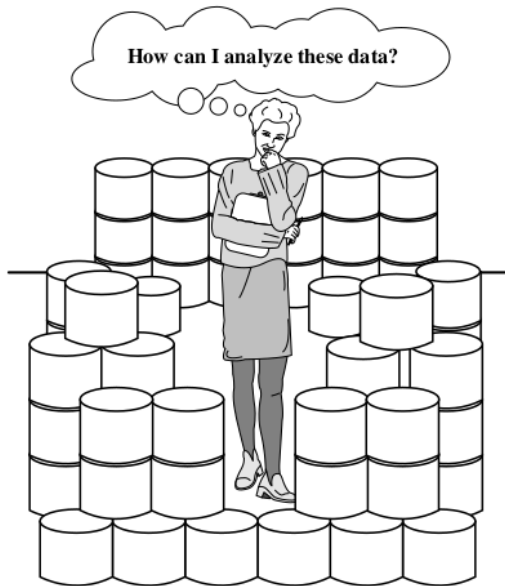


PERTEMUAN 1

PENGANTAR DATA MINING

1. Kenapa Perlu Data Mining?

"data rich but information poor"



Contoh:

- Data DNA - 3 juta pasangan DNA pada Manusia
- data transaksi penjualan
- Data di sosial media

2. Apa Itu Data Mining?

Data Mining → Menambang data → Pengetahuan baru

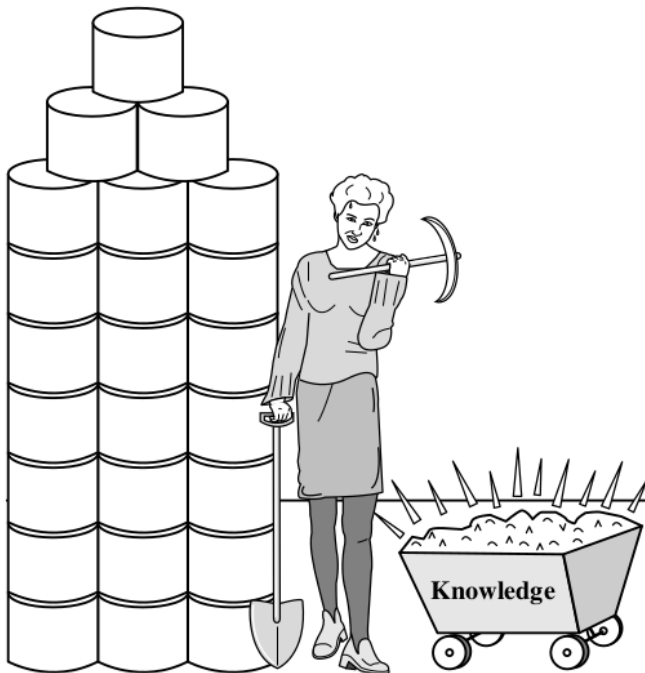
Data Mining → Knowledge Mining

KDD = *Knowledge Discovery in Database*

Data mining → cara mengekstrak **data** yang besar dengan **metode** tertentu untuk menghasilkan sebuah **pengetahuan** baru.

Knowledge → pengetahuan

Data → dalam jumlah besar



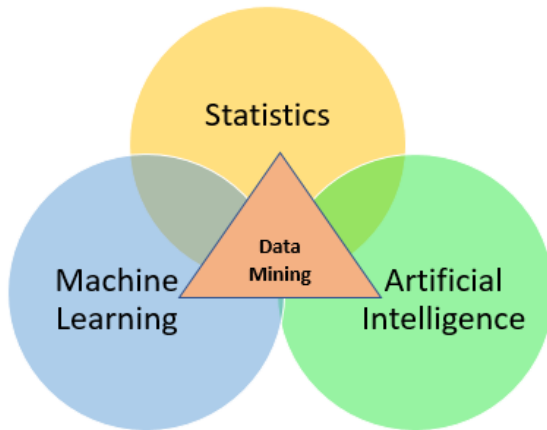
Contoh Data Mining:

ARM → *Association Rule Mining* → mencari aturan asosiasi antar item

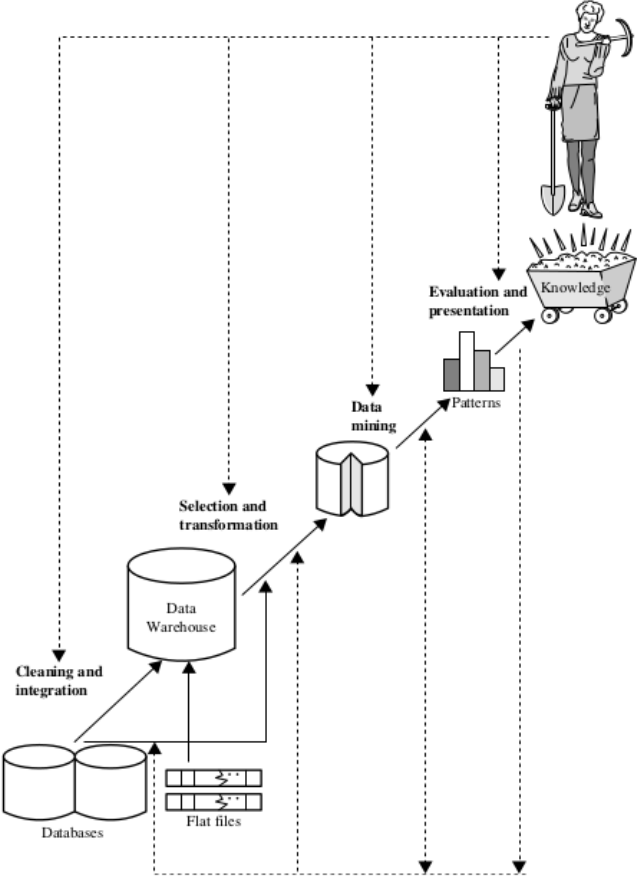
Contoh di mini market → data transaksi

-- Orang beli teh maka beli gula

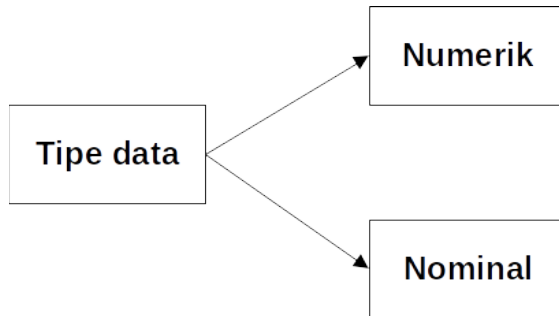
3. Ilmu yang berkaitan dengan data mining?



4. Proses Knowledge Discovery from Data



5. Data, Atribut/Variabel, Class/Label/Target



Data → fakta mentah yang belum ada maknanya.

Tipe data:

- Numerik: ipk, jarak, umur, tinggi badan
- Nominal: jenis kelamin, hobi, kelas, agama

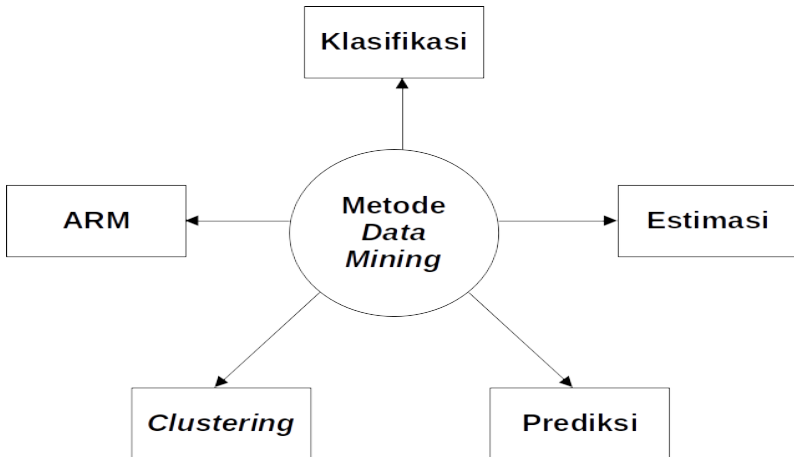
Asal sekolah	Nilai UN	Jenis Kelamin	Lulus Tepat Waktu
SMK Hasanah	68	Laki-laki	Tepat waktu
SMK Taruna	85	Laki-laki	Tepat waktu
SMA 1 Pujud	87	Laki-laki	Tidak tepat waktu

Asal sekolah, nilai UN, jenis kelamin → **Atribut/Fitur**

Lulus tepat waktu → **kelas/class/target/label**

Nilai/value

6. Metode-Metode Data Mining



1 3 3 6 4 5 1 3 3 6 4 5 1 ?

H H T H H T H

6.1. Klasifikasi

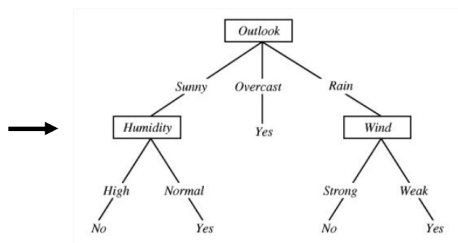
Ciri-ciri:

- Atribut: boleh nominal atau numerik
- Class: nominal

Contoh kasus:

- Klasifikasi seorang mahasiswa akan lulus tepat waktu atau tidak
- Klasifikasi seorang mahasiswa berhak menerima beasiswa atau tidak

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No



Contoh algoritma:

- id3
- KNN

6.2. Estimasi

Ciri-ciri:

- Atribut: boleh nominal atau numerik
- Kelas: numerik

Cycle time (ns)	Main memory (Kb)		Cache (Kb)	Channels		Performance
	MMIN	MMA		CHMIN	CHMAX	
MYCT	MMIN	MMA	CACH	CHMIN	CHMAX	PRP
125	256	6000	256	16	128	198
29	8000	32000	32	8	32	269
480	512	8000	32	0	0	67
480	1000	4000	0	0	0	45

- **Linear regression function**

$$\text{PRP} = -55.9 + 0.0489 \text{ MYCT} + 0.0153 \text{ MMIN} + 0.0056 \text{ MMA} \\ + 0.6410 \text{ CACH} - 0.2700 \text{ CHMIN} + 1.480 \text{ CHMAX}$$

Contoh kasus:

- Estimasi umur pohon dari diameter batang
- Estimasi performa CPU

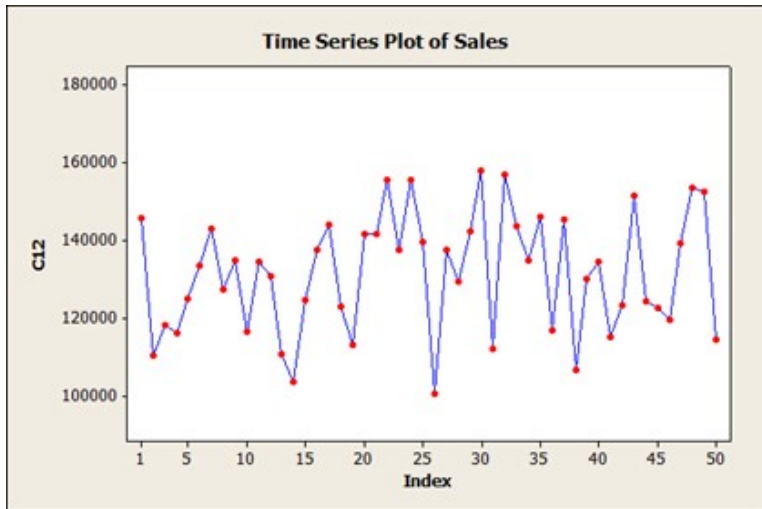
Contoh algoritma:

- Regresi linear
- Backpropagation

6.3. Prediksi

Ciri-ciri:

- Datanya harus time series



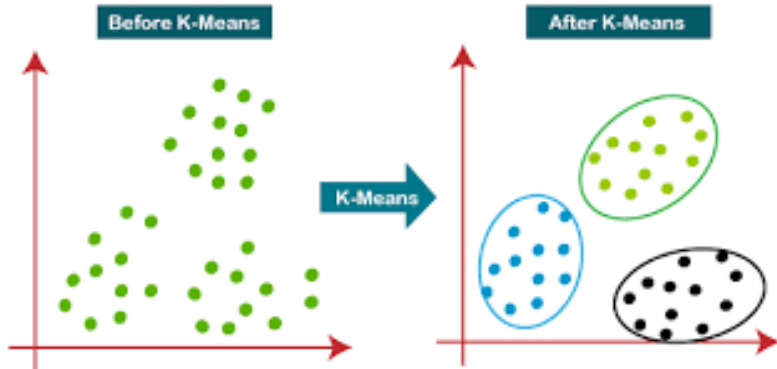
Contoh kasus:

- Prediksi curah hujan
data curah hujan:
1 3 5 4 7 1 3 5 x
Berapa kira-kiran nilai x?
- Prediksi jumlah penjualan

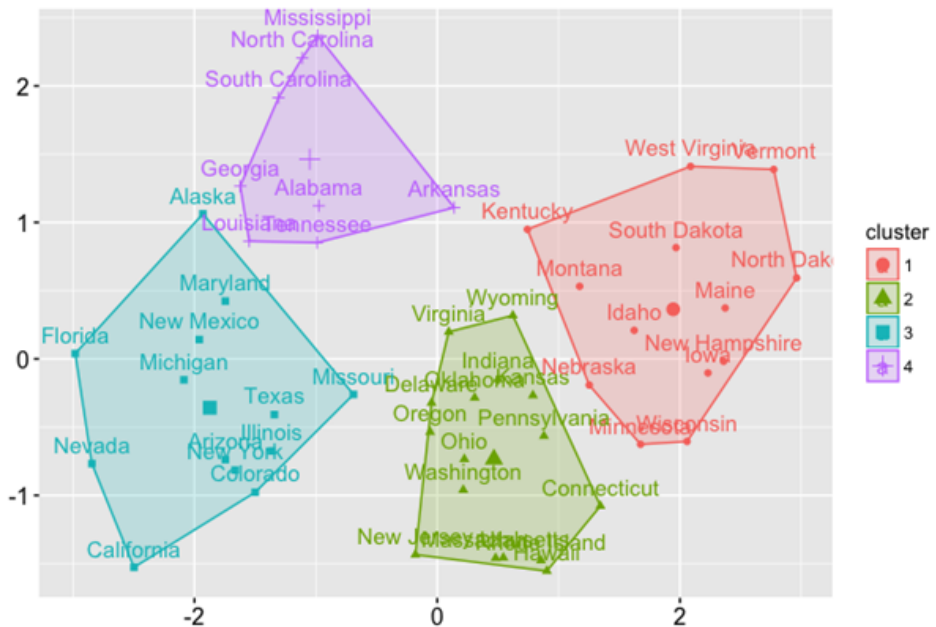
6.4. Clustering

Ciri-ciri:

- Tidak kelas/target
- Pengelompokan berdasarkan kedekatan/kemiripan antar data



Cluster plot



Contoh kasus:

- Pengklasteran data rakyat miskin

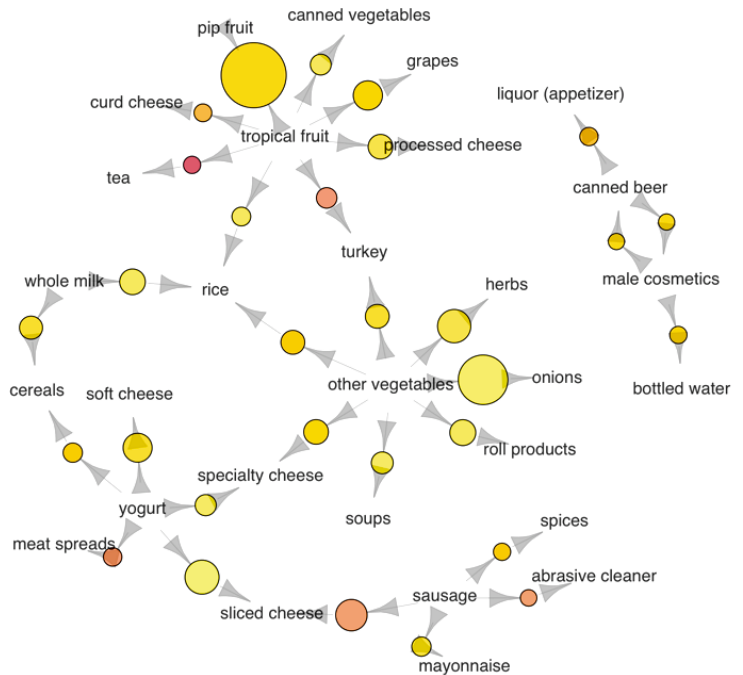
Contoh algoritma:

- K-Means
- SOM
- DBScan

6.5. Association Rule Mining (ARM)

Ciri-ciri:

- Hubungan antar item
- Market basket analysis



Contoh Kasus:

- Hubungan antar item yang dibeli di supermarket
- Hubungan antara hari dan item yang dibeli di supermarket

Contoh Algoritma:

- Apriori
- FP-Growth
- ECLAT

7. Machine Learning

