



www.esaunggul.ac.id

**DATA MINING
PERTEMUAN-1
NOVIANDI**

PRODI MIK | FAKULTAS ILMU-ILMU KESEHATAN

VISI UNIVERSITAS ESA UNGGUL

Menjadi perguruan tinggi **kelas dunia** berbasis **intelektualitas, kreatifitas dan kewirausahaan**, yang unggul dalam mutu pengelolaan dan hasil pelaksanaan Tridarma Perguruan Tinggi

VISI PS MIK UNIVERSITAS ESA UNGGUL

Menjadi penyelenggara Program Studi Manajemen Informasi Kesehatan berbasis **intelektualitas, kreatifitas dan kewirausahaan**, yang unggul dalam mutu pengelolaan dan hasil pelaksanaan Tridarma Perguruan Tinggi.

MISI PS MIK UNIVERSITAS ESA UNGGUL

1. Menghasilkan Sarjana Manajemen Informasi Kesehatan dengan keunggulan prodi yang professional, bermoral tinggi, mandiri, inovatif dan mampu berkontribusi dalam pengembangan teknologi informasi dan komunikasi bidang kesehatan.
2. Menyelenggarakan proses belajar mengajar yang bermutu, efektif, efisien, akuntabel dan berkelanjutan berbasis teknologi informasi dan komunikasi.
3. Menyelenggarakan Tri Dharma perguruan tinggi yang berkualitas.

KEUNGGULAN PS MIK

Sarjana Manajemen Informasi Kesehatan Universitas Esa Unggul handal dalam **analisis data** dan **tata kelola informasi kesehatan** untuk **peningkatan mutu dan efisiensi pelayanan serta keselamatan pasien** sesuai kebutuhan lokal, nasional dan global.

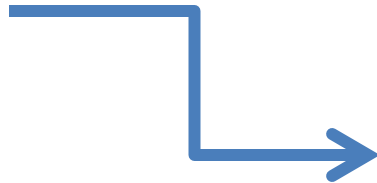
PROFIL LULUSAN PS MIK

1. Analis Data dan Manajer Informasi Kesehatan
2. Spesialis Koding Klinis
3. Manajer Unit Kerja MIK (RMIK)
4. Spesialis *Clinical Documentation Improvement* (CDI)
5. Inisiator Perancang dan Pengembang *Electronic Health Records* (EHR) atau *Electronic Medical Records* (EMR).

KETERKAITAN MK DG PROFIL LULUSAN PS MIK

MATA KULIAH

MIK 445 : Rekam Kesehatan
Elektronik 1 (2 sks)



PROFIL LULUSAN

1. Analis Data dan Manajer Informasi Kesehatan
2. Spesialis Koding Klinis
3. Manajer Unit Kerja MIK (RMIK)
4. Spesialis *Clinical Documentation Improvement* (CDI)
5. **Inisiator Perancang dan Pengembang *Electronic Health Records* (EHR) atau *Electronic Medical Records* (EMR).**

Materi Sebelum UTS

01. Pengertian dan Konsep Dasar Data Mining
02. Eksplorasi data, data warehouse dan OLAP
03. Data cleaning, Missing value, noisy data
04. Data integration dan transformation
05. Feature selection
06. Feature extraction
07. Association rule



Materi Setelah UTS

08. Unsupervised learning

09. Decision tree

10. Naïve bayes

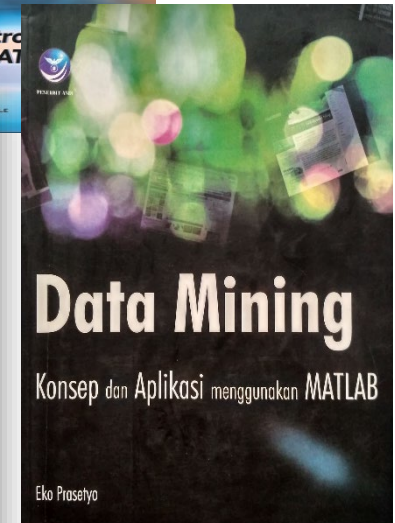
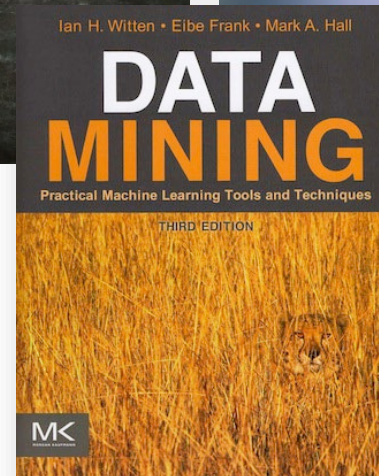
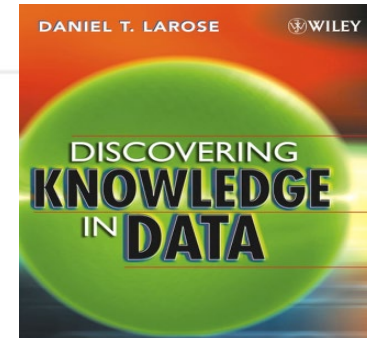
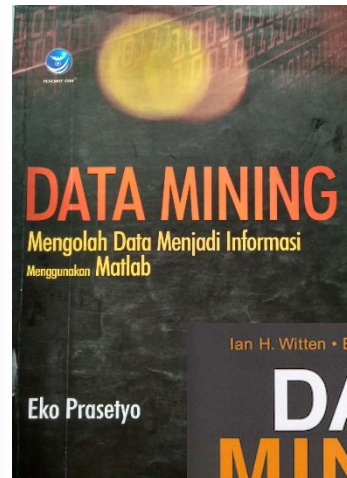
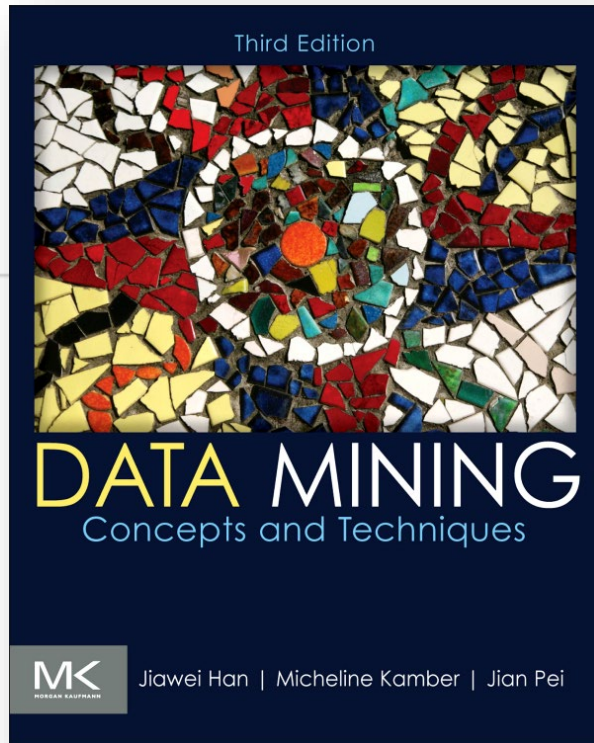
11. Clustering

12. Klasifikasi

13. Text mining

14. Presentasi dan diskusi tugas kelompok

Buku Bacaan



- ❑ Masalah eksplorasi data

Ukuran data yang selalu mengalir pada jaringan komputer dan menghasilkan jumlah data yang sangat besar bahkan mencapai Terabytes atau peta bytes = 1 juta gigabytes.



- ❑ Kebutuhan untuk berkompetisi dengan strategi yang tepat menjadi lebih tinggi.

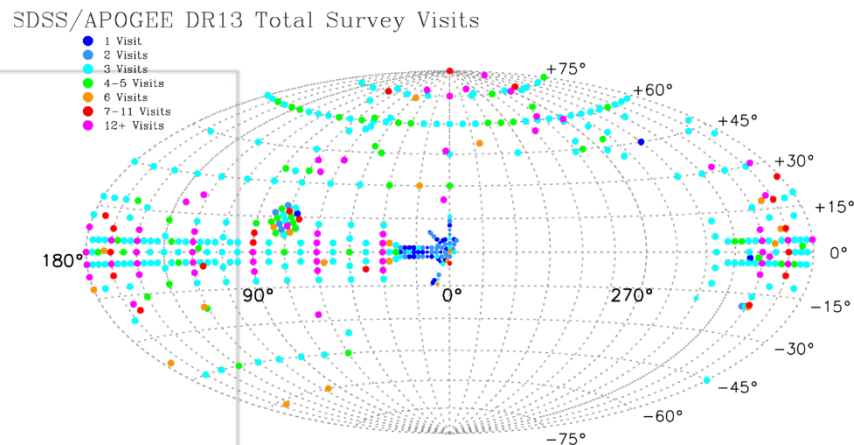
Beberapa kegiatan manusia yang memproduksi data:

- Mengakses *world wide web* / *Log kunjungan Web*.
- Riset science dan engineering / Akuisisi data dalam penelitian-penelitian. seperti; astronomi, kesehatan (menghasilkan rekam medis), dll.
- Transaksi penjualan, baik transaksi penjualan online maupun transaksi penjualan di supermarket.
- Transaksi perbankan dan kartu kredit.
- dll

Pertumbuhan Data

Astronomi

Menurut **Sloan Digital Sky Survey**, besar data yang di hasilkan setiap malam nya adalah 200 GB yang merupakan data fotometrik berupa bintang, galaksi ,dll



(<http://www.sdss.org>, 2016)

kilobyte (kB)	10^3
megabyte (MB)	10^6
gigabyte (GB)	10^9
terabyte (TB)	10^{12}
petabyte (PB)	10^{15}
exabyte (EB)	10^{18}
zettabyte (ZB)	10^{21}
yottabyte (YB)	10^{24}

Large Synoptic Survey Telescope

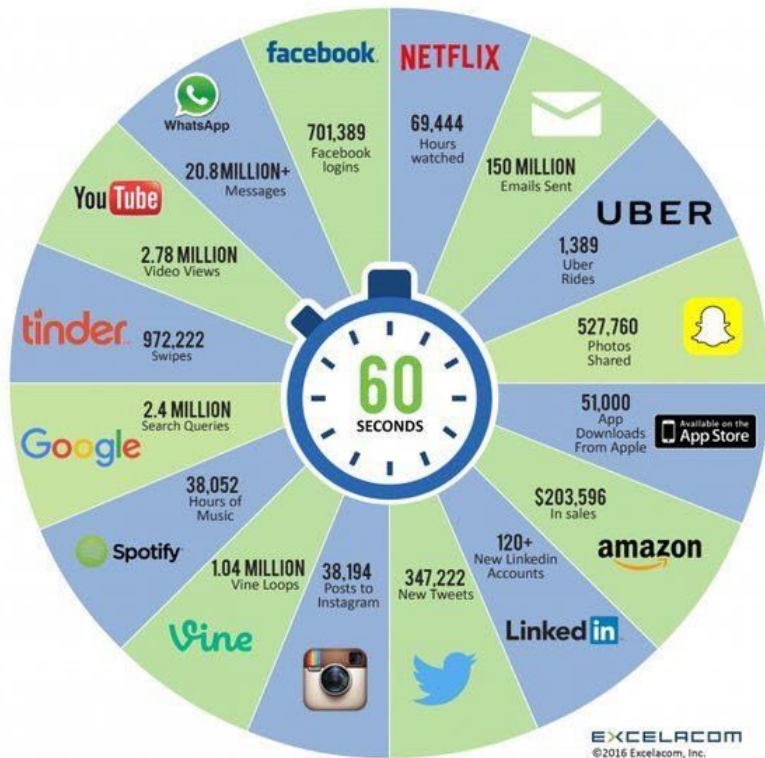
- Chile, 2016
- Akan mendapatkan data 140TB setiap lima hari

Datangnya Tsunami Data

- **Mobile Electronics** market
 - 5Billion mobile phones in use in 2010
 - 150Million tablets was sold in 2012 (IDC)
 - 200Million is global notebooks shipments (Pengiriman) in 2012 (Digitimes Research)
- **Web and Social Networks** generates amount of data
 - Google processes 100 PB per day, 3 million servers
 - Facebook has 300 PB of user data per day
 - Youtube has 1000PB video storage
 - 235 TBs data collected by the US Library of Congress

Perubahan Culture dan Perilaku

2016 What happens in an INTERNET MINUTE?



2017 This Is What Happens In An Internet Minute



(Insight, Big Data Trends for Media, 2017)

Mengapa Melakukan Data Mining??

Dilihat dari sudut pandang komersil

- ✓ Data *over load*
 - Web data, e-commerce, e-banking
 - Grocery stores
 - Bank/Credit Card Transaction
- ✓ Teknologi komputer menjadi lebih murah dan powerful
- ✓ Tekanan kompetisi semakin kuat
 - Dalam bisnis, dapat memberikan layanan yang lebih baik
contoh: *Customer Relationship Management*

Mengapa Melakukan Data Mining??

Dilihat dari sudut pandang keilmuan

- ✓ Data dikumpulkan dan disimpan dengan kecepatan tinggi (GB/hour)
 - Remote sensor pada satelite
 - Data spasial dalam GIS
 - Simulasi keilmuan
- ✓ Teknik tradisional tidak mampu melakukan analisis data yang berukuran besar
- ✓ Data Mining membantu ilmuan dalam
 - Klasifikasi dan segmentasi data
 - Pemodelan
 - Clustering

Apa itu Data Mining?

B	C	D	E	F	G	H
STATUS KELAMIN	STATUS MAHASISWA	UMUR	STATUS NIKAH	IPS 1	IPS 2	IPS 3
PEREMPUAN	BEKERJA	28	BELUM MENIKAH	2,76	2,8	3,2
MALE PEREMPUAN	MAHASISWA	32	BELUM MENIKAH	3	3,3	3,84
MALE PEREMPUAN	BEKERJA	29	BELUM MENIKAH	3,5	3,5	3,7
PEREMPUAN	MAHASISWA	37	BELUM MENIKAH	3,17	3,42	3,61
MALE PEREMPUAN	BEKERJA	29	BELUM MENIKAH	2,9	2,89	3,3
MALE PEREMPUAN	MAHASISWA	27	BELUM MENIKAH	2,95	2,82	3,09
MALE PEREMPUAN	MAHASISWA	26	BELUM MENIKAH	2,76	3,34	2,6
MALE PEREMPUAN	MAHASISWA	27	BELUM MENIKAH	3,42	2,89	2,32
MALE PEREMPUAN	BEKERJA	25	MENIKAH	2,6	2,54	3,32
MALE PEREMPUAN	BEKERJA	28	BELUM MENIKAH	2,71	2,55	1,77

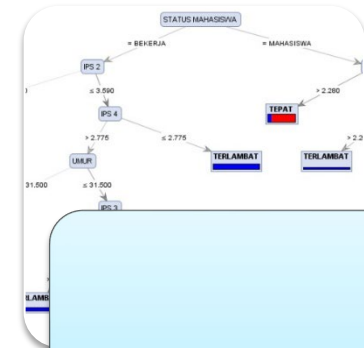
Himpunan Data

$$f(x) dx = \lim_{n \rightarrow \infty} \frac{b-a}{n} \sum_{k=1}^n f\left(a + \frac{b-a}{n} \cdot k\right)$$

$$= \left(-m_p \tilde{x} \sin(\Phi) \right) \left[l - \frac{r^2}{4l} + r \cos(\omega t) + \frac{r}{4l} \cos(2\omega t) \right]$$

$$= R_1 e^{\left(-\zeta + \sqrt{\zeta^2 - 1} \right) \omega_p t} \left(-\zeta - \sqrt{\zeta^2 - 1} \right) \omega_p t$$

Metode Data Mining



Pengetahuan

(romi@romisatriawahono.net)

Pengertian Data Mining

- Disiplin ilmu mengolah himpunan **data yang sangat besar** menjadi suatu **pengetahuan, rumus, pola** sehingga nanti dapat dimanfaatkan manusia untuk memprediksi kejadian kedepan dari kehidupan manusia.
- Melakukan **ekstraksi** untuk mendapatkan **informasi penting** yang sifatnya **implisit** dan sebelumnya tidak diketahui, dari suatu data (*Witten et al., 2011*)
- Ekstraksi informasi atau pola yang menarik (non-trivial, implicit, previously unknown dan potentially useful) dalam basis data berukuran besar (Han et al. 2011)

Nama lain Data Mining

- Knowledge Discovery in Database (KDD)
- Knowledge extraction
- Pattern analysis
- Information harvesting
- Business intelligence

Ekstraksi dari **Data** ke **Pengetahuan**

1. **Data**: fakta yang terekam dan tidak membawa arti
2. **Pengetahuan**: pola, rumus, aturan atau model yang muncul dari data



Apa itu data ???



Apa itu informasi ???



Apa itu pengetahuan ???

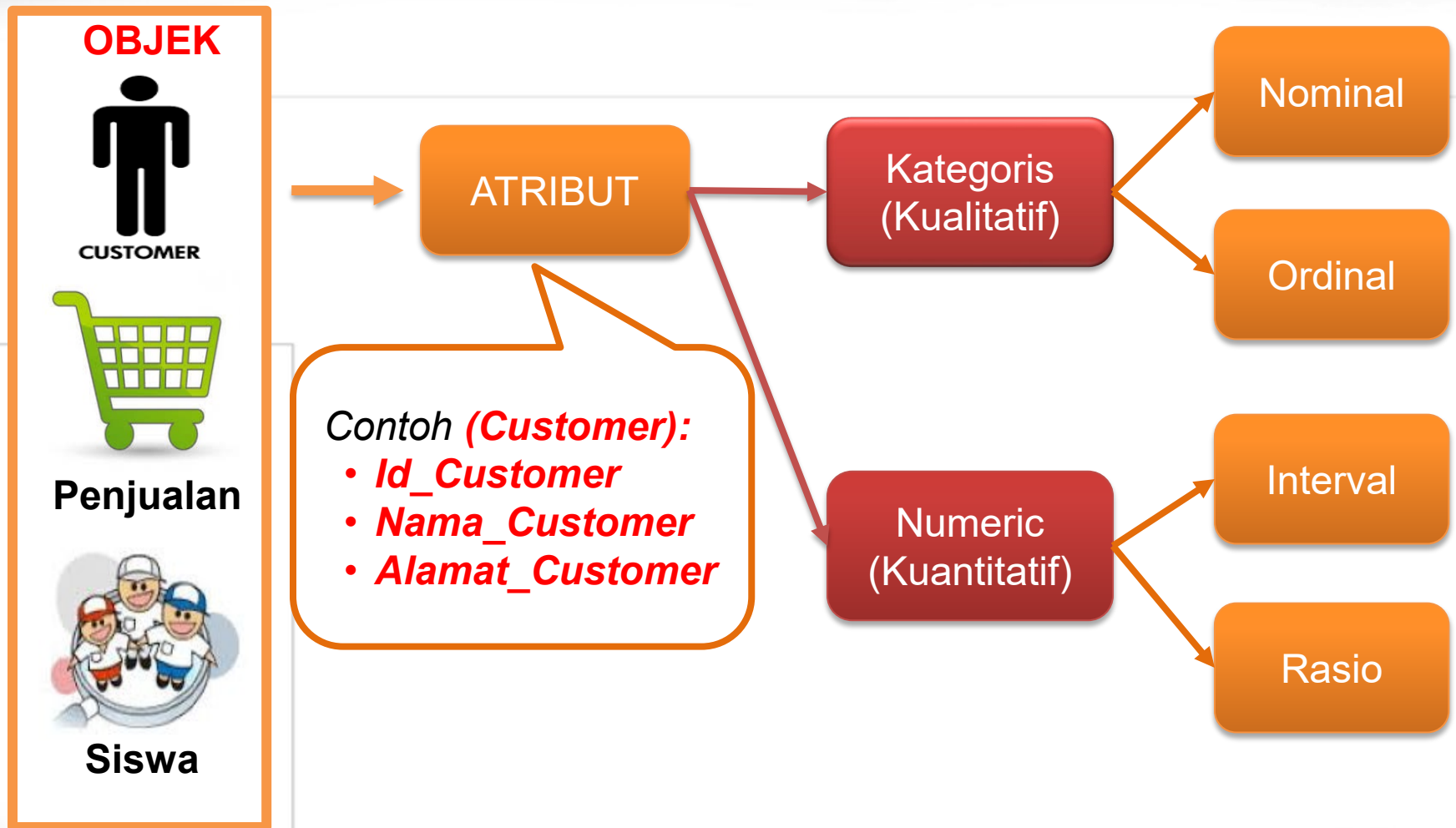
Data – Informasi - Pengetahuan



Segala **fakta**, **angka** atau **teks** yang dapat diproses oleh komputer.

Jenis data dilihat dari akumulasi pertumbuhan jumlah data dan basis data yang berbeda adalah:

- Data operasional atau transaksional
- Data Non operational
- Meta data



Atribut Diskrit dan Atribut Kontinyu

❑ Atribut Diskrit

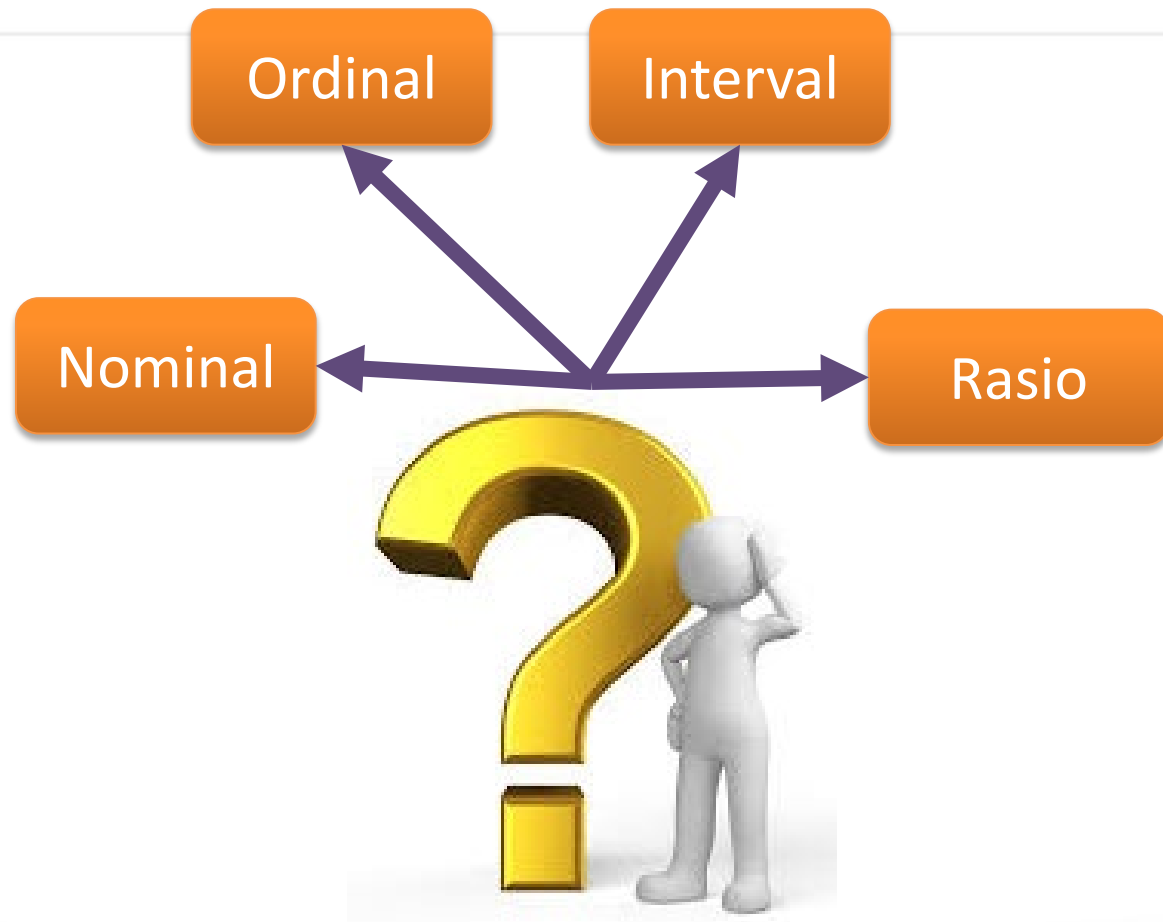
- ✓ Hanya memiliki seperangkat nilai yang terbatas atau terhitung tak terbatas.

Contoh: kode pos, hitungan, atau kumpulan kata dalam kumpulan dokumen

- ✓ Sering diwakili sebagai variabel integer.
Catatan: atribut biner adalah kasus khusus atribut diskrit

❑ Atribut Kontinyu

- ✓ Memiliki bilangan real sebagai nilai atribut
Contoh: suhu, tinggi, atau berat.
- ✓ Praktis, nilai sebenarnya hanya bisa diukur dan diwakili dengan menggunakan jumlah digit yang terbatas.
- ✓ Atribut kontinyu biasanya diwakili sebagai variabel floating-point.



Tiga Jenis Data Set Berdasarkan Golongannya

1. Record

- ✓ Matrik data
- ✓ Data transaksi
- ✓ Data dokumen

2. Graph

- ✓ Word Wide Web
- ✓ Struktur molekul

3. Ordered Data

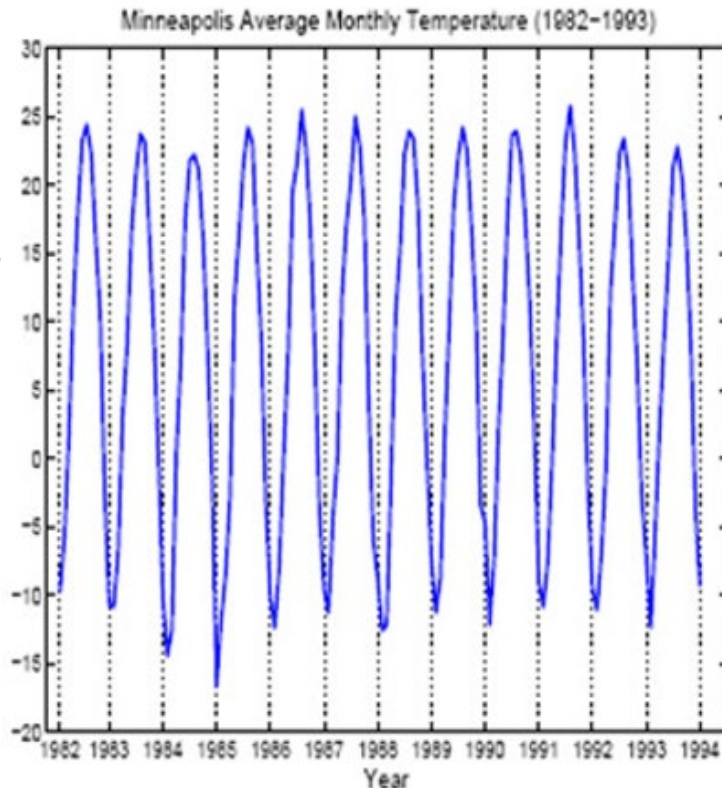
- ✓ Data spasial
- ✓ Data temporal
- ✓ Data sekuensial
- ✓ Data urutan genetic (*genetic sequence*)

Data Transaksi

- Setiap *record* (transaksi) melibatkan satu set item yang biasanya menyertakan nomor identitas transaksi.
 - Misalnya;
kumpulan produk yang dibeli oleh seorang *customer* selama satu kali belanja dan dikategorikan satu kali transaksi atau disebut juga dengan *market basket*.

TID	Items
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

Data Temporal



Data temporal adalah data yang objeknya memiliki atribut yang mewakili pengukuran yang diambil dari waktu ke waktu.

Misalnya, kumpulan data keuangan adalah deret waktu yang memberi harga harian berbagai saham.

Seri waktu adalah urutan pengukuran beberapa atribut

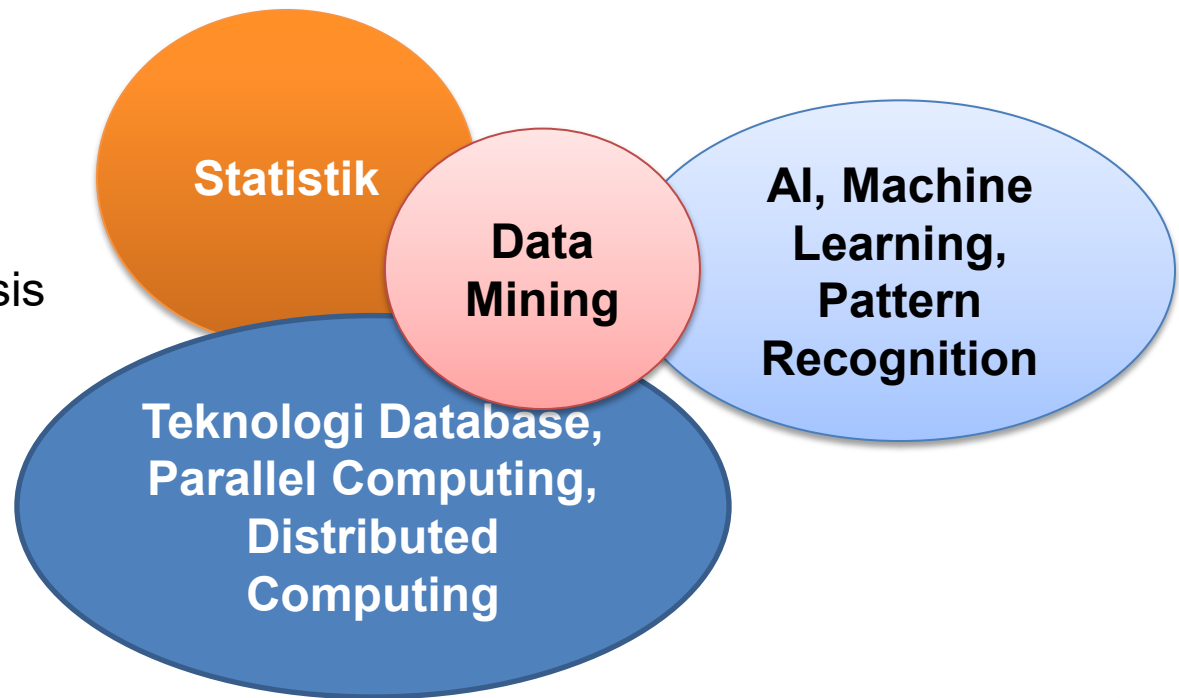
Misalnya, harga saham atau curah hujan, diambil pada (biasanya reguler) pada waktunya.)

Bukan Data Mining	Data Mining
✓ Pencarian informasi tertentu di Internet.	✓ Pengelompokkan informasi yang mirip dalam konteks tertentu pada hasil pencarian.
✓ Pencarian data medis untuk menganalisis catatan penyakit pasien.	✓ Peneliti medis mencari cara pengelompokkan data penyakit pasien berdasarkan data diagnosis, umur dan alamat
✓ Pembuatan laporan tahunan penjualan perusahaan	✓ Pemanfaatan data penjualan perusahaan untuk mendapatkan pola prediksi stok yang sebaiknya disediakan pada tahun berikutnya.

Hubungan Data Mining dengan Bidang Ilmu Lain

Kesamaan bidang **Data Mining** dengan bidang **Statistik**

- ✓ Penyampelan
- ✓ Estimasi
- ✓ Pengujian Hipotesis



Hubungan Data Mining dengan Bidang Ilmu Lain

Bidang lain yang mempengaruhi **Data Mining** adalah bidang **Basisdata**, misalkan:

Data Mining

Teknologi Database,
Parallel Computing,
Distributed
Computing

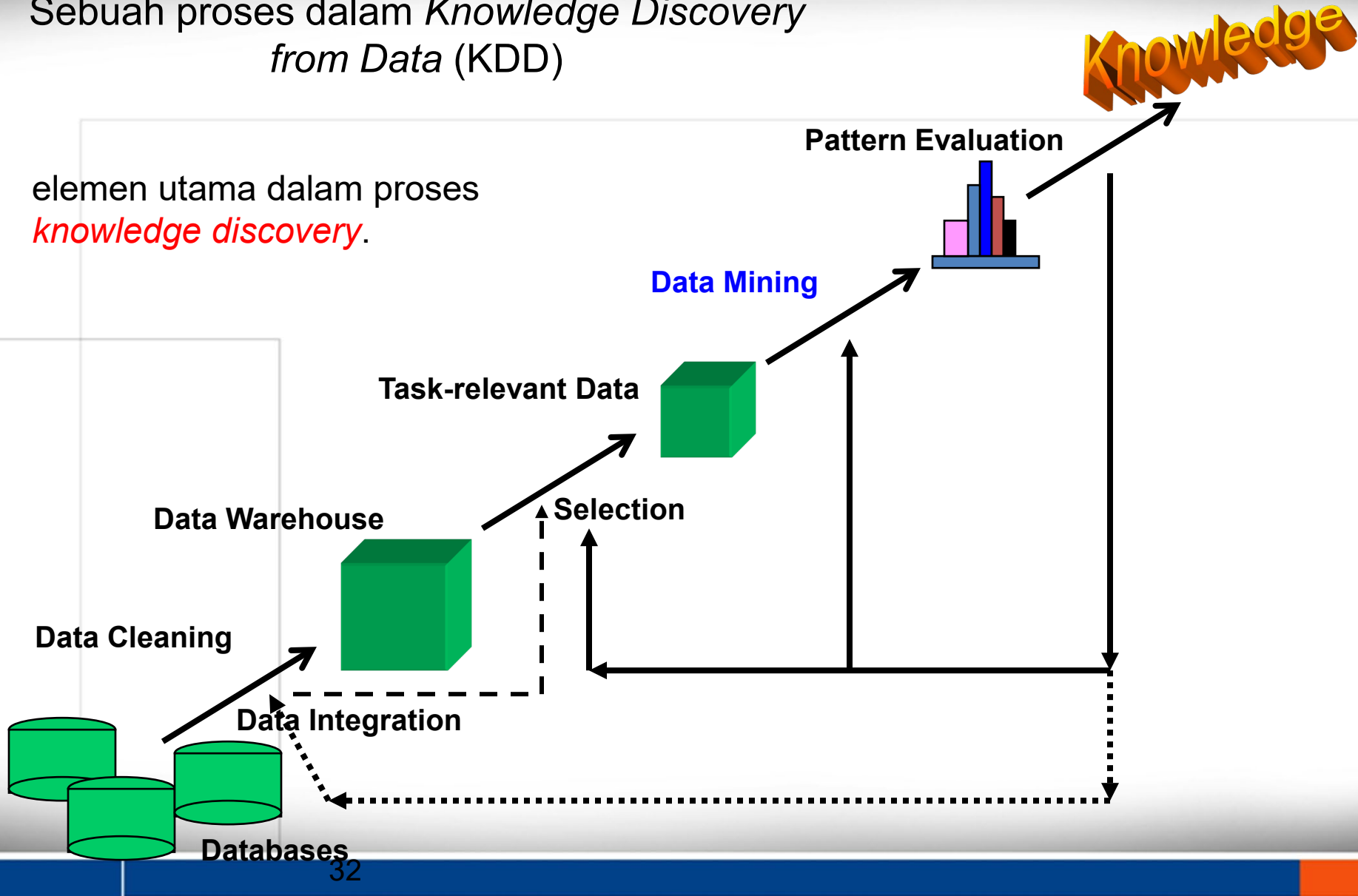
- ✓ Mendukung penyediaan penyimpanan yang efisien
- ✓ pengindeksan
- ✓ Pemrosesan *query*

Parallel computing sering digunakan untuk memberikan kinerja yang tinggi untuk ukuran set data yang besar.

Distributed Computing digunakan untuk menangani masalah ketika data tidak dapat disimpan di satu tempat.

Sebuah proses dalam *Knowledge Discovery from Data* (KDD)

elemen utama dalam proses
knowledge discovery.

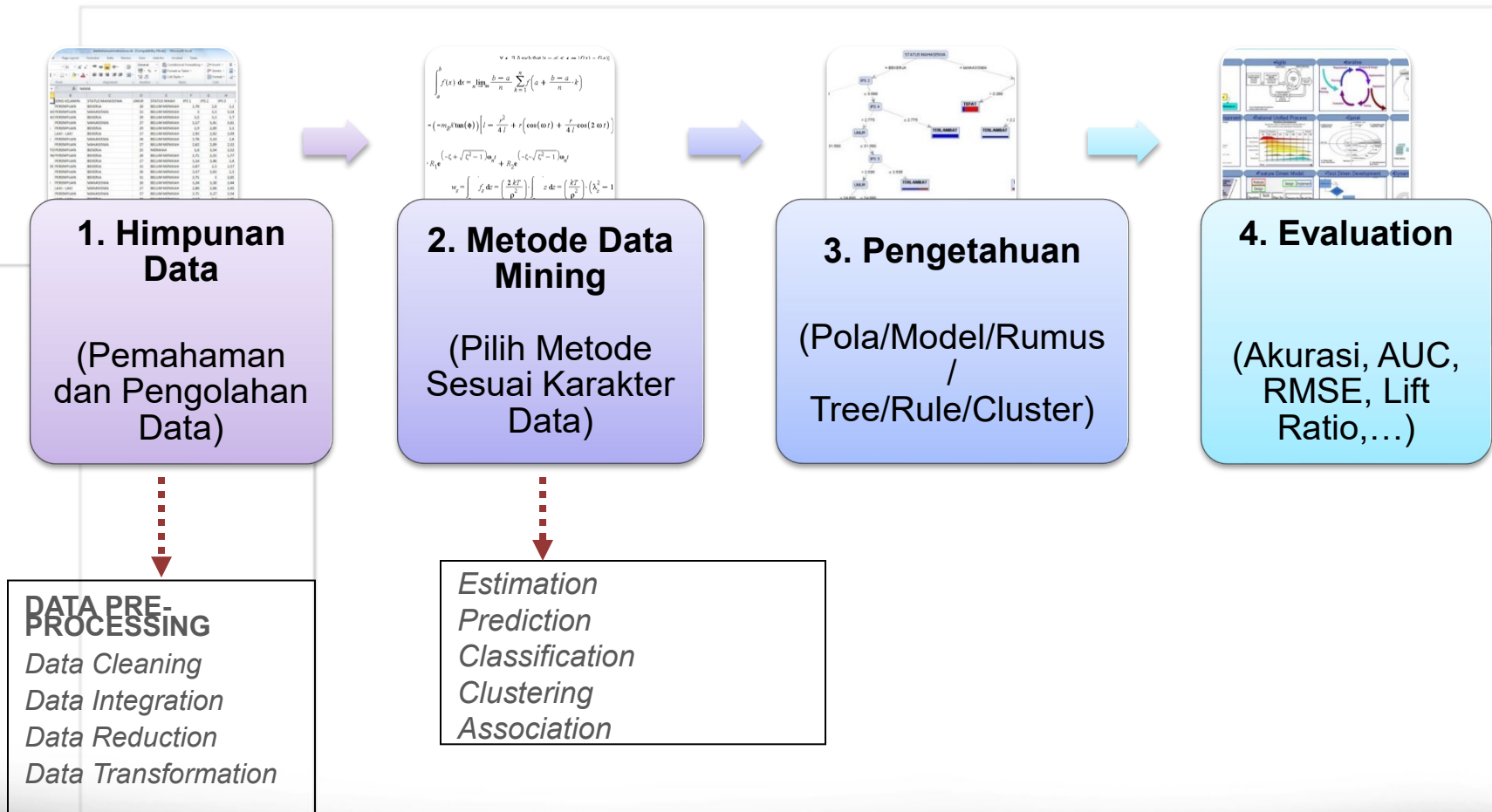


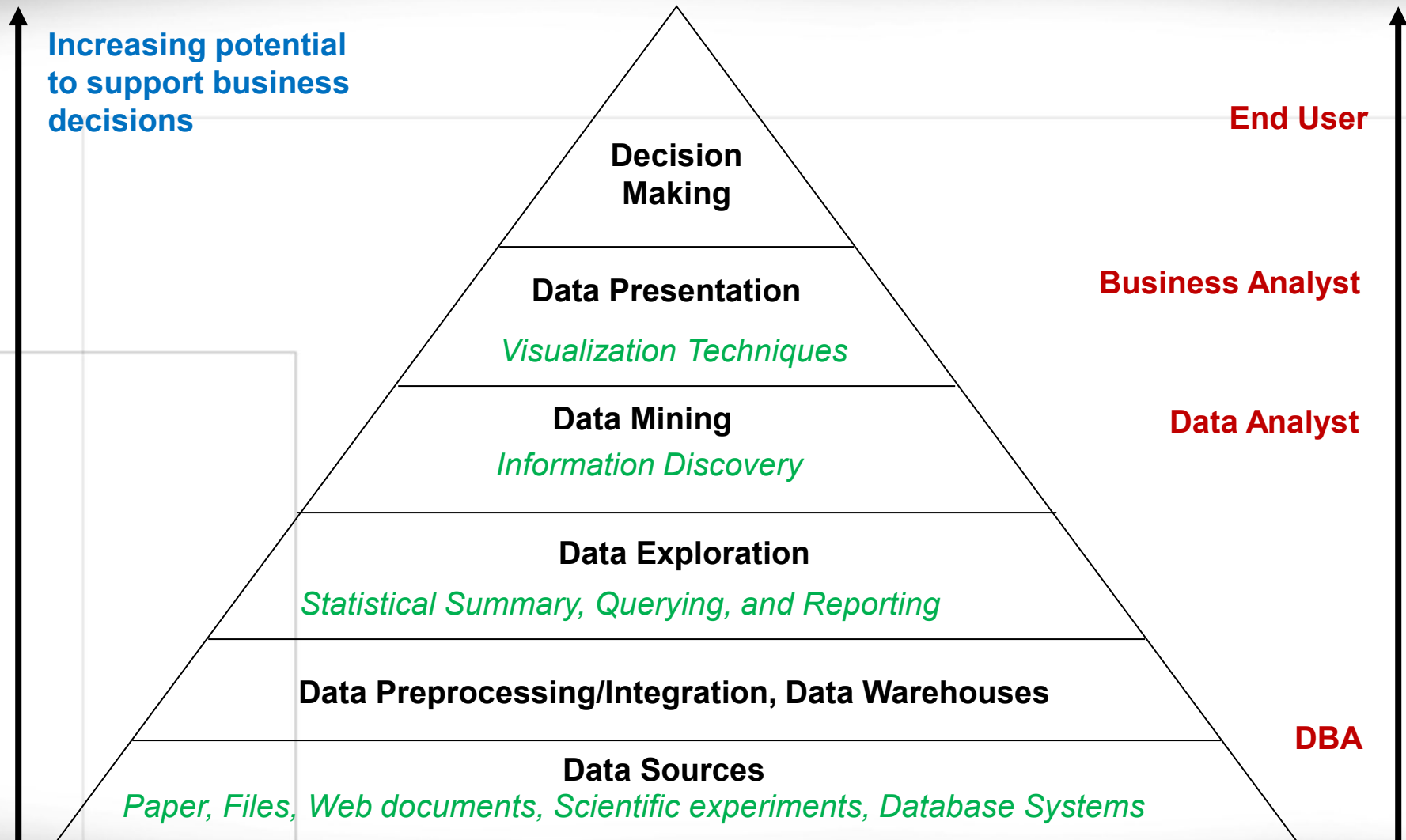
Proses dalam KDD

1. **Pembersihan data:** menghilangkan *noise* dan data yang tidak konsisten.
2. **Pengintegrasian data:** data digabungkan dari berbagai sumber.
3. **Seleksi data:** data yang relevan dengan proses analisis diambil dari basis data.
4. **Transformasi data:** data ditransformasikan atau digabungkan ke dalam bentuk yang sesuai untuk di-*mine* dengan cara dilakukan peringkasan atau operasi agregasi.
5. **Data mining:** merupakan proses yang penting dalam KDD dimana metode-metode cerdas diaplikasikan untuk mengekstrak pola-pola data.
6. **Evaluasi pola:** untuk mengidentifikasi pola-pola yang menarik yang merepresentasikan pengetahuan berdasarkan suatu ukuran kemenarikan.
7. **Presentasi pengetahuan:** merepresentasikan pengetahuan yang telah digali kepada pengguna.

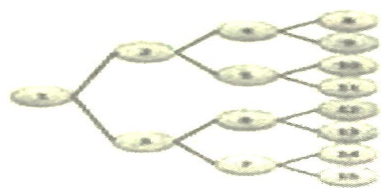
Proses dalam KDD

Typical View dari Machine Learning dan Statistic





Pekerjaan dalam Data Mining



Predictive Modelling



Cluster Analysis

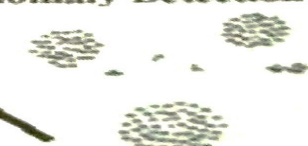
Data

Nama hewan	Penutup kulit	Melahirkan	Berat	Kelas
Ular	Sisik	Ya	10	Reptil
Tikus	Bulu	Ya	0.8	Mamalia
Kambing	Rambut	Ya	21	Mamalia
Sapi	Rambut	Ya	120	Mamalia
Kadal	Sisik	Tidak	0.4	Reptil
Kucing	Rambut	Ya	1.5	Mamalia
Bekicot	Cangkang	Tidak	0.2	Reptil
Harimau	Rambut	Ya	43	Mamalia
Rusa	Rambut	Ya	45	Mamalia
Kura-kura	Cangkang	Tidak	-	Reptil
Musang	Rambut	Ya	15	Mamalia

Association Analysis



Anomaly Detection



Pekerjaan Utama Data Mining
Eko Prasetyo, 2014

Tugas-tugas dalam Data Mining

➤ Metode Prediksi

Menggunakan beberapa variabel (atribut) untuk memprediksi nilai yang tidak diketahui atau nilai yang akan datang dari variabel (atribut) lain.

➤ Metode Deskripsi

Menemukan pola-pola (korelasi, *trend*, *cluster*, *trayektori*, dan anomali) yang meringkas hubungan dalam data.