

# Data Mining

Masalah-Masalah Dalam Data Mining

# Metode Data Mining

- › Prediksi (Prediction Method)
  - Memprediksi nilai yang akan datang berdasarkan variabel-variabel yang sudah ada
  - Menentukan target (variabel terikat) dan penjelas (variabel bebas)
- › Deskripsi (Description Method)
  - Menemukan pola yang dapat diprediksi manusia untuk mendeskripsikan data

## Jenis Permasalahan

- › Klasifikasi → Prediktif
- › Forecasting → Prediktif
- › Clustering → Deskriptif
- › Association Rule Discovery → Deskriptif
- › Sequential Pattern Discovery → Deskriptif
- › Deviation Detection → Prediktif

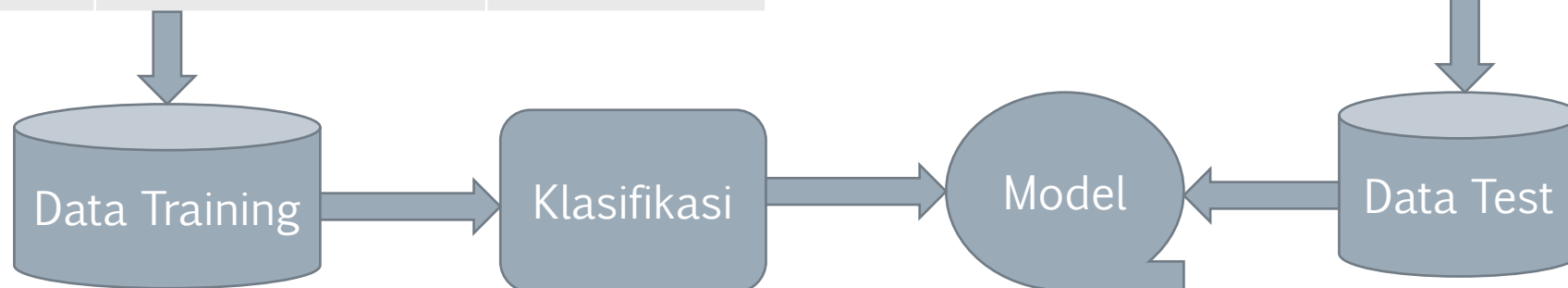
# Klasifikasi

- › Klasifikasi → sebuah proses menemukan definisi kesamaan karakteristik dalam suatu kelompok atau kelas (class)
- › Sekumpulan record untuk data training
  - Masing-masing record berisi sekumpulan atribut, dimana masing-masing atribut berisi kelas output
- › Tentukan model untuk masing-masing class atribut sebagai fungsi nilai dari atribut lain
- › Tentukan tujuan berdasarkan record yang tidak terlihat sebelumnya yang berasal dari class terukur untuk menghasilkan output yang akurat
  - Test digunakan untuk menentukan tingkat akurasi model
  - Data set biasanya dibagi menjadi ke dalam training data test set, dimana data training digunakan untuk membangun model dan test set digunakan sebagai validasi

# Contoh Klasifikasi

No	Lancar	Status	Pendapatan
1	Ya	Belum menikah	1.250.000
2	Tidak	Menikah	1.000.000
3	Tidak	Belum menikah	700.000
4	Ya	Menikah	1.200.000
5	Tidak	Cerai	950.000
6	Ya	Menikah	600.000
7	Ya	Cerai	2.200.000
8	Tidak	Belum menikah	850.000
9	Tidak	Menikah	750.000
10	Tidak	Belum menikah	900.000

Lancar	Status	Pendapatan
?	Belum menikah	750.000
?	Menikah	500.000
?	Menikah	1.500.000
?	Cerai	900.000
?	Belum menikah	400.000
?	Menikah	800.000



# Penerapan-Klasifikasi

## › Direct marketing

- Tujuan → untuk mengurangi biaya dengan mengirimkan informasi katalog produk terbaru kepada customer target
- Pendekatan
  - › Menggunakan data produk dengan kemiripan yang sudah pernah ditawarkan atau sudah pernah dibeli sebelumnya
  - › Tentukan atribut yang mempengaruhi keputusan customer dalam membeli atau tidak membeli produk
  - › Kumpulkan data demografis, gaya hidup, dan informasi lain yang dibutuhkan perusahaan
  - › Gunakan informasi sebagai atribut input untuk membangun model

# Penerapan-Klasifikasi

## › Persaingan customer

- Tujuan → memprediksi apakah konsumen memiliki kecenderungan untuk berpindah pada competitor
- Pendekatan :
  - › Gunakan detail record transaksi yang berhubungan dengan konsumen lama dan baru untuk memperoleh atribut
    - Seberapa sering customer menghubungi, dimana dia menghubungi, waktu paling sering, status finansial dll
  - › Labelkan loyalitas customer (Loyal atau Non Loyal)
  - › Temukan model untuk mengetahui loyalitas customer

# Forecasting (Regresi)

- › Prediksikan nilai dari sekumpulan variabel continues yang diberikan berdasarkan nilai variabel lain dengan mengasumsikan ketergantungan model
- › Contoh :
  - Memprediksi volume penjualan produk baru berdasarkan pemasukan iklan
  - Memprediksi kecepatan angin sebagai fungsi temperature, kelembaban, tekanan udara dsb
  - Prediksi waktu untuk stock



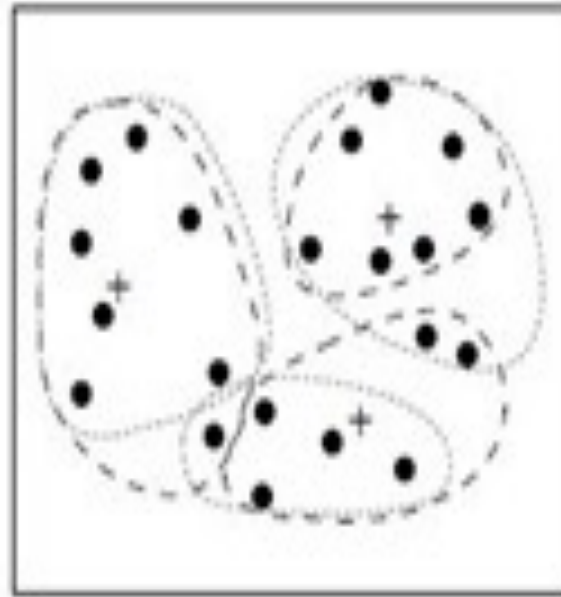
# Clustering

- › Menemukan cluster dari sekumpulan set data point dimana masing-masing set memiliki sekumpulan atribut yang diukur menggunakan pengukuran yang sama
  - Data point dalam satu cluster hampir mirip satu sama lain baik di dalam maupun di luar cluster
  - Data point dalam cluster yang terpisah sedikit memiliki kesamaan satu sama lain baik di dalam maupun di luar cluster
- › Contoh :
  - Memprediksi volume penjualan produk baru berdasarkan pemasukan iklan
  - Memprediksi kecepatan angin sebagai fungsi temperature, kelembaban, tekanan udara dsb
  - Prediksi waktu untuk stock

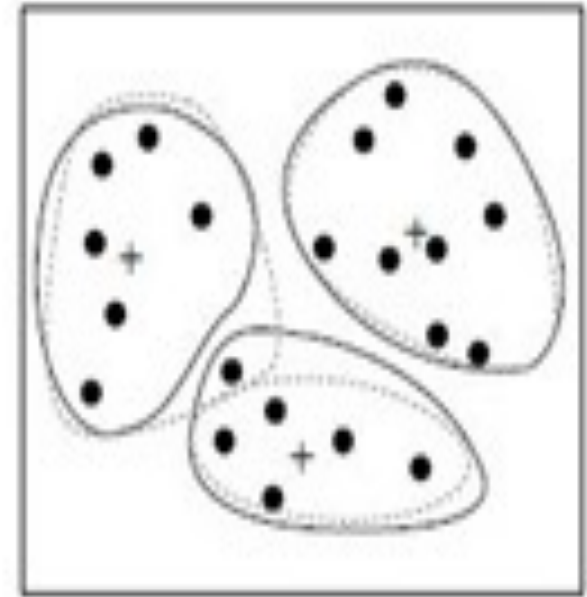
# Ilustrasi Clustering



(a) *cluster Awal*



(b) Proses Iterasi



(c) *cluster Akhir*

# Penerapan Clustering

## › Segmentasi pasar

### – Tujuan :

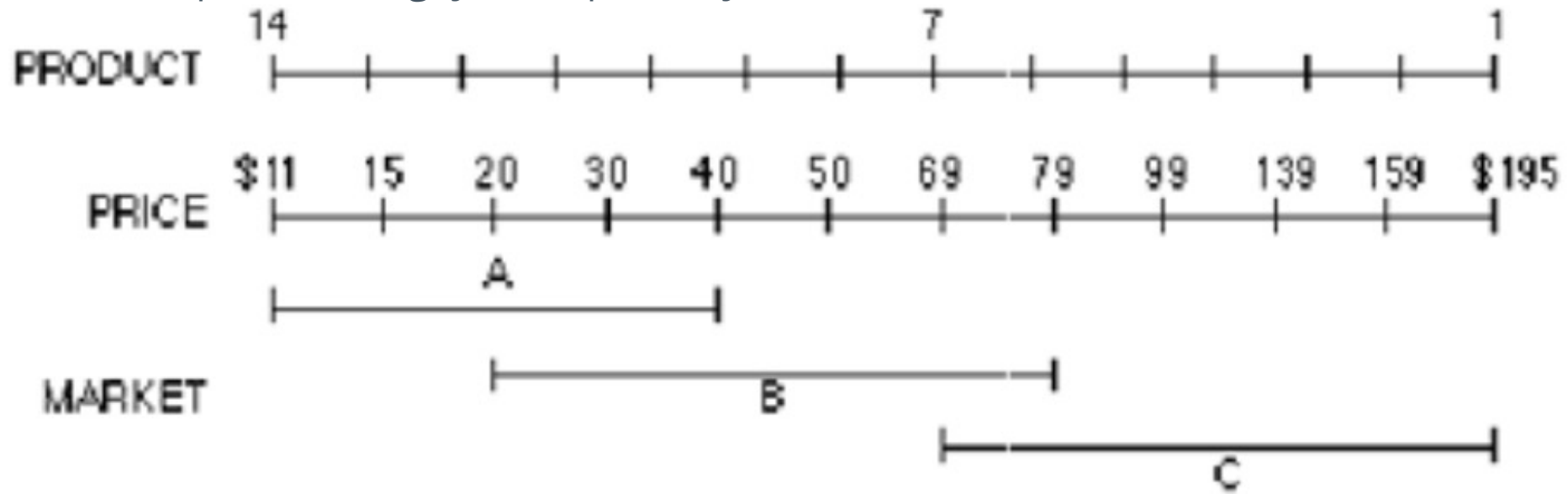
- › Membagi pasar kedalam beberapa kelompok customer

### – Pendekatan :

- › Kumpulkan atribut yang berbeda dari customer berdasarkan letak geografis dan perbedaan gaya hidup (lifestyle)

- › Temukan cluster dari kelompok customer yang hamper sama

- › Ukur kualitas clustering dengan mengobservasi pola pembelian customer dalam cluster yang sama dengan pola pembelian dalam cluster yang berbeda




# Penerapan Clustering

## › Pengelompokan Dokumen

- Tujuan :
  - › Menemukan kelompok dokumen yang mirip satu sama lain berdasarkan tingkat kepentingan dokumen
- Pendekatan :
  - › Identifikasi ketentuan (term), keyword yang paling sering muncul dalam dokumen
  - › Dari pengukuran kesamaan berdasarkan frekuensi keseringan munculnya sebuah term, keyword untuk proses cluster

## – Manfaat :

- › Pengambilan informasi dapat memanfaatkan cluster untuk menghubungkan dokumen baru dengan dokumen lama atau melakukan pencarian berdasarkan term



Article	Key words
1	Dollar: 1, industry: 4, country: 2, loan: 3...
2	Machinery: 2, labor: 3, market: 4, industry: 2...
3	Patient: 4, symptom: 2, drug: 2, heath: 2...
4	Medical: 2, cost: 3, patient: 2, health: 3...
?	

Terima Kasih