



PROGRAM STUDI
TEKNIK INFORMATIKA – S1
FAKULTAS ILMU KOMPUTER
UNIVERSITAS DIAN NUSWANTORO

MATA KULIAH
DATA MINING



[Technology vector created by sentavio - www.freepik.com](https://www.freepik.com/vectors/technology)

DATA MINING

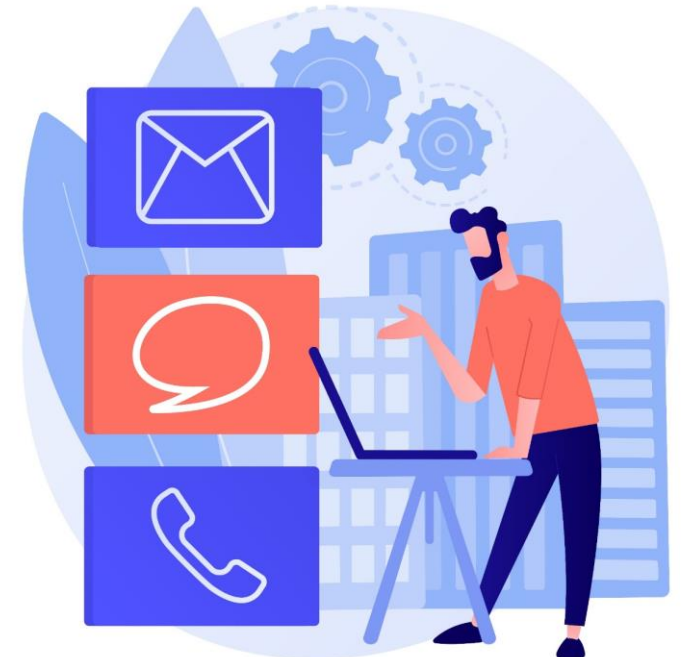
“Pengantar Data Mining”

TIM PENGAMPU DOSEN DATA MINING

2023

Kontak Dosen

- Junta Zeniarja, M.Kom
- Email: junta@dsn.dinus.ac.id
- Youtube : <https://www.youtube.com/JuntaZeniarja>
- Scholar : <http://bit.do/JuntaScholar>



CAPAIAN PEMBELAJARAN

Capaian Pembelajaran Mata Kuliah

M1	Mahasiswa mampu menjelaskan tentang konsep dasar data mining dan aplikasinya di bidang Teknologi Informasi.
M2	Mahasiswa mampu menjelaskan tentang konsep data untuk data mining, peran utama data mining dan proses data mining dengan CRISP-DM (Cross Industry Standard Process for Data Mining).
M3	Mahasiswa mampu merumuskan tahapan dalam pemrosesan awal data (preprocessing data) dan mengimplementasikannya ke dalam pemrograman python.
M4	Mahasiswa mampu menjelaskan konsep metode learning pada data mining dan hubungannya dengan 5 peran utama data mining (estimasi, prediksi, klasifikasi, klastering dan asosiasi).
M5	Mahasiswa mampu memahami dan menjelaskan konsep klasifikasi dengan algoritma Naïve Bayes serta mengimplementasikannya ke dalam pemrograman python.
M6	Mahasiswa mampu memahami dan menjelaskan konsep klasifikasi dengan algoritma K-Nearest Neighbor (K-NN) serta mengimplementasikannya ke dalam pemrograman python.
M7	Mahasiswa mampu memahami dan menjelaskan konsep klasifikasi dengan algoritma Decision Tree serta mengimplementasikannya ke dalam pemrograman python.
M8	Mahasiswa mampu menjelaskan konsep dasar tentang ANN (Artificial Neural Network) dan Deep Learning serta mengimplementasikannya ke dalam pemrograman python.
M9	Mahasiswa mampu memahami dan menjelaskan konsep klastering, jenis – jenis klastering serta contoh penerapannya.
M10	Mahasiswa mampu memahami dan menjelaskan konsep klastering dengan algoritma K-Means serta mengimplementasikannya ke dalam pemrograman python.
M11	Mahasiswa mampu memahami dan menjelaskan konsep klastering dengan algoritma AHC (Agglomerative Hierarchical Clustering) serta mengimplementasikannya ke dalam pemrograman python.
M12	Mahasiswa mampu memahami dan menjelaskan konsep dasar regresi serta mengimplementasikannya ke dalam pemrograman python.
M13	Mahasiswa mampu memahami dan menjelaskan konsep asosiasi dengan algoritma Apriori serta mengimplementasikannya ke dalam pemrograman python.
M14	Mahasiswa mampu memahami dan menjelaskan konsep validasi atau pengujian model serta mengimplementasikannya ke dalam pemrograman python.

MATERI PERKULIAHAN

Materi Pra UTS

- | | |
|----|--|
| #1 | Pengantar Data Mining |
| #2 | Data utk Data Mining (Jenis2 Data, Pengukuran Data, Nilai dan Atribut) |
| #3 | Preprosesing Data (Data Cleaning, Missing Value, Transformasi Data, koding python) |
| #4 | Metode Learning (Disiplin Data Mining, Supervised & Unsupervised, Klasifikasi,Prediksi, Estimasi, Klastering,dan Asosiasi) |
| #5 | Klasifikasi dengan Naive Bayes + Python |
| #6 | Klasifikasi dengan KNN + Python |
| #7 | Klasifikasi Decision Tree + Python |

Evaluasi Tengah Semester (UTS)

Materi Pasca UTS

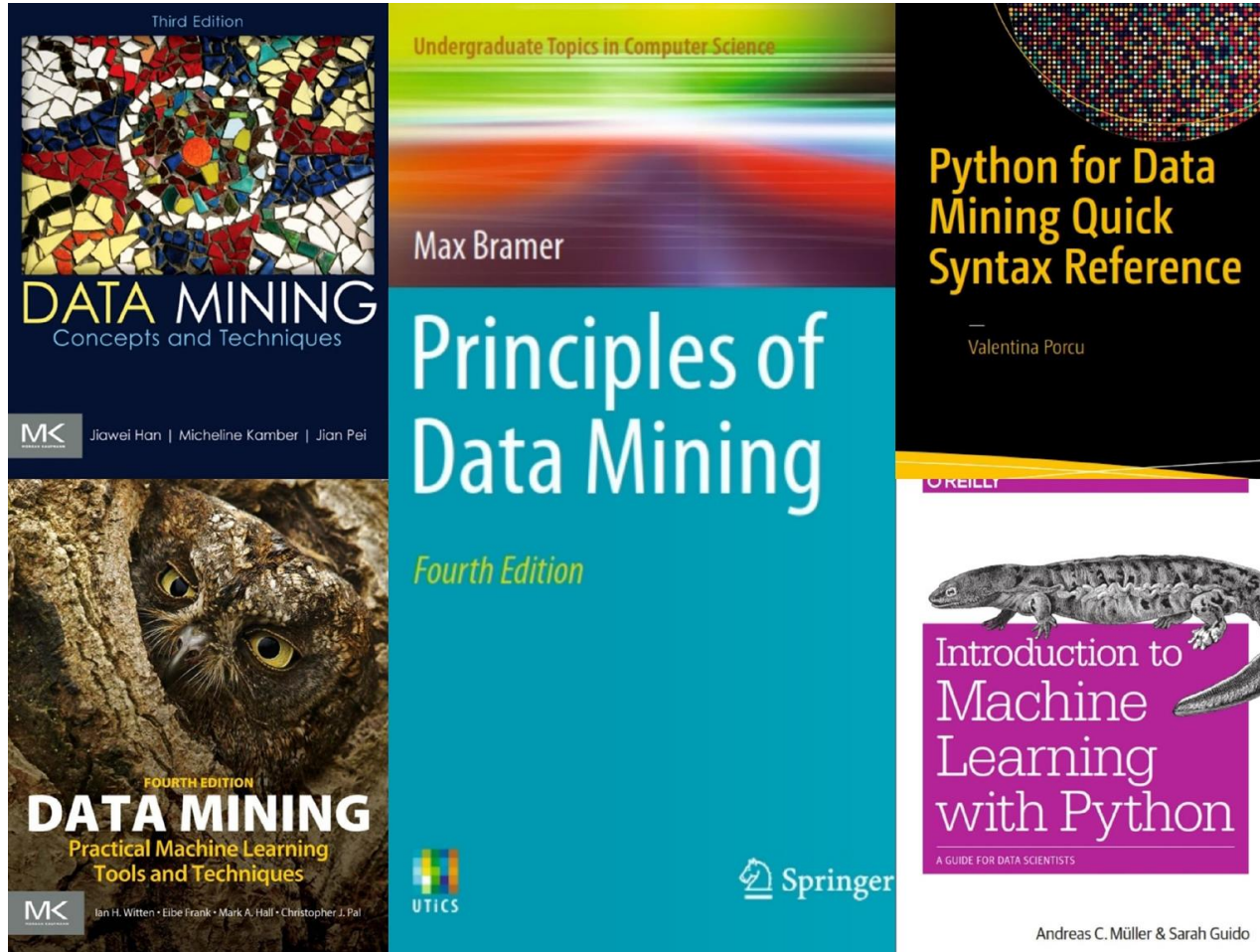
- | | |
|-----|--|
| #8 | ANN & Deep Learning + Python |
| #9 | Klastering (Teknik Klaster, Metode Partisi, Metode Hirarkis) |
| #10 | Metode Partisi (K-Means Klastering + Python) |
| #11 | Metode Hirarkis (HAC + Python) |
| #12 | Regresi (Sederhana dan Multivariate) + Python |
| #13 | Asosiasi + Apriori / FP-Growth + Python |
| #14 | Validasi dan Pengujian Model |

Evaluasi Akhir Semester (UAS)

KONTRAK PERKULIAHAN

- **Kehadiran** : 75 % (Absensi melalui media Forum Diskusi KULINO)
- **Bobot Penilaian**
 - TUGAS : 40% (Akumulasi BOBOT Kehadiran, Keaktifan, Sikap dan Tugas)
 - UTS : 25%
 - UAS : 35%
- **Perkuliahahan** : Synchronous (Google Meet), Asynchronous (Kulino), Youtube dan Media Interaktif Grup WA Kelas.
- **Tugas Akhir Makul**
 - Pembuatan Produk Data Mining menggunakan Python

SUMBER PUSTAKA



- Jiawei Han, Micheline Kamber, Jian Pei, Data mining : concepts and techniques – 3rd ed, Elsevier, 2012.
- Ian H. Witten, Eibe Frank, Mark A. Hall, Christopel J. Pal, Data Mining: Practical Machine Learning Tools and Techniques - 4th ed, Elsevier, 2017.
- Andreas C. Muller & Sarah Guido, Introduction to Machine Learning with Python, O'Reilly, 2017.
- Valentina Porcu, Python for Data Mining Quick Syntax Reference, Apress, 2018.
- Max Bramer, Principles of Data Mining – Undergraduate Topics in Computer Science – 4th ed, Springer, 2020.
- Sumber gambar: www.freepik.com



Pengantar Data Mining

Pertemuan 1

MINING vs WAREHOUSING

Mining



Warehousing

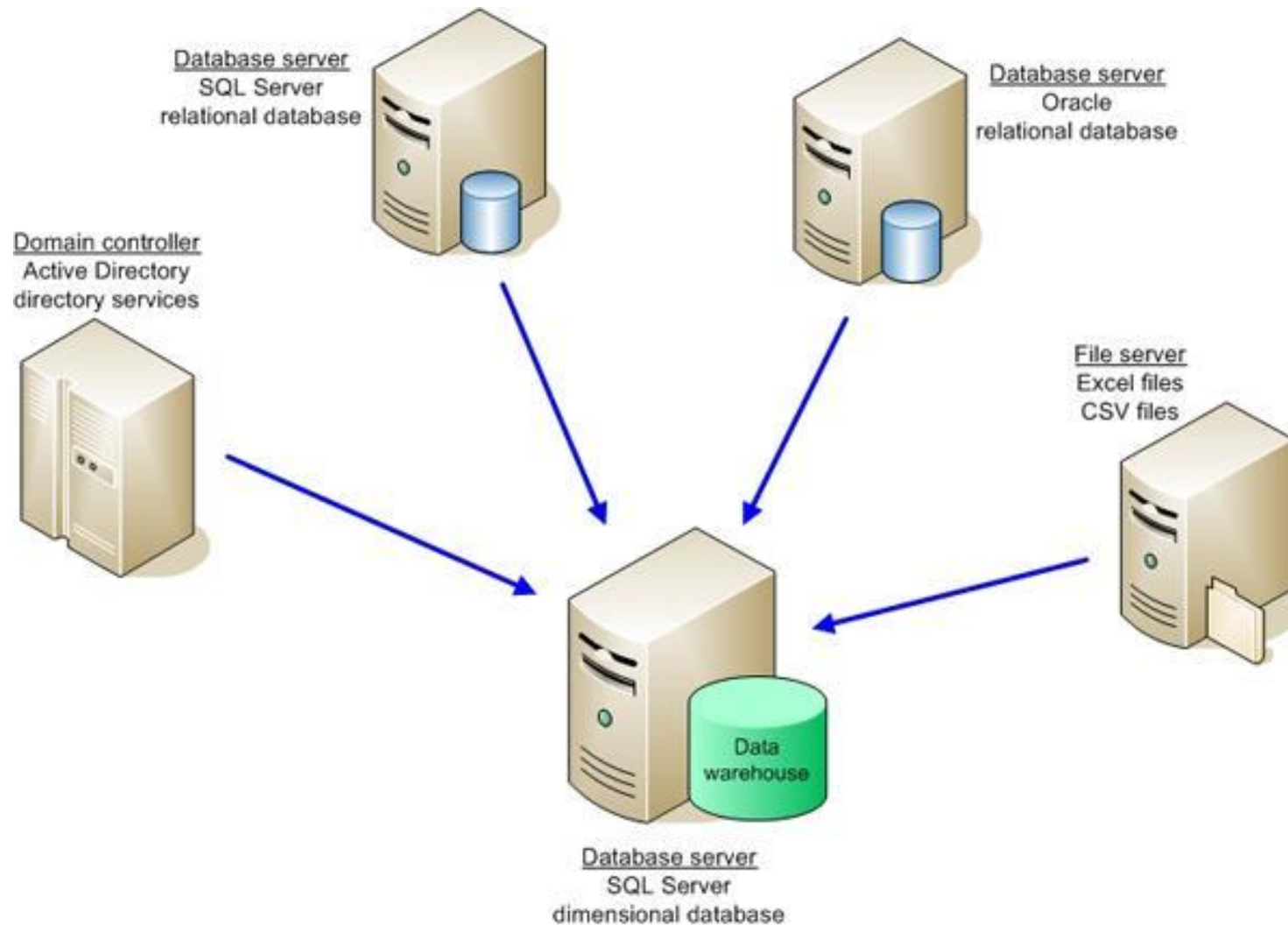


The World of Data





Heterogeneous Data



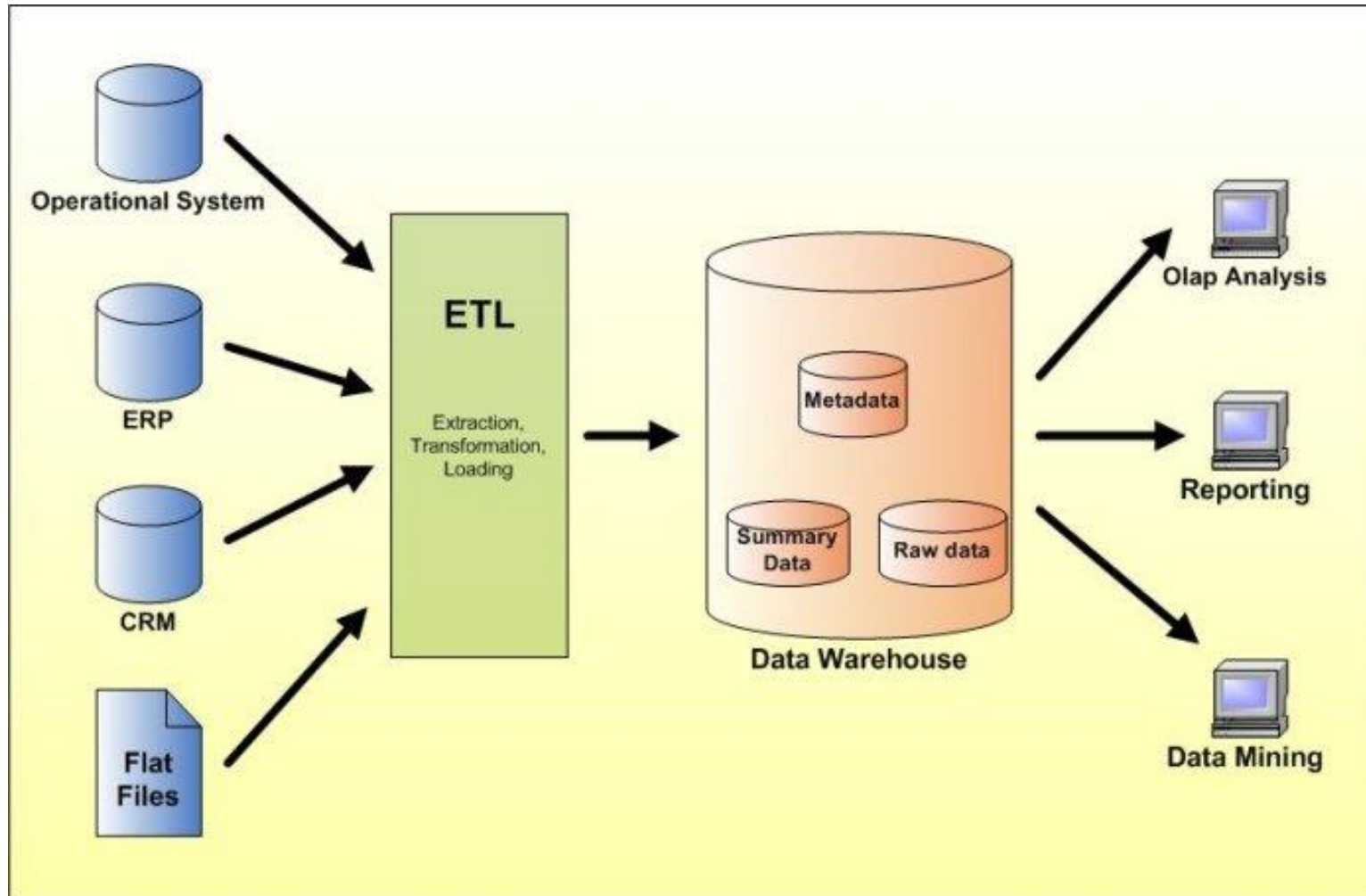
Data Rich, Information Poor



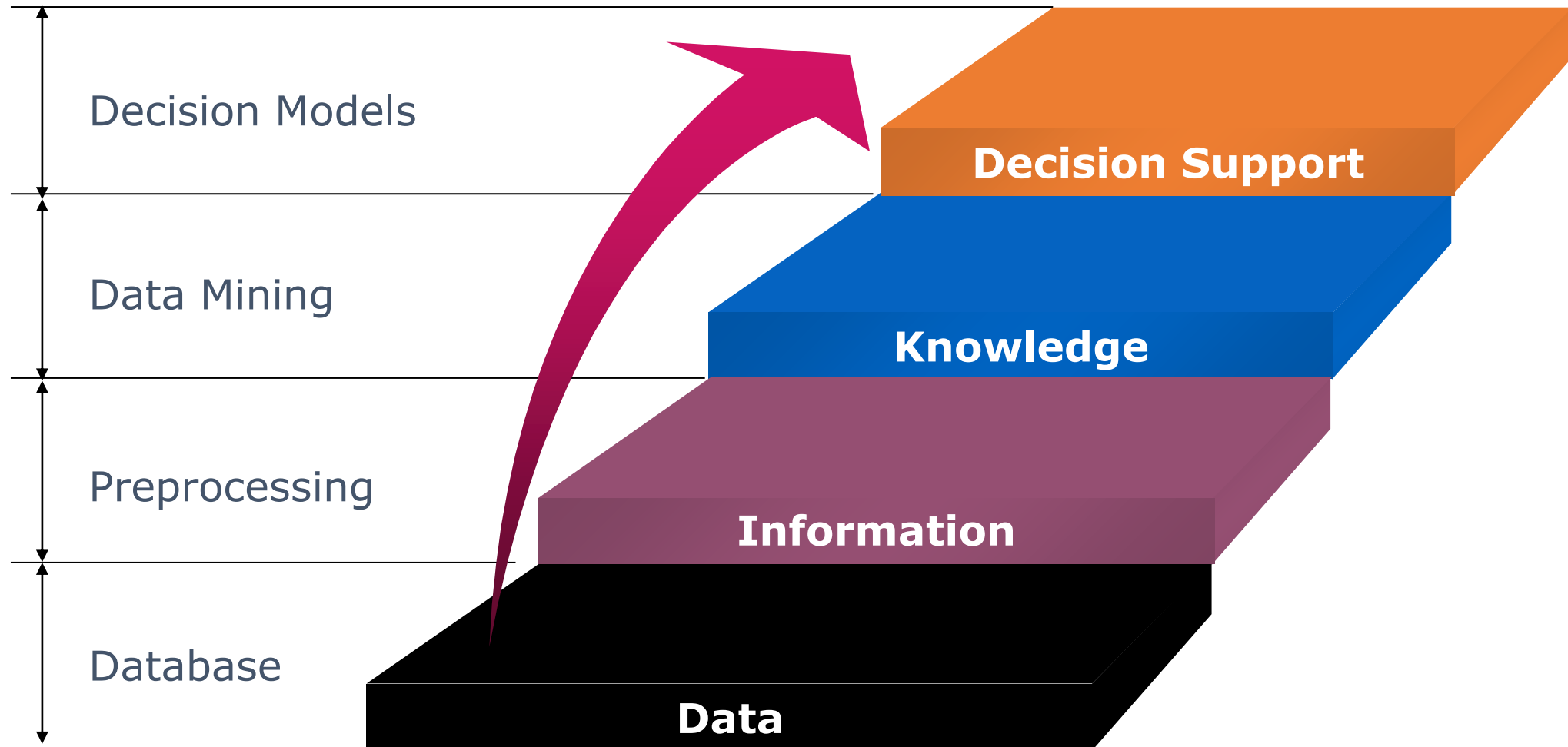
Business Intelligence



Data Integration & Analysis



From Data To Intelligence



It is all about data ...

Retail

|

Financial Institutions

Healthcare

WWW

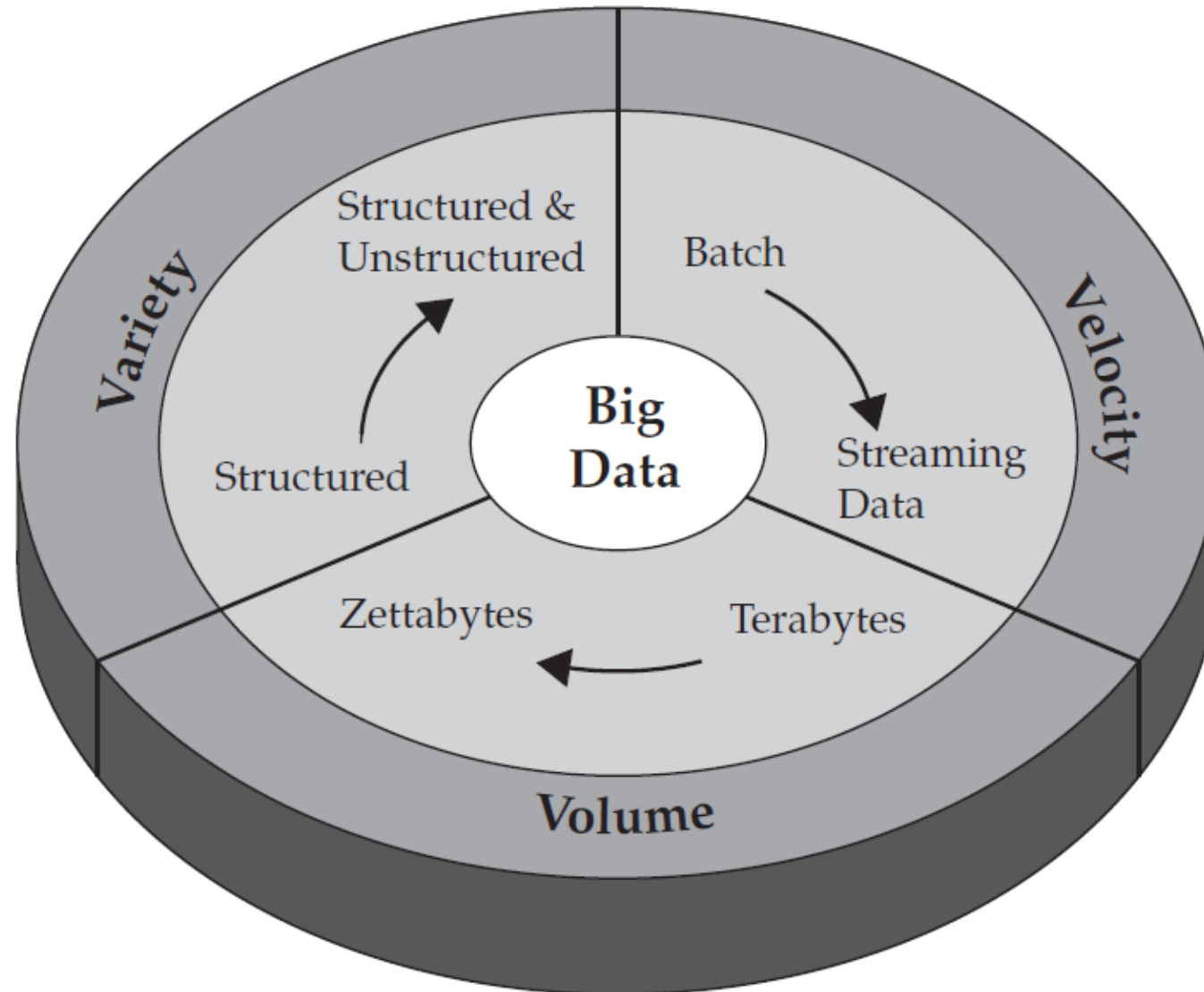
Consulting Companies

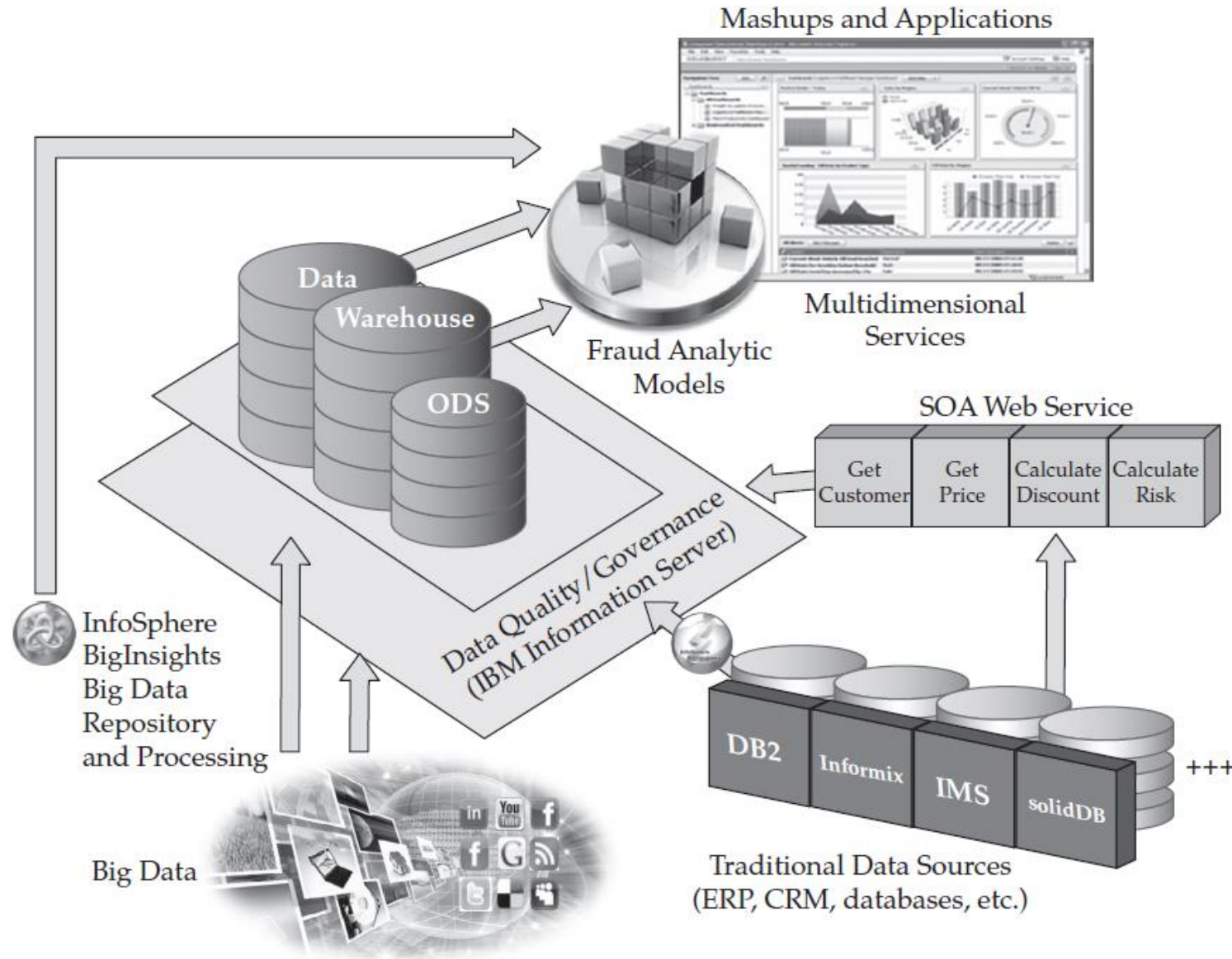
Government

Bioinformatics

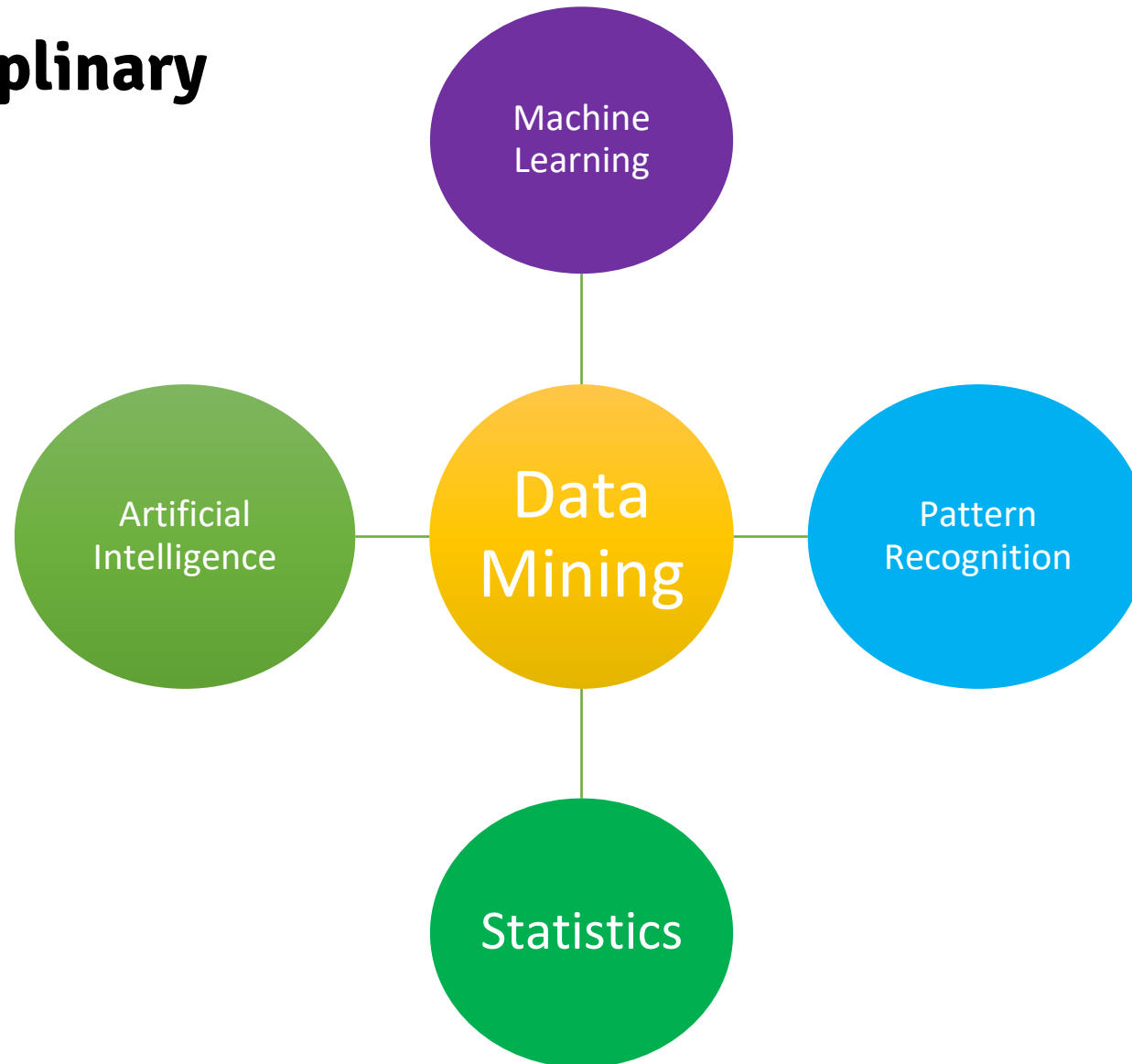
Telecommunications

Big Data

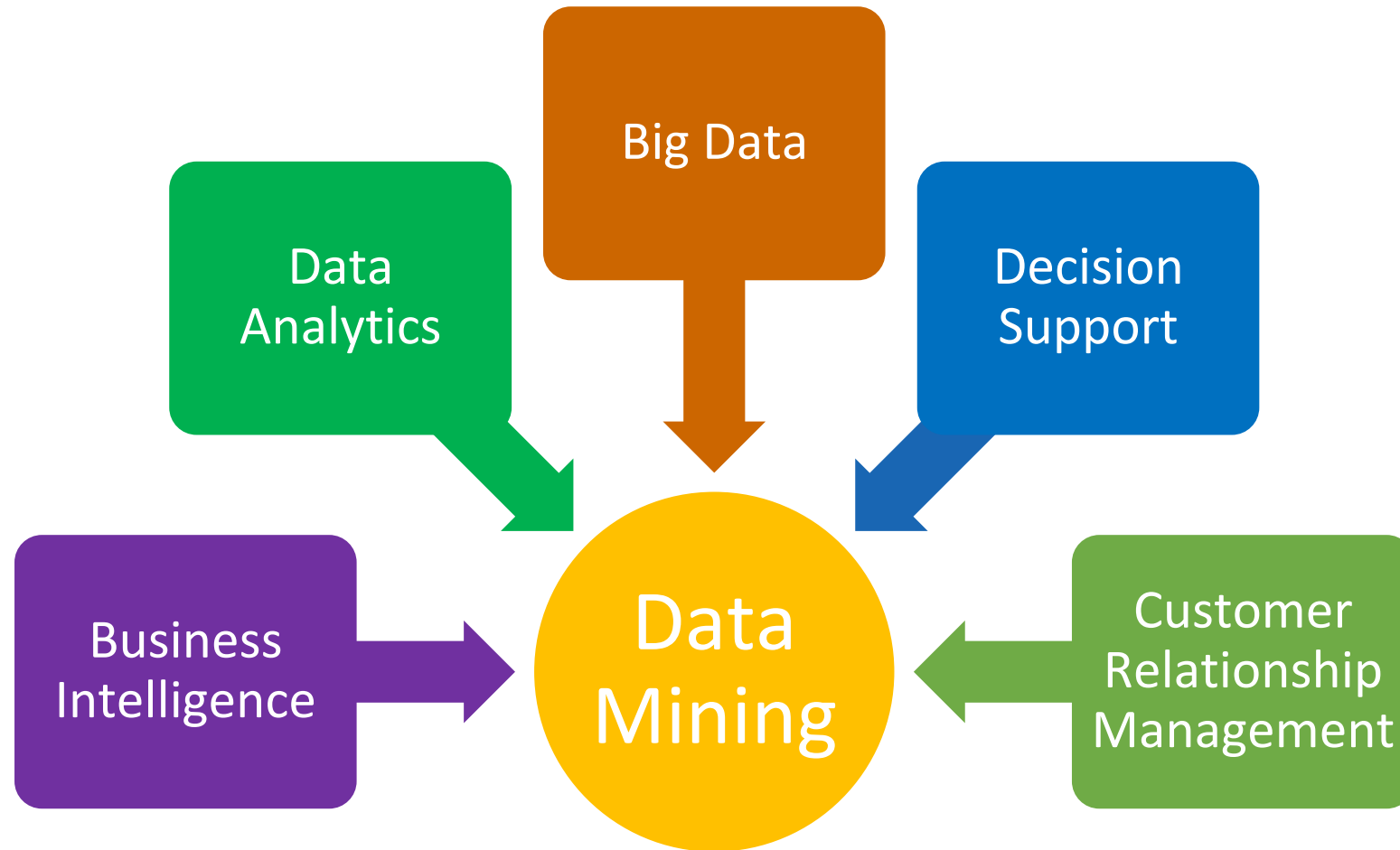




Interdisciplinary



Ubiquitous





Apa itu Data Mining?

Apa itu Data Mining?



- Disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data yang besar.
 - **Data:** fakta yang terekam dan tidak membawa arti.
 - **Informasi:** penjelasan, rangkuman, rekap dan statistic dari data.
 - **Pengetahuan:** pola, aturan atau model yang muncul dari data.
- Sehingga Data mining sering disebut Knowledge Discovery in Database (KDD).
- Konsep Transformasi:

Data ? Informasi ? Pengetahuan ? Kebijakan

Data

- Tidak membawa arti, merupakan kumpulan dari fakta-fakta tentang suatu kejadian.
- Suatu catatan terstruktur dari suatu transaksi.
- Merupakan materi penting dalam membentuk informasi.

Informasi

- Hasil olahan dari data setelah direkap atau dirangkum.
- Hasil statistik dari data.
- Sekumpulan data atau fakta yang telah diproses dan diolah sedemikian rupa sehingga menghasilkan sesuatu yang bisa dipahami dan memberikan manfaat bagi penerimanya.

Pengetahuan

- Gabungan dari suatu pengalaman, nilai, informasi kontekstual dan juga pandangan pakar yang memberikan suatu framework untuk mengevaluasi dan menciptakan pengalaman baru dan informasi (*Thomas H. Davenport, Laurence Prusak*).
- Bisa berupa solusi pemecahan suatu masalah, petunjuk suatu pekerjaan dan ini bisa ditingkatkan nilainya, dipelajari dan juga bisa diajarkan kepada yang lain.

Data => Informasi => Pengetahuan => Kebijakan

- Data Kehadiran Karyawan

Id Karyawan	Tanggal	Waktu Datang	Waktu Pulang
ID0001	03/03/2021	07:10	15:45
ID0002	03/03/2021	07:25	15:35
ID0003	03/03/2021	07:56	16:01
ID0004	03/03/2021	08:01	15:16
ID0005	03/03/2021	07:07	16:34

Data => Informasi => Pengetahuan => Kebijakan [2]

- Informasi Akumulasi Bulanan Kehadiran Karyawan

ID Karyawan	Masuk	Alpa	Cuti	Sakit	Telat
ID0001	25				
ID0002	19	5		1	
ID0003	11	4	10		
ID0004	10	7		4	4
ID0005	9	1		10	5

Data => Informasi => Pengetahuan => Kebijakan [3]

- Pola Kebiasaan Kehadiran Mingguan Karyawan

	Senin	Selasa	Rabu	Kamis	Jum'at
Terlambat	8	1	0	1	5
Pulang Awal	0	1	0	2	9
Izin	4	0	1	0	4
Alpa	3	0	2	0	3

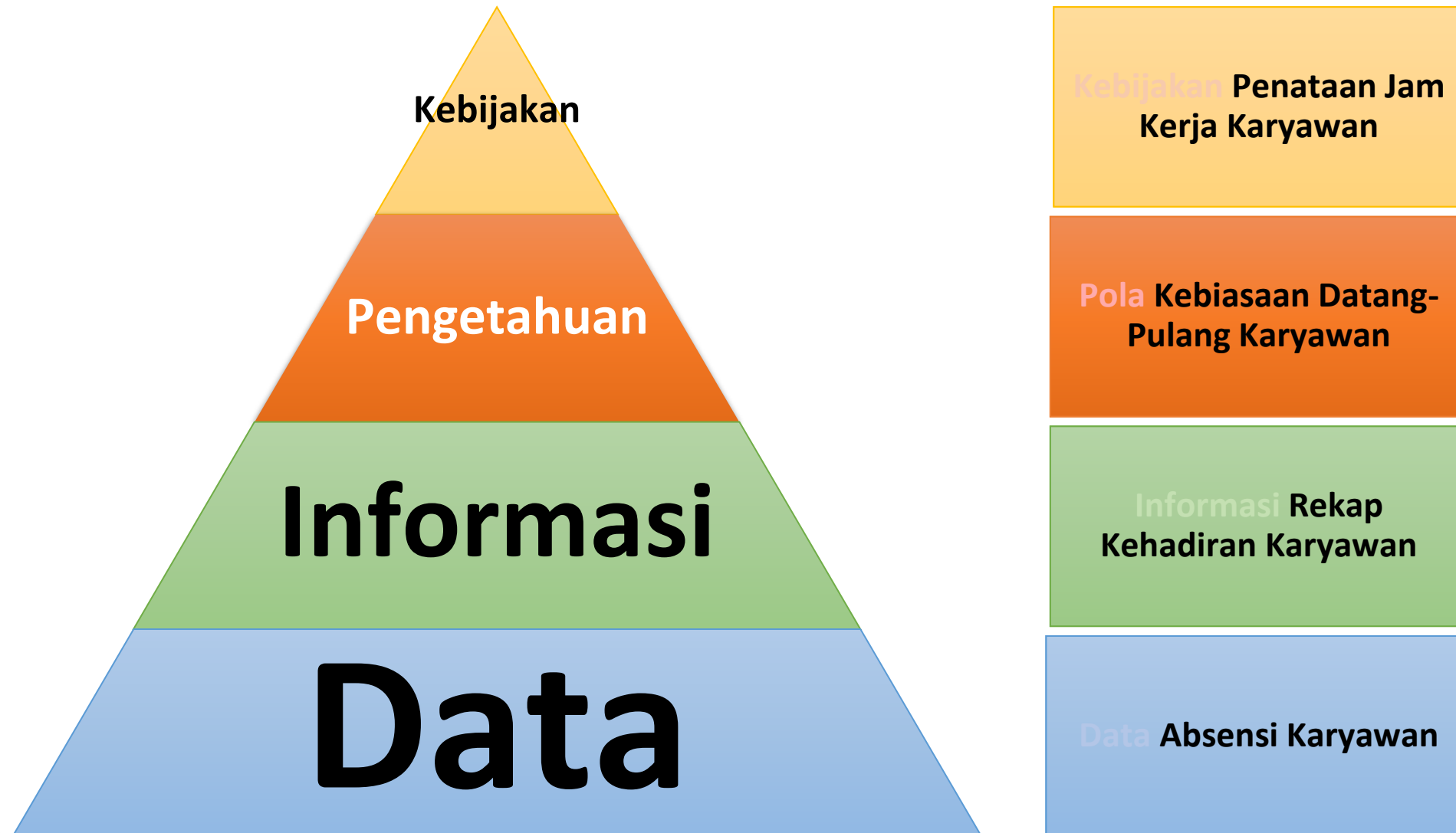
Data => Informasi => Pengetahuan => Kebijakan [4]

- Pengetahuan tentang kebiasaan karyawan dalam jam datang / pulang kerja.
- Pengetahuan tentang bagaimana teknik untuk meningkatkan kehadiran karyawan ? **kebijakan**

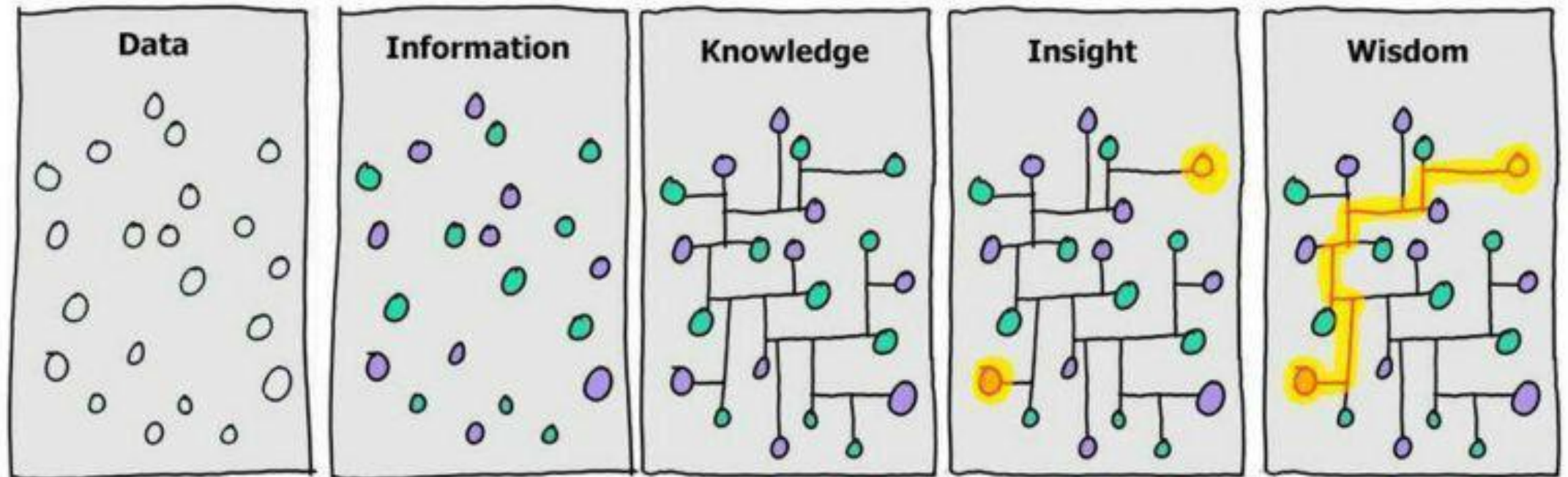
Data => Informasi => Pengetahuan => Kebijakan [5]

- Kebijakan penataan jam kerja karyawan khusus untuk hari Senin dan Jum'at.
- Peraturan Jam Kerja :
 - Hari Senin dimulai jam 10:00.
 - Hari Jumat diakhiri jam 14:00.
 - Sisa jam kerja dikompensasi ke hari lain.

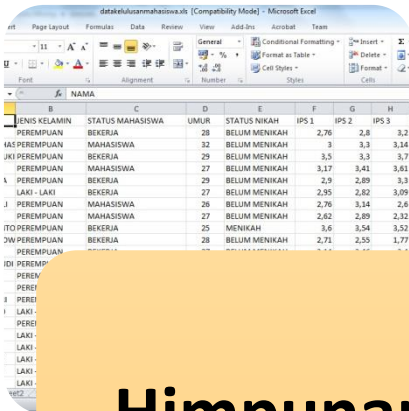
Data => Informasi => Pengetahuan => Kebijakan [6]



Data => Informasi => Pengetahuan => Kebijakan [7]



Konsep Proses Data Mining

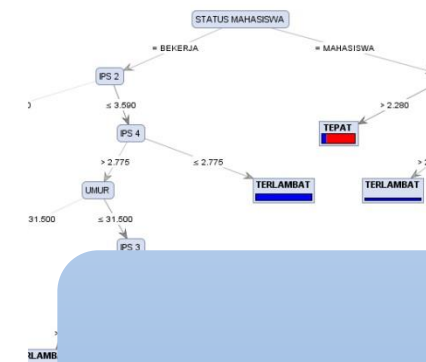


	B	C	D	E	F	G	H
	JENIS KELAMIN	STATUS MAHASISWA	UMUR	STATUS NIKAH	IPS 1	IPS 2	IPS 3
	PEREMPUAN	BEKERJA	28	BELUM MENIKAH	2,76	2,8	3,2
	MAHASISWA	MAHASISWA	32	BELUM MENIKAH	3	3,3	3,14
	PEREMPUAN	BEKERJA	29	BELUM MENIKAH	3,5	3,3	3,7
	MAHASISWA	MAHASISWA	27	BELUM MENIKAH	3,17	3,41	3,81
	PEREMPUAN	BEKERJA	29	BELUM MENIKAH	2,9	2,89	3,3
	LAKI - LAKI	BEKERJA	27	BELUM MENIKAH	2,95	2,82	3,09
	MAHASISWA	MAHASISWA	26	BELUM MENIKAH	2,76	3,14	2,6
	PEREMPUAN	BEKERJA	27	BELUM MENIKAH	2,62	2,89	2,32
	MAHASISWA	MAHASISWA	25	MENIKAH	3,6	3,54	3,52
	PEREMPUAN	BEKERJA	28	BELUM MENIKAH	2,71	2,95	1,77

Himpunan
Data

$$f(x) dx = \lim_{n \rightarrow \infty} \frac{b-a}{n} \sum_{k=1}^n f\left(a + \frac{b-a}{n} \cdot k\right)$$
$$-\left(-m g \tilde{x} \tan(\phi)\right) \left[l - \frac{r^2}{4l} + r \left(\cos(\omega t) + \frac{r}{4l} \cos(2\omega t) \right) \right]$$
$$= R_1 e^{\left(-\zeta + \sqrt{\zeta^2 - 1}\right) \omega t} \left(-\zeta - \sqrt{\zeta^2 - 1}\right) \omega t$$

Metode Data
Mining



Pengetahuan

Pengertian Data Mining

- *Techniques for finding patterns in data, patterns that provide insight or enable fast and accurate decision making (Witten et al., 2017).*
- Kegiatan mengekstrak informasi atau pengetahuan (*knowledge*) penting dari suatu set data berukuran besar dengan menggunakan teknik tertentu (*Santosa, 2018*).
- *The non-trivial extraction of implicit, previously unknown and potentially useful information from data (Bramer, 2020).*

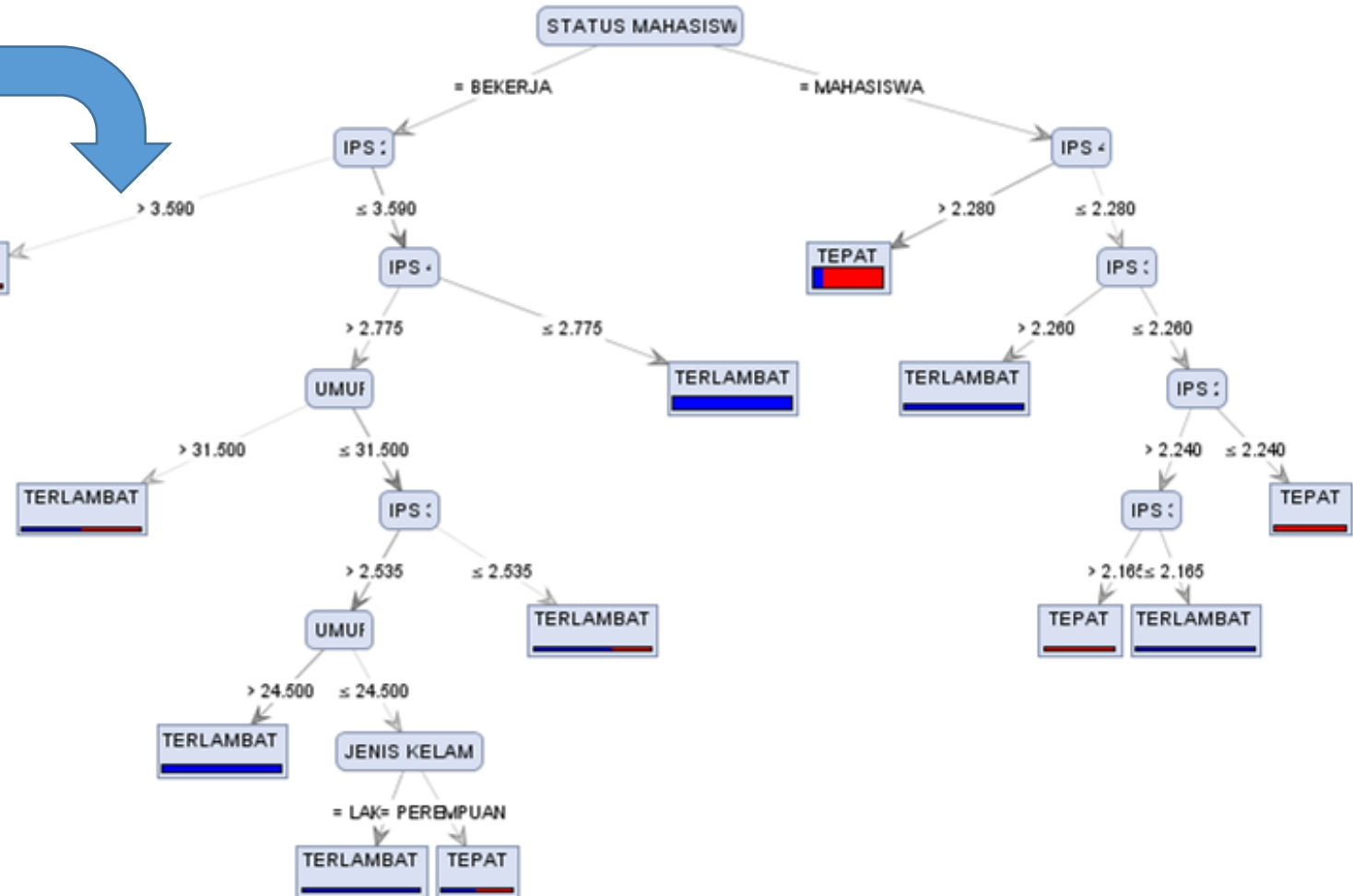
Contoh Data di Akademik

- Puluhan ribu data mahasiswa di kampus yang diambil dari Sistem Informasi Akademik.
- Apakah pernah kita ubah menjadi pengetahuan yang lebih bermanfaat? TIDAK!
- Seperti apa pengetahuan itu? Rumus, Pola, Aturan?

NIM	Gender	Nilai UN	Asal Sekolah	IPS1	IPS2	IPS3	IPS 4	...	Lulus Tepat Waktu
10001	L	28	SMAN 2	3.3	3.6	2.89	2.9		Ya
10002	P	27	SMA DK	4.0	3.2	3.8	3.7		Tidak
10003	P	24	SMAN 1	2.7	3.4	4.0	3.5		Tidak
10004	L	26.4	SMAN 3	3.2	2.7	3.6	3.4		Ya
...									
...									
11000	L	23.4	SMAN 5	3.3	2.8	3.1	3.2		Ya

Prediksi Kelulusan Mahasiswa

NIM	Gender	Nilai UN	Asal Sekolah	IPS1	IPS2	IPS3	IPS 4	...	Lulus Tepat Waktu
10001	L	28	SMAN 2	3.3	3.6	2.89	2.9		Ya
10002	P	27	SMA DK	4.0	3.2	3.8	3.7		Tidak
10003	P	24	SMAN 1	2.7	3.4	4.0	3.5		Tidak
10004	L	26.4	SMAN 3	3.2	2.7	3.6	3.4		Ya
...									
...									
11000	L	23.4	SMAN 5	3.3	2.8	3.1	3.2		Ya



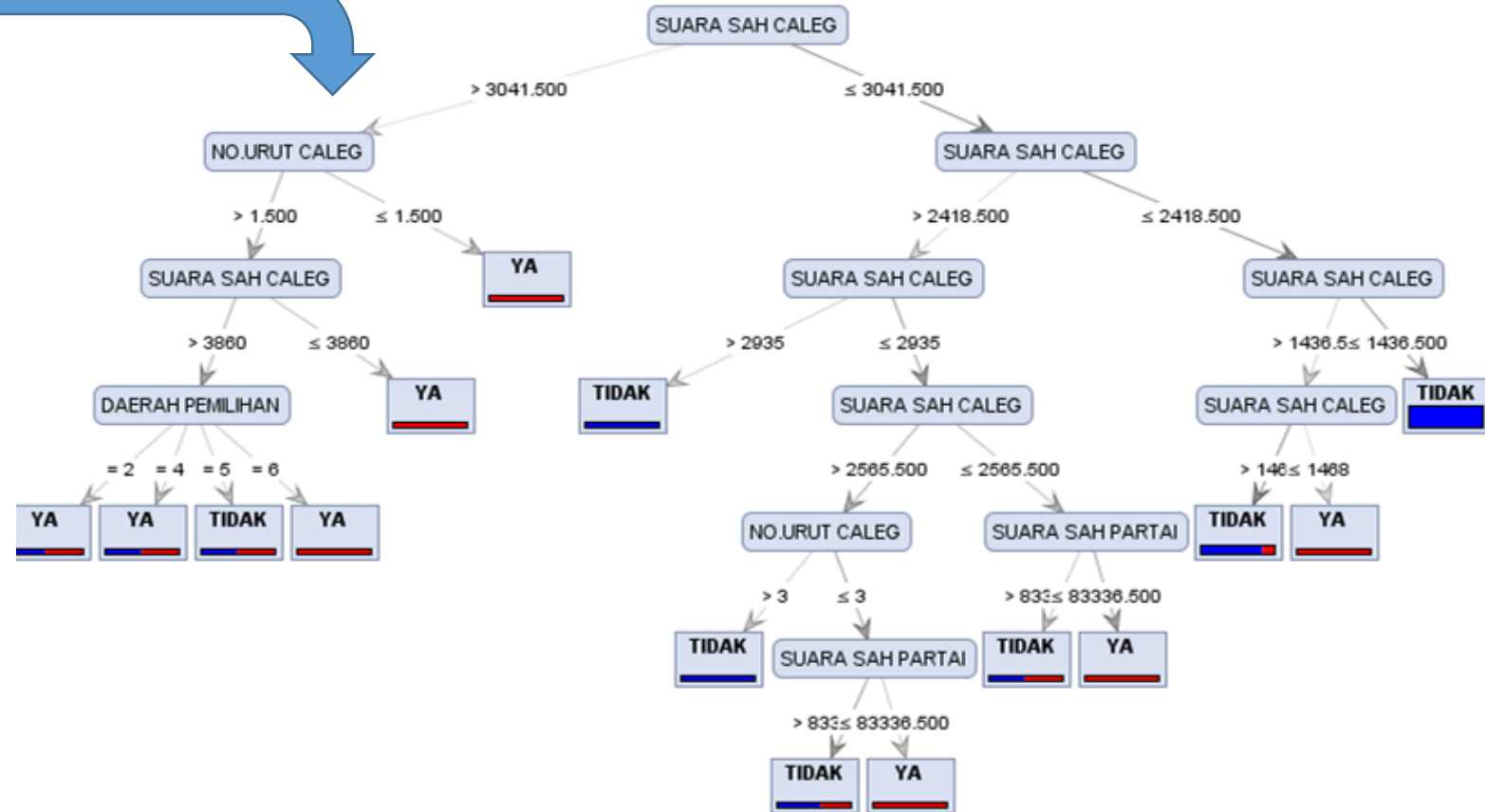
Contoh Data di KPU

- Puluhan ribu data calon anggota legislatif di KPU.
- Apakah pernah kita ubah menjadi pengetahuan yang lebih bermanfaat? TIDAK!

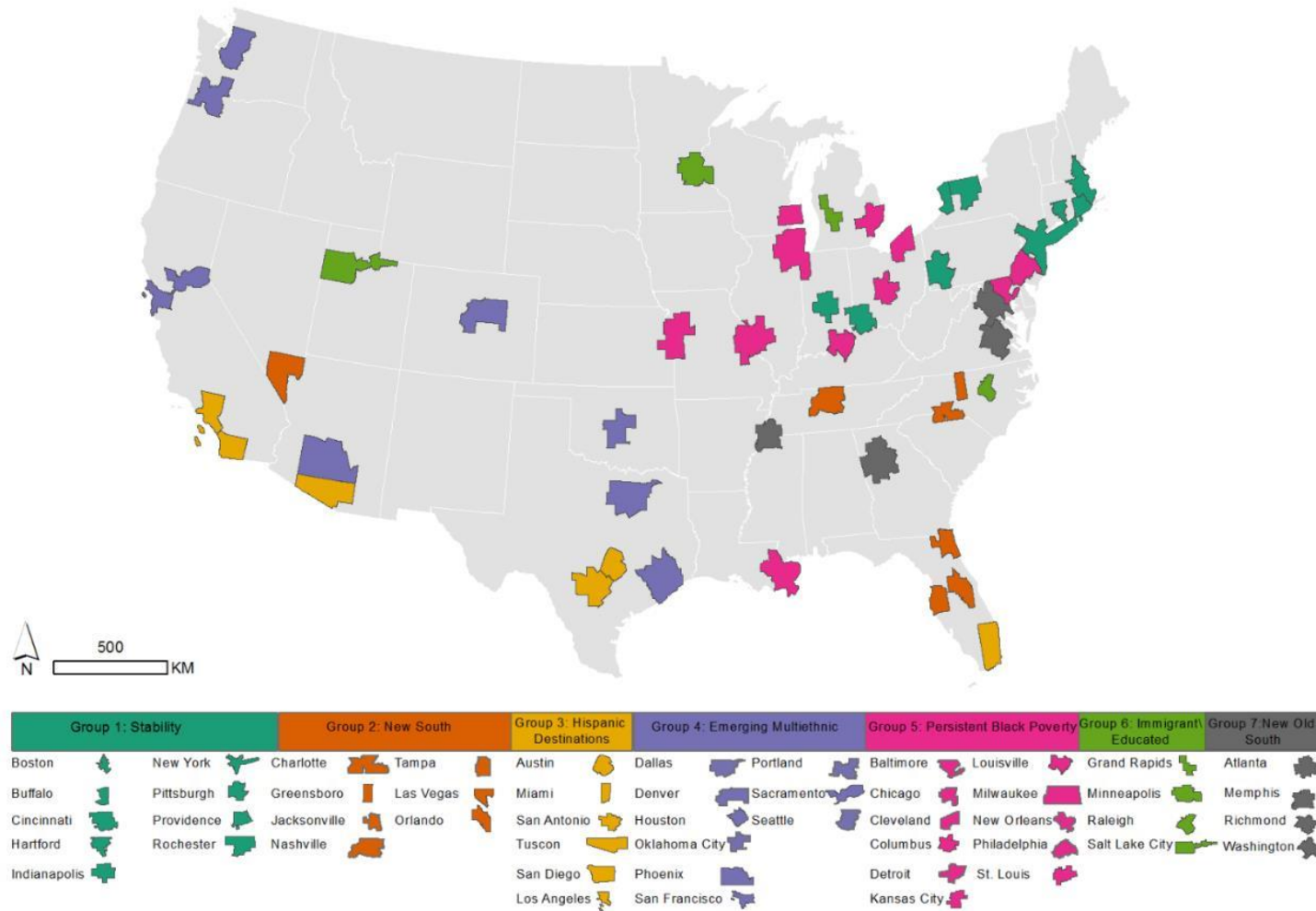
NAMA PARTAI POLITIK	NAMA CALON LEGESLATIF	JENIS KELAMIN	KECAMATAN	SUARA SAH PARTAI	DAERAH PEMILIHAN	SUARA SAH CALEG	TERPILIH ATAU TIDAK
HANURA	TOTO SUKISNO,BSc	L	LEBAKSIU	18578	1	594	TIDAK
HANURA	EDI PURYANTO,SH	L	SLAWI	18578	1	943	TIDAK
PKB	ELI RETNOWATI,SH	P	SLAWI	18578	1	1730	TIDAK
PKB	SAHYUDIN	L	DUKUHWARU	18578	1	2508	YA
GOLKAR	H.FAJAR SIGIT KUSUMAJAYA,SH	L	SLAWI	18578	2	923	TIDAK
GOLKAR	SUMIRAH	P	TARUB	18578	2	308	TIDAK
GOLKAR	DARYOTO	L	TARUB	18578	2	54	TIDAK
PKS	KHAPIP APRONI,S.Pdi	L	BOJONG	18578	3	1682	TIDAK
PKS	ENDANG SUCI RAHAYU	P	JATINEGARA	18578	3	918	TIDAK
PDI-P	KH.CHAFIDZ ISA MUFTI ,LC	L	SLAWI	18578	3	87	TIDAK
.....

Prediksi Calon Legislatif DKI Jakarta

NAMA PARTAI POLITIK	NAMA CALON LEGESLATIF	JENIS KELAMIN	KECAMATAN	SUARA SAH PARTAI	DAERAH PEMILIHAN	SUARA SAH CALEG	TERPILIH ATAU TIDAK
HANURA	TOTO SUKISNO,BSc	L	LEBAKSIU	18578	1	594	TIDAK
HANURA	EDI PURYANTO,SH	L	SLAWI	18578	1	943	TIDAK
PKB	ELI RETNOWATI,SH	P	SLAWI	18578	1	1730	TIDAK
PKB	SAHYUDIN	L	DUKUHWARU	18578	1	2508	YA
GOLKAR	H.FAJAR SIGIT KUSUMAJAYA,SH	L	SLAWI	18578	2	923	TIDAK
GOLKAR	SUMIRAH	P	TARUB	18578	2	308	TIDAK
GOLKAR	DARYOTO	L	TARUB	18578	2	54	TIDAK
PKS	KHAPIP APRONI,S.Pdi	L	BOJONG	18578	3	1682	TIDAK
PKS	ENDANG SUCI RAHAYU	P	JATINEGARA	18578	3	918	TIDAK
PDI-P	KH.CHAFIDZ ISA MUFTI ,LC	L	SLAWI	18578	3	87	TIDAK



Klasterisasi Tingkat Kemiskinan



Pola Aturan Asosiasi di Amazon.com

Systems Analysis and Design (Shelly Cashman Series) 11th Edition

by Scott Tilley (Author), Harry J. Rosenblatt (Author)

★★★★☆ 28 customer reviews

Look inside



ISBN-13: 978-1305494602

ISBN-10: 9781305494602

Why is ISBN important?

Sell yours for a Gift Card
We'll buy it for up to **\$7.49**
[Learn More](#)

Trade in now



Have one to sell?

Sell on Amazon

Add to List

Share



Hardcover
\$19.04 - \$110.00

Other Sellers
from \$19.04

☐ Rent

☐ Buy used

☒ **Buy new**

In Stock.

Sold by **Tome Dealers and Fulfilled**

This item ships to **Indonesia**. Want to see more shipping options? [Learn more](#)

More Buying Choices

50 New from \$110.00 / 9 Used from \$19.04

prime student

College student? Get more

Discover a practical, streamlined, updated and emerging technologies and approach. 11F offers a well-organized, streamlined and making content easy to comprehend and ensure content is current and relevant. In

Customers who bought this item also bought



Principles of Marketing
(17th Edition)
Philip T. Kotler
★★★★☆ 14
Hardcover
\$199.99



Operations and Supply Chain Management
Robert Jacobs
★★★★☆ 201
Hardcover
\$154.60

Sponsored products related to this item



PMP Exam Prep Simplified:
Based on PMBOK® Guide
Sixth Edition
Andrew Ramdayal
★★★★☆ 28
Paperback
\$36.09 **prime**



React Design Patterns and Best Practices: Build easy to scale modular applications ...
Michele Bertoli
★★★★☆ 5
Paperback
\$41.27 **prime**



Node.js Design Patterns - Second Edition: Master best practices to build modular an...
Mario Casciaro
Get the best out of Node.js by mastering its most powerful components and patterns to create modular and scalable applications with ease
★★★★☆ 52
Paperback



Vue.js 2 Design Patterns and Best Practices: Build enterprise-ready, modular Vue.js...
Paul Haddad
★★★★☆ 4
Paperback
\$44.99 **prime**



Mastering Node.js - Second Edition: Build robust and scalable real-time server-side...
Sandro Pasquali
Expert techniques for building fast servers and scalable, real-time network applications with minimal effort, rewritten for Node.js 8 and Node.js 9.
★★★★☆ 1



Learning Bootstrap 4 - Second Edition
Matt Lambert
★★★★☆ 4
Paperback
\$37.71 **prime**

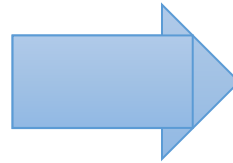


Hands-On Neural Network Programming with C#: Add powerful neural network capabilities...
Matt R. Cole
Just released!
Paperback
\$39.99 **prime**

From Stupid Apps to Smart Apps

Stupid Applications

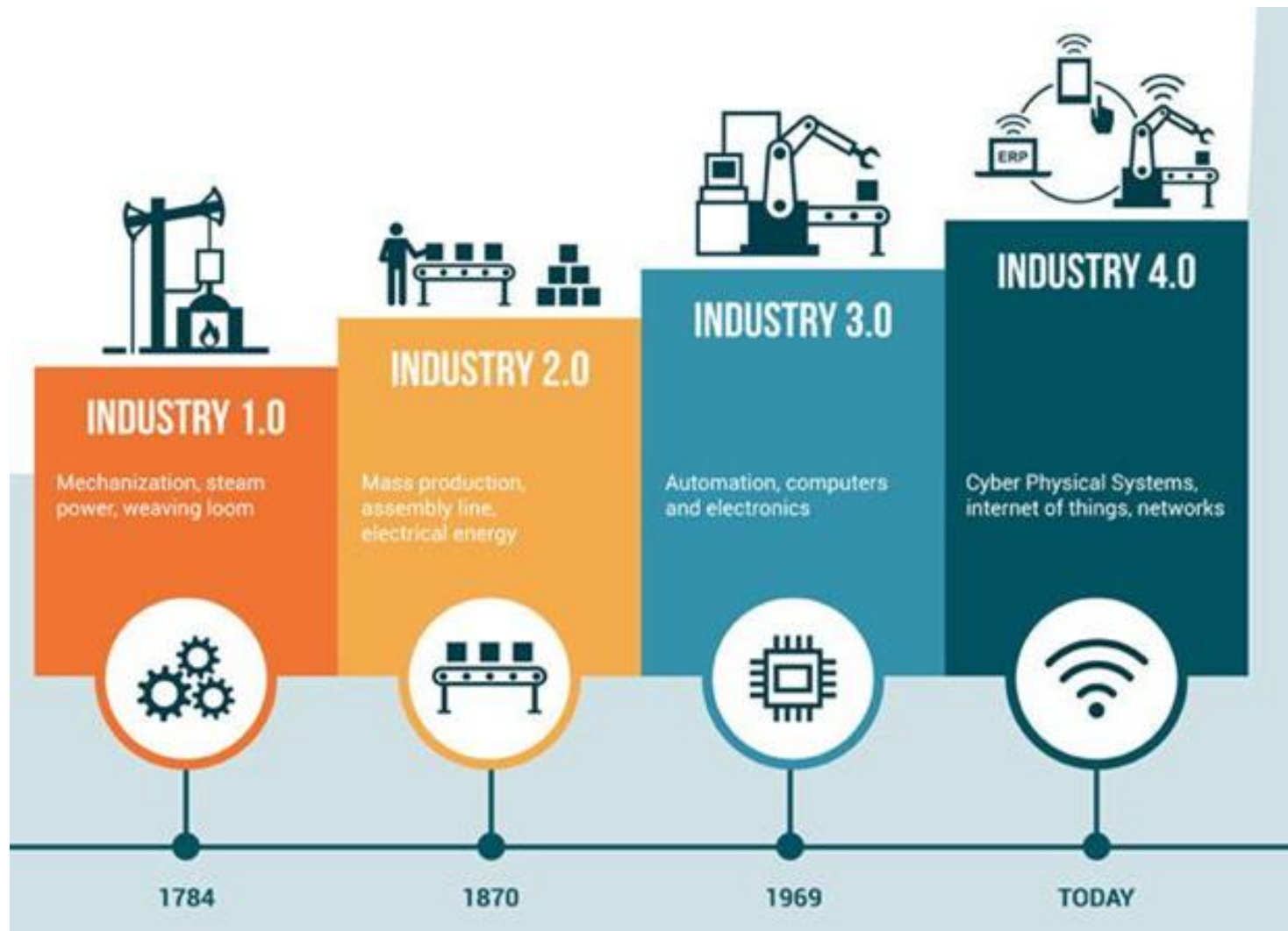
- Sistem Informasi Akademik.
- Sistem Pencatatan Pemilu.
- Sistem Laporan Kekayaan Pejabat.
- Sistem Pencatatan Kredit.



Smart Applications

- Sistem Prediksi Kelulusan Mahasiswa.
- Sistem Prediksi Hasil Pemilu.
- Sistem Prediksi Koruptor.
- Sistem Penentu Kelayakan Kredit.

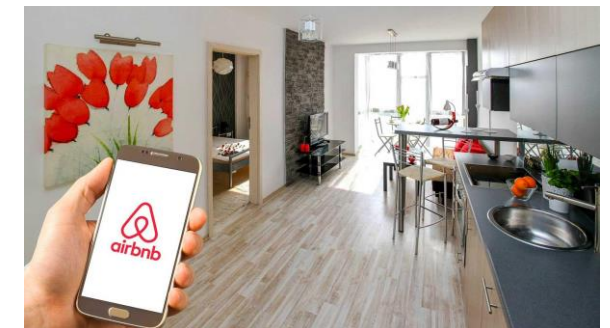
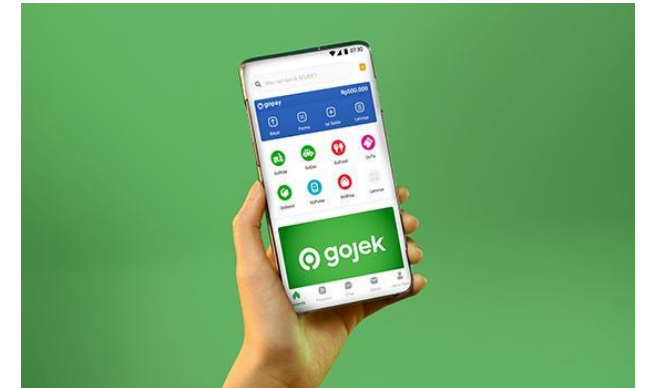
Revolusi Industri 4.0



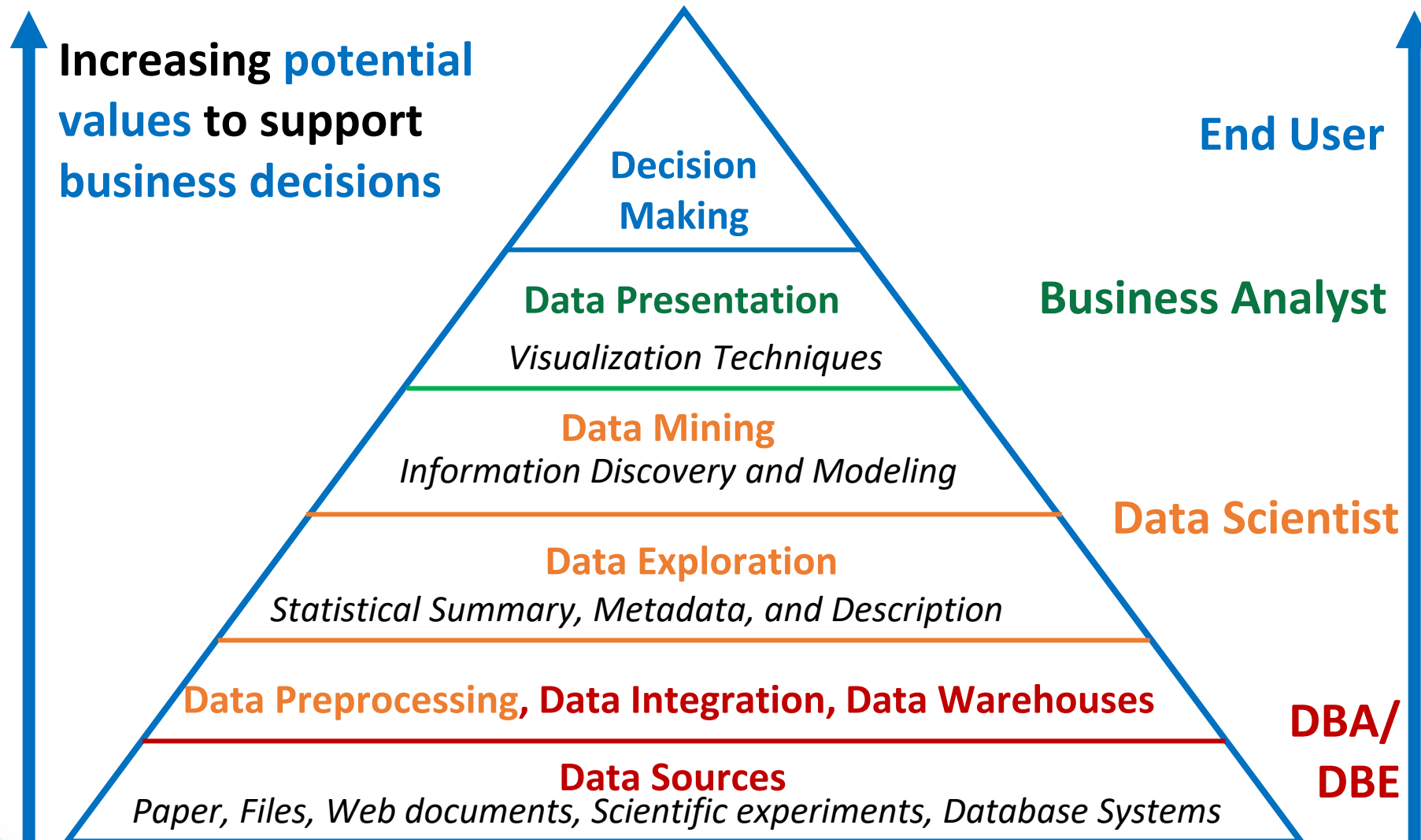
Perusahaan Pengolah Pengetahuan



- Uber - the world's largest taxi company, owns no vehicles.
- Google - world's largest media/advertising company, creates no content.
- Alibaba - the most valuable retailer, has no inventory.
- Airbnb - the world's largest accommodation provider, owns no real estate.
- Gojek - perusahaan angkutan umum, tanpa memiliki kendaraan.



Data Mining Tasks and Roles in General



Latihan Soal (Kuis)

- Jelaskan pemahaman Anda tentang Data Mining?
- Sebutkan dan jelaskan implementasi dari Data Mining?
- Bagaimana menurut Anda peran Data Mining dalam menyelesaikan permasalahan terkait dengan data saat ini?

Referensi

1. Jiawei Han, Micheline Kamber, Jian Pei, Data mining : concepts and techniques – 3rd ed, Elsevier, 2012.
2. Ian H. Witten, Frank Eibe, Mark A. Hall, Data mining: Practical Machine Learning Tools and Techniques 4th Edition, *Elsevier*, 2017.
3. Budi Santosa, Ardian Umam, Data Mining dan Big Data Analytics, Penebar Media Pustaka, 2018.
4. Max Bramer, Principles of Data Mining – Undergraduate Topics in Computer Science – 4th ed, Springer, 2020.
5. Romi Satrio Wahono, *Lecture Notes – Data Mining*, diakses 3 Maret 2021, <<https://romisatriawahono.net/dm/>>.
6. Sumber gambar: www.freepik.com.



THANKS

ANY QUESTIONS?

