b>BAB 10 ALAT DAN PLATFORM VISUALISASI DATA.....</b>	<b>154</b>
Pendahuluan .....	155
Pengertian Alat dan <i>Platform</i> Visualisasi Data Menurut Ahli. ....	155
Pentingnya Visualisasi Data Dalam Pengambilan Keputusan .....	156
Prinsip-Prinsip Dasar Visualisasi Data.....	158
Klasifikasi Alat Visualisasi Data Berdasarkan Fungsi dan Kompleksitas.....	159
Implementasi Alat dan <i>Platform</i> Visualisasi Data .....	161

Daftar Pustaka.....	165
Profil Penulis.....	167
<b>BAB 11 VISUALISASI DATA UNTUK ANALISIS SPASIAL.....</b>	<b>168</b>
Analisis dan Data Spasial.....	169
Tipe dan Sumber Data Spasial.....	169
Format Penyimpanan Data Spasial .....	171
Kegunaan dan Jenis Visualisasi Data Spasial.....	172
Alat dan <i>Software</i> Visualisasi Data Spasial .....	175
Cara Kerja Analisis Spasial .....	178
Analisis Data.....	178
Contoh Kasus Penggunaan Visualisasi Data Spasial .....	179
Daftar Pustaka.....	181
Profil Penulis.....	183
<b>BAB 12 VISUALISASI DATA WAKTU (<i>TIME SERIES</i>).....</b>	<b>184</b>
Pengertian Visualisasi Data Deret Waktu ( <i>Time Series</i> ) .....	185
Jenis-Jenis Visualisasi <i>Time Series</i> .....	185
<i>Tools</i> Visualisasi <i>Time Series</i> dan <i>Power BI</i> .....	186
Menghubungkan dan Mengkonfigurasi Sumber Data.....	189
Mengkonfigurasi Data.....	192
Tren Data Dengan <i>Line Chart</i> .....	193
Metrik Dengan Clustered Bar Chart.....	195
Membandingkan Metrik Dengan <i>Combo Chart</i> .....	195
Daftar Pustaka.....	197
Profil Penulis.....	198
<b>BAB 13 VISUALISASI DATA UNTUK <i>MACHINE LEARNING</i>.....</b>	<b>199</b>
Pentingnya Visualisasi Data Dalam <i>Machine Learning</i> .....	200
<i>Tools</i> dan <i>Library</i> Visualisasi <i>Machine Learning</i> .....	201
Visualisasi Data Sebelum <i>Training Model</i> .....	202
Visualisasi Setelah <i>Training Model</i> .....	206
Kesimpulan.....	211
Daftar Pustaka.....	212
Profil Penulis.....	213
<b>BAB 14 ETIKA DAN KEAMANAN DALAM PEMODELAN DATA... 214</b>	
Pendahuluan .....	215
Pengertian Etika Dalam Pemodelan Data .....	215
Pengertian Keamanan Dalam Pemodelan Data.....	216

Etika Dalam Pemodelan Data.....	216
Isu Etika Umum Dalam Teknologi dan Pemodelan Data .....	218
Keamanan Dalam Pemodelan Data .....	220
Regulasi Standar Etika dan Keamanan Dalam Pemodelan Data.....	221
Daftar Pustaka.....	224
Profil Penulis.....	226
<b>BAB 15 INTEGRASI MODEL DATA KE SISTEM PRODUKSI .....</b>	<b>227</b>
Pendahuluan .....	228
Tahap Utama Integrasi Model.....	229
Pengembangan dan Penyimpanan Model.....	229
Pembuatan API Model .....	231
<i>Deployment Model</i> .....	233
Integrasi Dengan Sistem Aplikasi .....	233
<i>Monitoring dan Logging</i> .....	236
Otomatisasi dan MLOps.....	238
Pemeliharaan dan Pembaharuan Model .....	241
Daftar Pustaka.....	243
Profil Penulis.....	246
<b>BAB 16 PEMODELAN DAN VISUALISASI DI SEKTOR KESEHATAN .....</b>	<b>247</b>
Transformasi <i>Digital</i> dan Industri Kesehatan 4.0 .....	248
Peran Data Dalam Pelayanan Kesehatan Modern.....	248
Pentingnya Pemodelan dan Visualisasi Data.....	249
Jenis dan Sumber Data Kesehatan (EHR, Citra Medis, dan Sensor Vital ).....	250
Etika dan Privasi Data Kesehatan .....	252
Proses Akuisisi, Penyimpanan, dan Manajemen Data.....	253
Prinsip Desain Visualisasi Untuk Data Medis.....	254
Alat dan Teknik Visualisasi ( <i>Tableau, Power BI, Python, R</i> ) ....	255
Visualisasi Data Klinis dan Citra Medis .....	256
Dasar-Dasar <i>Computer Vision</i> dan Prapemrosesan Citra .....	257
Segmentasi, Klasifikasi, dan Deteksi Objek Pada Citra Medis	258
Daftar Pustaka.....	260
Profil Penulis.....	261

## **BAB 17 PEMODELAN DAN VISUALISASI DI SEKTOR *E-COMMERCE***

.....	<b>262</b>
Pendahuluan .....	263
Pemodelan Data dan Visualisasi Interaktif Untuk Optimalisasi Strategi <i>E-Commerce</i> .....	264
Di Balik Klik dan <i>Scroll</i> : Mengenal Sumber-Sumber Data <i>E-Commerce</i> .....	267
Visualisasi Data Interaktif Sebagai Alat Navigasi Strategi .....	270
Pemetaan Tren Konsumen dan Optimasi Penjualan Melalui Teknik Pemodelan Modern.....	272
Tantangan dan Solusi.....	273
<i>Smart Visualization for Smart Commerce</i> : Membuka Rahasia Perilaku Konsumen Digital .....	274
Visualisasi Sebagai Jembatan Loyalitas Konsumen Dalam Ekosistem <i>E-Commerce</i> .....	277
Preferensi Visualisasi Konsumen: Menyesuaikan Format Visual dengan Ekspektasi <i>Digital</i> .....	280
Simpulan Strategis .....	284
Daftar Pustaka.....	285
Profil Penulis .....	287



# **BAB 1**

# **KONSEP DASAR**

# **PEMODELAN DAN**

# **VISUALISASI DATA**

---

**Fifit Alfiah, S.Kom., M.Kom.**  
Universitas Raharja



## Pendahuluan

Konsep dasar pemodelan dan visualisasi data adalah bagian penting dalam proses analisis data untuk membantu memahami, menggambarkan, dan mengkomunikasikan informasi secara lebih efektif. Konsep dasar pemodelan dan visualisasi data berpusat pada transformasi data mentah menjadi informasi yang bermakna dan mudah dipahami.

Pemodelan data melibatkan proses pengorganisasian dan strukturisasi data untuk menggambarkan relasi antar variabel dan pola yang mendasarinya. Ini bisa dilakukan melalui berbagai metode statistik dan *machine learning*. Sementara itu, visualisasi data menggunakan representasi grafis, seperti grafik, diagram, dan peta, untuk mengkomunikasikan wawasan data kepada audiens.

Tujuannya adalah untuk mengungkap tren, anomali, dan korelasi yang mungkin tersembunyi dalam data, sehingga memudahkan pengambilan keputusan yang lebih baik dan lebih terinformasi. Dengan kombinasi yang tepat antara pemodelan yang solid dan visualisasi yang efektif, data dapat diubah menjadi aset strategis yang bernilai bagi organisasi.

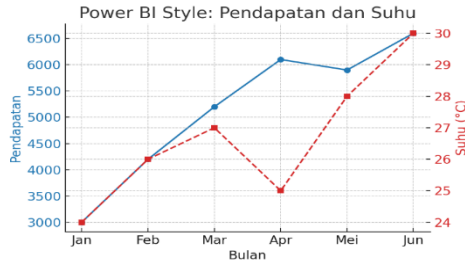
## Konsep Dasar Pemodelan Data

Konsep dasar pemodelan data adalah proses menciptakan representasi visual atau diagramatik dari sistem data informasi yang kompleks. Tujuannya adalah untuk memahami, mendokumentasikan, dan mengkomunikasikan struktur data, relasi antar data, dan batasan-batasan yang berlaku dalam sebuah sistem. Pemodelan data merupakan langkah penting dalam pengembangan basis data, sistem informasi, dan aplikasi perangkat lunak lainnya.

Dengan pemodelan data yang baik, pengembang dan *stakeholder* dapat memiliki pemahaman yang sama mengenai data yang akan disimpan, diproses, dan digunakan, sehingga meminimalkan kesalahan dan meningkatkan efisiensi pengembangan. Inti dari pemodelan data adalah identifikasi entitas (objek atau konsep yang ingin direpresentasikan), atribut (karakteristik atau properti entitas), dan relasi (hubungan antara entitas).

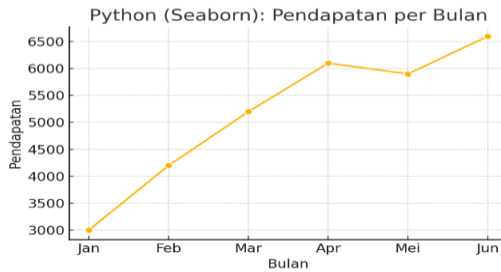
Pemodelan data juga melibatkan penentuan tipe data untuk setiap atribut dan penerapan batasan-batasan integritas data untuk

### 3. **Power BI**: kuat untuk integrasi data korporat.



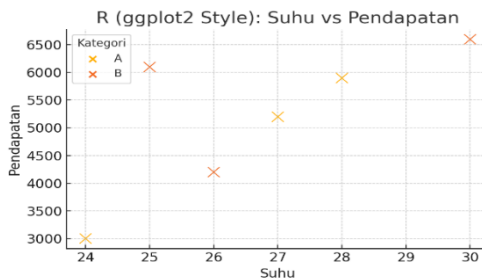
**Gambar 1.8: Power BI**  
Sumber: Diolah Penulis.

### 4. **Python**: fleksibel dengan pustaka *Matplotlib*, *Seaborn*, *Plotly*.



**Gambar 1.9: Python**  
Sumber: Diolah Penulis.

### 5. **R**: visualisasi statistik yang kuat dengan *ggplot2*.



**Gambar 1.10: R**  
Sumber: Diolah Penulis.

Pemilihan *tools* disesuaikan dengan kompleksitas data dan kebutuhan pengguna.

## Daftar Pustaka

- Healy, K. (2018). *Data Visualization: A Practical Introduction*. Princeton University Press.
- Hua, Z. (2025). Data Visualization And Prediction Model Analysis. *ResearchGate*.  
[https://www.researchgate.net/publication/388315254\\_Data\\_Visualization\\_and\\_Prediction\\_Model\\_Analysis](https://www.researchgate.net/publication/388315254_Data_Visualization_and_Prediction_Model_Analysis).
- Irfansyah, A. (2022, July 8). *Alasan Pentingnya Visualisasi Data Untuk Perusahaan Anda Eduparx Blog*.  
<https://eduparx.id/blog/insight/pentingnya-visualisasi-data-untuk-perusahaan/>.
- Jamilatul, A. (2022, September 22). *Visualisasi Data: Definisi, Manfaat, dan Caranya. Pacmann*. <https://pacmann.io/blog/visualisasi-data-definisi-manfaat-dan-caranya>.
- Jondri, J., & Rohmawati, A. A. (2021). Pelatihan Visualisasi Dan Analisis Data Menggunakan Tableau Di SMKN 3 Bandung. *Charity: Jurnal Pengabdian Masyarakat*, 4(2),  
<https://doi.org/10.25124/charity.v4i2.3435>.
- Kusramdani, R. (2025). Visualisasi Data Untuk Data Warehouse: Memahami dan Mengoptimalkan Informasi. *ResearchGate*.  
[https://www.researchgate.net/publication/388860619\\_Visualisasi\\_Data\\_untuk\\_Pemula](https://www.researchgate.net/publication/388860619_Visualisasi_Data_untuk_Pemula).
- Kusramdani, R. (2025). Visualisasi Data Untuk Pemula. *ResearchGate*.  
[https://www.researchgate.net/publication/388860619\\_Visualisasi\\_Data\\_untuk\\_Pemula](https://www.researchgate.net/publication/388860619_Visualisasi_Data_untuk_Pemula).
- Lavalle, A., Maté, A., & Trujillo, J. (2024). Requirements-Driven Visualizations For Big Data Analytics: A Model-Driven Approach. *arXiv preprint*. <https://arxiv.org/abs/2402.07914>.
- Lestari, V. N. (2021). *Pemilihan Article Visualisasi*.  
<https://eksplorasidata.mipa.ugm.ac.id/2021/08/20/pemilihan-visualisasi-data/>.
- Ma, K.-L. (2020). A Deep Generative Model For Graph Layout. *IEEE Transactions on Visualization and Computer Graphics*, 26 (1), 665–675.

- McKinney, W. (2018). *Python For Data Analysis: Data Wrangling With Pandas, NumPy, and IPython (2nd ed.)*. O'Reilly Media.
- Ouyang, W. (2024). Data Visualization In Big Data Analysis: Applications And Future Trends. *Journal of Computer and Communications*, 12 (11), 76–85. <https://doi.org/10.4236/jcc.2024.1211005>.
- Ratnaningsih, E., & Hakim, A. (2021). *Visualisasi Dan Analisis Data*. Universitas Negeri Medan. <https://digilib.unimed.ac.id/id/eprint/53010/1/Book.pdf>
- Romie Priyastama. (2017). *Buku Sakti Kuasai SPSS Pengolahan Data dan Analisis Data*. Yogyakarta: Penerbit Start Up.
- Sulianta, D. (2024). Visualisasi Data Dalam Pengembangan AI. *ResearchGate*. [https://www.researchgate.net/publication/388860619\\_Visualisasi\\_Data\\_untuk\\_Pemula](https://www.researchgate.net/publication/388860619_Visualisasi_Data_untuk_Pemula).
- Wang, J., Liu, S., & Zhang, W. (2023). Visual Analytics For Machine Learning: A Data Perspective Survey. *arXiv preprint*. <https://arxiv.org/abs/2307.07712>.
- Wu, E. (2024). Database Theory + X: Database Visualization. *Arxiv Preprint*. <https://arxiv.org/abs/2412.04101>.
- Yilmaz, R., & Yilmaz, M. (2023). Advanced Large Language Models And Visualization Tools For Data Analytics Education. *Frontiers in Education*, 8, 1418006. <https://www.frontiersin.org/articles/10.3389/feduc.2024.1418006/full>.

## PROFIL PENULIS



### **Fifit Alfiah, M.Kom.**

Ketertarikan penulis terhadap ilmu komputer dimulai pada tahun 2011 yaitu semenjak melanjutkan pendidikan ke Perguruan Tinggi dan berhasil menyelesaikan Studi S1 di Prodi Teknik Informatika Universitas Raharja pada tahun 2015, selama menempuh Pendidikan Studi S1 penulis juga membangun komunitas *Software Engineering* pada tahun 2012 bersama rekan-rekan mahasiswa di Prodi Teknik Informatika. Setelah rehat beberapa bulan pada tahun 2015 kemudian penulis melanjutkan Pendidikan Pasca Sarjana (S2) pada Januari 2016 dan 2 (Dua) tahun kemudian, penulis menyelesaikan studi Pasca Sarjana (S2) di prodi Rekayasa Komputasi Terapan Program Pasca Sarjana Ilmu Komputer Universitas Budi Luhur. Penulis memiliki ketertarikan dibidang *Decision Support System*, *Image Processing* dan *Web Technology*. Guna mewujudkan karir sebagai dosen profesional, penulis pun aktif sebagai peneliti di bidang kepakarannya tersebut dengan telah menghasilkan puluhan artikel ilmiah di berbagai jurnal nasional dan internasional. Selain peneliti, penulis juga aktif menulis buku dengan harapan dapat memberikan kontribusi positif bagi bangsa dan negara yang sangat tercinta ini. Atas dedikasi dan kerja keras Penulis juga telah memiliki Sertifikasi Pendidik sebagai dosen profesional yang diakui oleh kemenristekdikti pada tahun 2023.

Email Penulis: [fifitalfiah@raharja.info](mailto:fifitalfiah@raharja.info).



# **BAB 2**

## **PERSIAPAN DATA (*DATA PREPARATION*)**

---

**Agung Yuliyanto Nugroho, M.Kom., M.Par.**  
Universitas Cendekia Mitra Indonesia





**Gambar 2.1: Diagram Alur Proses Persiapan Data (*Data Preparation*) Dalam Pemodelan Dan Visualisasi Data**

Sumber: Diolah Penulis.

### **Pengumpulan Data (*Data Collection*)**

Pengumpulan Data (*Data Collection*) adalah tahap awal dalam proses persiapan data yang bertujuan untuk memperoleh data mentah dari berbagai sumber yang relevan. Data dapat dikumpulkan melalui berbagai metode, baik secara manual maupun otomatis, tergantung pada kebutuhan dan jenis data yang diinginkan.

Sumber data bisa berasal dari internal organisasi (seperti *database* perusahaan, laporan keuangan, sistem informasi), maupun eksternal (seperti data publik, API, media sosial, atau hasil survei). Metode pengumpulan juga dapat mencakup observasi, wawancara, kuesioner, sensor *digital*, hingga *scraping data* dari situs *web*. Data dikumpulkan dari berbagai sumber yang relevan, seperti:

1. *Database* Transaksi: data penjualan, harga, jumlah produk yang terjual, tanggal transaksi, dan informasi pelanggan

## 2. Pemeriksaan Konsistensi

Memastikan bahwa data yang terhubung di seluruh sistem atau sumber data adalah konsisten. Misalnya, nama produk yang muncul di dua tabel berbeda harus seragam dan tidak ada perbedaan ejaan atau penulisan.

## 3. Pemeriksaan Kelengkapan

Memastikan bahwa tidak ada data yang hilang atau tidak terisi, terutama pada kolom-kolom yang penting untuk analisis atau pemodelan. Jika ada nilai yang hilang, perlu diambil tindakan seperti pengisian atau penghapusan.

## 4. Pemeriksaan Rentang Nilai

Validasi dilakukan untuk memeriksa apakah nilai dalam dataset berada dalam rentang yang diharapkan. Misalnya, usia seseorang harus berada antara 0 dan 120 tahun, atau penghasilan harus berada dalam rentang yang realistis.

## 5. Pemeriksaan Duplikasi

Memeriksa apakah ada data yang terduplikasi dan memutuskan apakah duplikasi tersebut perlu dipertahankan atau dihapus.

Melalui validasi data, kesalahan atau inkonsistensi yang mungkin terlewatkan pada tahap sebelumnya dapat ditemukan dan diperbaiki, memastikan data yang digunakan untuk analisis memiliki kualitas tinggi dan memberikan hasil yang valid dan dapat dipercaya. Kemudian data yang telah dipersiapkan perlu divalidasi untuk memastikan kualitasnya dengan cara:

1. **Memeriksa Keakuratan Data:** memastikan bahwa data yang digabungkan benar dan konsisten. Misalnya, memeriksa apakah harga produk di database sesuai dengan harga yang tercatat dalam transaksi.
2. **Validasi Rentang Nilai:** memeriksa apakah harga dan jumlah penjualan berada dalam rentang yang realistis.

Proses persiapan data mencakup beberapa langkah seperti pembersihan data (menghapus duplikat, memperbaiki kesalahan penulisan, mengisi atau menangani data yang hilang), transformasi data (mengubah format data, melakukan normalisasi, atau encoding data kategorikal), serta integrasi data dari berbagai sumber menjadi satu kesatuan yang utuh.

Selain itu, proses ini juga dapat melibatkan reduksi data, yaitu menyederhanakan jumlah atribut atau baris data untuk memfokuskan hanya pada informasi yang relevan. Melalui tahap ini, kualitas data ditingkatkan sehingga hasil analisis atau visualisasi yang dilakukan selanjutnya menjadi lebih akurat dan dapat diandalkan. Dengan demikian, persiapan data merupakan pondasi penting dalam setiap proses berbasis data.

## Daftar Pustaka

- Few, S. (2009). *Now You See It: Simple Visualization Techniques For Quantitative Analysis*. Analytics Press.
- Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: Concepts And Techniques (3rd ed.)*. Morgan Kaufmann.
- Hevner, A. R., & Chatterjee, S. (2010). *Design Science Research In Information Systems: Theory And Practice*. Springer.
- Kelleher, J. D., & Tierney, B. (2018). *Data Science*. The MIT Press.
- McKinney, W. (2017). *Python For Data Analysis: Data Wrangling With Pandas, NumPy, And IPython (2nd ed.)*. O'Reilly Media.
- Provost, F., & Fawcett, T. (2013). *Data Science For Business: What You Need To Know About Data Mining And Data-Analytic Thinking*. O'Reilly Media.
- Shmueli, G., Bruce, P. C., Gedeck, P., & Patel, N. R. (2020). *Data Mining For Business Analytics: Concepts, Techniques, And Applications In R*. Wiley.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.
- Ware, C. (2012). *Information Visualization: Perception For Design (3rd ed.)*. Morgan Kaufmann.
- Yau, N. (2013). *Data Points: Visualization That Means Something*. Wiley.

## PROFIL PENULIS




### **Agung Yuliyanto Nugroho, M.Kom., M.Par.**

Ketertarikan penulis terhadap Ilmu Komputer dimulai pada tahun 2015 silam. Hal tersebut membuat penulis melanjutkan pendidikan ke Perguruan Tinggi dan berhasil menyelesaikan studi S1 di prodi Teknik Informatika Universitas Teknologi Yogyakarta pada tahun 2018. Dua tahun kemudian, penulis menyelesaikan studi S2 di

Prodi Teknik Informatika Program Pasca Sarjana Universitas Amikom Yogyakarta dan juga Prodi Magister Pariwisata di Sekolah Tinggi Pariwisata Ambarukmo Yogyakarta. Atas dedikasi dan kerja keras dalam membuat suatu karya, Republik Indonesia Kementerian Hukum dan Hak Asasi Manusia sudah mencatat ada kurang lebih 100 karya yang sudah tercatat di surat pencatatan ciptaan sebagai salah satu kontribusi dalam melindungi hak kekayaan intelektual.

Email Penulis: [agungboiler11@gmail.com](mailto:agungboiler11@gmail.com).



**BAB 3**  
**EKSPLORASI DATA**  
**(*EXPLORATORY DATA***  
***ANALYSIS/EDA*)**

---

**Ade Setiadi, S.Kom., M.Kom.**  
Universitas Raharja



Melalui EDA, kami mengubah data mentah menjadi wawasan yang dapat ditindaklanjuti. Visualisasi seperti histogram, plot sebar, dan plot kotak sangat berharga untuk mengungkapkan distribusi data, hubungan antar variabel, dan keberadaan outlier. Ringkasan statistik seperti rata-rata, median, standar deviasi, dan kuartil memberikan perspektif numerik tentang kecenderungan dan variabilitas sentral data. Dengan menggabungkan pendekatan ini, EDA memungkinkan kami untuk membersihkan dan menyiapkan data secara efektif, memilih fitur yang sesuai untuk pemodelan, dan pada akhirnya membangun model prediktif yang lebih kuat dan andal.

Tanpa EDA yang menyeluruh, kami berisiko menarik kesimpulan yang tidak akurat atau membangun model berdasarkan asumsi yang salah, menyoroti pentingnya hal yang penting dalam alur kerja ilmu data. EDA memungkinkan analisis untuk mendeteksi *outlier*, *missing value*, korelasi antar variabel, serta membantu membentuk hipotesis awal. Dengan menggunakan teknik statistik dan visualisasi, EDA menjadi pondasi penting dalam proses *data science* dan *machine learning*.

### **Tujuan dan Manfaat *Exploratory Data Analysis* (EDA)**

Tujuan utama *Exploratory Data Analysis* (EDA) adalah untuk memahami data secara mendalam sebelum melakukan analisis yang lebih mendalam atau membangun model prediktif. EDA membantu kita mengungkap struktur data, mengidentifikasi pola, anomali, dan hubungan antar variabel.

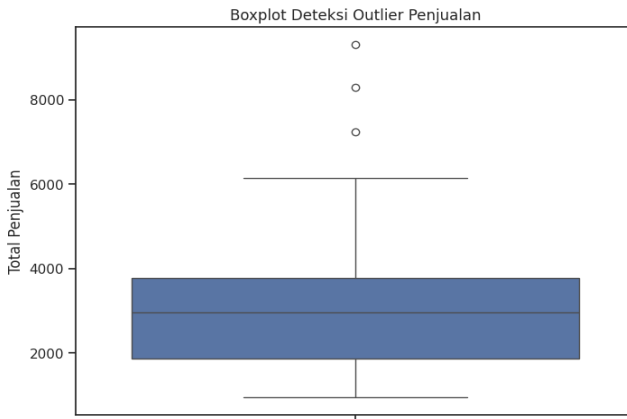
Proses ini melibatkan visualisasi data, perhitungan statistik deskriptif, dan teknik lain untuk mendapatkan wawasan (*insight*) tentang data yang dimiliki. Dengan memahami data secara menyeluruh, kita dapat membuat keputusan yang lebih tepat dalam langkah-langkah analisis selanjutnya, seperti:

1. Memahami distribusi data.
2. Menemukan pola atau tren.
3. Mendeteksi *outlier* dan anomali.
4. Memastikan kualitas dan kebersihan data.
5. Membantu pemilihan fitur penting.

Manfaat dilakukannya EDA sangatlah beragam. Pertama, EDA membantu dalam pembersihan data dengan mengidentifikasi *missing*

hubungan linier secara visual. Koefisien korelasi yang dicetak sebelumnya memberikan ukuran numerik dari kekuatan dan arah hubungan ini.

#### 4. *Outlier*: deteksi lonjakan tidak wajar



**Gambar 3.9: Boxplot Deteksi *Outlier* Penjualan**

Sumber: Diolah Penulis.

Grafik ini menampilkan *boxplot* untuk kolom 'Penjualan'. *Boxplot* memberikan ringkasan visual dari distribusi data melalui kuartil.

- Kotak (*Box*): menunjukkan rentang interkuartil (IQR), yaitu area antara kuartil pertama (Q1) dan kuartil ketiga (Q3). *Median* (Q2) ditandai di dalam kotak.
- "Kumis" (*Whiskers*): memanjang dari kotak hingga nilai minimum dan maksimum dalam  $1.5 \cdot \text{IQR}$  dari tepi kotak.
- Titik-titik di luar "kumis": Titik-titik ini mewakili nilai-nilai yang dianggap sebagai *outlier* potensial. Ini adalah nilai-nilai yang lebih kecil dari  $Q1 - 1.5 \cdot \text{IQR}$  atau lebih besar dari  $Q3 + 1.5 \cdot \text{IQR}$ .

Dalam konteks data penjualan, *outlier* ini bisa menunjukkan hari-hari dengan penjualan yang sangat tinggi atau sangat rendah yang mungkin disebabkan oleh faktor-faktor tertentu (seperti promosi besar atau masalah teknis). Hasil: ditemukan tren musiman dan pengaruh promosi terhadap lonjakan penjualan.

## Daftar Pustaka

- Alcantara, T., & Calvo, H. (2022). Exploratory Data Analysis for the Automatic Detection of Question Paraphrasing in Collaborative Environments. In *Lecture Notes in Computer Science* (Vol. 13456, pp. 123–134). Springer. ResearchGate.
- Ariza, P., Piñeres-Melo, M. A., & Barcelo, E. (2025). Exploratory Data Analysis for Malware Detection in Cybersecurity. *International Journal of Cybersecurity Research*, 12(2), 99–115. ResearchGate.
- Behrens, J. T., & Yu, C. H. (2003). Exploratory Data Analysis. In *The SAGE Handbook of Quantitative Methodology For The Social Sciences* (pp. 33–48). SAGE Publications.
- Camizuli, E., & Carranza, E. J. M. (2018). Exploratory Data Analysis (EDA). In *Encyclopedia of Archaeological Sciences*. Wiley. <https://doi.org/10.1002/9781119188230.saseas0271> ResearchGate.
- Chakri, P., Pratap, S., & Kumar, S. (2023). An Exploratory Data Analysis Approach for Analyzing Financial Accounting Data Using Machine Learning. *Decision Analytics Journal*, 7, Article 100212. <https://doi.org/10.1016/j.dajour.2023.100212>.
- De Castro, L. N. (2024). *Exploratory Data Analysis: Descriptive Analysis, Visualization, and Dashboard Design*. Springer.
- Dhany, H. W., Sutarman, S., & Izhari, F. (2023). Exploratory Data Analysis (EDA) methods for healthcare classification. *Journal of Intelligent Decision Support System (IDSS)*, 6(4), 209–215. <https://doi.org/10.35335/idss.v6i4.165>.
- Manatkar, A., Patel, D., Patel, H., & Manwani, N. (2024). ILAEDA: An Imitation Learning Based Approach for Automatic Exploratory Data Analysis. *arXiv preprint arXiv:2410.11276*. arxiv.org.
- Omoniyi, S., & Ajayi, O. (2024). Exploratory Data Analytics of Air Pollutant Data for Air Quality Management Application. *International Journal of Technology and Systems*, 9(2), 82–107. <https://doi.org/10.47604/ijts.2790>.

- Putri Silmina, E., & Najmah, S. (2025). Analisis Faktor Sosial yang Mempengaruhi Percobaan Penggunaan Narkoba Menggunakan Exploratory Data Analysis (EDA) dan Linear Regression. *Jurnal Ilmu Sosial dan Humaniora*, 10(1), 55–68. ResearchGate.
- Rahmany, M., Zin, A. M., & Sundararajan, E. (2020). Comparing Tools Provided by Python and R for Exploratory Data Analysis. *Journal of Data Science Tools*, 5(3), 45–58. ResearchGate.
- Shaik, H., Rao, A. S., & Vardhan, B. V. (2021). Role of Exploratory Data Analysis in Data Science. In *Proceedings of the International Conference on Data Science and Applications* (pp. 25–30). ResearchGate.
- Singh, G., & Sharma, S. (2025). Enhancing precision agriculture through cloud-based transformative crop recommendation model. *Journal of Agricultural Informatics*, 16(1), 45–60. ResearchGate.
- Velleman, P. F. (2012). Exploratory Data Analysis. In H. Cooper (Ed.), *APA Handbook of Research Methods in Psychology: Vol. 3. Data Analysis and Research Publication* (pp. 51–70). American Psychological Association.
- Wootton, D., Fox, A. R., Peck, E., & Satyanarayan, A. (2024). Charting EDA: Characterizing Interactive Visualization Use in Computational Notebooks with a Mixed-Methods Formalism. *arXiv preprint arXiv:2409.10450*. arxiv.org.
- Yusnita, I., Sumarmo, U., & Fitriani, N. (2025). Exploratory Data Analysis on Mathematics Learning Difficulties for Junior High School Students/MTs on Solving Story Problems in Terms of Gender. *Jurnal Pendidikan Matematika*, 9(2), 112–125. ResearchGate.
- Zhu, J.-P., Niu, B., Cai, P., Ni, Z., Wan, J., Xu, K., ... & Liu, Q. (2024). Towards Automated Cross-domain Exploratory Data Analysis Through Large Language Models. *arXiv preprint arXiv:2412.07214*. arxiv.org.

## PROFIL PENULIS



### **Ade Setiadi, S.Kom., M.Kom.**

Ketertarikan penulis terhadap ilmu komputer dimulai pada usia 10 tahun. Hal tersebut membuat penulis memilih untuk masuk kuliah Strata 1 ke STMIK Raharja dengan memilih Jurusan Teknik Informatika dengan konsentrasi *Software Engineering* (SE) lulus dengan IPK 3,67 dan menjadi wisudawan terbaik jurusan Teknik Informatika. Penulis kemudian melanjutkan pendidikan ke Strata 2 di Universitas Budi Luhur dengan Jurusan Magister Ilmu Komputer dengan konsentrasi Rekayasa Komputasi Terapan (RKT) dan berhasil lulus dengan IPK 3,44 pada tahun 2018. Penulis mewujudkan karir sebagai dosen pada Universitas Raharja mulai dari 2017. Kini penulis telah menjadi dosen profesional yaitu telah lulus sertifikasi dosen pada tahun 2023 serta memiliki jabatan fungsional Lektor dalam mata kuliah Pemrograman dan *Web Programming*, penulis pun aktif sebagai peneliti di bidang kepakarannya tersebut dengan menulis jurnal nasional/jurnal internasional setiap semesternya. Aktif menjadi pembicara pada kegiatan pengabdian masyarakat ke Sekolah Menengah Atas Sederajat setiap semesternya. Penulis aktif menjadi Reviewer Jurnal Nasional, Selain itu penulis aktif sebagai di anggota profesi yaitu sebagai anggota Asosiasi Dosen Indonesia (ADI) sejak tahun 2022.

Email Penulis: [adesetiadi@raharja.info](mailto:adesetiadi@raharja.info).



# **BAB 4**

## **PEMILIHAN DAN REKAYASA FITUR**

---

**Indo Intan, S.T., M.T.**  
Universitas Dipa Makassar



## Pentingnya Fitur Dalam *Machine Learning*

Dalam proses pembangunan model *machine learning*, data adalah bahan baku utama yang akan digunakan untuk pelatihan dan prediksi. Namun, data mentah yang dikumpulkan dari dunia nyata umumnya belum siap langsung digunakan. Di sinilah fitur memainkan peran vital. Fitur (*features*) (Ding et al., 2025) adalah representasi terstruktur dari data mentah yang digunakan sebagai input bagi algoritma pembelajaran mesin.

Kualitas dan relevansi fitur yang digunakan sangat menentukan seberapa baik model dapat mempelajari pola dari data dan melakukan generalisasi terhadap data baru (Moslemi, 2023). Selama bertahun-tahun, berbagai penelitian dan pengalaman praktis telah menunjukkan bahwa kualitas fitur seringkali lebih penting daripada jenis algoritma yang digunakan.

Bahkan model sederhana seperti regresi linier bisa mengungguli model kompleks jika didukung oleh fitur yang informatif dan representatif. Oleh karena itu, kemampuan untuk memilih fitur yang tepat dan merekayasa fitur baru dari data mentah menjadi salah satu keterampilan paling kritis dalam *machine learning* dan *data science* secara umum (HAQUE, 2022).

## Peran Fitur Terhadap Kinerja Model

Fitur berfungsi sebagai jembatan antara data dan model. Algoritma *machine learning* tidak memahami konteks data dalam bentuk aslinya, tetapi mereka mengenali pola dari angka dan struktur yang terdefinisi dengan jelas. Fitur yang relevan dan bersih akan membantu model untuk (Barbieri et al., 2024):

1. Meningkatkan Akurasi Prediksi: karena fitur yang baik membawa sinyal yang kuat terhadap target.
2. Mempercepat Proses Pelatihan: karena model tidak perlu memproses informasi yang tidak berguna.
3. Menghindari *Overfitting*: dengan mengeliminasi fitur yang redundan atau tidak informatif.
4. Meningkatkan Interpretabilitas, terutama dalam model-model yang sifatnya transparan seperti regresi atau pohon keputusan.

Sebaliknya, fitur yang buruk, seperti yang mengandung *noise*, *missing value* ekstrem, atau informasi tidak relevan bisa menurunkan

## Analisis dan Diskusi Hasil

Dari tiga studi kasus di atas, dapat disimpulkan bahwa:

1. Pemilihan fitur penting untuk mengurangi kompleksitas model, menghindari *overfitting*, dan meningkatkan akurasi.
2. Rekayasa fitur seringkali lebih berdampak daripada memilih algoritma model itu sendiri, terutama pada data tabular dan teks.
3. Pada data citra, metode deep learning seperti CNN dapat secara otomatis mengekstrak fitur penting, namun untuk model klasik, rekayasa fitur masih diperlukan.
4. Hasil akhir sangat dipengaruhi oleh pemahaman domain dan kreativitas dalam merancang fitur baru yang representatif.

## Tantangan dan *Best Practices*

### 1. *Overfitting* Karena Rekayasa Fitur Berlebihan

*Overfitting* (Orlando & Bueno, 2025) terjadi ketika model terlalu kompleks atau terlalu disesuaikan dengan data pelatihan. Dalam konteks rekayasa fitur, hal ini bisa muncul saat fitur baru terlalu spesifik terhadap pola yang hanya ada di data pelatihan, tetapi tidak umum di data nyata.

Contohnya, membuat fitur kombinasi dari beberapa kolom tanpa dasar statistik atau domain yang jelas.

*Best practice:*

- a. Lakukan validasi silang (*cross-validation*) untuk menguji keandalan fitur baru.
- b. Gunakan regularisasi pada model untuk mengurangi kompleksitas.
- c. Hindari fitur yang mengandung kebisingan tinggi tanpa kontribusi signifikan terhadap akurasi.

### 2. *Data Leakage* Akibat Fitur Informasi Masa Depan

*Data leakage* (Yunling et al., 2024) terjadi ketika fitur yang seharusnya tidak tersedia saat prediksi digunakan dalam pelatihan model, menyebabkan akurasi yang tidak realistis. Ini sering terjadi saat rekayasa fitur mengakses data masa depan (*future information*), seperti label target atau data hasil dari peristiwa di masa depan.

Contoh: Menggunakan saldo akhir bulan untuk memprediksi apakah seseorang akan gagal bayar di bulan tersebut. *Best practice*:

- a. Memisahkan data pelatihan dan pengujian secara ketat berdasarkan waktu (jika *applicable*).
- b. Menggabungkan *dataset* dengan detail; memastikan tidak ada informasi yang salah.
- c. Mengaudit setiap fitur baru untuk mengecek kemungkinan adanya kebocoran informasi.

### 3. **Reproducibility** Dalam Proses Rekayasa Fitur

*Reproducibility* (Ameli et al., 2024) adalah kemampuan untuk mereplikasi proses rekayasa fitur dan mendapatkan hasil yang sama di waktu lain atau oleh tim lain.

Tantangan muncul saat proses tidak terdokumentasi dengan baik atau menggunakan teknik manual dan tidak terotomatisasi.

*Best practice*:

- a. Gunakan *pipeline* otomatis (misalnya dengan ``sklearn.pipeline``, ``Featuretools``, atau ``mlflow``).
- b. Dokumentasikan setiap transformasi fitur secara detail.
- c. Simpan kode dan konfigurasi versi yang digunakan selama proses pelatihan.

### 4. **Feature Drift** Dalam Sistem Produksi

*Feature drift* (Shyaa et al., 2024) mengacu pada perubahan distribusi fitur dari waktu ke waktu antara data pelatihan dan data produksi. Hal ini dapat menyebabkan penurunan performa model karena model belajar dari pola lama yang tidak lagi relevan.

Contoh: Fitur "jumlah kunjungan ke situs" yang menurun drastis karena perubahan perilaku pengguna. *Best practice*:

- a. Memantau distribusi fitur secara berkala dalam sistem produksi.
- b. Menggunakan alat pemantauan seperti *Evidently*, *WhyLogs*, atau *Vertex AI Monitoring*.
- c. Melakukan *retraining* atau *update* model secara periodik saat *drift* terdeteksi.

## Implementasi Model:

### 1. Regresi

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.datasets import fetch_california_housing
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.ensemble import RandomForestRegressor
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
import joblib

# 1. Load data
data = fetch_california_housing()
X = pd.DataFrame(data.data, columns=data.feature_names)
y = pd.Series(data.target)

# Visualisasi korelasi antar fitur
plt.figure(figsize=(10, 8))
sns.heatmap(X.corr(), annot=True, cmap="YlGnBu", fmt=".2f")
plt.title("Korelasi antar fitur")
plt.tight_layout()
plt.show()

# Split data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
# Standardisasi
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

# 2. Model
model = RandomForestRegressor(random_state=42)

# 3. Hyperparameter tuning
param_grid = {
    'n_estimators': [50, 100],
    'max_depth': [None, 10],
    'min_samples_split': [2, 5],
    'min_samples_leaf': [1, 2]}

grid_search = GridSearchCV(model, param_grid,
cv=3, n_jobs=-1, verbose=1)
grid_search.fit(X_train_scaled, y_train)

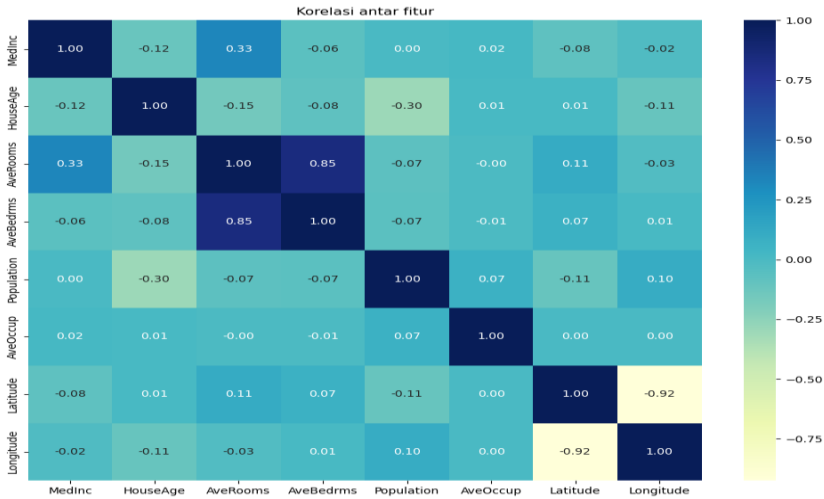
# 4. Model terbaik
best_model = grid_search.best_estimator_
print("Best Parameters:",
grid_search.best_params_)

# 5. Simpan model
joblib.dump(best_model, 'best_regressor.pkl')
print("Model terbaik disimpan sebagai
'best_regressor.pkl'")

# 6. Prediksi
y_pred = best_model.predict(X_test_scaled)

# 7. Evaluasi
mse = mean_squared_error(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
```

```
print("\nEvaluasi Model Regresi:")
print(f"Mean Squared Error (MSE): {mse:.3f}")
print(f"Mean Absolute Error (MAE): {mae:.3f}")
print(f"R-squared (R2): {r2:.3f}")
```



## 2. Klasifikasi *Machine Learning*

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.datasets import load_breast_cancer
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.ensemble import RandomForestClassifier
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
import joblib
```

# 1. Load dan visualisasi data

```

data = load_breast_cancer()
X      =      pd.DataFrame(data.data,
columns=data.feature_names)
y = pd.Series(data.target)

# Korelasi antar fitur
plt.figure(figsize=(12, 10))
sns.heatmap(X.corr(),      cmap="coolwarm",
linewidths=0.5)
plt.title("Korelasi antar fitur")
plt.tight_layout()
plt.show()

# 2. Split dan Standardisasi
X_train, X_test, y_train, y_test =
train_test_split(X, y, test_size=0.2,
random_state=42)

scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

# 3. Model
model = RandomForestClassifier(random_state=42)

# 4. Grid Search (Pencarian Hyperparameter
Terbaik)
param_grid = {
    'n_estimators': [50, 100],
    'max_depth': [None, 5, 10],
    'min_samples_split': [2, 4],
    'min_samples_leaf': [1, 2]}
grid_search = GridSearchCV(model, param_grid,
cv=5, n_jobs=-1, verbose=1)
grid_search.fit(X_train_scaled, y_train)

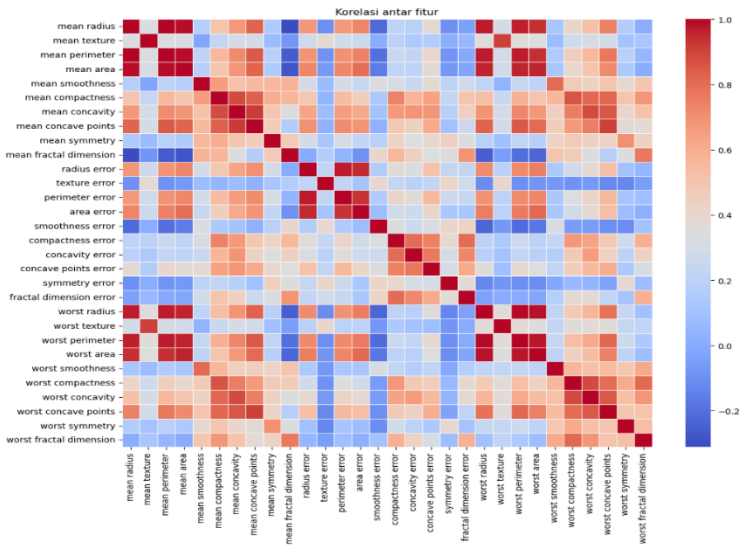
print("Best      Parameters:",
grid_search.best_params_)

```

```
# 5. Simpan model terbaik
best_model = grid_search.best_estimator_
joblib.dump(best_model, "best_classifier.pkl")
print("Model terbaik disimpan sebagai
'best_classifier.pkl'")

# 6. Prediksi
y_pred = best_model.predict(X_test_scaled)

# 7. Evaluasi performa
print("\nClassification Report:\n",
classification_report(y_test, y_pred))
print("Confusion Matrix:\n",
confusion_matrix(y_test, y_pred))
print("Accuracy:", accuracy_score(y_test,
y_pred))
```



### 3. Klasifikasi *Deep Learning* (*Deep Neural Network*)

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```

from sklearn.datasets import load_breast_cancer
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import classification_report, confusion_matrix

import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Dropout

# 1. Load dataset dan visualisasi korelasi fitur
data = load_breast_cancer()
X = pd.DataFrame(data.data, columns=data.feature_names)
y = pd.Series(data.target)

# Korelasi antar fitur
plt.figure(figsize=(12, 10))
sns.heatmap(X.corr(), cmap='coolwarm', linewidths=0.5)
plt.title("Korelasi antar fitur")
plt.tight_layout()
plt.show()

# 2. Split dan standarisasi data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)

```

```

X_test_scaled = scaler.transform(X_test)

# 3. Bangun model deep learning (Keras)
model = Sequential([
    Dense(32, activation='relu',
input_shape=(X_train_scaled.shape[1],)),
    Dropout(0.2),
    Dense(16, activation='relu'),
    Dense(1, activation='sigmoid') # Output
layer untuk klasifikasi biner
])
model.compile(optimizer='adam',
loss='binary_crossentropy',
metrics=['accuracy'])

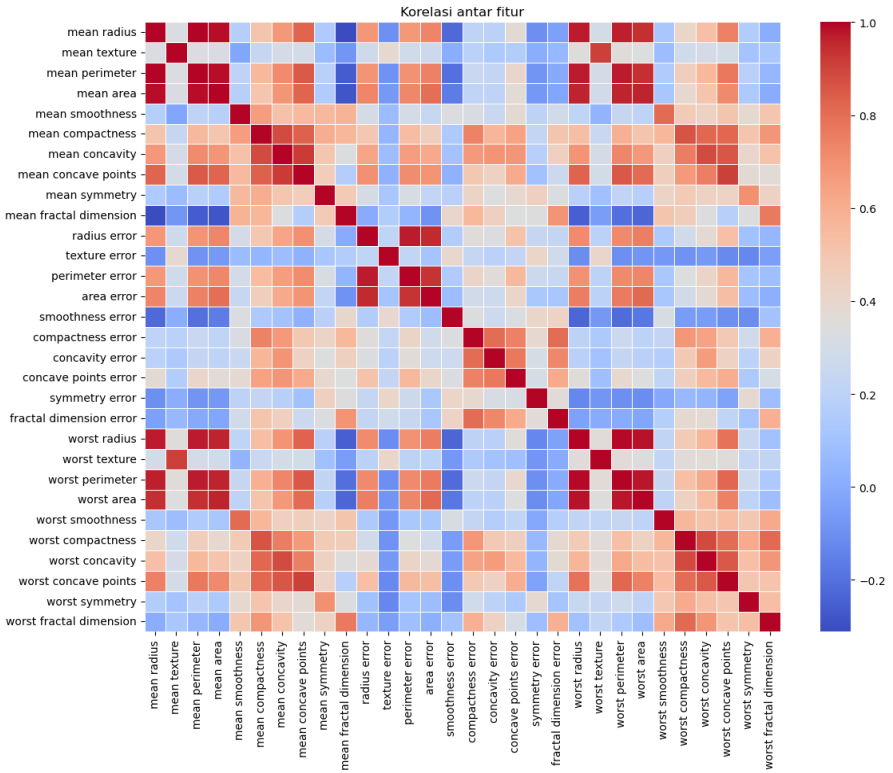
# 4. Training model
history = model.fit(X_train_scaled, y_train,
epochs=50, batch_size=16, validation_split=0.1,
verbose=0)

# 5. Prediksi
y_pred_prob = model.predict(X_test_scaled)
y_pred = (y_pred_prob >
0.5).astype(int).flatten()

# 6. Evaluasi performa
print("Classification Report:\n")
print(classification_report(y_test, y_pred,
target_names=data.target_names))

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

```



4/4 ————— 0s 14ms/step

Classification Report:

	precision	recall	f1-score	support
malignant	0.98	0.95	0.96	43
benign	0.97	0.99	0.98	71
accuracy			0.97	114
macro avg	0.97	0.97	0.97	114
weighted avg	0.97	0.97	0.97	114

Confusion Matrix:

```
[[41  2]
 [ 1 70]]
```

## Daftar Pustaka

- Ameli, A., Peña-Castillo, L., & Usefi, H. (2024). Assessing The Reproducibility Of Machine-Learning-Based Biomarker Discovery In Parkinson's Disease. *Computers in Biology and Medicine*, 174(March).  
<https://doi.org/10.1016/j.compbimed.2024.108407>.
- Barbieri, M. C., Grisci, B. I., & Dorn, M. (2024). Analysis And Comparison of Feature Selection Methods Towards Performance And Stability. *Expert Systems with Applications*, 249, 123667.  
<https://doi.org/https://doi.org/10.1016/j.eswa.2024.123667>.
- Boeschoten, S., Catal, C., Tekinerdogan, B., Lommen, A., & Blokland, M. (2023). The Automation of The Development of Classification Models And Improvement of Model Quality Using Feature Engineering Techniques. *Expert Systems with Applications*, 213(September 2022).  
<https://doi.org/10.1016/j.eswa.2022.118912>.
- Chaudhuri, A. (2024). Search Space Division Method For Wrapper Feature Selection on High-Dimensional Data Classification. *Knowledge-Based Systems*, 291, 111578.  
<https://doi.org/https://doi.org/10.1016/j.knosys.2024.111578>.
- Chen, H., Yue, W., Bin, G., Jiang, Q., Shao, W., & She, C. (2024). Filter Methods Comparation For Incremental Capacity Analysis In Lithium-Ion Batteries Health Prediction. *Journal of Energy Storage*, 101, 113878.  
<https://doi.org/https://doi.org/10.1016/j.est.2024.113878>.
- Ding, X., Cui, M., & Wang, K. (2025). Automatic Ensemble Feature Selection For Multi-View Data. *Neurocomputing*, 645, 130437.  
<https://doi.org/https://doi.org/10.1016/j.neucom.2025.130437>.
- Gharaba, S., Shalem, A., Paz, O., Mughtar, N., Wolf, L., & Weil, M. (2024). Aberrant Migration Features In Primary Skin Fibroblasts Of Huntington's Disease Patients Hold Potential For Unraveling Disease Progression Using An Image Based Machine Learning Tool. *Computers in Biology and Medicine*, 180, 108970.  
<https://doi.org/https://doi.org/10.1016/j.compbimed.2024.108970>.
- Gumilang, P. B., Listyorini, S., & Ngatno, N. (2022). Pengaruh Perceived Quality dan Brand Experience Terhadap Keputusan Pembelian

- Motor Yamaha Mio (Studi pada Yamaha Flagship Shop (FSS) Kota Semarang). *Jurnal Ilmu Administrasi Bisnis*, 11(3), 420–432. <https://doi.org/10.14710/jiab.2022.34919>.
- HAQUE, A. (2022). *Feature Engineering & Selection For Explainable Models: A Second Course For Data Scientists*. LULU COM.
- Moslemi, A. (2023). A Tutorial-Based Survey on Feature Selection: Recent Advancements on Feature Selection. *Engineering Applications of Artificial Intelligence*, 126, 107136. <https://doi.org/https://doi.org/10.1016/j.engappai.2023.10713>.
- Mujeeb, A., Dai, W., Erdt, M., & Sourin, A. (2019). One Class Based Feature Learning Approach For Defect Detection Using Deep autoencoders. *Advanced Engineering Informatics*, 42, 100933. <https://doi.org/https://doi.org/10.1016/j.aei.2019.100933>.
- Orlando, C., & Bueno, P. (2025). *Journal Pre-proof*. <https://doi.org/10.1016/j.jacl.2025.04.197>.
- Qiu, S., He, J., Wang, Y., & E, B. (2025). A Feature Selection Method for Software Defect Prediction Based on Improved Beluga Whale Optimization Algorithm. *Computers, Materials and Continua*, 83(3), 4879–4898. <https://doi.org/https://doi.org/10.32604/cmc.2025.061532>.
- Shyaa, M. A., Ibrahim, N. F., Zainol, Z., Abdullah, R., Anbar, M., & Alzubaidi, L. (2024). Evolving Cybersecurity Frontiers: A Comprehensive Survey on Concept Drift And Feature Dynamics Aware Machine And Deep Learning In Intrusion Detection Systems. *Engineering Applications of Artificial Intelligence*, 137(January). <https://doi.org/10.1016/j.engappai.2024.109143>.
- Sreekumari, S., Bhalla, R., & Singh, G. (2025). Feature Selection and Model Evaluation for Heart Disease Prediction Using Ensemble Methods. *Procedia Computer Science*, 259, 1282–1295. <https://doi.org/10.1016/j.procs.2025.04.083>.
- Taha, Z. Y. (n.d.). *Optimizing Feature Selection with Genetic Algorithms: A Review of Methods and Applications*.
- Team, M. (2025). *Scikit-Learn*. Yayasan Kita Menulis. <https://scikit-learn.org/stable/index.html>.
- Wang, J., Deng, J., & Liu, D. (2025). Deep Prior Embedding Method For Electrical Impedance Tomography. *Neural Networks*, 188, 107419.

- <https://doi.org/https://doi.org/10.1016/j.neunet.2025.107419>.
- Wei, K., Zhao, R., Kou, H., Chen, P., Cao, Y., Zheng, Y., & Deng, L. (2025). Dimensionality Reduction Of Rolling Bearing Fault Data Based On Graph-Embedded Semi-Supervised Deep Auto-Encoders. *Engineering Applications of Artificial Intelligence*, 152, 110689. <https://doi.org/https://doi.org/10.1016/j.engappai.2025.11068>.
- Winter, C. De, FrasinCAR, F., Peuter, B. De, Matsiako, V., & Ido, E. (2025). *Knowledge-Based Systems Automated Feature Engineering For Automated Machine Learning*. 321(February).
- Wu, Y., & Huang, Z. (2024). Feature Selection Considering Feature Relevance, Redundancy And Interactivity For Neighbourhood Decision Systems. *Neurocomputing*, 596, 128092. <https://doi.org/https://doi.org/10.1016/j.neucom.2024.128092>.
- Xu, J., Liu, X., Gu, Z., & Xiao, G. (2025). A Rapid Cross-Validation Computing For Three-Way Decisions In Imbalanced Data. *Information Sciences*, 707(February). <https://doi.org/10.1016/j.ins.2025.122016>.
- Yunling, W., Gao, C., Huang, Y., Fu, L., & Yu, Y. (2024). Less Leakage and More Precise: Efficient Wildcard Keyword Search Over Encrypted Data. *JournalPre-Proof*, 1-11. <https://doi.org/10.1016/j.hcc.2025.100297>.
- Zhang, X., Fang, S., Jin, Y., Huang, Y., Wang, S., Wang, J., & Xu, Y. (2025). Comprehensive Performance Evaluation of Valuable Medical Equipment Based On Cloud Modelling And Combined Weighting Methodologies. *Engineering Applications of Artificial Intelligence*, 154, 110969. <https://doi.org/https://doi.org/10.1016/j.engappai.2025.11096>.
- Zhang, Z., Li, Y., Huang, J., Yu, Y., Qin, R., Su, Y., Wen, G., Cheng, W., & Chen, X. (2025). Class-Imbalanced Pattern Recognition In Pipeline Weld Cracks Damage Via Feature Characterization And Sample Enhancement. *Measurement*, 253, 117558. <https://doi.org/https://doi.org/10.1016/j.measurement.2025.117558>.
- Zito, F., Talbi, E. G., Cavallaro, C., Cutello, V., & Pavone, M. (2025). Metaheuristics In Automated Machine Learning: Strategies for Optimization. *Intelligent Systems With Applications*, 26(September 2024). <https://doi.org/10.1016/j.iswa.2025.200532>.

## PROFIL PENULIS




### **Indo Intan, S.T., M.T.**

Penulis menyelesaikan S1 di Prodi Teknik Telekomunikasi Jurusan Teknik Elektro tahun 2002 silam di Universitas Hasanuddin. Selanjutnya pada jenjang S2 penulis berhasil mempertahankan tesis di bidang Kecerdasan Buatan khususnya Image Processing dan Pengenalan Pola. Terdaftar sebagai dosen di Universitas Dipa Makassar sejak tahun 2005 pada Prodi Informatika. Selain mengajar, penulis aktif melakukan kegiatan penelitian maupun pengabdian masyarakat. Selama beberapa kali memperoleh dana hibah pada penelitian dosen pemula, penelitian fundamental reguler, penelitian terapan, dan pengabdian masyarakat.

Buku yang penulis pernah diterbitkan di antaranya: Statistik, Teman Belajar dan Analisis Mahasiswa (2019); Dasar-dasar Pembelajaran Mesin (*Foundations of Machine Learning*) (2023); *Machine Learning: Konsep, Algoritma, dan Implementasi* (2024); *Deep Learning: Teori, Algoritma, dan Aplikasi* (2025); *Machine Learning: Teori, Model, dan Implementasi Sebuah Pendekatan Komprehensif* (2025). Selama beberapa tahun ini kajian riset berfokus pada analisis citra medis yang bermitra dengan rumah sakit demi menghadirkan inovasi yang bisa mengeksplorasi potensi antara dunia kedokteran dengan dunia informatika basis ilmu komputer. *From zero to hero*, kalimat ini yang masih membutuhkan pembuktian untuk sebuah pencapaian yang optimal. Penulis terus belajar, meriset, dan menulis agar bisa memberikan kontribusi bermanfaat bagi yang memerlukannya. Semoga buku ini bisa membuka wawasan kita bersama, bahwa eksplorasi ilmu melalui buku akan melahirkan “energi baru” untuk menjadi bekal mengajar, meriset, dan memberikan kontribusi di tengah masyarakat sebagai jariah pahala.

Email Penulis: [indo.intan@undipa.ac.id](mailto:indo.intan@undipa.ac.id). *Youtube: Indo Intan Channel - YouTube.*



**BAB 5**  
**PEMODELAN LAYANAN**  
**VISUALISASI DATA AWAN**  
**UNTUK PEMANTAUAN**  
**DATA MULTISENSOR**

---

**Wiranti Sri Utami, S.Kom., M.T.I.**  
Universitas Cendekia Abditama



buatan untuk otomatisasi pengenalan pola dan anomali dalam data, sehingga membantu pengguna fokus pada informasi penting (Zhou et al., 2023). Pemahaman konsep dasar ini menjadi pondasi penting sebelum membahas implementasi layanan visualisasi data multisensor berbasis *cloud* yang akan dijelaskan pada bagian berikutnya.

## **Komponen Utama Arsitektur Berbasis *Cloud***

Arsitektur layanan visualisasi data berbasis *cloud* dirancang untuk mengelola aliran data yang besar dan kompleks dari berbagai sensor secara efisien, serta menyediakan tampilan visual yang mudah dipahami oleh pengguna.

Arsitektur ini umumnya terdiri dari beberapa lapisan (*layers*) yang saling berinteraksi untuk mendukung proses pengumpulan, penyimpanan, pengolahan, analisis, hingga penyajian suatu data (He et al., 2020).

Arsitektur berbasis *cloud* memiliki komponen utama untuk yang meliputi beberapa elemen penting yang dapat saling bekerja sama dalam menyediakan layanan komputasi berbasis *cloud*. Berikut adalah komponen utama arsitektur berbasis *cloud*:

### **1. Lapisan Sensor dan Perangkat *Edge* (*Edge Layer*)**

Pada lapisan sensor dan perangkat *edge*, tugas utamanya adalah mengumpulkan data dari berbagai sensor seperti suhu, tekanan, kelembapan, dan lainnya.

Dalam beberapa situasi, teknologi *edge computing* dimanfaatkan untuk melakukan pra-pemrosesan data sebelum data tersebut dikirim ke *cloud*, dengan tujuan yaitu untuk mempercepat waktu respons antara permintaan (*request*) dan tanggapan (*response*) dalam sistem *digital*, serta mengurangi beban jaringan pada layanan *cloud* (Zhou et al., 2020).

### **2. Lapisan *Transport* dan *Gateway***

Pada lapisan *transport*, dapat menggunakan protokol komunikasi seperti MQTT, HTTP, atau CoAP untuk mengirimkan data dari perangkat *edge* ke *cloud*. Kemudian lapisan *gateway* dapat melakukan penyaringan atau kompresi suatu data.

Studi kasus yang diangkat dalam bab ini menunjukkan bahwa implementasi layanan visualisasi data berbasis *cloud* bukan hanya memungkinkan pemantauan lingkungan secara *real-time*, namun dapat memberikan landasan untuk integrasi teknologi lanjutan seperti kecerdasan buatan (AI) dan pembelajaran mesin (ML). Hasil evaluasi sistem menunjukkan bahwa solusi ini mampu memberikan kinerja yang baik, skalabilitas tinggi, serta keamanan yang memadai dalam pengelolaan data sensitif.

Meskipun masih terdapat beberapa tantangan, seperti keterbatasan konektivitas jaringan dan kompleksitas dalam integrasi teknologi kecerdasan buatan, pendekatan ini tetap memiliki peluang yang sangat baik dalam mengembangkan sistem pemantauan cerdas di masa depan. Dengan pemilihan *platform cloud* dan *tools* visualisasi yang sesuai, serta perencanaan arsitektur sistem yang matang, organisasi dapat membangun solusi pemantauan berbasis data yang akurat, informatif, dan mudah diakses oleh berbagai pihak.

Secara keseluruhan, pemodelan layanan visualisasi data berbasis *cloud* untuk pemantauan data multisensor menjadi pendekatan strategis dalam mendukung pengambilan keputusan berbasis data dan pengelolaan sumber daya yang lebih efisien di era *digital*.

## Daftar Pustaka

- Alkadi, A., Almekhlafi, A., & Alzahrani, A. (2022). Scalable And Secure Architecture For IoT Data Visualization In Cloud Environments. *Sensors*, 22(8), 2815. <https://doi.org/10.3390/s22082815>.
- Almeida, L., Rodrigues, J., & Ferreira, P. (2021). Interactive Dashboards For Real-Time Multisensor Data Monitoring: A Review. *Sensors*, 21(4), 1234. <https://doi.org/10.3390/s21041234>.
- Banerjee, M., & De, D. (2021). IoT Sensor Data Storage And Visualization Using TimescaleDB. *Internet of Things and Cyber Physical Systems*, 1, 100001. <https://doi.org/10.1016/j.iotcps.2021.100001>.
- Ghorbel, A., Jemai, A., & Kachouri, A. (2020). IoT Security Using Lightweight Cryptographic Algorithms For Healthcare Applications. *Procedia Computer Science*, 175, 204–209. <https://doi.org/10.1016/j.procs.2020.07.029>.
- Gonzalez, R., Lim, Y. J., & Nguyen, T. (2023). Building Scalable Cloud-Based IoT Dashboards For Sensor Data Visualization. *Sensors*, 23(4), 1781. <https://doi.org/10.3390/s23041781>.
- He, Y., Chen, C., & Liu, Z. (2020). A Layered Architecture For Real-Time Iot Data Analytics And Visualization In Cloud Environment. *Future Generation Computer Systems*, 108, 533–545. <https://doi.org/10.1016/j.future.2020.02.039>.
- Iqbal, M., ur Rehman, M. H., & Khan, S. U. (2021). Cloud Data Storage And Management For IoT: A Systematic Review. *Journal of Cloud Computing*, 10(1), 21. <https://doi.org/10.1186/s13677-021-00235-5>.
- Kaur, R., & Awasthi, R. (2021). Cloud-based IoT Architecture For Real-Time Applications. *Journal of Ambient Intelligence and Humanized Computing*, 12(9), 9191–9202. <https://doi.org/10.1007/s12652-020-02737-3>
- Khan, S., Ahmed, S., & Rehman, S. (2022). Real-Time Data Visualization Techniques For Streaming Sensor Data: A Comprehensive Review.

- IEEE Access*, *10*, 45678–45692.  
<https://doi.org/10.1109/ACCESS.2022.3178379>.
- Kumar, R., Sharma, A., & Singh, P. (2021). Comparative Analysis Of Cloud Platforms For IoT Applications: AWS, Azure and GCP. *International Journal of Cloud Applications and Computing*, *11*(3), 1–14. <https://doi.org/10.4018/IJCAC.2021070101>.
- Liu, X., Yang, J., Chen, M., & Zhang, L. (2021). Cloud-Based Multisensor Data Visualization For Urban Air Quality Monitoring. *IEEE Internet of Things Journal*, *8*(10), 7890–7901. <https://doi.org/10.1109/JIOT.2021.3065789>.
- Patel, D., & Desai, A. (2021). Real-Time Data Visualization Using Grafana and MQTT in IoT Systems. *International Journal of Advanced Computer Science and Applications*, *12*(7), 401–407. <https://doi.org/10.14569/IJACSA.2021.0120751>.
- Patel, D., & Singh, R. (2020). Spatial-Temporal Visualization of IoT Sensor Data For Smart City Applications. *Journal of Urban Technology*, *27*(2), 45–63. <https://doi.org/10.1080/10630732.2020.1750498>.
- Rahman, M. A., Rahman, M. M., & Hossain, M. A. (2023). A Review of Iot Middleware Platforms For Smart Applications. *Journal of Sensor and Actuator Networks*, *12*(1), 10. <https://doi.org/10.3390/jsan12010010>.
- Singh, A., Kumar, N., & Kaur, J. (2021). Challenges And Solutions In Multisensor Data Integration And Visualization. *International Journal of Computer Applications*, *175*(5), 12–20. <https://doi.org/10.5120/ijca2021921545>.
- Sun, L., Wang, Q., & Zhou, J. (2022). Cloud-Based Big Data Analytics Framework For IoT Sensor Streams. *Information Fusion*, *83*, 1–13. <https://doi.org/10.1016/j.inffus.2022.02.001>.
- Wang, H., Chen, Y., & Li, K. (2022). IoT Data Processing And Visualization Using AWS IoT Core And Lambda For Smart City

- Applications. *Sustainable Cities and Society*, 75, 103300. <https://doi.org/10.1016/j.scs.2021.103300>.
- Wang, H., Chen, Y., & Li, M. (2020). Data Fusion And Visualization In Heterogeneous Sensor Networks: A Survey. *IEEE Sensors Journal*, 20(14), 7790–7802. <https://doi.org/10.1109/JSEN.2020.2984123>.
- Wang, Y., Zhang, T., & Chen, J. (2021). Secure And Intelligent Cloud Computing For IoT-Based Smart Systems: A Survey. *IEEE Internet of Things Journal*, 8(4), 2345–2366. <https://doi.org/10.1109/JIOT.2020.3017196>.
- Zhang, Q., Huang, Y., & Xu, B. (2023). Real-Time Air Pollution Monitoring And Prediction Based On Cloud Computing And Machine Learning. *Environmental Modelling & Software*, 158, 105488. <https://doi.org/10.1016/j.envsoft.2022.105488>.
- Zhao, Y., Li, Y., Wu, J., & Zhou, M. (2022). Lightweight Communication Protocols For The Internet Of Things: A Survey. *IEEE Communications Surveys & Tutorials*, 24(2), 1010–1040. <https://doi.org/10.1109/COMST.2022.3151380>.
- Zhou, J., Li, F., & Zhang, R. (2020). Edge-Cloud Collaborative Architecture For Real-Time Multisensor Data Visualization. *IEEE Access*, 8, 93527–93538. <https://doi.org/10.1109/ACCESS.2020.2995421>.

## PROFIL PENULIS



### **Wiranti Sri Utami, S.Kom., M.T.I.**

Minat penulis terhadap dunia komputer dimulai sejak tahun 2005 ketika menempuh pendidikan di SMAN 6 Kota Tangerang pada jurusan IPA. Pada tahun 2008, penulis melanjutkan studi di program Professional IT CEP-CCIT FTUI dengan fokus konsentrasi pada *Software Engineering* dan berhasil menyelesaikan program tersebut pada tahun 2010. Selanjutnya, pada tahun 2026 penulis meraih gelar Sarjana Teknik Informatika dari STMIK Raharja dan kemudian melanjutkan pendidikan Magister Teknik Informatika di Universitas Raharja dengan konsentrasi *Business Intelligence*. Saat ini, penulis mengajar beberapa mata kuliah seperti Algoritma dan Pemrograman, Pemrograman Visual, Sistem Berbasis Pengetahuan, *Wireless/Mobile Computing*, serta Rekayasa Perangkat Lunak (RPL). Sebagai bagian dari pengembangan karir sebagai dosen, penulis aktif menulis artikel yang diterbitkan di jurnal nasional serta menyusun buku, dengan tujuan agar pengetahuan yang dimiliki penulis dapat memberikan manfaat luas sesuai dengan bidang keahlian.

Email Penulis: [whiranty68@gmail.com](mailto:whiranty68@gmail.com).



# **BAB 6**

# **PENGOLAHAN DAN**

# **VISUALISASI DATA**

---

**Hamzah Nurrifqi Fakhri F., S.Kom., M.Kom.**  
Universitas Garut



## Pendahuluan

Setiap harinya, data dihasilkan dari berbagai sumber, seperti kegiatan transaksi bisnis, interaksi media sosial, perangkat elektronik dan/atau sensor perangkat pintar. Fenomena ini menghasilkan volume data yang menumpuk dari waktu ke waktu. Dalam laporan yang berjudul *Data Age 2025: The Digitization of the World from Edge to Core*, diprediksi bahwa pertumbuhan data pada tahun 2025 akan mencapai 175 *zettabyte* atau setara dengan 175 triliun *terabyte*.

Jumlah tersebut lebih besar 75% dibandingkan total populasi manusia di bumi, setidaknya setiap satu orang akan menghasilkan satu data setiap 18 detik (Reinsel et al., 2018). Dengan jumlah data yang terus meningkat, tantangan utama adalah bagaimana mengolah dan memanfaatkan data tersebut secara efektif agar dapat menghasilkan wawasan yang bernilai.

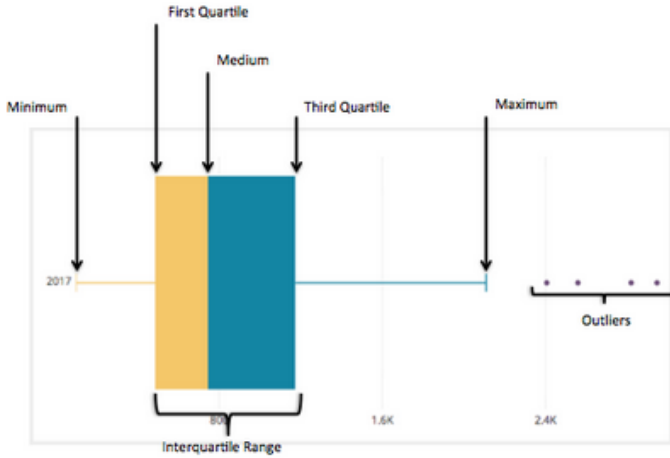
Untuk menjawab tantangan tersebut, diperlukan metode untuk mengekstraksi data menjadi informasi yang bermakna dan wawasan yang dapat ditindaklanjuti. Keputusan yang diambil dari hasil ekstraksi data dapat menjadi landasan penting dalam berbagai bidang termasuk strategi bisnis, ilmu pengetahuan, kebijakan, hingga pendidikan.

Penguasaan teknik pengolahan dan visualisasi data menjadi aspek yang perlu dipelajari oleh individu maupun organisasi agar data dapat dimanfaatkan secara optimal. Pengolahan dan visualisasi data merupakan dua hal yang berkaitan dalam alur kerja analisis data. Pengolahan data bertugas untuk membersihkan, mengubah, dan mengintegrasikan data mentah hingga tahap analisis (Santoso, 2019).

Sementara itu, visualisasi data memungkinkan penyajian informasi dari hasil analisis data disajikan dalam format visual untuk mempermudah pemahaman terhadap data. Selain itu, visualisasi data dapat membantu mengidentifikasi tren dan pola yang ada dalam data (Muharni & Candra, 2022).

## Konsep Dasar Pengolahan Data

Di era *digital* saat ini, kemajuan teknologi berbasis komputer telah menghasilkan data dalam jumlah yang sangat besar. Namun, data mentah tersebut belum dapat langsung dimanfaatkan karena masih



**Gambar 6.9: Contoh *Box Plot***

Sumber: Monga, 2018.

Setiap jenis visualisasi memiliki keunggulan masing-masing tergantung pada konteks data dan pesan apa yang hendak disampaikan. Pemilihan jenis visualisasi yang tepat sangat penting agar informasi yang disajikan mudah dipahami dan mampu mendukung pengambilan keputusan yang efektif.

### **Kesimpulan**

Pengolahan dan visualisasi data merupakan dua aktivitas penting untuk mengoptimalkan pemanfaatan data di era *digital*. Pengolahan data melibatkan serangkaian teknik untuk membersihkan, mentransformasi, dan mengintegrasikan data mentah sehingga berubah menjadi informasi yang bermakna. Sementara itu, visualisasi data mengubah informasi tersebut menjadi format grafis yang mudah dipahami, memudahkan identifikasi pola, tren, dan *insight* yang tersembunyi.

## Daftar Pustaka

- Agung, A., & Yuesti, A. (2017). *Metodologi Penelitian Kuantitatif dan Kualitatif (Vol. 11, Issue 1)*. AB Publisher.
- Amna, Wahyudin, S., Putra, T. A. E., Wahidin, A. J., Syukrilla, W. A., Wardhani, A. K., Heryana, N., Indiyani, T., & Santoso, L. W. (2023). *Data Mining* (D. Ediana & A. Yanto (eds.); 1st ed.). PT. Global Eksekutif Teknologi.
- Anggarawal, C. C. (2015). *Data Mining: The Text Book*. Springer International Publishing Switzerland.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37–53.
- Monga, R. (2018). *Most important question while working on Data Visualization-Which Graph/Chart Type Should I Use?* <https://www.linkedin.com/pulse/most-important-question-while-working-data-which-graphchart-monga>.
- Muharni, S., & Candra, A. (2022). *Visualisasi Data Menggunakan Data Studio* (A.-N. Zarkasyi & H. Almumtaza (eds.); 1). CV. Literasi Nusantara Abadi.
- Mukhlis, I. R., Hayam, U., Perbanas, W., Pipin, S. J., & Mikroskil, U. (2024). *BIG DATA (Mengenal Big Data & Implementasinya di Berbagai Bidang)* (Sepriano & Y. Agusdi (eds.); 1st ed.). Sonpedia Publishing Indonesia.
- Open Data Jabar. (2021). *Jumlah Kelahiran Bayi Berdasarkan Status Kelahiran dan Jenis Kelamin di Jawa Barat*. Open Data Jabar. <https://opendata.jabarprov.go.id/id/dataset/jumlah-kelahiran-bayi-berdasarkan-status-kelahiran-dan-jenis-kelamin-di-jawa-barat>.
- Reinsel, D., Gantz, J., & Rydning, J. (2018). *The Digitization of The World -From Edge to Core*.
- Santoso, J. T. (2019). *Kecerdasan Buatan (Artificial Intelligence)* (M. Sholikan (ed.); 1st ed., Vol. 1). Yayasan Prima Agus Teknik.

Santoso, J. T. (2020). *Analisis Big Data (J. T. Santoso (ed.))*. Yayasan Prima Agus Teknik.

Wali, M., Efitra, Sudipa, I., Heriyani, A., Hendriyani, C., Rahman, R., Santika, P. P., Indarto, S. L., Tanwir, Ibrahim, M. B., Iskandar, A., Kertati, I., Nainggolan, H., & Sepriano. (2023). *Penerapan & Implementasi Big Data di Berbagai Sektor (Efitra & A. Juansa (eds.))*; 1st ed., Vol. 1). PT. Sonpedia Publishing Indonesia.

Yuan, Y. (2010). Multiple Imputation For Missing Data: Concepts And New Development. *SAS Institute Inc, Rockville, MD*, 1–13.

## PROFIL PENULIS



### **Hamzah Nurriqfi Fakhri F., S.Kom., M.Kom.**

Lahir di Bandung pada 21 Desember 1996 dan memiliki ketertarikan di bidang ilmu komputer sejak menempuh pendidikan di SMK Ciledug Al-Musaddadiyah Garut. Lulus pada tahun 2015, saya memilih jurusan Teknik Komputer & Jaringan (TKJ) sebagai awal perjalanan akademik di dunia teknologi. Selanjutnya, saya melanjutkan studi S1 di Sekolah Tinggi Teknologi Garut (STTG) dengan program studi Teknik Informatika pada tahun 2016 dan menyelesaikan pendidikan tersebut pada pertengahan tahun 2020. Untuk memperdalam keahlian, saya melanjutkan studi pascasarjana di STMIK LIKMI Bandung pada tahun 2023 dengan fokus pada Sistem Informasi dan lulus pada awal tahun 2024. Dalam perjalanan karir, saya pernah berkontribusi sebagai Data Analis dan Pengelola Data di Dinas Sosial Kabupaten Garut dari Januari 2021 hingga Januari 2025. Selama bertugas, saya bertanggung jawab dalam pengelolaan aplikasi SIKS-NG. Selain itu, saya dipercaya secara langsung oleh pimpinan daerah untuk menganalisis dan menyiapkan data bagi calon penerima bantuan sosial. Saya juga aktif sebagai narasumber dalam berbagai bimbingan teknis bagi petugas pengelola data di tingkat desa hingga kecamatan. Saat ini, saya baru mulai berkarir sebagai dosen di sebuah perguruan tinggi swasta. Meskipun belum memiliki banyak pengalaman menulis, saya memiliki semangat tinggi untuk terus belajar, memperluas wawasan akademik, serta mendalami perkembangan di bidang analisis data, dan teknologi informasi.

Email Penulis: hamzah.nurriqfi2@gmail.com.



**BAB 7**  
***BIG DATA DAN FRAMEWORK***  
**PEMROSESAN DATA**

---

**Dr. Ir. Norbertus Tri Suswanto Saptadi, S.Kom., M.T., M.M., IPM.**  
Universitas Atma Jaya Makassar



## Pendahuluan *Big Data*

*Big Data* merujuk pada kumpulan data yang sangat besar, kompleks, dan tumbuh dengan cepat, sehingga sulit dikelola menggunakan alat manajemen data tradisional (Virgiawan, Absharina and Informasi, 2025). Konsep *Big Data* mencakup karakteristik utama yang dikenal sebagai 5V, yaitu: *Volume* (jumlah data yang sangat besar), *Velocity* (kecepatan data dihasilkan dan diproses), *Variety* (keragaman format data), *Veracity* (tingkat keakuratan data), dan *Value* (nilai yang bisa diekstrak dari data).

Melalui kemajuan teknologi *digital*, data kini dihasilkan dari berbagai sumber seperti media sosial, perangkat IoT, sistem transaksi bisnis, dan layanan *digital* lainnya. Sejarah perkembangan teknologi *Big Data* berakar dari kemajuan teknologi komputasi dan internet. Di awal tahun 2000-an, perusahaan teknologi seperti *Google* dan *Yahoo* mulai berupaya mengembangkan infrastruktur untuk mengelola data dalam skala besar, yang kemudian melahirkan teknologi seperti *MapReduce* dan *Hadoop* (Siregar and Harahap, 2025).

Perkembangan *cloud computing* dan penyimpanan terdistribusi turut serta mendorong kemampuan organisasi untuk menyimpan dan menganalisis data dalam jumlah besar tanpa harus berinvestasi besar pada perangkat keras fisik. Di era *digital* saat ini, *Big Data* memainkan peran penting dalam pengambilan keputusan berbasis data (*data-driven decision making*). Organisasi di berbagai sektor, mulai dari kesehatan, keuangan, pendidikan, hingga pemerintahan berupaya memanfaatkan *Big Data* untuk menggali pola tersembunyi, memperkirakan tren masa depan, dan meningkatkan efisiensi dalam operasional (Yuli *et al.*, 2024).

Kemampuan untuk memproses dan menganalisis data dalam skala besar memberikan keunggulan kompetitif bagi banyak perusahaan dan membuka peluang inovasi baru di berbagai bidang. Tantangan dalam pengelolaan *big data* tidak dapat diabaikan. Tantangan meliputi kebutuhan infrastruktur memadai, keterampilan teknis tenaga kerja, keamanan data, serta kepatuhan terhadap regulasi privasi. Pemahaman mendalam mengenai konsep dasar *Big Data*, teknologi pendukungnya, dan praktik terbaik implementasi sangat penting untuk memastikan bahwa data besar tidak hanya menjadi beban, tetapi juga menjadi aset strategis organisasi.

AI dan ML memungkinkan analisis yang lebih canggih dan keputusan berbasis data yang lebih akurat. Model-model yang dibangun dari data besar kini dapat memberikan prediksi yang lebih baik dan menyarankan tindakan secara otomatis. Semakin banyaknya data yang tersedia, metode pembelajaran mesin dapat terus berkembang untuk memberikan wawasan yang lebih dalam dan personalisasi yang lebih tinggi di berbagai sektor, dari *e-commerce* hingga kesehatan.

Dorongan kuat untuk mengadopsi *Edge Computing* dalam ekosistem *Big Data*. *Edge computing* memungkinkan pengolahan data lebih dekat dengan sumbernya, mengurangi latensi, dan menghemat *bandwidth* yang diperlukan untuk mentransfer data ke *cloud*. Hal ini sangat relevan dengan aplikasi IoT (*Internet of Things*) di mana data yang dihasilkan oleh perangkat pintar perlu diproses dengan cepat dan efisien, seperti pada kendaraan otonom dan sistem pengawasan cerdas. *Edge computing* diprediksi memainkan peran yang lebih besar dalam ekosistem *big data* yang semakin terdesentralisasi.

Masa depan *big data* dipengaruhi kemajuan dalam komputasi kuantum, yang berpotensi merevolusi cara memproses data besar. Dengan kemampuan komputasi yang jauh lebih tinggi dibandingkan komputer tradisional, komputasi kuantum mempercepat analisis data yang sangat kompleks, seperti dalam riset ilmiah atau simulasi molekuler. Meskipun teknologi masih dalam tahap awal, dampaknya terhadap *big data* di masa depan diperkirakan akan sangat besar, membuka kemungkinan baru dalam pemrosesan dan analitik data.

## Kesimpulan

*Big data* telah menjadi elemen penting dalam dunia teknologi informasi dan bisnis modern. *Volume*, kecepatan, dan keragaman data yang terus berkembang memberikan tantangan sekaligus peluang yang signifikan bagi organisasi untuk mengambil keputusan berbasis data yang lebih baik dan tentu lebih cepat.

Kemampuan untuk mengumpulkan, menyimpan, dan menganalisis data dalam jumlah besar ini telah memungkinkan perusahaan untuk memahami lebih dalam perilaku konsumen, meramalkan tren pasar, dan upaya mengoptimalkan operasi. *Big data* tidak hanya akan mengubah cara organisasi beroperasi, tetapi

menciptakan inovasi berbagai industri, dari kesehatan hingga manufaktur, dan bahkan sektor publik.

Salah satu aspek penting dalam memanfaatkan *big data* adalah adanya teknologi penyimpanan dan manajemen data yang mampu mendukung *volume* data yang sangat besar. Teknologi seperti *Hadoop*, *NoSQL databases*, dan penyimpanan berbasis *cloud* memungkinkan data untuk disimpan secara efisien dan diakses dengan cepat.

Namun, lebih dari itu, organisasi memiliki kerangka kerja pemrosesan data yang tepat, seperti *Apache Spark* dan *Apache Flink*, untuk mengelola aliran data yang cepat dan melakukan pemrosesan analitik secara *real-time*. Teknologi ini memungkinkan pengolahan data yang lebih cepat dan memberikan hasil yang dapat segera digunakan untuk pengambilan keputusan.

Integrasi antara *big data* dan *machine learning* semakin memperkuat kemampuan analitik organisasi. *Machine learning* memberikan kekuatan bagi organisasi untuk tidak hanya mengolah data, tetapi untuk mengekstrak wawasan prediktif dan menemukan pola tersembunyi yang sebelumnya sulit dikenali. Penggunaan *machine learning* dalam *Big Data* memungkinkan otomatisasi banyak proses, dari deteksi penipuan hingga personalisasi konten.

Kedua teknologi, ketika digabungkan, membuka peluang baru dalam pengambilan keputusan berbasis data yang lebih akurat dan efisien. Keamanan dan etika menjadi dua aspek yang tidak dapat diabaikan dalam pengelolaan *big data*. Pengumpulan data dalam jumlah besar, khususnya data pribadi dan sensitif, menimbulkan risiko pelanggaran privasi dan kebocoran data.

Organisasi harus memastikan dalam mematuhi regulasi yang ada, seperti GDPR atau CCPA, serta menerapkan kebijakan dan alat keamanan yang memadai. Di sisi lain, etika dalam pengelolaan data juga penting untuk memastikan bahwa data tidak digunakan dengan cara yang merugikan individu atau kelompok tertentu, serta memberikan transparansi kepada pemilik data mengenai bagaimana data akan digunakan.

Masa depan perkembangan *big data* dan pemrosesan data akan terus berkembang seiring dengan kemajuan teknologi. Tren seperti pemrosesan data *real-time*, adopsi *machine learning* yang lebih luas,

dan integrasi *edge computing* akan semakin terus meningkatkan dengan kemampuan organisasi dalam mengelola dan memanfaatkan penggunaan data besar.

Teknologi seperti komputasi kuantum juga diprediksi akan membawa perubahan besar dalam pemrosesan *big data* di masa depan, membuka kemungkinan baru untuk analitik yang lebih kompleks dan efisien. *Big data* akan terus memainkan peran utama dalam transformasi *digital* di berbagai sektor industri.

Secara keseluruhan, *big data* bukan hanya tentang suatu proses pengumpulan data dalam jumlah besar semata, tetapi juga tentang bagaimana data tersebut akan dikelola, dianalisis, dan diterjemahkan untuk menjadikan wawasan yang dapat mendorong keputusan yang lebih baik dan akurat.

Organisasi yang berhasil dalam mengelola dan memanfaatkan *Big Data* dengan cara yang aman, etis, dan efisien akan memiliki keunggulan kompetitif yang signifikan di era *digital* saat ini. Melalui teknologi yang terus berkembang, potensi *big data* untuk mendorong inovasi dan menciptakan nilai baru hanya akan semakin besar di masa yang akan datang.

## Daftar Pustaka

- Ayu Hapsari, N.F. (2020). Big Data dan Pemanfaatannya di Perpustakaan, *Jurnal Ilmu Perpustakaan (Jiper)*, 2(1), pp. 24–32. Available at: <https://doi.org/10.31764/jiper.v2i1.2219>.
- Christian, S.B. and Fajriah, R. (2021). Aplikasi Sistem Informasi Inventaris Perusahaan Untuk Mendukung Manajemen Procurement, *JUST IT: Jurnal Sistem Informasi, Teknologi Informatika dan Komputer*, 11(1), pp. 62–71.
- Event-data, P.E.T.L.U. et al. (2022). Analisis Performa Kecepatan Pemrosesan Big Data Dengan Menggunakan Micro Batching Dalam Tahap Load, *Jurnal Nasional Teknologi dan Aplikasinya*, 1(1), pp. 659–668.
- Gede, K. et al. (2024). Pengolahan Big Data dengan Sharding Database dan Kappa Architecture untuk Data Time-Series, *Jurnal Elektronik Ilmu Komputer Udayana*, 13(1), pp. 43–54.
- Hakim, G., Kaimuddin Haris, O. and Mohammad, M. (2023). Analisis Perbandingan Hukum Mengenai Regulasi Perlindungan Data Pribadi Antara Uni Eropa dan Indonesia Comparative Analysis of Laws Concerning Personal Data Protection Regulations Between the European Union and Indonesia, *Halu Oleo Legal Research* |, 5(2), pp. 443–453. Available at: <https://ojs.uajy.ac.id/index.php/jik/article/view/682>.
- Hernandez-Almazan, J.A. et al. (2022). A Framework to Build a Big Data Ecosystem Oriented to the Collaborative Networked Organization, *Applied Sciences (Switzerland)*, 12(22). Available at: <https://doi.org/10.3390/app122211494>.
- Malik Ibrahim, D., Primananda, R. and Data, M. (2018). Perbandingan Performa Database Apache HBase dan Apache Cassandra sebagai Media Penyimpanan Data Sensor Internet of Things, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(8), pp. 2943–2949. Available at: <http://j-ptiik.ub.ac.id>.
- Nusantara, P. (2018). Data Management Model In Electronic Archives Management: Implementation Of Data Management Body of

- Knowledge, *Jurnal kearsipan*, 13(1), pp. 55–75.
- Perwiratama, R. (2024). Desain Infrastruktur: Kubernetes dan Hadoop sebagai Penyimpanan Data Terdistribusi, *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, 4(2), pp. 1–11.
- Rahmala, A., Ardhito, D. and Riska, L. (2022). Analitik Big Data: Social Media Mining, *Journal Cerita*, 8(2), pp. 164–170. Available at: <https://doi.org/10.33050/cerita.v8i2.2464>.
- Romadhona, L.A. *et al.* (2022). Hadoop-MapReduce Pada YARN Framework, *NetPLG (Journal of Network and Computer)*, 1(2), pp. 91–101.
- Siahaan, D.A. and Raihan. (2024). Manajemen Proyek Big Data: Tantangan dan Strategi Dalam Mengelola Proyek Analisis Data Besar pada Organisasi, *SINTESIA: Jurnal Sistem dan Teknologi Informasi Indonesia*, 3(2), pp. 53–60.
- Siregar, E.K.B. and Harahap, L.S. (2025). Tinjauan Literatur Peran Big Data dalam Pengembangan Konten Digital, *Jurnal Media Akademik (JMA)*, 3(1), pp. 1–12.
- Surahman, Y. and Saptono, H. (2018). Evaluasi Kinerja HDFS sebagai Infrastruktur Pembangunan Big Data, *Jurnal Informatika Terpadu*, 4(2), pp. 63–70. Available at: <https://journal.nurulfikri.ac.id/index.php/JIT>.
- Syafi, M.I., Bhawiyuga, A. and Data, M. (2019). Analisis Perbandingan Kinerja File System GlusterFS dan HDFS dengan Skenario Distribusi Striped dan Replicated, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 3(1), pp. 212–219.
- Tumbel, C.Z., Sitepu, H. and Hutagalung, M. (2017). Analisis Big Data Berbasis Stream Processing Menggunakan Apache Spark, *Jurnal Telematika*, 11(1), p. 6. Available at: <https://doi.org/10.61769/telematika.v11i1.145>.
- Virgiawan, A.K., Absharina, E.D. and Informasi, P.S. (2025). Peran Big Data dalam Meningkatkan Daya Saing Bisnis di Era Digital, *Jurnal Sistem Informasi dan Teknik Komputer*, 10(1).

- Widodo. (2024). Database Manajemen System (DBMS) Basic Concepts, *Jurnal Prodi MPI Idaaratul 'Ulum*, 6(1), pp. 32–50.
- Yuli, S. *et al.* (2024). Application Of Data-Driven Management For More Effective Decision Making: A Perspective, *Management Studies and Entrepreneurship Journal*, 5(2), pp. 5260–5268.

## PROFIL PENULIS



### **Dr. Ir. Norbertus Tri Suswanto Saptadi, S.Kom., M.T., M.M., IPM.**

Penulis lahir di Cirebon, Jawa Barat, tanggal 7 Juni 1975. Memiliki Jabatan Fungsional Lektor Kepala, Pembina Tingkat I (IV/b). Berpendidikan Sarjana Komputer (S.Kom.) di Universitas Teknologi Digital Indonesia (UTDI) tahun 1998, Magister Manajemen (M.M.) di Universitas Hasanuddin (UNHAS) tahun 2004, Magister Teknologi Informasi (M.T.) di Universitas Gadjah Mada (UGM) tahun 2007, Insinyur (Ir.) di Pendidikan Profesi Insinyur UNHAS tahun 2020, Insinyur Profesional Madya (IPM.) di Persatuan Insinyur Indonesia (PII) tahun 2021, Doktor (Dr.) di Fakultas Teknik UNHAS tahun 2023, Kursus Kader Pimpinan (Suskapin) XXVI Menwa RI tahun 1997, dan Program Pendidikan Reguler Angkatan (PPRA) LX Lemhannas RI tahun 2020.

Menjadi tenaga pengajar (Dosen) pada Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Atma Jaya Makassar (UAJM). Peraih Poster terbaik DPRM Dikti tahun 2016. Dosen berprestasi IKDKI tahun 2020, 2021, dan 2024. Pernah menjabat Kepala UPT Komputer, Kepala BAPSI, Wakil Dekan FT, Dekan FT dan FTI, Wakil Rektor III, Ketua Penjaminan Mutu. Tim PAK Dosen dan Asesor BKD UAJM. *Reviewer International Conference* dan Jurnal SINTA. Pemenang Hibah Kemdikbud Penelitian Dosen Pemula, Bersaing, Fundamental, dan Strategi Nasional.

Penulis artikel media massa Tribun Timur, Koinonia, Bisnis Sulawesi, *Sesawi.net*, *Mirifica.net*, *HidupKatolikCom*, *OMKNet*, *KatolikanaTV*, *Jalan Hidup Katolik*, dll. Penulis Buku di Kanisius, Sada Kurnia Pustaka, Aksara Sastra Media, *Future Science*, *HEI Publishing*, *Mifandi Mandiri Digital*, *Rey Media Grafika*, *Widina Salemba*, *Andi*, dan *Cendikia Mulia Mandiri*. Aktifis organisasi IKA Lemhannas RI LX, IARMI, DPP ISKA, BAPOMI Sulsel, LP3KD Sulsel, IKDKI SulSelTraBar, Komkep KAMS, Komsos KAMS, PUKAT KAMS, TPP KAMS, FMKI KAMS, UPS KAMS, Pengurus Kebun Sawit Laimbo, FDI, PII Makassar, INAPR, Dewan Keuangan Paroki dan Program Ayo Sekolah Mariso, Animator *Laudato Si'*, dll.



# **BAB 8**

## **PRINSIP DASAR**

### **VISUALISASI DATA**

---

**Tugiman, S.Kom., M.Kom., IPM.**  
Universitas Medika Suherman



## Pentingnya Prinsip Dasar Visualisasi Data

Visualisasi data adalah proses mengubah data mentah menjadi bentuk visual yang mudah dimengerti oleh manusia. Di era informasi saat ini, yang ditandai oleh tingginya kompleksitas data, visualisasi memainkan peranan penting dalam menyampaikan informasi secara cepat dan efektif.

Namun demikian, tidak semua visualisasi berhasil menjalankan fungsinya dengan baik. Banyak visualisasi yang justru menyesatkan karena tidak mengikuti prinsip-prinsip dasar yang penting dalam menyajikan informasi secara akurat dan terpercaya (Kelleher & Wagener, n.d.).

Prinsip dasar dalam visualisasi data meliputi aspek seperti kejelasan, kesederhanaan, dan juga etika penyajian. Prinsip-prinsip ini bertujuan untuk membantu desainer visualisasi agar terhindar dari kesalahan umum, seperti penggunaan grafik yang tidak sesuai, penyampaian informasi yang menyesatkan, atau pemakaian elemen visual yang berlebihan dan tidak relevan (Cairo, n.d.-a).

Menerapkan prinsip-prinsip ini dapat meningkatkan pemahaman pengguna terhadap data dan sekaligus menjamin bahwa informasi disampaikan secara etis. Salah satu tokoh utama dalam bidang ini, Edward Tufte, memperkenalkan konsep *data-ink ratio*, yang mendorong penggunaan elemen grafis seminimal mungkin asalkan makna data tetap utuh (Tufte, n.d.-b). Konsep ini menegaskan bahwa tujuan utama visualisasi adalah menyampaikan data, bukan sekadar mempercantik tampilan. Di sisi lain, Stephen Few menekankan bahwa desain visual harus mendorong munculnya pemahaman mendalam, bukan sekadar menampilkan informasi secara pasif (Few, n.d.-b).

Menerapkan prinsip-prinsip dasar visualisasi memiliki pengaruh langsung terhadap kualitas pengambilan keputusan. Visualisasi yang efektif mampu menunjukkan pola tersembunyi, mengenali anomali, dan menyampaikan tren dengan cara yang mudah dipahami, bahkan oleh orang-orang tanpa latar belakang teknis (Schwabish, n.d.).

Dengan pertumbuhan data yang sangat pesat di berbagai sektor seperti bisnis, kesehatan, pendidikan, dan pemerintahan, kemampuan memahami dan menerapkan prinsip-prinsip dasar visualisasi data menjadi kompetensi penting dalam abad ke-21.

yang kompleks. Sebaliknya, apabila *audiens* terdiri dari individu dengan keahlian teknis, maka visualisasi dapat menampilkan informasi yang lebih mendetail dengan tingkat *kompleksitas* yang lebih tinggi (Dewabrata, n.d.).

Selain itu, pemilihan jenis *visualization* juga harus mempertimbangkan preferensi serta kebiasaan dari *audiens*. Contohnya, penggunaan *pictograms* atau ikon visual dapat lebih menarik bagi *audiens* yang cenderung responsif terhadap elemen visual, sedangkan *bar charts* atau *line graphs* mungkin lebih cocok untuk mereka yang terbiasa membaca representasi data numerik.

Dengan menyelaraskan visualisasi data berdasarkan karakteristik *audiens*, maka informasi akan lebih mudah dicerna dan memiliki relevansi yang tinggi, sehingga dapat memperkuat proses pengambilan keputusan yang berbasis data.

#### 4. Penggunaan Agregasi dan Penyaringan

Menyaring dan mengelompokkan data secara tepat akan membantu menyajikan informasi secara ringkas dan efisien, tanpa menghilangkan makna utamanya. Hal ini mendukung pemahaman yang lebih cepat dan jelas. Dalam konteks visualisasi data, penerapan teknik agregasi dan penyaringan menjadi elemen kunci untuk menyajikan informasi secara lebih ringkas dan fokus.

Agregasi berperan dalam merangkum data menjadi informasi yang lebih bermakna, seperti menghitung jumlah total, nilai rata-rata, atau angka tertinggi dari suatu variabel. Sementara itu, penyaringan digunakan untuk menyeleksi data berdasarkan kriteria tertentu, sehingga perhatian analisis dapat diarahkan hanya pada bagian-bagian data yang paling relevan (Darmanto et al., n.d.).

Contohnya, dalam menganalisis data penjualan properti, teknik agregasi dapat digunakan untuk memperoleh rata-rata harga jual per wilayah, sedangkan proses penyaringan memungkinkan analisis data berdasarkan periode waktu tertentu atau jenis properti tertentu. Pendekatan ini membantu dalam menampilkan pola atau tren yang tidak mudah terlihat jika hanya melihat data dalam bentuk mentah.

Pemanfaatan perangkat visualisasi seperti *Tableau* dan *Looker Studio* turut mendukung proses ini dengan menyediakan berbagai fitur yang memudahkan pengguna dalam melakukan agregasi dan penyaringan data secara interaktif. Hal ini tidak hanya mempercepat pengolahan informasi, tetapi juga memperkuat akurasi dan ketepatan dalam pengambilan keputusan berbasis data .

## 5. Eliminasi Informasi yang Tidak Perlu

Hindari menambahkan elemen visual atau data sekunder yang tidak memberikan kontribusi terhadap pemahaman. Elemen dekoratif yang berlebihan atau pengulangan data justru bisa mengaburkan pesan utama. Dengan menerapkan prinsip relevansi secara konsisten, visualisasi data akan menjadi alat komunikasi yang lebih efektif, terfokus, dan mudah dipahami oleh berbagai *audiens*.

## Daftar Pustaka

- A. Haryanto, H. S. (n.d.). Prinsip-Prinsip Visualisasi Data dalam Representasi Informasi. *Jurnal Teknologi Informasi Dan Komunikasi*, 12(1), 21–30.
- A. R. Ramadhan, F. D. A. (n.d.). Perancangan Dashboard Visualisasi Data Penjualan Menggunakan Metode Data-Driven Decision Making (DDDM). *Jurnal Teknologi Informasi Dan Komputer (JTIC)*, 6(2), 133–140.
- A. Saputra, M. N. (n.d.). Desain Visualisasi Data Berbasis Profil Pengguna. *Jurnal Sistem Informasi Dan Komputer*, 10(2), 15–23.
- Cairo, A. (n.d.-a). *The Functional Art: An Introduction to Information Graphics and Visualization*. New Riders.
- Cairo, A. (n.d.-b). *The Truthful Art: Data, Charts, and Maps for Communication*.
- Darmanto, Y., Koryanto, L., & Faizah, N. M. (n.d.). Perancangan Aplikasi Visualisasi Database Produksi dengan Metode Agregasi menggunakan Looker Studio dan Google Sheets Berbasis Web Studi Kasus di PT. ANTAM. *Tbk, Journal Innovations Computer Science*, 2(2), 50–62.,  
<https://journal.kawanad.com/index.php/jics/article/view/141>
- Dewabrata, M. (n.d.). Kuasai Teknik Visualisasi Data yang Efektif dengan Memahami 3 Prinsip Dasar Ini. *Medium*.  
<https://medium.com/@mikael.dewabrata/buat-visualisasi-data-yang-efektif-dengan-memahami-3-prinsip-dasar-ini-0ddb6b5663ae>.
- Few, S. (n.d.-a). *Information Dashboard Design: Displaying Data for At-a-Glance Monitoring (2nd ed.)*. Analytics Press.
- Few, S. (n.d.-b). *Show Me the Numbers: Designing Tables and Graphs to Enlighten (2nd ed.)*. Analytics Press.
- Haikal. (n.d.). 3 Strategi Yang Dapat Membantu Audiens Anda Menjadi Lebih Nyaman Dalam Interpretasi Visualisasi Data. In *LLDIKTI Wilayah XIII*. <https://lldikti13.kemdikbud.go.id/2021/11/02/3-strategi-yang-dapat-membantu-audiens-anda-menjadi-lebih->

nyaman-dalam-interpretasi-visualisasi-data/.

- Jurnali, A. (n.d.). Visualisasi Data untuk Pengambilan Keputusan. *Jurnal Teknologi Informasi Dan Komputer*, 6(2), 113–120.
- Kelleher, C., & Wagener, T. (n.d.). Ten Guidelines For Effective Data Visualization In Scientific Publications. *Environ. Modell. Softw*, 26(6), 822–827.
- Liputan6.com. (2022). *Kehadiran 3,6 Juta Toko Kelontong Bawa Berkah Bagi Ekonomi Indonesia*. Liputan6.Com. <https://www.liputan6.com/bisnis/read/4895543/kehadiran-36-juta-toko-kelontong-bawa-berkah-bagi-ekonomi-indonesia>.
- N. Lestari, D. P. (n.d.). Visualisasi Data Interaktif Berbasis Konteks untuk Pengambilan Keputusan. *Jurnal Sistem Informasi*, 10(1), 12–20.
- Norman, D. (n.d.). *The Design of Everyday Things* (Revised). MIT Press.
- R. F. Lestari, H. S. (n.d.). Penerapan Prinsip Audience-Centric dalam Visualisasi Data Kesehatan. *Jurnal Teknologi Dan Sistem Informasi*, 12(1), 25–32.
- R. Kurniawan, D. P. (n.d.). Konsistensi Desain Visual pada Dashboard Interaktif untuk Analisis Data. *Jurnal Sistem Informasi*, 9(1), 55–63,.
- R. M. Hardiyanto, H. S. P. (n.d.). Implementasi Visualisasi Data pada Sistem Monitoring Kinerja Karyawan Menggunakan Dashboard Interaktif. *Jurnal Teknologi Dan Sistem Komputer*, 9(1), 15–22,.
- R. Pratama, A. R. (n.d.). Penggunaan Visualisasi Data dalam Mendukung Pengambilan Keputusan. *Jurnal Sistem Informasi*, 8(2), 45–52.
- Schwabish, J. (n.d.). *Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks*. Columbia Univ. Press.
- Tufte, E. R. (n.d.-a). *Envisioning Information*. Graphics Press.
- Tufte, E. R. (n.d.-b). *The Visual Display of Quantitative Information* (2nd ed.). Graphics Press.
- Ware, C. (n.d.). *Information Visualization: Perception for Design* (3rd ed.). Morgan Kaufmann.

## PROFIL PENULIS



### **Tugiman, S.Kom., M.Kom., IPM.**

Tugiman, S.Kom, M.Kom, lahir di Boyolali, 15 September 1968. Lulus Diploma I dari Institut Manajemen Komputer Indonesia (IMKI) Surakarta tahun 1991. Menamatkan Sarjana Jurusan Sistem Informasi di Universitas Budi Luhur tahun 2014. Kemudian melanjutkan studi Program Pascasarjana Jurusan Sistem Informasi di Universitas Budi Luhur dan selesai tahun 2016. Saat ini saya sebagai Dekan Fakultas Sosial dan Teknologi Universitas Medika Suherman, mengajar di Universitas Buddhi Dharma Tangerang, dan sebagai konsultan di sebuah Rumah Sakit. Organisasi yang diikuti saat ini adalah sebagai Pengurus Asosiasi Perguruan Tinggi Manajemen Ritel Indonesia (APTMRI), sebagai pengurus DPW Ikatan Ahli Informatika Indonesia (IAII) Banten, sebagai pengurus pusat Badan Kejuruan Informatika (BKI) Persatuan Insinyur Indonesia (PII), sebagai anggota Gugus Tugas Penyusunan Standar Layanan Insinyur-Kementerian Teknis (Kementerian Kesehatan), sebagai anggota APTIKOM, sebagai anggota Asosiasi Dosen Indonesia (ADI). Adapun mata kuliah yang pernah diampu adalah Rekayasa Perangkat Lunak, Manajemen Proyek, Analisa dan Perancangan Sistem Informasi, Audit Sistem Informasi, E-Bisnis, *E-Commerce*, IT Budgeting, Manajemen Operasi, Strategi Pemasaran, Testing dan Implementasi, dan Manajemen Sumber Daya Manusia. Selain itu juga pernah mengerjakan beberapa aplikasi yang dipakai di UMKM dan Rumah Sakit.

Email Penulis: [tugiman0311@gmail.com](mailto:tugiman0311@gmail.com).




# **BAB 9**

# **JENIS-JENIS**

# **VISUALISASI DATA**

---

**Mohammad Badrul, S.Kom., M.Kom.**  
Universitas Bina Sarana Informatika



## Pendahuluan

Visualisasi Data merupakan upaya menyampaikan informasi yang terkandung pada data agar lebih mudah dipahami oleh orang lain. Dengan kata lain visualisasi data merupakan proses mengubah data hasil analisis menjadi gambaran visual berupa bagan, peta, grafik atau lainnya (Sartono & Hidayatulloh, 2021).

Salah satu cara yang dilakukan analisis data yang memungkinkan untuk memahami, menyajikan dan menyampaikan data ke pengguna non teknis supaya mereka mudah untuk membaca, memahami data tersebut sehingga menjadi informasi yang lebih efektif adalah Visualisasi Data. Ada banyak jenis-jenis visualisasi data yang bisa kita temukan dari berbagai sumber (Wiyono, 2023).

Masing-masing jenis visualisasi data digunakan untuk memberikan informasi utama yang berbeda-beda. Kami sadar bahwa teknik dan jenis-jenis visualisasi masih terus berkembang dan banyak diagram-diagram jenis baru yang diusulkan oleh para praktisi dan ilmuwan sehingga jenis diagram yang kami bahas pada buku ini mungkin kurang lengkap. Beberapa jenis visualisasi data antara lain seperti *line plot*, *Scatter plot*, Histogram, *Bar chart* dan *Pie chart*.

### ***Line Plot***

Salah satu jenis visualisasi data yang paling dasar, namun sangat efektif adalah diagram *line plot*. Diagram *line plot* adalah teknik visualisasi data yang digunakan untuk menampilkan informasi melalui rangkaian titik data yang terhubung oleh garis lurus (Pratiwi, 2024). Diagram ini berguna untuk menyajikan tren dan pola dalam data dari waktu ke waktu. Analisis dapat membandingkan variabel, melihat pola dan fluktuasi dalam data, dan membuat proyeksi dari data. Tipe data kontinu merupakan tipe data yang digunakan di *line plot* pada umumnya:

#### **1. Karakteristik Utama Diagram *Line Plot***

Diagram *line plot* memiliki beberapa karakteristik yang membedakannya dari jenis grafik lainnya (Hyman et al., 2024):

- a. Sumbu X dan Y: sumbu horizontal (X) biasanya merepresentasikan waktu atau kategori, sedangkan sumbu vertikal (Y) menunjukkan nilai data.

### 3. Stacked Bar Chart

Selain mengetahui penjualan total, kita juga ingin tahu berapa banyak dari penjualan tersebut yang berasal dari penjualan *online* dan *offline*. Dapat ditambahkan data sintetis untuk skenario tersebut dan visualisasikan dengan *Stacked Bar Chart*.

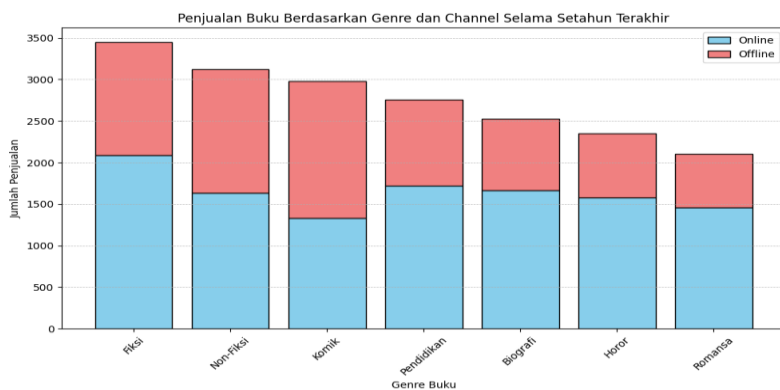
```
# Menambahkan data sintetis untuk penjualan online dan offline (Stacked Bar Chart)
df_penjualan['Online'] = df_penjualan['Penjualan'] * np.random.uniform(0.4, 0.7, len(df_penjualan))
df_penjualan['Offline'] = df_penjualan['Penjualan'] - df_penjualan['Online']

# Membuat stacked bar chart untuk penjualan buku berdasarkan genre
plt.figure(figsize=(10, 6))
plt.bar(df_penjualan['Genre'], df_penjualan['Online'], label='Online',
        color='skyblue', edgecolor='black')
plt.bar(df_penjualan['Genre'], df_penjualan['Offline'], label='Offline',
        bottom=df_penjualan['Online'], color='lightcoral', edgecolor='black')
plt.title('Penjualan Buku Berdasarkan Genre dan Channel Selama Setahun Terakhir')
plt.xlabel('Genre Buku')
plt.ylabel('Jumlah Penjualan')
plt.xticks(rotation=45)
plt.grid(True, which='both', linestyle='--', linewidth=0.5, axis='y')
plt.legend()
plt.tight_layout()
plt.show()
```

**Gambar 9.12: Source Code Stacked Bar Chart**

Sumber: Pradita, 2023.

Dari *source code* di atas akan dihasilkan *Stacked Bar Chart* seperti dibawah ini yang terdapat dua dimensi sekaligus yaitu total Penjualan untuk Setiap Genre: diwakili oleh tinggi total dari setiap batang. Bagian biru mewakili penjualan *online*, sementara bagian merah mewakili penjualan *offline*.



**Gambar 9.13: Stacked Bar Chart Penjualan Buku**

Sumber: Pradita, 2023.

## Diagram Lingkaran

Diagram lingkaran atau *Pie chart* merupakan grafik yang paling sederhana dan sering digunakan dalam memvisualisasikan komposisi dari data. *Pie chart* cocok digunakan untuk data diskrit, nominal, ordinal (atau dengan kata lain data yang bersifat kategorik)(Pahlephi, 2022).

Diagram lingkaran mewakili angka dalam persentase, dan jumlah total semua segmen harus sama dengan 100%. Yang perlu diperhatikan dalam pembuatan *pie chart* adalah jangan memasukkan terlalu banyak kategori agar dapat terlihat perbedaan antar komposisinya. Kemudian urutkan komposisi/*slice* dari *pie chart* sesuai dengan ukurannya (dari kecil ke besar atau sebaliknya).

### 1. Membuat *Pie Chart* Di *Python*

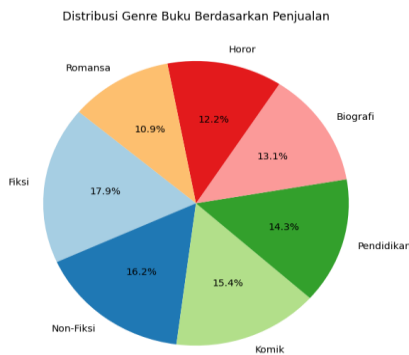
Pemilik toko buku ingin mengetahui proporsi genre buku yang dijual. *Pie Chart* dapat memberikan gambaran visual yang cepat tentang distribusi *genre* buku. Kita bisa gunakan data penjualan buku yang sebelumnya sudah dibuat.

```
# Membuat Pie Chart untuk distribusi genre buku berdasarkan penjualan
plt.figure(figsize=(10, 7))
plt.pie(df_penjualan['Penjualan'], labels=df_penjualan['Genre'],
autopct='%1.1f%%', startangle=140, colors=plt.cm.Paired.colors)
plt.title('Distribusi Genre Buku Berdasarkan Penjualan')
plt.show()
```

**Gambar 9.14: Source Code Pie Chart**

Sumber: Pradita, 2023.

Dari *source code* di atas akan didapatkan *pie chart* seperti berikut:



**Gambar 9.15: Pie Chart Distribusi Genre Penjualan Buku**

Sumber: Pradita, 2023

Dari *Pie Chart* di atas, pemilik toko dapat melihat distribusi penjualan buku berdasarkan *genre*. Setiap potongan lingkaran mewakili suatu *genre*, dan ukurannya menunjukkan proporsi penjualannya dalam total penjualan buku. Label dan persentase di setiap potongan membantu pemilik toko memahami proporsi relatif *genre* tersebut.

## Daftar Pustaka

- Aribowo, B. (2023). *Eksplorasi dan Visualisasi Data Menggunakan Python Jilid 1*.
- Enterprise, J. (2024a). *Analisis dan Visualisasi Data Dengan Power BI*. Elex Media Komputindo.
- Enterprise, J. (2024b). *Python Untuk Analisis dan Visualisasi Data*. Elex Media Komputindo.
- Hyman, J. A., Massaron, L., McFedries, P., Mueller, J. P., Reichental, J., Schmuller, J., Simon, A. R., & Taylor, A. G. (2024). *Data Analytics & Visualization All-in-One For Dummies*. For Dummies.
- Isnanto, B. A. (2023). *Diagram Batang: Definisi, Jenis, Contoh, dan Cara Membuatnya*. <https://www.detik.com/edu/detikpedia/d-6982357/diagram-batang-definisi-jenis-contoh-dan-cara-membuatnya>.
- Pahlephi, R. D. (2022). *Diagram Lingkaran: Jenis, Rumus, Cara Menghitung, dan Contohnya*. <https://www.detik.com/bali/berita/d-6485054/diagram-lingkaran-jenis-rumus-cara-menghitung-dan-contohnya>.
- Pradita, G. A. (2023). *Modul Ajar Data Visualisasi STIKOM Bali*.
- Pratiwi, D. (2024). *Visualisasi Data Teori dan Penerapannya*. Uwais Inspirasi Indonesia.
- Randani, F., & Utami, I. Q. (2022). *Pengantar Data Science*. PT. Bumi Aksara.
- Sartono, B., & Hidayatulloh, A. (2021). *Visualisasi Data yang Efektif*. PT. Penerbit IPB Press.
- Wiyono, H. (2023). *Dasar-Dasar Visualisasi Data Dengan Tableau*. Balai Pengembangan Kompetensi PUPR.

## PROFIL PENULIS



### **Mohammad Badrul, S.Kom., M.Kom.**

Penulis lahir dari keluarga petani sederhana di daerah Bangkalan tahun 1984. Dari kecil penulis sudah di didik di lingkungan agamis dan sederhana. Setelah lulus dari Sekolah Menengah Pertama dan Sekolah Menengah Atas di lingkungan Pesantren, Sejak di Sekolah Menengah Atas, Penulis memiliki ketertarikan di bidang komputer. Oleh sebab itu penulis melanjutkan program Diploma III atau D3 di Akademi Manajemen Informatika dan Komputer (AMIK) Bina Sarana Informatika di Jakarta dengan Jurusan Komputerisasi Akuntansi. Lulus dari Program Diploma III tahun 2007, Penulis melanjutkan Program Strata Satu (S1) di Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Nusa Mandiri di Jakarta Jurusan Sistem Informasi. Setelah lulus program Strata Satu (S1) tahun 2009, Penulis melanjutkan program Strata Dua (S2) di Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Nusa Mandiri Jakarta jurusan Ilmu Komputer tahun 2010 dan lulus tahun 2012. Saat ini penulis aktif mengajar di beberapa kampus di Jakarta khususnya Jurusan Sistem Informasi, Informatika dan Teknologi Informasi dengan konsentrasi di bidang Data Mining, Data Science, dan Sistem Penunjang Keputusan. Dan untuk mewujudkan karir sebagai dosen profesional, penulis pun aktif sebagai peneliti di bidang kepakarannya tersebut. Beberapa penelitian yang telah dilakukan didanai oleh internal perguruan tinggi dan juga Kemenristek DIKTI.

Email Penulis: [mohammad.mbl@bsi.ac.id](mailto:mohammad.mbl@bsi.ac.id).



# **BAB 10**

## **ALAT DAN *PLATFORM***

### **VISUALISASI DATA**

---

**Max Teja Ajie Cipta Widiyanto, S.Kom., M.Kom.**  
Institut Teknologi PLN



## Pendahuluan

Visualisasi data merupakan proses representasi data dalam bentuk grafis atau visual, seperti grafik, diagram, peta, dan *dashboard*, untuk memudahkan pemahaman, analisis, dan pengambilan keputusan. Dengan berkembangnya teknologi, berbagai alat dan *platform* visualisasi data telah muncul, menawarkan fitur yang beragam mulai dari yang sederhana hingga kompleks.

Visualisasi data dapat dipahami juga sebagai komponen kunci dalam analisis data modern, memungkinkan pengguna untuk memahami pola, tren, dan wawasan yang tersembunyi dalam *dataset* yang kompleks. Menurut para ahli, visualisasi data tidak hanya sekadar alat representasi grafis, tetapi juga sarana untuk meningkatkan *cognitive perception* (persepsi kognitif) dan *decision-making* (pengambilan keputusan).

Alat visualisasi data dapat dibedakan berdasarkan tujuan penggunaannya, seperti untuk analisis bisnis, penelitian akademik, atau pelaporan *real-time*. Beberapa platform menyediakan fitur *drag-and-drop* yang mudah digunakan oleh pemula, sementara lainnya memungkinkan kustomisasi tingkat lanjut dengan pemrograman. Perbedaan alat vs. *platform*: alat visualisasi: lebih spesifik, bisa berupa *library* pemrograman (*Matplotlib*, *Seaborn*) atau *software* tunggal (*Excel*). *Platform* visualisasi: lebih komprehensif, menggabungkan alat analisis, penyimpanan data, dan fitur kolaborasi (contoh: *Tableau Server*, *Looker Studio*).

## Pengertian Alat dan Platform Visualisasi Data Menurut Ahli

### 1. Alat Visualisasi Data

Menurut Edward Tufte (2001), alat visualisasi data adalah perangkat lunak atau teknik grafis yang digunakan untuk mengubah data mentah menjadi representasi visual (seperti grafik, diagram, atau peta) guna meningkatkan pemahaman manusia terhadap pola, tren, dan anomali dalam data. Tufte menekankan bahwa alat yang baik harus memaksimalkan kejelasan dan efisiensi komunikasi data ("*data-ink ratio*").

Stephen Few (2009) mendefinisikan alat visualisasi data sebagai sarana untuk menampilkan informasi kuantitatif secara visual dengan cara yang memudahkan analisis, mengurangi

ambiguitas, dan mendukung pengambilan keputusan yang lebih baik. Contohnya termasuk *Tableau*, *Power BI*, dan *tools* berbasis statistik seperti *R* dan *Python*.

## 2. Definisi Platform Visualisasi Data

Alberto Cairo (2016) menyatakan bahwa *platform* visualisasi data adalah lingkungan terintegrasi (baik berbasis *web* atau *desktop*) yang menyediakan fitur lengkap untuk mengolah, menganalisis, dan memvisualisasikan data dalam skala besar.

*Platform* seperti *Google Data Studio*, *Tableau Public*, dan *Microsoft Power BI* memungkinkan kolaborasi, pembuatan *dashboard* interaktif, dan penyajian data secara *real-time*. Hadley Wickham (2010) menambahkan bahwa *platform* visualisasi modern harus mendukung:

- a. *Grammar of Graphics* (seperti dalam *ggplot2* di *R*),
- b. Interaktivitas (seperti *Plotly* dan *D3.js*),
- c. Kemampuan bekerja dengan *big data* (seperti *Apache Superset*).

Menurut beberapa pendapat ahli di atas maka dapat disimpulkan bahwa alat visualisasi data fokus pada transformasi data ke bentuk visual, sementara *platform* visualisasi data mencakup ekosistem lengkap untuk analisis, kolaborasi, dan penyajian data. Pemilihannya tergantung pada kompleksitas data, kebutuhan interaktivitas, dan keterampilan pengguna.

## Pentingnya Visualisasi Data Dalam Pengambilan Keputusan

Visualisasi data memainkan peran kritis dalam proses pengambilan keputusan karena membantu manusia memahami informasi kompleks dengan cepat, mengidentifikasi pola, dan menemukan wawasan yang tidak terlihat dalam data mentah. Berikut adalah beberapa alasan utama mengapa visualisasi sangat penting dalam pengambilan keputusan, didukung oleh pendapat ahli dan referensi terkait:

### 1. Mempercepat Pemahaman Data

Otak manusia memproses informasi visual 60.000 kali lebih cepat daripada teks atau angka mentah (Sousa, 2011). Visualisasi seperti grafik, diagram, dan peta panas memungkinkan pengambil keputusan untuk:

- 2) *Tableau*: ekstrak data dengan *hyper*.
- 3) *Python*: gunakan *Dask* atau *Modin*.

**4. Studi Kasus Nyata**

- a. *Retail Analytics (Unilever)*
  - 1) Solusi: *Tableau + Alteryx*.
  - 2) Hasil: 30% peningkatan kecepatan analisis inventori (Gartner, 2023. *Retail Analytics Case Studies*).
- b. Kesehatan Masyarakat (WHO COVID-19)
  - 1) *Tools: R Shiny + Leaflet*.
  - 2) *Output*: Peta sebaran real-time (WHO, 2022. *Data Visualization in Public Health*).

**5. Troubleshooting Umum**

**Tabel 10.4: Troubleshooting Umum**

Masalah	Solusi
Visual lambat	Optimasi <i>query</i> , agregasi
Warna tidak <i>accessible</i>	Gunakan <i>ColorBrewer</i> palet
<i>User engagement</i> rendah	Tambah interaktivitas

Sumber: Diolah Penulis.

**6. Emerging Trends 2024**

- a. *Augmented Analytics*:
  - 1) *Tableau GPT (AI-generated insights)*.
  - 2) *Power BI Copilot*.
- b. *Embedded Analytics*:
  - 1) Integrasi *dashboard* ke SaaS.
  - 2) Contoh: *Shopify + Looker*.
- c. *Real-time Streaming*:
  - 1) *Apache Kafka + Grafana*.
  - 2) *IoT sensor visualization*.

*Checklist Implementasi*:

- 1. Audit kebutuhan *stakeholder*.

2. *Proof-of-concept* dengan *sample data*.
3. Pelatihan *end-user*.
4. Dokumentasi *style guide*.
5. Mekanisme *feedback*.

Tips Ahli: "Mulailah dengan MVP (*Minimum Viable Product*) sebelum mengembangkan solusi kompleks" (Cairo, 2021).

## Daftar Pustaka

- Cairo, A. (2016). *The Truthful Art. New Riders.*
- Cairo, A. (2016). *The Truthful Art: Data, Charts, and Maps for Communication.* New Riders.
- Cairo, A. (2019). *How Charts Lie: Getting Smarter About Visual Information.* W.W. Norton & Company.
- Evergreen, S. (2019). *Effective Data Visualization.*
- Few (2019). *Data Visualization Best Practices.*
- Few, S. (2009). *Now You See It: Simple Visualization Techniques for Quantitative Analysis.* Analytics Press.
- Few, S. (2009). *Now You See It.*
- Few, S. (2012). *Show Me the Numbers.* Analytics Press.
- Few, S. (2012). *Show Me the Numbers: Designing Tables and Graphs to Enlighten.* Analytics Press.
- Gartner. (2023). *Retail Analytics Case Studies.*
- Knaflic, C. N. (2015). *Storytelling with Data.*
- LaValle, S., et al. (2016). *Data Visualization and the Decision-Making Process.* MIT Sloan Management Review.
- McKinsey. (2022). *Data Visualization in Digital Transformation.*
- Microsoft. (2023). *How Data Visualization Improves Business Decisions.* Power BI Documentation.
- Microsoft. (2023). *Power BI Implementation Guide*
- Murray, S. (2017). *Interactive Data Visualization for the Web.* O'Reilly.
- Sousa, D. A. (2011). *How the Brain Learns.* Corwin Press.
- Tableau (2023). *Blueprints for Enterprise Deployment.*
- Tufte, E. (2001). *The Visual Display of Quantitative Information.* Graphics Press.

Tufte, E. R. (1983). *The Visual Display of Quantitative Information*.

Tufte, E. R. (2001). *The Visual Display of Quantitative Information*.  
Graphics Press.

Ware, C. (2020). *Information Visualization: Perception for Design*.

WHO. (2022). *Data Visualization in Public Health*.

Wickham, H. (2010). A Layered Grammar of Graphics. *Journal of Computational and Graphical Statistics*.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*.

Wilkinson, L. (2005). *The Grammar of Graphics*. Springer.

## PROFIL PENULIS




### **Max Teja Ajie Cipta Widiyanto, S.Kom., M.Kom.**

Minat penulis terhadap dunia ilmu komputer mulai tumbuh sejak tahun 2006. Ketertarikan tersebut mendorong penulis untuk menempuh pendidikan tinggi Sarjana di Universitas Dian Nuswantoro, memilih pada Fakultas ilmu komputer dengan program studi pada bidang Teknik Informatika (TI).

Penulis berhasil menyelesaikan pendidikan jenjang Sarjana (S1) dan lulus pada tahun 2011. Sebelum memperdalam pemahaman dan kompetensinya di bidang teknologi informasi, penulis pernah Mengajar di SMK Ashabul Kahfi Semarang Teknik Komputer Jaringan tahun 2011-2015, kemudian melanjutkan studi ke jenjang Magister (S2) di Universitas Dian Nuswantoro, dan pada tahun 2017 berhasil meraih gelar Magister Ilmu Komputer, pada program studi Teknik Informatika. Saat ini, penulis aktif sebagai dosen tetap di lingkungan pendidikan tinggi, tepatnya pada Program Studi Teknik Informatika di bawah naungan LLDIKTI Wilayah III, bertugas di Institut Teknologi PLN.

Dalam kapasitasnya sebagai tenaga pengajar, penulis mengampu berbagai mata kuliah yang berfokus pada pengembangan kompetensi teknis dan manajerial mahasiswa, antara lain: Teknik *Digital*, Mikroprosesor, *Embedded System*, Interaksi Manusia Komputer, Matematika Diskrit, Komunikasi Data dan *Internet of Things (IoT)*. Selain mengajar, penulis juga aktif melakukan penelitian dan publikasi ilmiah dalam bidang keilmuannya. Karya-karya ilmiah penulis dapat ditemukan dan diakses melalui portal akademik seperti SINTA (*Science and Technology Index*) dan *Google Scholar*. Dengan latar belakang akademik yang kuat dan pengalaman mengajar yang luas, penulis terus berkomitmen untuk berkontribusi dalam pengembangan ilmu pengetahuan dan teknologi, serta membimbing generasi muda dalam menghadapi tantangan dunia digital yang semakin kompleks.



# **BAB 11**

# **VISUALISASI DATA**

# **UNTUK ANALISIS**

# **SPASIAL**

---

**Erna Astriyani, S.Kom., M.TI.**  
Universitas Raharja



## Analisis dan Data Spasial

Analisis spasial merupakan sekumpulan metode yang digunakan untuk mengolah, mengevaluasi, dan menafsirkan data yang memiliki acuan lokasi guna menghasilkan informasi baru berdasarkan hubungan spasial antar data (Longley et al., 2015). Ada dua jenis data utama dalam Analisis Spasial: data vektor (titik, garis, poligon) dan data raster (citra satelit, elevasi). Alat-alat seperti GIS dan FineBI mempermudah proses pengolahan serta penyajian data spasial, sehingga memungkinkan analisis yang lebih efisien.

Data spasial merupakan informasi geografis yang digunakan untuk menjelaskan posisi dan ciri-ciri fisik suatu area. Banyak orang yang menggunakan istilah data spasial, padahal yang sebenarnya mereka maksud adalah data geospasial. data geospasial merupakan data spasial yang telah mengalami proses georeferensi, yaitu proses pengaitan data dengan lokasi yang akurat di permukaan bumi menggunakan koordinat geografis.

(Otto Huisman dan Rolf A. De: 2009). Prosedur ini sering diterapkan untuk memastikan bahwa data spasial (seperti peta, gambar satelit, atau foto udara) sesuai dengan sistem koordinat bumi, agar data tersebut dapat digunakan bersamaan dengan data spasial lainnya dalam Sistem Informasi Geografis (GIS). Data spasial biasanya disimpan dalam format yang mendukung analisis spasial serta visualisasi, seperti koordinat geografis dan data topologi.

## Tipe dan Sumber Data Spasial

### 1. Data Vektor

Data vektor dalam geografis diwakili sebagai titik, garis, dan poligon. Setiap elemen tersebut memiliki posisi dan dilengkapi dengan atribut tambahan:

- a. Titik: digunakan untuk menunjukkan tempat yang akurat dari fitur geografi seperti lokasi kota, alamat, atau fasilitas.
- b. Garis: menjabarkan karakteristik linier seperti jalan, sungai, atau rel kereta api.
- c. Poligon: digunakan untuk menunjukkan wilayah seperti batas administratif, luas hutan, atau kawasan pemukiman.

#### **4. Google Earth Engine (Pemrosesan Citra Satelit Berbasis Cloud) Google Earth Engine (GEE)**

Adalah *platform cloud computing* yang dikembangkan oleh Google untuk mengelola dan menganalisis data geospasial, terutama citra satelit dalam skala besar dan jangka panjang secara global. *Platform* ini memberikan kemampuan kepada ilmuwan, akademisi, dan pengambil kebijakan untuk menjalankan analisis spasial berskala besar tanpa ketergantungan pada perangkat keras lokal berkapasitas tinggi.

### **Cara Kerja Analisis Spasial**

Berikut cara kerja Analisis Spasial yang dapat diketahui adalah:

#### **1. Pengumpulan Data**

Langkah awal dalam proses analisis spasial adalah pengumpulan data, yang harus memastikan bahwa informasi yang diperoleh relevan dan akurat terhadap tujuan analisis yang akan dilakukan.

#### **2. Sumber Data Geografis (Citra Satelit, Sensor, Survey)**

Citra satelit memberikan gambaran visual mengenai permukaan bumi, termasuk perubahan lahan atau pola tumbuhan. Sensor, mirip dengan perangkat *Internet of Things (IoT)*, berfungsi untuk membantu Anda dalam mengumpulkan data secara langsung, seperti suhu atau kualitas udara. *Survey* lapangan menyediakan informasi khusus yang tidak dapat ditemukan dari sumber lainnya.

#### **3. Pengolahan Data Mentah**

Data mentah seringkali tidak siap untuk digunakan. Anda harus membersihkan dan memproses data tersebut, misalnya menghapus data yang tidak relevan atau memperbaiki kesalahan. Selain itu, proses ini melibatkan penyesuaian data ke format yang sesuai untuk analisis. Data menjadi lebih terstruktur dan siap untuk digunakan dalam analisis ruang dengan pengolahan yang tepat.

### **Analisis Data**

Setelah data telah dikumpulkan dan diolah, Anda dapat memulai analisis guna menemukan pola dan hubungan dalam data spasial. Maka perlu melakukan beberapa analisis data sebagai berikut:

## 1. Analisis Pola Spasial

Analisis pola spasial memungkinkan Anda untuk memahami penyebaran fenomena di suatu area. Anda bisa mengenali daerah dengan tingkat konsentrasi yang tinggi atau rendah dari suatu fenomena. Sebagai contoh, Anda dapat menggambarkan wilayah dengan tingkat pencemaran udara yang tinggi. Melalui analisis ini, Anda dapat menemukan pola tersembunyi dalam data yang umum.

## 2. Analisis Jarak dan Hubungan

Analisis jarak dan hubungan memungkinkan anda menilai keterkaitan antara berbagai lokasi. Anda dapat menentukan jarak antara dua lokasi geografis atau mempelajari keterkaitan antara tempat-tempat tertentu. Sebagai contoh, Anda dapat menentukan posisi yang paling tepat untuk fasilitas umum dengan mempertimbangkan jaraknya dari kelompok populasi yang dituju. Analisis ini memberikan pemahaman yang mendetail tentang cara lokasi saling berpengaruh satu sama lain.

## Contoh Kasus Penggunaan Visualisasi Data Spasial

1. Mitigasi Bencana: mengidentifikasi wilayah rawan banjir dengan memanfaatkan data curah hujan dan ketinggian permukaan tanah.
2. Perencanaan Layanan Kesehatan: memetakan lokasi fasilitas kesehatan sebagai dasar dalam menyusun rencana penyediaan layanan medis.
3. Pemantauan Perubahan Tutupan Lahan: melacak perubahan kawasan hutan melalui citra satelit guna memantau aktivitas deforestasi.
4. Manajemen Perangkat IoT (*Internet of Things*): mendata dan memetakan perangkat IoT seperti sensor cuaca, CCTV lalu lintas, dan alat deteksi kualitas udara secara real-time dalam sistem kota pintar.
5. Analisis Lokasi Potensial Usaha: investor menggunakan data spasial untuk menentukan area strategis yang sesuai bagi pengembangan UMKM, pasar, atau kawasan industri baru.
6. Perencanaan Tata Ruang Kota: pemerintah kota memanfaatkan visualisasi spasial untuk merancang zonasi ruang, jalur transportasi publik, dan pengembangan wilayah pemukiman.

7. Optimalisasi Jalur Distribusi: perusahaan logistik seperti J&T dan SiCepat menggunakan data spasial untuk menyusun rute pengiriman barang yang lebih efisien dan hemat waktu.

## Daftar Pustaka

- Badruzaman, M., & Raharjo, B. (2018). *Penerapan Sistem Informasi Geografis untuk Analisis Spasial*. Yogyakarta: Andi Publisher.
- De Smith, M. J., Goodchild, M. F., & Longley, P. A. (2020). *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*. Winchelsea Press.
- Esri. (2023). *What is ArcGIS?* Retrieved from <https://www.esri.com/en-us/arcgis/about-arcgis/overview>.
- Fanruan. (2024). Spatial Analysis. <https://blog.unmaha.ac.id/mengenal-tableau-dari-kelebihan-dan-kekurangannya-dalam-memvisualisasikan-data/>.
- Google Earth Engine. (2023). *Google Earth Engine: Data Catalog & Documentation*. Retrieved from <https://earthengine.google.com/>.
- Gorr, W. L., & Kurland, K. S. (2020). *GIS Tutorial for ArcGIS Pro 2.8*. Redlands, CA: Esri Press.
- Huisman, O., & de By, R. A. (2009). Principles of Geographic Information Systems: An Introductory Textbook. (ITC Educational Textbook Series; Vol. 1). *International Institute for Geo-Information Science and Earth Observation*.
- Kraak, M. J., & Ormeling, F. (2020). *Cartography: Visualization of Geospatial Data (4th ed.)*. Routledge.
- Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2021). *Geographic Information Science and Systems (5th ed.)*. Wiley.
- Mapbox. (2023). *Design and Use Custom Maps*. Retrieved from <https://www.mapbox.com/>.
- QGIS Documentation. (2024). *User Guide*. Retrieved from <https://docs.qgis.org/>.
- Tableau. (2023). *Tableau Mapping and Mapbox Integration*. Retrieved from <https://www.tableau.com/learn/articles/mapping-data-mapbox>.

- Tomlinson, R. (2013). *Thinking About GIS: Geographic Information System Planning for Managers*. Esri Press.
- Wicaksono, P., & Sofyan, A. (2019). *Penginderaan Jauh dan Sistem Informasi Geografis untuk Lingkungan*. Malang: UB Press.
- Wijaya, H., & Kurniawan, B. (2022). Analisis Spasial Menggunakan QGIS Untuk Pemetaan Risiko Bencana. *Jurnal Geomatika dan Lingkungan*, 4(1), 45–56. <https://doi.org/10.31289/jgl.v4i1.8971>.

## PROFIL PENULIS



### **Erna Astriyani, S. Kom., M. TI.**

Ketertarikan penulis terhadap Dunia Komputer dan Teknologi Informasi sejak tahun 2010 silam. Hal tersebut membuat penulis memilih untuk melanjutkan pendidikan ke Jenjang Sarjana. Penulis menempuh pendidikan Sarjana pada Program Studi Sistem Informasi pada STMIK Raharja tahun 2014. Semangat untuk mendalami Ilmu Komputer Penulis Melanjutkan Studi Jenjang

Pasca Sarjana Program Studi Teknik Informatika Konsentrasi *Business Intelligence* Tahun 2018. Penulis Mulai mengajar menjadi Dosen Sejak Tahun 2019 dan Menjabat sebagai Sekretaris Prodi Sistem Informasi hingga saat ini. Dalam Karir Dosen penulis aktif mengampu Mata Kuliah yang berkaitan dengan pengembangan sistem informasi dan teknologi Komputer. Selain Tugas menjadi dosen penulis juga aktif dalam menulis Jurnal Ilmiah baik Nasional maupun Internasional. Penulis juga pernah menjadi Redaksi Jurnal SENSI Tahun 2015 dan aktif menjadi *Reviewer* Jurnal Ilmiah. Buku Manajemen Proyek merupakan karya ilmiah penulis sebagai Pengembangan Ilmu yang diterapkan. Penulis berharap dapat berkontribusi dalam pengembangan ilmu komputer dan bermanfaat bagi masyarakat luas. Untuk mewujudkan karir sebagai dosen profesional, penulis aktif sebagai peneliti di bidang kepakarannya tersebut.

Email Penulis: [erna.astriyani@raharja.info](mailto:erna.astriyani@raharja.info).



# **BAB 12**

## **VISUALISASI DATA**

### **WAKTU (*TIME SERIES*)**

---

**Patria Adhistian, S.T., M.M.**  
Universitas Pamulang



## Pengertian Visualisasi Data Deret Waktu (*Time Series*)

Visualisasi data deret waktu (*time series*) merupakan salah satu teknik analisis penting untuk menampilkan pola perubahan data terhadap waktu secara kronologis. Dalam konteks bisnis, ekonomi, dan manufaktur, data deret waktu sering kali digunakan untuk memantau tren penjualan, permintaan produk, fluktuasi harga, dan perencanaan produksi.

Dengan memvisualisasikan data deret waktu, pengambil keputusan dapat lebih mudah mengidentifikasi tren (*trend*), pola musiman (*seasonality*), siklus (*cycles*), serta anomali (*outliers*) dalam data historis. Visualisasi ini menjadi pendukung utama dalam proses *forecasting* (peramalan) dan pengendalian proses. Tujuan utama dari visualisasi deret waktu adalah:

1. Menyajikan perubahan data terhadap waktu secara intuitif dan komunikatif.
2. Mendeteksi pola jangka pendek maupun jangka panjang.
3. Mengidentifikasi fluktuasi musiman atau kejadian tidak biasa.
4. Membandingkan hasil aktual dengan hasil peramalan.
5. Menyampaikan hasil analisis kepada stakeholder secara efektif.

Karakteristik Data *Time Series*: data deret waktu memiliki komponen utama yang dapat dikenali melalui visualisasi: (1) *Trend* arah pergerakan data dalam jangka panjang (naik/turun), (2) *Seasonality*: pola berulang dalam periode waktu tertentu (bulanan, tahunan), (3) *Cyclic*: fluktuasi yang tidak beraturan tapi berlangsung dalam jangka waktu panjang. (4) *Noise/Irregular*: variasi acak yang tidak memiliki pola khusus.

## Jenis-Jenis Visualisasi *Time Series*

Beberapa jenis grafik yang umum digunakan dalam visualisasi data *time series*:

### 1. *Line Chart* (Grafik Garis)

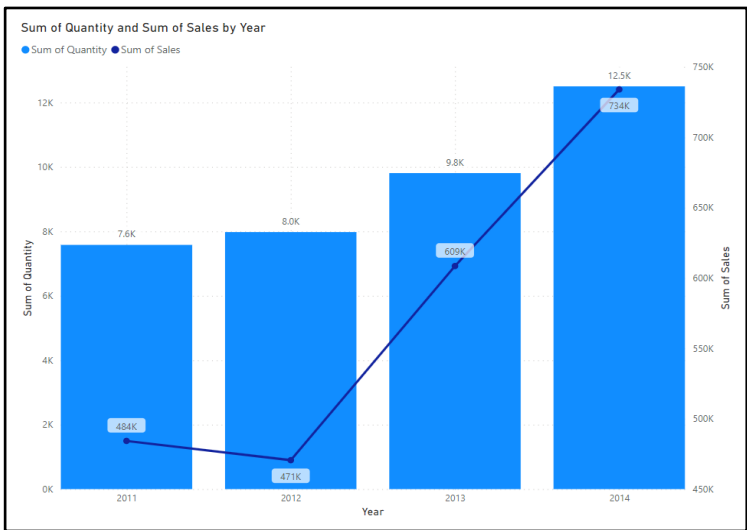
- a. Menampilkan nilai data berdasarkan sumbu waktu (*x-axis*).
- b. Sangat efektif untuk melihat tren naik-turun dari waktu ke waktu.

### 2. *Moving Average Plot*

Menghaluskan fluktuasi jangka pendek agar tren jangka panjang lebih terlihat.

*Combo Chart* berguna untuk menunjukkan bagaimana metrik dengan satuan pengukuran yang berbeda dibandingkan satu sama lain. Contohnya, kita ingin membandingkan total '*Quantity*' terjual dengan total '*Sales*' untuk setiap tahunnya produk:

1. Pilih salah satu visualisasi *Combo Chart* yang kita inginkan.
2. Masukkan salah satu metrik ke *Column Axis* dan satunya lagi ke *Line Axis*.
3. Masukkan kolom kategori ke *X-axis*.
4. (Opsional) Pergi ke *Tab Format Visual* lalu aktifkan *Data Labels* dan *Markers* untuk mempermudah melihat nilai pada titik data di visualisasi.



**Gambar 13.12: Graph Power BI**

Sumber: *Power BI Desktop*.

Dengan *Combo Chart* kita dapat memvisualisasikan dua metrik dengan skala pengukuran yang berbeda dalam satu grafik yang sama. Di atas dapat kita lihat bahwa *Power BI* akan membuatkan *axis* untuk masing-masing metrik sesuai dengan skalanya. Dari contoh *Combo Chart* tersebut kita juga mendapatkan *insight* bahwa meningkatnya total produk terjual tidak menjamin naiknya penjualan perusahaan secara keseluruhan.

## Daftar Pustaka

- Jayalaksono, H. S. D. (2024). *Power BI for Business Intelligence Coursebook*. Algoritma Data Science School.
- Krishnan, V. (2017). *Research Data Analysis With Power BI, 11th International CALIBER-2017 [Preprint]*. Chennai, Tamil Nadu-600025: Anna University, Chennai, Tamil Nadu.
- Molke, A., Bhagat, R. dan Gahat, V. (2024). Empowering Insights: The Power of Data Visualization With Power BI, *SSGM Journal of Science And Engineering*, 2(1), hal. 77–81.
- Sahaya, A.P.N. *et al.* (2024). Powering Sales Insights: A Comparative Analysis of Data Visualization Tools, Microsoft Power BI vs Tableau, in *2024 9th International Conference on Business and Industrial Research (ICBIR)*. IEEE, hal. 1042–1047.
- Setiawan, M.F.F. dan Winursito, Y.C. (2025). Operational Data Analysis and Visualization of PT XYZ Using Business Intelligence Approach with Microsoft Power BI, *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, 4(2), hal. 648–653.
- Singh, G. *et al.* (2023). Data Visualization For Developing Effective Performance Dashboard With Power BI, in *2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*. IEEE, hal. 968–973.


## PROFIL PENULIS



### **Patria Adhastian, S.T., M.M.**

Patria Adhastian, ST, MM adalah seorang profesional di bidang teknologi dan manajemen strategis dengan pengalaman lebih dari 15 tahun di industri telekomunikasi dan *digital*. Lulus sebagai Sarjana Teknik (ST) dari jurusan Teknik Informatika, ia kemudian melanjutkan studi dan meraih gelar Magister Manajemen (M.M.) dengan konsentrasi Manajemen. Kombinasi keahlian teknis dan wawasan bisnis menjadikannya sosok yang mampu menjembatani kebutuhan teknologi dengan arah strategis perusahaan. Dengan latar belakang sebagai insinyur, Patria dikenal sebagai sosok yang analitis dan *detail-oriented*, namun juga mampu berpikir strategis dalam merancang solusi yang berdampak luas. Karirnya ditandai dengan keterlibatannya dalam berbagai program integrasi sistem, optimalisasi proses bisnis, serta inisiatif efisiensi operasional yang mendukung tujuan perusahaan jangka panjang. Dan untuk mewujudkan karir sebagai dosen profesional, penulis pun aktif sebagai peneliti di bidang kepakarannya tersebut. Beberapa penelitian yang telah dilakukan didanai oleh internal perguruan tinggi dan juga Kemenristek DIKTI.


Email Penulis: dosen01529@unpam.ac.id.



**BAB 13**  
**VISUALISASI DATA**  
**UNTUK *MACHINE***  
***LEARNING***

---

**Siti Mutmainah, M.Kom.**  
Universitas Muhammadiyah Bima



mengilustrasikan hasil model atau temuan secara persuasif. Grafik yang disajikan memudahkan *stakeholder* sehingga dapat mengidentifikasi tren dan pola yang tersembunyi dalam data melalui visualisasi. Selain itu, data *storytelling* yang dipadukan dengan visual dapat memperkuat pesan, sesuai rekomendasi literatur komunikasi data.

## **Tools dan Library Visualisasi Machine Learning**

### **1. Matplotlib**

*Matplotlib* merupakan library visualisasi dasar *python*, library ini digunakan untuk *plotting* 2D di *Python* dalam sains data dan *machine learning*. Dokumentasi resminya menyatakan bahwa “*Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python*” ([pypi.org](http://pypi.org), t.t.). *Matplotlib* memungkinkan untuk membuat *plot* garis, *scatter*, *histogram*, serta visualisasi hasil model ML seperti kurva *loss* dan *confusion matrix*.

### **2. Seaborn**

*Seaborn* merupakan library visualisasi statistik dan korelasi fitur atau *pairplot* untuk identifikasi fitur yang berkorelasi. Library ini dibangun dengan *Matplotlib* untuk visualisasi statistik. Dokumentasi resminya menyatakan bahwa “*Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics*” ([seaborn.pydata.org](http://seaborn.pydata.org), t.t.). *Seaborn* dilengkapi fungsi-fungsi siap pakai seperti *heatmap* untuk visualisasi matriks korelasi antar fitur, *pairplot scatter matrix* antar fitur, *distplot*, *boxplot*, *violin plot*, dan grafik regresi.

### **3. Plotly**

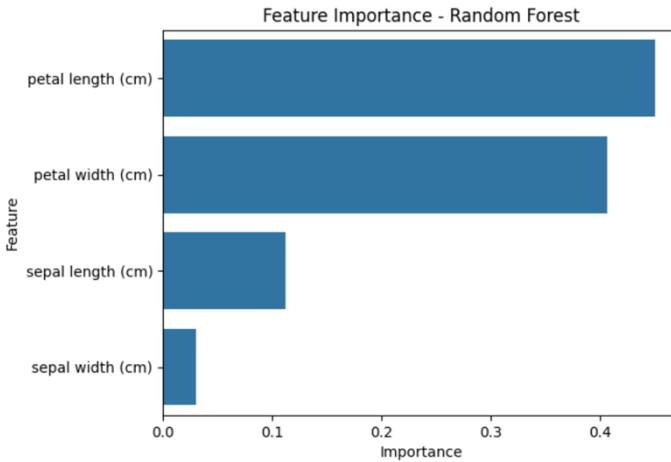
*Plotly* merupakan library interaktif *python* yang mendukung grafik interaktif seperti *heatmaps* dinamis, grafik 3D, maupun *dashboard*. Visualisasi interaktif ini dapat membantu eksplorasi dan presentasi data/model secara lebih dinamis dibandingkan *plot* statis. Dokumentasi resminya menyatakan bahwa “*Plotly’s Python graphing library makes interactive, publication-quality graphs*” ([plotly.com](http://plotly.com), t.t.).

```

feat_df = feat_df.sort_values(by='Importance',
                              ascending=False)

sns.barplot(x='Importance', y='Feature', data=feat_df)
plt.title("Feature Importance - Random Forest")
plt.show()

```



**Gambar 13.8: Feature Importance**  
 Sumber: Diolah Penulis.

```

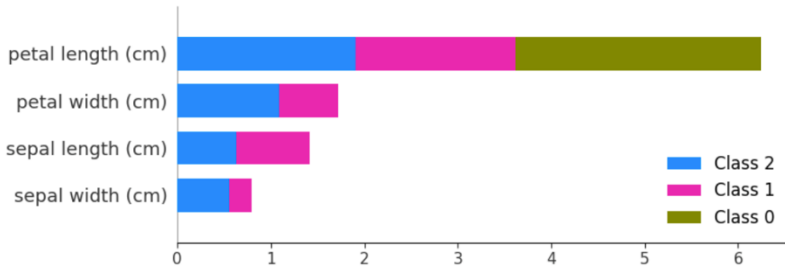
# SHAP Visualisasi
import shap
import xgboost as xgb

# XGBoost model
xgb_model = xgb.XGBClassifier()
xgb_model.fit(X_train, y_train)

# SHAP Explainer
explainer = shap.Explainer(xgb_model)
shap_values = explainer(X_test)

# Summary plot
shap.summary_plot(shap_values, X_test,
                  feature_names=iris.feature_names)

```



**Gambar 13.9: SHAP**

Sumber: Diolah Penulis.

## Kesimpulan

Visualisasi data berperan penting pada keseluruhan proses machine learning, mulai dari tahap eksplorasi data hingga evaluasi dan interpretasi hasil model. Melalui visualisasi, data kompleks dapat direpresentasikan dalam bentuk yang lebih intuitif dan mudah dipahami, baik oleh praktisi teknis maupun pemangku kepentingan non-teknis. Pada tahap sebelum training model, visualisasi seperti *histogram*, *heatmap*, dan *pairplot* membantu memahami sebaran data, deteksi nilai ekstrim, dan melihat korelasi antar fitur.

Selama proses training, visualisasi metrik seperti loss dan akurasi membantu dalam mengamati performa model secara dinamis, sekaligus mendeteksi potensi masalah seperti *overfitting* atau *underfitting*. Setelah model di-*training*, berbagai bentuk visualisasi seperti *confusion matrix*, kurva ROC, dan kurva *precision-recall* dapat dimanfaatkan untuk mengevaluasi kinerja model dalam mengklasifikasikan data.

Selain itu, teknik interpretasi model seperti SHAP dan LIME memungkinkan untuk memahami kontribusi setiap fitur terhadap hasil prediksi, sehingga dapat meningkatkan visibilitas dan kepercayaan terhadap model yang digunakan. Visualisasi bukan hanya sekedar alat teknis, tetapi juga merupakan elemen strategis dalam mengembangkan sistem *machine learning* yang efektif, informatif, dan mudah dikomunikasikan. Visualisasi yang memiliki peran menyeluruh ini menjadikannya sebagai salah satu pilar utama dalam implementasi *machine learning*.

## Daftar Pustaka

Arxiv.org. (t.t.). *Why Should I Trust You?": Explaining the Predictions of Any Classifier & A Unified Approach to Interpreting Model Predictions*. <https://arxiv.org/abs/>. <https://arxiv.org/abs/>.

Plotly.com. (t.t.). *Plotly Open Source Graphing Library For Python*. [plotly.com](https://plotly.com/python/). <https://plotly.com/python/>.

Pypi.org. (t.t.). *Matplotlib*. [pypi.org](https://pypi.org/project/matplotlib). <https://pypi.org/project/matplotlib>.

Seaborn.pydata.org. (t.t.). *Seaborn: Statistical Data Visualization*. <https://seaborn.pydata.org/>. <https://seaborn.pydata.org/>.

Tensorflow.org. (t.t.). *TensorBoard: Toolkit visualisasi TensorFlow*. [tensorflow.org](https://www.tensorflow.org/tensorboard). [www.tensorflow.org/tensorboard](https://www.tensorflow.org/tensorboard).

## PROFIL PENULIS



### **Siti Mutmainah, S.Kom., M.Kom.**

Penulis menempuh pendidikan sarjana (S1) di jurusan Informatika tahun 2020 pada Universitas AMIKOM Yogyakarta. Setelah menyelesaikan gelar sarjana, penulis melanjutkan pendidikan di tingkat Magister (S2) Program Studi Informatika pada Universitas Islam Indonesia dan lulus pada tahun 2023. Keahlian penulis terfokus pada bidang Data Sains seperti pengolahan data, analisis *statistic* dan pemodelan prediksi teknik *machine learning* dan *deep learning* menggunakan bahasa pemrograman *Python*. Saat ini, penulis berkarir sebagai seorang dosen pada Universitas Muhammadiyah Bima. Selain kegiatan mengajar di perguruan tinggi, penulis juga terlibat dalam kegiatan penelitian dan pengabdian kepada masyarakat. Penulis, penulis berusaha untuk memberikan kontribusi positif kepada masyarakat melalui penelitian yang inovatif dan pengaplikasian teknologi informasi.

Email Penulis: [siti.mutmainah9810@gmail.com](mailto:siti.mutmainah9810@gmail.com).



# **BAB 14**

# **ETIKA DAN KEAMANAN**

# **DALAM PEMODELAN**

# **DATA**

---

**Nasril Sany, S.Kom., M.Kom.**  
Institut Teknologi PLN



## Pendahuluan

Pemodelan data merupakan proses penting dalam pengelolaan informasi untuk mendukung pengambilan keputusan, analisis bisnis, dan pengembangan sistem. Namun, seiring dengan meningkatnya volume dan kompleksitas data, isu etika dan keamanan menjadi aspek kritis yang harus diperhatikan, penting untuk mempertimbangkan aspek etika dan keamanan untuk melindungi privasi, mencegah penyalahgunaan, dan memastikan kepatuhan terhadap regulasi.

Pemodelan data melibatkan pengumpulan, pengolahan, dan penyimpanan data yang mungkin bersifat sensitif atau pribadi. Tanpa prinsip etika yang jelas, terdapat risiko penyalahgunaan data, diskriminasi algoritmik, atau pelanggaran privasi. Etika data menekankan pada: (1) transparansi (bagaimana data digunakan dan diproses), (2) akuntabilitas (tanggung jawab dalam pengelolaan data), (3) keadilan (menghindari bias dalam model data). Keamanan data memastikan bahwa informasi tetap terlindungi dari ancaman seperti peretasan, kebocoran data, atau akses tidak sah. Dalam pemodelan data, langkah-langkah seperti enkripsi, anonymisasi, dan kontrol akses harus diterapkan untuk memenuhi standar keamanan seperti ISO 27001 atau GDPR.

## Pengertian Etika Dalam Pemodelan Data

### 1. Floridi & Taddeo (2016)

Etika data adalah bidang yang mempelajari prinsip moral yang mengatur praktik pengumpulan, penyimpanan, dan penggunaan data, termasuk dalam pemodelan data. Menekankan pentingnya keadilan, transparansi, dan tanggung jawab penggunaan algoritma dan data. *Data ethics is concerned with moral problems related to data practices, including the generation, recording, curation, processing, dissemination, sharing, and use of data* (Floridi & Taddeo, 2016). Etika data berkaitan dengan masalah moral yang terkait praktik data, termasuk pembuatan, pencatatan, kurasi, pemrosesan, penyebaran, pembagian, dan penggunaan data.

### 2. Kitchin (2014)

Kitchin menyatakan bahwa dalam era big data, pemodelan data harus mempertimbangkan dampaknya terhadap individu dan

- 1) Standar internasional untuk manajemen keamanan informasi.
  - 2) Mencakup kebijakan enkripsi, kontrol akses, dan manajemen risiko.
- b. NIST *Privacy Framework*
- 1) Dibuat oleh *National Institute of Standards and Technology* (AS).
  - 2) Membantu organisasi mengelola risiko privasi data.
- c. PCI DSS (*Payment Card Industry Data Security Standard*)
- 1) Wajib bagi perusahaan yang menangani data kartu kredit.
  - 2) Mengharuskan enkripsi, audit keamanan, dan kontrol akses ketat.

**Tabel 14.4: Standar Etika Dalam Pemodelan Data**

<b>Standar Etika</b>	<b>Prinsip Inti</b>
<b><i>Fairness</i></b>	Tidak diskriminatif, bebas bias
<b><i>Transparency</i></b>	Proses pemodelan dapat dijelaskan dan dipahami
<b><i>Accountability</i></b>	Tanggung jawab atas dampak model
<b><i>Privacy &amp; Consent</i></b>	Data digunakan sesuai izin
<b><i>Explainability</i></b>	Model harus dapat dijelaskan kepada pemangku kepentingan
<b><i>Data Minimization</i></b>	Mengumpulkan data secukupnya sesuai kebutuhan saja

Sumber: Diolah Penulis.

## 5. Praktik Baik yang Sesuai Regulasi dan Etika

- a. Melakukan *Data Protection Impact Assessment* (DPIA) sebelum pembangunan model.
- b. Menyediakan opsi persetujuan dan *opt-out* kepada pengguna data.
- c. Menerapkan enkripsi dan kontrol akses pada data sensitif.
- d. Menggunakan audit etika untuk mengevaluasi dampak model.
- e. Menerapkan prinsip "*privacy by design*" dan "*ethics by design*".

## Daftar Pustaka

- Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and Machine Learning*. fairmlbook.org.
- California Civil Code § 1798.100 et seq. (CCPA).
- Cheswick, W. R., Bellovin, S. M., & Rubin, A. D. (2003). *Firewalls and Internet Security: Repelling the Wily Hacker (2nd ed.)*. Addison-Wesley.
- DAMA International. (2017). *DAMA-DMBOK: Data Management Body of Knowledge (2nd ed.)*.
- Doshi-Velez, F., & Kim, B. (2017). *Towards A Rigorous Science of Interpretable Machine Learning*. arXiv.
- Dwork, C. (2006). *Differential Privacy*. ICALP.
- European Union. (2016). *General Data Protection Regulation (GDPR)*.
- Fanelli, D. (2009). *How Many Scientists Fabricate And Falsify Research? A Systematic Review And Meta-Analysis*. PLoS ONE, 4(5).
- Resnik, D. B. (2015). *What Is Ethics In Research & Why Is It Important?* NIH Office of Research Integrity.
- Floridi, L. (2013). *The Ethics of Information*. Oxford University Press.
- Floridi, L. (2018). *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford University Press.
- GDPR. (2018). *General Data Protection Regulation (EU) 2016/679*.
- Goodfellow, I., McDaniel, P., & Papernot, N. (2018). *Making Machine Learning Robust Against Adversarial Inputs*. Communications of the ACM, 61(7), 56–66.
- <https://www.oecd.org/going-digital/ai/principles/>.
- ISO/IEC 27001. (2022). Information Security Management Systems.
- ISO/IEC 27001:2022. *Information Security, Cybersecurity, And Privacy Protection*.
- Jobin, A., et al. (2019). *The Global Landscape of AI Ethics Guidelines*. Nature Machine Intelligence.
- Johnson, D. G. (2001). *Computer Ethics. 3rd Edition*. Prentice Hall.

- Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures And Their Consequences*. SAGE Publications.
- Laudon & Laudon. (2020). *Management Information Systems: Managing The Digital Firm, 16th Edition*. New Jersey: Pearson Education.
- Lundberg, S. M., & Lee, S. I. (2017). *A Unified Approach to Interpreting Model Predictions*. NeurIPS.
- Mehrabi, N., et al. (2021). *A Survey on Bias and Fairness in Machine Learning*. ACM Computing Surveys.
- NIST Privacy Framework (Version 1.0, 2020).
- NIST SP 800-53 Rev. 5 (2020) *Security and Privacy Controls for Information Systems and Organizations*.  
<https://csrc.nist.gov/publications/detail/sp/800-53/rev-5/final>.
- Noble, S. U. (2018). *Algorithms of Oppression*. NYU Press.
- OECD. (2019). *OECD Principles on Artificial Intelligence*.
- OECD. (2019). *Principles on Data Governance*.
- PCI Security Standards Council.
- Regulation (EU) 2016/679 (GDPR).
- Republik Indonesia. (2022). *Undang-Undang No. 27 Tahun 2022 Tentang Perlindungan Data Pribadi*.
- Richards, N. M., & King, J. H. (2014). Big Data Ethics. *Wake Forest Law Review*, 49(2), 393–432.
- Solove, D. J. (2008). *Understanding Privacy*. Harvard University Press.
- Regulation (EU) 2016/679-General Data Protection Regulation (GDPR).
- Undang-Undang Republik Indonesia No. 27 Tahun 2022 tentang Perlindungan Data Pribadi.

## PROFIL PENULIS




### **Nasril Sany, S.Kom., M.Kom.**

Minat penulis terhadap dunia ilmu komputer mulai tumbuh sejak tahun 1999. Ketertarikan tersebut mendorong penulis untuk menempuh pendidikan tinggi di STMIK Raharja, memilih Jurusan Teknik Informatika (TI) dengan konsentrasi pada bidang *Software Engineering* (SE). Penulis berhasil menyelesaikan pendidikan jenjang Sarjana (S1) dan lulus pada tahun 2005.

Guna memperdalam pemahaman dan kompetensinya di bidang teknologi informasi, penulis melanjutkan studi ke jenjang Magister (S2) di Universitas Budi Luhur, dan pada tahun 2017 berhasil meraih gelar Magister Ilmu Komputer, dengan peminatan pada Teknologi Sistem Informasi.

Saat ini, penulis aktif sebagai dosen tetap di lingkungan pendidikan tinggi, tepatnya pada Program Studi Sistem Informasi di bawah naungan LLDIKTI Wilayah III, bertugas di Institut Teknologi PLN. Dalam kapasitasnya sebagai tenaga pengajar, penulis mampu berbagai mata kuliah yang berfokus pada pengembangan kompetensi teknis dan manajerial mahasiswa, antara lain: Konsep Sistem Informasi, Dasar Pemrograman, Sistem Informasi Manajemen, Struktur Data, Sistem Basis Data, Sistem Pendukung Keputusan, Jaringan Komputer dan UI/UX Desain. Selain mengajar, penulis juga aktif melakukan penelitian dan publikasi ilmiah dalam bidang keilmuannya. Karya-karya ilmiah penulis dapat ditemukan dan diakses melalui portal akademik seperti SINTA (*Science and Technology Index*) dan *Google Scholar*. Dengan latar belakang akademik yang kuat dan pengalaman mengajar yang luas, penulis terus berkomitmen untuk berkontribusi dalam pengembangan ilmu pengetahuan dan teknologi, serta membimbing generasi muda dalam menghadapi tantangan dunia digital yang semakin kompleks.

Email Penulis: [nasrilsanypenulis@gmail.com](mailto:nasrilsanypenulis@gmail.com).



# **BAB 15**

## **INTEGRASI MODEL DATA KE SISTEM PRODUKSI**

---

**Muhamad Yusup, S.Kom., M.Kom.**  
Universitas Raharja



yang kompatibel, pengembangan *Application Programming Interface* (API), *deployment* model ke dalam lingkungan produksi, hingga strategi pemeliharaan dan *monitoring* model jangka panjang.

Menurut Liang et al. (2024), otomasi dalam pelatihan dan deployment model memainkan peran penting dalam efisiensi sistem produksi berbasis *machine learning*. Tantangan utama yang dihadapi dalam integrasi ini mencakup kestabilan infrastruktur, pengelolaan siklus hidup model, dan kolaborasi lintas tim (Windmann et al., 2024). Selain itu, pendekatan MLOps terbukti mampu menyelaraskan antara pengembangan dan operasi dalam sistem berbasis AI, yang sangat penting dalam konteks industri modern (Zimelewicz et al., 2024).

Dalam implementasinya, sistem produksi cerdas memerlukan strategi *deployment* yang adaptif dan berbasis data *real-time* agar dapat menyesuaikan performa model dengan kebutuhan operasional secara dinamis (Ren et al., 2024). Oleh karena itu, integrasi model ke dalam sistem produksi bukan hanya persoalan teknis semata, melainkan juga membutuhkan pendekatan holistik yang mencakup pemrograman, manajemen sistem, dan rekayasa perangkat lunak secara terpadu (Tjandrarini et al., 2024; Hidayah et al., 2024).

## Tahap Utama Integrasi Model

Proses integrasi model ke sistem produksi dapat diringkas ke dalam beberapa tahap utama:

1. Pengembangan dan penyimpanan model.
2. Pembuatan layanan API.
3. *Deployment* model ke *server* produksi.
4. Integrasi model dengan sistem aplikasi.
5. *Monitoring* performa dan *logging*.
6. Otomatisasi proses dengan MLOps.
7. Pemeliharaan dan pembaruan model

Setiap tahapan ini saling terkait dan harus dirancang dengan prinsip efisiensi, skalabilitas dan keandalan.

## Pengembangan dan Penyimpanan Model

Tahap awal dari integrasi model ke sistem produksi adalah pengembangan dan penyimpanan model. Proses ini mencakup beberapa langkah penting diantaranya adalah:

### 1. Penyebab Penurunan Performa Model (*Model Drift*)

*Model drift* adalah kondisi saat model tidak lagi mencerminkan kondisi data terkini. Penyebabnya adalah meliputi:

- a. *Data drift*: distribusi fitur input berubah dari data pelatihan.
- b. *Concept Drift*: hubungan antara fitur dan label berubah dari waktu ke waktu.
- c. *Target Leakage*: ketidaksengajaan memasukkan informasi masa depan ke dalam pelatihan model.

### 2. Evaluasi performa Berkala

Model perlu dievaluasi secara rutin menggunakan metrik yang sesuai, seperti:

**Tabel 15.2: Tabel Evaluasi Performa Berkala**

Type Model	Metriks Evaluasi Umum
Klasifikasi	<i>Accuracy, Precision, Recall, F1-score, AUC-ROC</i>
Regresi	MAE, MSE, RMSE, $R^2$
Probabilistik	<i>Log Loss, Brier Score</i>

Sumber: Diolah Penulis.

Ambil sampel data terbaru dari produksi (*real-time data*). Gunakan label aktual (jika tersedia) untuk mengukur performa. Bandingkan dengan *baseline* saat model pertama kali diluncurkan. Catat hasil evaluasi dan waktu pelaksanaannya. Contoh otomasi: evaluasi performa tiap 30 hari via pipeline MLOps.

### 3. Retraining Model

Jika ditemukan degradasi performa signifikan, maka retraining diperlukan. Langkah-langkah *retraining*: pengumpulan data baru: data dari sistem produksi dikumpulkan dan dibersihkan. Analisis data: pastikan tidak ada anomali atau perubahan struktur. Pelatihan ulang model: model dilatih kembali dengan data terbaru (bisa *full training* atau *incremental*). Validasi ulang: lakukan *cross-validation* untuk memastikan kualitas model baru. Deployment versi baru: gunakan sistem versioning untuk mengganti model lama.

## Daftar Pustaka

- Coleman, A. (2024). Integrating MLOps Pipelines With DevOps for Seamless Model Deployment And Continuous Delivery. *Australian Journal of Machine Learning Research & Applications*, 4(2), 87–94.
- Dmitriev, D., Ivanov, V., & Suvorov, R. (2021). Data Version Control (DVC): Git For Data Scientists. *DVC.org*. <https://dvc.org/doc/start>.
- Docker Inc. (n.d.). *Docker Overview*. <https://www.docker.com/resources/what-container/>.
- Elastic. (n.d.). *What is the ELK Stack?*. <https://www.elastic.co/what-is/elk-stack>.
- GitHub. (n.d.). *Git Large File Storage (LFS)*. <https://git-lfs.github.com/>.
- GitHub. (n.d.). *GitHub Actions for CI/CD*. <https://docs.github.com/en/actions>.
- GitLab. (n.d.). *GitLab CI/CD Pipelines*. <https://docs.gitlab.com/ee/ci/>.
- Grafana Labs. (n.d.). *Grafana: The Open Observability Platform*. <https://grafana.com/docs/>.
- Hidayah, H. K., Ekojono, E., & Sintiya, E. S. (2024). Implementasi Machine Learning Pada Sistem Informasi Pemeliharaan Transformator Daya. *Jurnal Informatika Polinema*, 10(4), 571–580. <https://doi.org/10.33795/jip.v10i4.6008>.
- Jenkins. (n.d.). *Jenkins for ML/AI Pipelines*. <https://www.jenkins.io/solutions/machine-learning/>.
- Johnson, A. (2024). Improving CI/CD Pipelines With MLOps-Oriented Automation For Machine Learning Models. *Journal of AI-Assisted Scientific Discovery*, 4(2), 59–72. <https://doi.org/10.60087/jaasd.vol4.n2.p176>.
- KFServing Contributors. (2022). *KFServing: Serverless Inference on Kubernetes*. <https://github.com/kubeflow/kfserving>.
- Kubernetes Authors. (n.d.). *Kubernetes Documentation*. <https://kubernetes.io/docs/home/>.
- Liang, P., Song, B., Zhan, X., Chen, Z., & Yuan, J. (2024). Automating The Training And Deployment of Models In MLOps By Integrating Systems With Machine Learning. *Applied and Computational*

- Engineering*, 76, 1–7. <https://doi.org/10.54254/2755-2721/76/20240690>.
- Liang, P., Song, B., Zhan, X., Chen, Z., & Yuan, J. (2024). Automating The Training And Deployment Of Models In MLOps By Integrating Systems With Machine Learning. *arXiv preprint arXiv:2405.09819*. <https://arxiv.org/abs/2405.09819>.
- Nogare, D., & Silveira, I. F. (2024). Experimentation, Deployment And Monitoring Machine Learning Models: Approaches For Applying MLOps. *arXiv*. <https://arxiv.org/abs/2408.11112>.
- Prometheus Authors. (n.d.). *Prometheus: Monitoring System & Time Series Database*. <https://prometheus.io/docs/introduction/overview/>.
- Ramkumar, A. (2024). Machine Learning Models In Production: A Systematic Framework For Scalable And Robust Deployment. *International Journal of Research in Computer Applications and Information Technology*, 7(2), 1608–1628. <https://doi.org/10.60087/ijrcait.vol7.n2.p125>.
- Ren, Y., He, Y., Zhang, X., Yen, A., & Li, G. P. (2024). A Cyber Manufacturing IoT System For Adaptive Machine Learning Model Deployment By Interactive Causality Enabled Self-Labeling. *arXiv preprint arXiv:2404.05976*. <https://arxiv.org/abs/2404.05976>.
- Seldon. (2023). Seldon Core: Deploying Machine Learning Models On Kubernetes. <https://docs.seldon.io/projects/seldon-core/en/latest/>.
- Singla, A., & Chavalmane, S. (2023). Automating Model Deployment: From Training To Production. *Journal of Knowledge Learning and Science Technology*, 2(3), 340–347. <https://doi.org/10.60087/jklst.vol2.n3.p347>.
- Smith, J. (2024). Deploying Real-Time Machine Learning Models In DevOps Environments: A Hybrid MLOps Approach. *Journal of AI-Assisted Scientific Discovery*, 4(2), 45–58. <https://doi.org/10.60087/jaasd.vol4.n2.p175>.
- Smith, J. (2024). Exploring The Role of Kubernetes In MLOps and DevOps For Containerized Machine Learning Model Management. *Hong Kong Journal of AI and Medicine*, 4(2), 88–94.

- Thompson, A. (2024). Implementing Scalable DevOps Pipelines For Machine Learning Model Monitoring And Performance Management. *Journal of Artificial Intelligence Research*, 4(2), 117–122.
- Tjandrarini, T., Tanuwijaya, H., & Purnawan, Y. (2024). Implementasi Sistem Informasi Manajemen Produksi Terintegrasi Menggunakan Metode Agile. *Jurnal Ekonomi Manajemen Sistem Informasi*, 5(6), 650–662. <https://doi.org/10.38035/jemsi.v5i6.2502>.
- Weights & Biases. (2023). *Experiment Tracking With W&B*. <https://wandb.ai/site>.
- Windmann, A., Wittenberg, P., Schieseck, M., & Niggemann, O. (2024). Artificial Intelligence in Industry 4.0: A Review Of Integration Challenges For Industrial Systems. *arXiv preprint arXiv:2405.18580*. <https://arxiv.org/abs/2405.18580>.
- Zaharia, M., Chen, A., Davidson, A., Ghodsi, A., Hong, M., Konwinski, A., ... & Gonzalez, J. (2018). Accelerating The Machine Learning Lifecycle With MLflow. *Databricks*. <https://mlflow.org/docs/latest/index.html>.
- Zimelewicz, E., Kalinowski, M., Mendez, D., Giray, G., Alves, A. P. S., Lavesson, N., ... & Wagner, S. (2024). ML-Enabled Systems Model Deployment And Monitoring: Status Quo And Problems. *arXiv preprint arXiv:2402.05333*. <https://arxiv.org/abs/2402.05333>.

## PROFIL PENULIS




### **Muhamad Yusup, S.Kom., M.Kom.**

Ketertarikan penulis terhadap bidang ilmu komputer dimulai sejak tahun 2007. Minat tersebut mendorong penulis untuk menempuh pendidikan di SMAN 4 Tangerang dengan memilih jurusan Ilmu Pengetahuan Alam (IPA) dan berhasil menyelesaikan pendidikan menengah pada tahun yang sama. Selanjutnya, penulis melanjutkan studi ke jenjang pendidikan tinggi dan berhasil meraih gelar Sarjana Komputer (S.Kom.) dari Program Studi Sistem Informasi di STMIK Raharja. Tidak berhenti sampai di situ, penulis kemudian melanjutkan pendidikan magister di Program Studi Ilmu Komputer Universitas Budi Luhur dan berhasil meraih gelar Magister Komputer (M.Kom.) pada tahun 2014.

Saat ini, penulis sedang menempuh pendidikan program doktor (S3) di Program Studi Ilmu Komputer Universitas Kristen Satya Wacana, Salatiga. Penulis memiliki kepakaran di bidang Sains Data melalui *Workshop Digitalent* Departemen Kominfo RI selama 3 bulan. Saat ini penulis aktif sebagai dosen pada Program Studi Bisnis *Digital* Fakultas Ekonomi dan Bisnis Universitas Raharja. Dalam perannya sebagai akademisi, penulis terus mengembangkan keilmuan melalui kegiatan penelitian, pengajaran, dan penulisan ilmiah yang berfokus pada bidang kepakarannya. Penulis memiliki komitmen untuk berkontribusi dalam pengembangan teknologi dan pendidikan di Indonesia, khususnya dalam era transformasi *digital*.

Email Penulis: [yusup@raharja.info](mailto:yusup@raharja.info).



# **BAB 16**

## **PEMODELAN DAN VISUALISASI DI SEKTOR KESEHATAN**

---

**Inda Rusdia Sofiani, S.T., M.Sc.**  
Universitas Muhammadiyah Malang



## **Transformasi *Digital* dan Industri Kesehatan 4.0**

Dalam beberapa dekade terakhir, sektor kesehatan mengalami perubahan fundamental yang dipacu oleh kemajuan teknologi digital. Revolusi Industri 4.0 yang ditandai oleh integrasi antara teknologi informasi, kecerdasan buatan, *Internet of Things* (IoT), dan *big data* telah membawa transformasi mendalam dalam penyelenggaraan layanan kesehatan. Konsep "Industri Kesehatan 4.0" merefleksikan pergeseran paradigma dari layanan kesehatan yang bersifat reaktif menuju pendekatan yang prediktif, preventif, personal, dan partisipatif. Transformasi digital ini membuka peluang besar bagi peningkatan efisiensi, kualitas, dan aksesibilitas pelayanan kesehatan. Penggunaan teknologi seperti *cloud computing*, sensor cerdas, dan analitik data memungkinkan tenaga kesehatan dan institusi layanan kesehatan untuk membuat keputusan yang lebih cepat dan tepat berdasarkan data yang tersedia secara *real-time*.

Dalam praktiknya, teknologi ini memfasilitasi sistem monitoring pasien jarak jauh, diagnosa berbasis citra digital dengan dukungan komputer, serta pengelolaan sumber daya rumah sakit secara lebih efisien. Sebagai contoh, pemanfaatan *computer vision* dalam pencitraan medis memungkinkan deteksi dini berbagai jenis penyakit seperti kanker payudara, tumor otak, dan penyakit jantung kongenital pada janin. Teknologi ini juga telah berperan penting dalam penanganan pandemi COVID-19, melalui analisis X-ray dan CT scan untuk mendeteksi infeksi paru-paru. Pendekatan ini tidak hanya meningkatkan kecepatan diagnosis, tetapi juga menurunkan beban kerja tenaga medis dan meminimalkan risiko infeksi silang.

Namun, transformasi ini juga membawa tantangan baru, seperti kebutuhan akan infrastruktur teknologi yang memadai, pelatihan tenaga kerja, serta regulasi mengenai etika dan privasi data pasien. Oleh karena itu, pemahaman yang menyeluruh tentang lanskap digital dalam sektor kesehatan menjadi hal yang sangat penting bagi semua pemangku kepentingan, baik dari sisi tenaga medis, pengambil kebijakan, maupun pengembang teknologi.

## **Peran Data Dalam Pelayanan Kesehatan Modern**

Data merupakan komponen sentral dalam penyelenggaraan layanan kesehatan modern. Dengan volume data yang semakin besar mulai

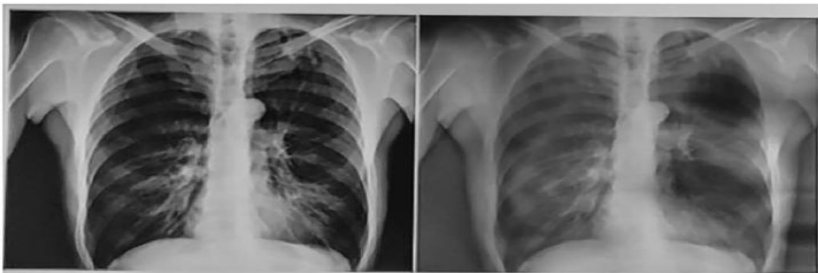
Fungsi prapemrosesan yang efektif sangat penting untuk menurunkan variasi non-klinis antar gambar, seperti perbedaan cahaya atau sudut pengambilan, yang dapat mempengaruhi hasil analisis. Hasil prapemrosesan yang baik akan memperkuat akurasi sistem *computer vision* dan mempercepat waktu pemrosesan.

### Segmentasi, Klasifikasi, dan Deteksi Objek Pada Citra Medis

Dalam analisis citra medis, terdapat tiga fungsi utama yang dilakukan oleh sistem *computer vision*: segmentasi, klasifikasi, dan deteksi objek. Segmentasi adalah proses memisahkan area atau struktur tertentu dari citra, seperti organ, jaringan, atau lesi.

Segmentasi yang akurat sangat penting untuk mengukur ukuran tumor, memodelkan struktur anatomi, atau merencanakan tindakan medis. Teknik segmentasi modern menggunakan metode seperti U-Net atau Mask R-CNN yang mampu menangkap detail spasial dengan presisi tinggi. Klasifikasi bertujuan mengkategorikan citra ke dalam kelas tertentu, seperti "normal" atau "abnormal", atau mengidentifikasi tipe penyakit tertentu.

Model klasifikasi sering menggunakan arsitektur *deep learning* seperti *Convolutional Neural Network* (CNN) yang telah terbukti efektif dalam berbagai tugas klasifikasi citra medis.



**Input: Chest X-Ray Image**

**Output: Pneumonia Positive**

**Gambar16.7: Citra X-Ray Hasil Deteksi Dan Klasifikasi Kondisi Kesehatan Paru Untuk Deteksi Abnormalitas**

Sumber: Rajagopal, n.d.

Deteksi objek adalah proses menemukan dan memberi anotasi pada objek-objek tertentu dalam citra, misalnya tumor, fraktur, atau kelainan bentuk. Teknologi seperti YOLO (*You Only Look Once*) dan

*Faster R-CNN* banyak digunakan untuk mendeteksi lokasi spesifik kelainan dalam satu waktu pandang, sangat berguna untuk analisis cepat oleh radiolog.

Ketiga pendekatan ini tidak hanya dapat bekerja secara terpisah, tetapi juga sering dikombinasikan dalam sistem *pipeline* untuk mendapatkan analisis citra yang menyeluruh. Misalnya, deteksi awal digunakan untuk menemukan area abnormal, kemudian dilakukan segmentasi dan klasifikasi lanjutan untuk penilaian klinis lebih mendalam.

## Daftar Pustaka

Chen, C. H. (2014). *Computer Vision in Medical Imaging*.

Gandhi, T., Bharracharyya, S., De, S., Konar, D., & Dey, S. (2020). *Advance Machine Vision Paradigms For Medical Image Analysis*.

Joshi, M. s., Ransom, E. R., Nash, D. B., & Ransom, S. B. (n.d.). *Healthcare Quality (S. B. Ransom (ed.); third edit)*. Health Administration Press, Chicago, Illinois.

Rajagopal, K. P. (n.d.). *Healthcare Industry 4.0*. Taylor & Francis Group.  
<https://doi.org/10.1201/9781003345411Typeset>.

## PROFIL PENULIS



### **Inda Rusdia Sofiani, S.T., M.Sc.**

Ketertarikan penulis terhadap bidang elektronika dan teknologi mekatronika tumbuh sejak masa perkuliahan di Universitas Brawijaya, Malang, tempat penulis menempuh pendidikan Sarjana Teknik Elektro dan lulus pada tahun 1999. Keingintahuan yang besar terhadap sistem cerdas dan otomasi membawanya melanjutkan studi ke jenjang Magister (S2) di bidang Mekatronika di FH Aachen, Jerman, dan lulus pada tahun 2009. Saat ini, penulis tengah menyelesaikan studi doktoral (S3) di bidang Teknik Elektro, Universitas Brawijaya. Penulis merupakan dosen tetap di Program Studi D3-Teknologi Elektronika, Universitas Muhammadiyah Malang, dan aktif mengajar berbagai mata kuliah seperti Mekatronika, Sistem Instrumentasi, Pengolahan Citra *Digital*, Robotika, dan Kecerdasan Buatan. Di samping aktivitas mengajar, penulis juga aktif dalam penelitian terkait pemanfaatan teknologi *computer vision* untuk deteksi dini penyakit, peningkatan kualitas citra medis, dan pemrosesan sinyal EEG. Beberapa penelitian penulis yang telah didanai oleh Universitas Muhammadiyah Malang antara lain mencakup deteksi citra paru-paru untuk identifikasi COVID-19, serta penghapusan *noise* dan refleksi *specular* pada gambar serviks. Komitmen terhadap riset yang aplikatif dan relevan dengan dunia kesehatan digital menjadikan penulis sebagai salah satu figur penting dalam pengembangan teknologi vokasi berbasis data dan AI.

Email Penulis: indarusdias05@umm.ac.id.



**BAB 17**  
**PEMODELAN DAN**  
**VISUALISASI DI SEKTOR**  
***E-COMMERCE***

---

**Aqil Farras, S.T., MMSI.**  
Universitas Muhammadiyah Riau



Seiring dengan meningkatnya kompleksitas dan persaingan dalam industri *e-commerce*, pemanfaatan data tidak lagi menjadi keunggulan tambahan, tetapi telah berubah menjadi kebutuhan mendasar.

Perusahaan yang mampu mengelola dan menginterpretasi data secara efektif akan lebih unggul dalam merespons perubahan permintaan pasar, mengurangi risiko operasional, serta menciptakan nilai tambah bagi pelanggan. Dalam konteks ini, pemodelan dan visualisasi bukan hanya sekadar alat teknis, tetapi telah menjadi bagian integral dari strategi bisnis yang berorientasi pada data (*data-driven strategy*).

Dengan berkembangnya teknologi seperti kecerdasan buatan, big data, dan *Internet of Things* (IoT), masa depan *e-commerce* sangat bergantung pada kemampuan aktor-aktor bisnis dalam memanfaatkan data secara cerdas dan terstruktur. Oleh karena itu, pemodelan dan visualisasi data bukan hanya relevan untuk saat ini, tetapi akan terus menjadi komponen strategis dalam menjawab tantangan dan memanfaatkan peluang di era *digital* yang terus berubah.

## **Pemodelan Data dan Visualisasi Interaktif Untuk Optimalisasi Strategi *E-Commerce***

Dalam ekosistem *e-commerce* yang sarat dinamika, strategi bisnis yang efektif sangat bergantung pada kemampuan dalam membaca dan merespons pola-pola yang terbentuk dari aktivitas pengguna. Di sinilah peran pemodelan data menjadi krusial.

Pemodelan merupakan proses representasi sistematis dari realitas bisnis ke dalam bentuk struktur data yang dapat dianalisis. Proses ini tidak hanya menyusun ulang data mentah agar lebih terstruktur, tetapi juga mengekstraksi makna dan pola dari data dalam skala besar.

### **1. Pemodelan Data Sebagai Pondasi Strategi**

Sistem ini memprediksi preferensi pengguna berdasarkan aktivitas pengguna lain yang serupa atau karakteristik produk yang telah diminati sebelumnya. Model ini tidak hanya meningkatkan kemungkinan transaksi, tetapi juga memperkuat loyalitas pelanggan melalui pengalaman belanja yang lebih personal.

meningkatkan CTR (*Click-Through Rate*) hingga 23% dibandingkan hanya menggunakan deskripsi teks biasa.

Sementara itu, grafik tren yang menunjukkan stok yang terus menurun atau harga yang akan naik dalam waktu dekat menciptakan rasa urgensi (*sense of urgency*) yang berujung pada peningkatan konversi. Hal ini menunjukkan bahwa visualisasi juga dapat mengaktifkan aspek psikologis dalam perilaku konsumen.

#### 4. Integrasi Preferensi Visual Dengan Teknologi AI

Banyak *platform* saat ini mulai mengintegrasikan AI untuk menyesuaikan tampilan visualisasi berdasarkan perilaku pengguna secara *real-time*. Misalnya, sistem dapat menampilkan infografik produk untuk pengguna baru yang datang dari media sosial, sedangkan pengguna lama yang mengunjungi lewat *email* mungkin akan disuguhi grafik tren harga atau ketersediaan stok.

Visualisasi berbasis AI ini memanfaatkan data interaksi historis untuk mengubah elemen visual sesuai profil pengguna, menciptakan pengalaman yang jauh lebih relevan dan personal. Pendekatan ini telah terbukti mengurangi bounce rate hingga 18% pada situs yang mengimplementasikannya secara konsisten.

#### 5. Strategi Optimalisasi Visualisasi Berdasarkan Preferensi

Agar dapat memanfaatkan potensi penuh dari visualisasi yang disesuaikan dengan preferensi pelanggan, pelaku *e-commerce* perlu menerapkan beberapa langkah berikut:

- a. Lakukan *A/B Testing* Visualisasi: uji dua atau lebih jenis visual untuk melihat mana yang paling efektif dalam menghasilkan konversi.
- b. Segmentasi Berdasarkan Perangkat & Demografi: gunakan visualisasi yang berbeda antara pengguna *mobile* dan *desktop* serta sesuaikan dengan kelompok usia.
- c. Gunakan *Heatmap* & *Scrollmap* Secara Berkala: untuk memvalidasi apakah elemen visual yang paling penting benar-benar dilihat dan diklik.
- d. Perbarui Visual Secara Berkala: visual yang sama terlalu lama dapat membuat pengguna tidak lagi memperhatikannya (*banner blindness*).

- e. Integrasikan Visual dengan Sistem Rekomendasi Produk: contoh, tampilkan grafik perbandingan produk serupa untuk membantu proses seleksi.

### **Simpulan Strategis**

Preferensi pelanggan terhadap jenis visualisasi dalam e-commerce sangat beragam dan bergantung pada banyak faktor seperti usia, perangkat, dan tujuan kunjungan. Memahami preferensi ini dan mengadaptasikannya ke dalam desain *visual platform* tidak hanya meningkatkan kenyamanan pengguna, tetapi juga memperbesar potensi konversi dan loyalitas. Penerapan strategi visual yang tepat, didukung oleh analitik yang akurat dan pendekatan berbasis data, menjadikan visualisasi bukan lagi pelengkap, melainkan komponen utama dalam desain pengalaman pelanggan modern.

## Daftar Pustaka

- Bhargava, H. K., & Chen, C.-Y. (2020). *Decision Support Systems For E-Commerce: A Research Agenda. Information Systems And E-Business Management*, 18, 1–20. <https://doi.org/10.1007/s10257-019-00456-2>.
- Chaffey, D. (2020). *Digital marketing: Strategy, Implementation And Practice (7th ed.)*. Pearson Education.
- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4), 1165–1188. <https://doi.org/10.2307/41703503>
- Davenport, T. H., & Harris, J. G. (2007). *Competing on Analytics: The New Science Of Winning*. Harvard Business Review Press.
- Gandomi, A., & Haider, M. (2015). Beyond The Hype: Big Data Concepts, Methods, And Analytics. *International Journal of Information Management*, 35(2), 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>.
- Google. (2022). *Think With Google: E-commerce Insights*. <https://www.thinkwithgoogle.com/>.
- Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: Concepts And Techniques (3rd ed.)*. Elsevier.
- Hariri, R. H., Fredericks, E. M., & Bowers, K. M. (2019). Uncertainty In Big Data Analytics: Survey, Opportunities, And Challenges. *Journal of Big Data*, 6(44). <https://doi.org/10.1186/s40537-019-0206-3>.
- Kelleher, J. D., & Tierney, B. (2018). *Data Science: An introduction*. MIT Press.
- Kotler, P., Kartajaya, H., & Setiawan, I. (2021). *Marketing 5.0: Technology For Humanity*. Wiley.
- Kumar, V., & Reinartz, W. (2016). *Creating Enduring Customer Value. Journal of Marketing*, 80(6), 36–68. <https://doi.org/10.1509/jm.15.0414>.
- McKinsey Global Institute. (2011). *Big Data: The Next Frontier For Innovation, Competition, And Productivity*. <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation>.

- Provost, F., & Fawcett, T. (2013). *Data Science For Business: What You Need To Know About Data Mining And Data-Analytic Thinking*. O'Reilly Media.
- Sharda, R., Delen, D., & Turban, E. (2020). *Analytics, Data Science, & Artificial Intelligence: Systems For Decision Support (11th ed.)*. Pearson.
- Statista. (2024). *E-commerce worldwide-Statistics & Facts*. <https://www.statista.com/topics/871/online-shopping/>.
- Tableau Software. (2023). *Data Visualization for E-commerce*. <https://www.tableau.com/solutions/ecommerce>.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2019). *Introduction To Data Mining (2nd ed.)*. Pearson.

## PROFIL PENULIS



### **Aqil Farras, S.T., MMSI.**

Aqil Farras lahir pada 23 Oktober di Jakarta, tumbuh sebagai anak yang penuh energi dan tantangan. Masa kecilnya tidak selalu mudah dikenal nakal dan sulit diatur namun hidup memberinya jalan untuk berubah. Pendidikan menengah di SMAN 1 Kuantan Mudik menjadi titik balik: di sanalah ia mengenal teknologi, dan dari situlah transformasi dimulai.

Berbekal tekad kuat, ia merantau ke ibu kota dan menuntut ilmu di Universitas Gunadarma, salah satu kampus terbaik di Indonesia. Dalam waktu lima tahun, ia menuntaskan studi S1 di bidang Rekayasa Perangkat Lunak dan S2 di Sistem Informasi Bisnis, perjalanan cepat yang mencerminkan semangat, kedisiplinan, dan visi besarnya. Hobi yang dulunya sekadar bermain komputer berubah menjadi profesi. Sebagai *developer*, Aqil tidak hanya membangun sistem, tapi juga membangun harapan: membantu masyarakat, terutama generasi muda dari daerah, agar tidak tertinggal dalam arus teknologi.

Hobi yang dulunya sekadar bermain komputer perlahan menjelma menjadi profesi. Sebagai *developer*, Aqil tidak hanya membangun sistem, tetapi juga membangun harapan: membantu masyarakat, terutama generasi muda dari daerah, agar tidak tertinggal dalam arus teknologi. Kini, ia mengajar sebagai dosen di sebuah perguruan tinggi di Pekanbaru, membagikan ilmu dan inspirasinya di bidang yang ia kuasai. Bagi para mahasiswanya, Aqil bukan sekadar pengajar, tetapi juga panutan yang menunjukkan bahwa perubahan itu nyata asal mau berusaha.

Saya adalah bukti hidup bahwa masa lalu bukanlah penghalang, dan teknologi bisa menjadi jembatan menuju masa depan yang lebih cerah. Dari anak "nakal" menjadi akademisi dan inovator, kisahnya menjadi pengingat bahwa mimpi besar tak pernah terlalu jauh untuk dijangkau. Kini, penulis mengajar sebagai dosen di sebuah perguruan tinggi di Pekanbaru, membagikan ilmu dan inspirasinya di bidang yang ia kuasai. Aqil Farras adalah bukti nyata bahwa perubahan adalah mungkin teknologi bisa menjadi jembatannya.

# PEMODELAN DAN VISUALISASI DATA

Di era informasi yang masif ini, data telah menjadi salah satu aset paling berharga. Namun, data mentah, seberapa pun melimpahnya, tidak akan memberikan nilai tanpa kemampuan untuk memahami dan mengomunikasikannya secara efektif. Inilah inti dari pemodelan dan visualisasi data merupakan sebuah disiplin yang menjembatani kesenjangan antara kumpulan angka dan wawasan yang dapat ditindaklanjuti. Buku ini hadir sebagai panduan komprehensif bagi siapa saja yang ingin menguasai seni dan sains di balik pemodelan serta visualisasi data, konsep dasar yang fundamental hingga teknik-teknik lanjutan yang transformatif. Anda akan belajar bagaimana memahami struktur intrinsik data, memilih model statistik atau pembelajaran mesin yang paling tepat untuk mengungkap pola dan hubungan yang tersembunyi, dan yang terpenting, bagaimana menyajikan temuan-temuan tersebut dalam bentuk visual yang tidak hanya menarik secara estetika, tetapi juga mudah dipahami, informatif, dan memiliki dampak nyata. Dalam setiap bab, kami telah berupaya keras untuk menyajikan materi ini dengan cara yang jelas, praktis, dan relevan dengan tantangan dunia nyata. Anda akan menemukan beragam contoh kasus, studi kasus mendalam, dan latihan praktis yang dirancang khusus untuk memperdalam pemahaman konseptual sekaligus mengasah keterampilan teknis. Topik utama yang dibahas pada buku ini adalah:

1. Konsep Dasar Pemodelan dan Visualisasi Data
2. Persiapan Data (*Data Preparation*)
3. Eksplorasi Data (*Exploratory Data Analysis/EDA*)
4. Pemilihan dan Rekayasa Fitur
5. Pemodelan Layanan Visualisasi Data Awan Untuk Pemantauan Data Multisensor
6. Pengolahan dan Visualisasi Data (PVD)
7. *Big Data* dan *Framework* Pemrosesan Data
8. Prinsip Dasar Visualisasi Data
9. Jenis-Jenis Visualisasi Data
10. Alat dan Platform Visualisasi Data
11. Visualisasi Data untuk Analisis Spasial
12. Visualisasi Data Waktu (*Time Series*)
12. Visualisasi Data untuk *Machine Learning*
14. Etika dan Keamanan dalam Pemodelan Data
15. Integrasi Model Data ke Sistem Produksi
16. Pemodelan dan Visualisasi di Sektor Kesehatan
17. Pemodelan dan Visualisasi di Sektor *E-Commerce*