



Konsep Dasar Data Mining Pada Era 4.0

Oleh: Dr. Indra, S.Kom , M.T.I

Sumber:

<https://www-users.cs.umn.edu/~kumar001/dmbook/index.php>

Outline

- Pengenalan Data Mining
- Alasan Penggunaan Data Mining
- Definisi Data mining
- Data Mining Task
- Praktikum Data Mining dengan Orange (Visual Programming)

Data Berskala Besar ada di Sekitar Kita

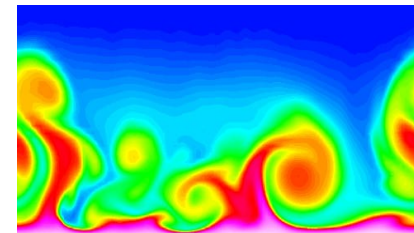
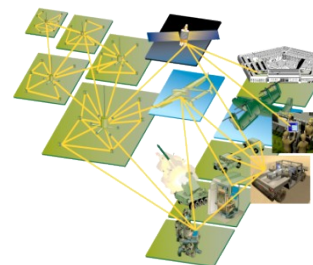
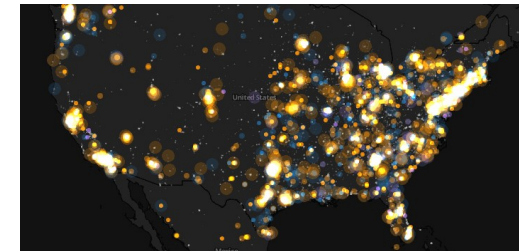
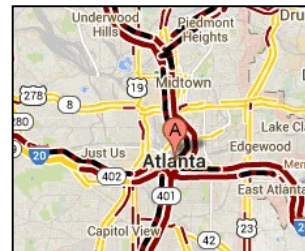
Telah berkembang dengan pesat pertumbuhan data yang sangat besar dalam database E-Commerce dan database ilmiah karena kemajuan teknologi menghasilkan data dalam jumlah besar dan majunya teknologi pengumpulan

- **New mantra**

Kumpulkan data apa pun yang Anda bisa kapanpun dan dimanapun.

- **Expectations**

Data yang terkumpul akan memiliki nilai tinggi untuk tujuan tertentu atau untuk tujuan yang tidak dibayangkan



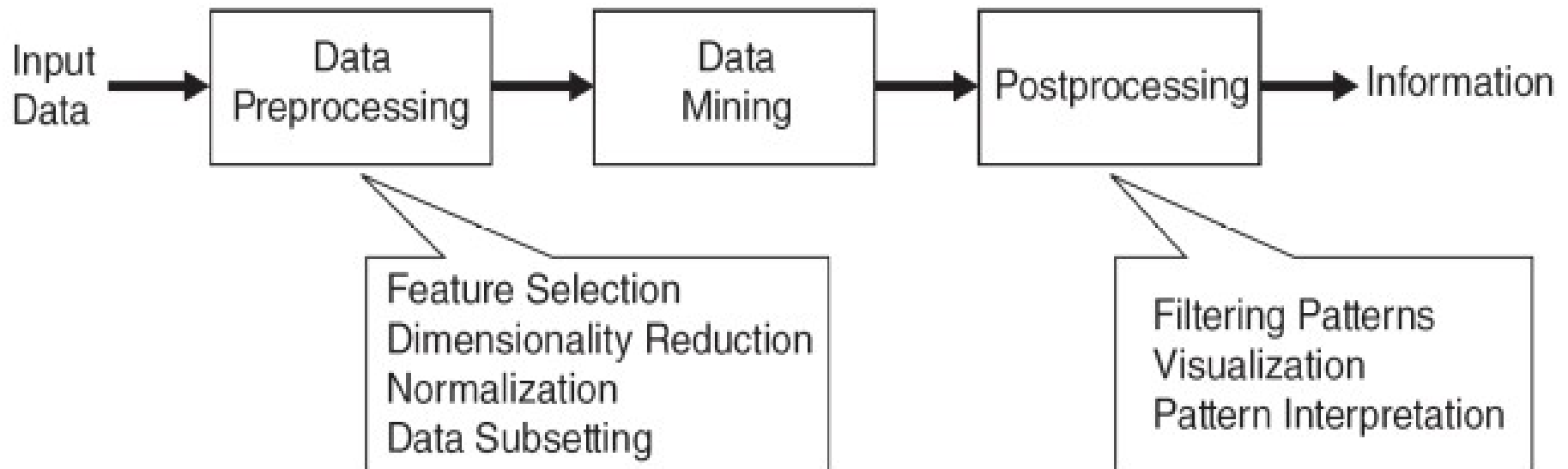
Mengapa ada Data Mining

- Data terkumpul dalam jumlah besar dan ditempatkan pada Pangkalan Data (Data Warehoused)
 - Web Data: Twitter, Google, Facebook, Instagram
 - E-Commerce: Amazon, Tokopedia, Bukalapak
- Komputer menjadi lebih murah dan teknologi lebih powerful
- Tingkat persaingan yang semakin tinggi
- Datamining membantu Ilmuwan untuk menganalisis dataset dalam jumlah besar dan membangun hipotesis

Definisi Data Mining (Tan et al, 2010)

- Ekstraksi informasi pada suatu data dengan ukuran besar, yang sebelumnya tidak diketahui, dan berpotensi memberikan informasi dari data tertentu
- Eksplorasi & analisis, dengan cara otomatis atau semi-otomatis, data dalam jumlah besar untuk menemukan pola yang bermakna

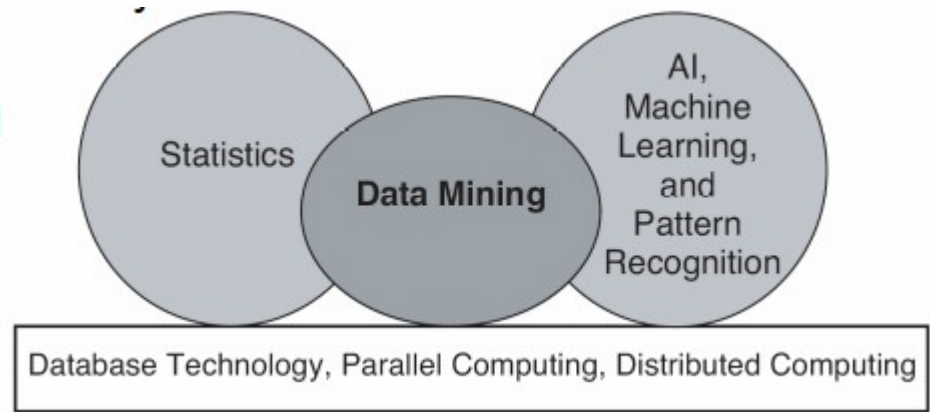
Definisi Data Mining



Asal Usul Penambangan Data

- Menarik ide dari pembelajaran mesin / AI, pengenalan pola, statistik, dan sistem basis data
- Teknik tradisional mungkin tidak cocok untuk data dengan kriteria:

- Skala besar
- Dimensi tinggi
- Heterogen
- Kompleks
- Didistribusikan



- Komponen utama dari bidang baru yang muncul dari ilmu data science dan penemuan berbasis data (KDD)

Data Mining Task

- **Metode Prediksi**

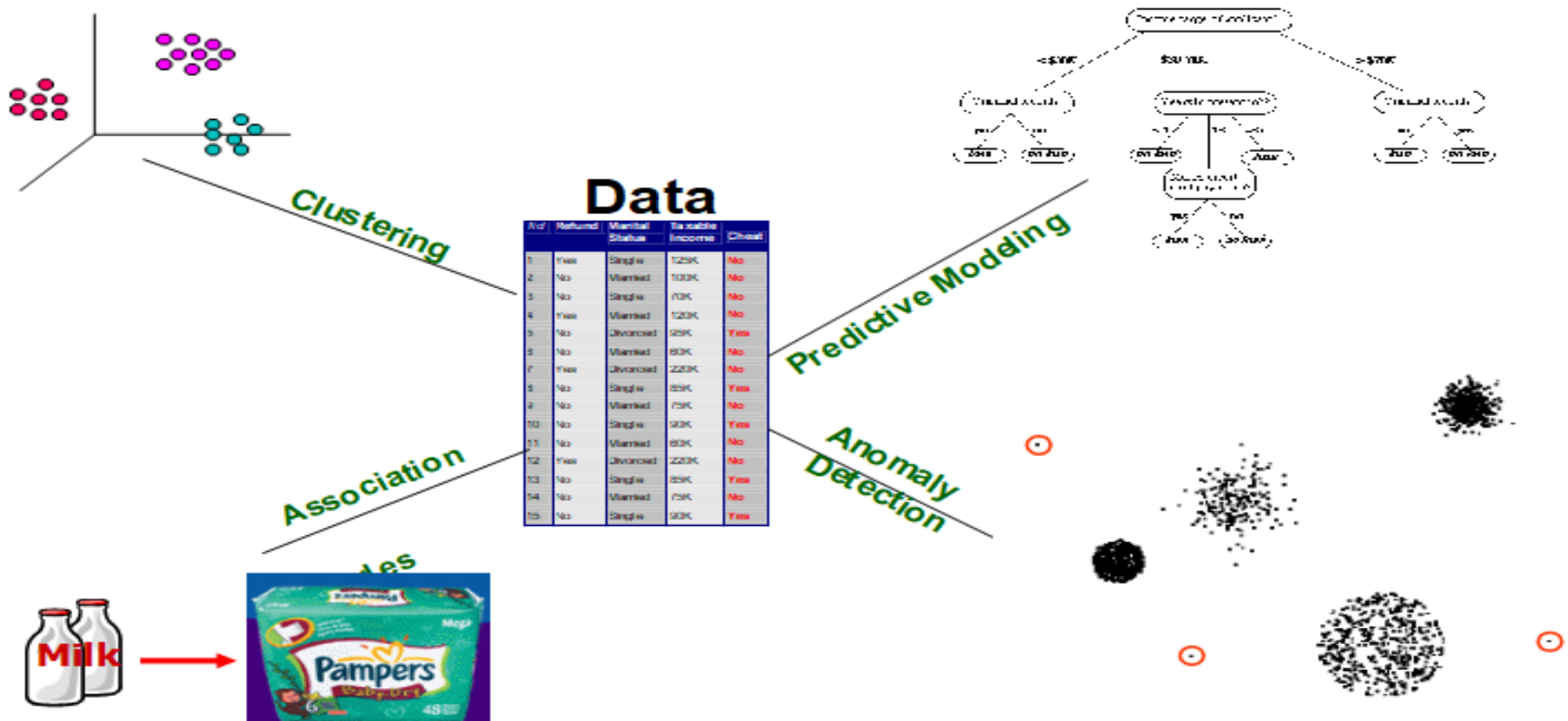
- Menggunakan beberapa variabel atau atribut untuk memprediksi nilai variabel atau atribut lain yang tidak diketahui untuk kebutuhan di masa mendatang.

- **Metode Deskripsi**

- Menemukan pola yang bisa ditafsirkan oleh manusia dan menggambarkan dataset yang digunakan.

Data Mining Task

Data Mining Tasks ...

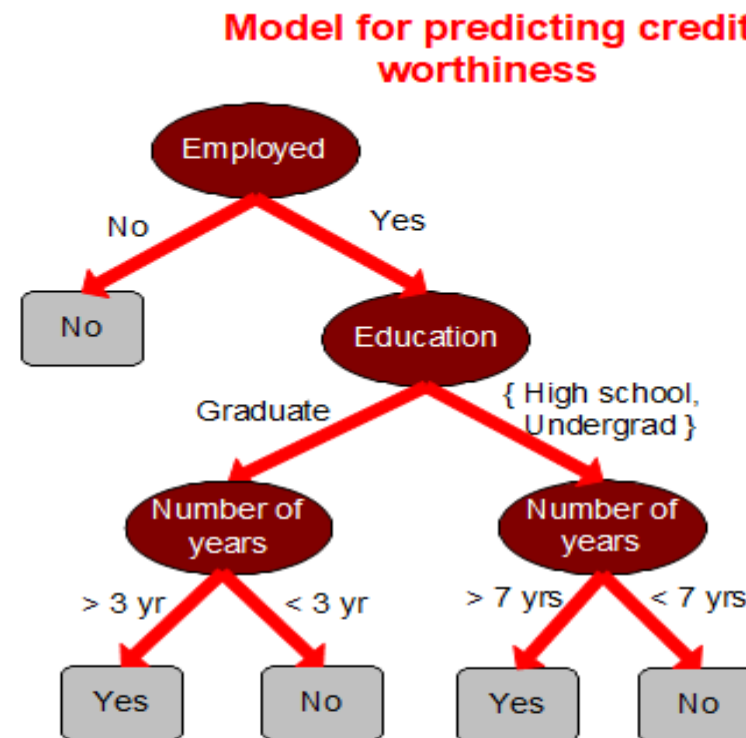


Model Prediksi: Klasifikasi

- Mencari model untuk kelas atribut sebagai target

Class

<i>Tid</i>	Employed	Level of Education	# years at present address	Credit Worthy
1	Yes	Graduate	5	Yes
2	Yes	High School	2	No
3	No	Undergrad	1	No
4	Yes	High School	10	Yes
...	...	...	...	...



01/17/2018

Introduction to Data Mining, 2nd Edition

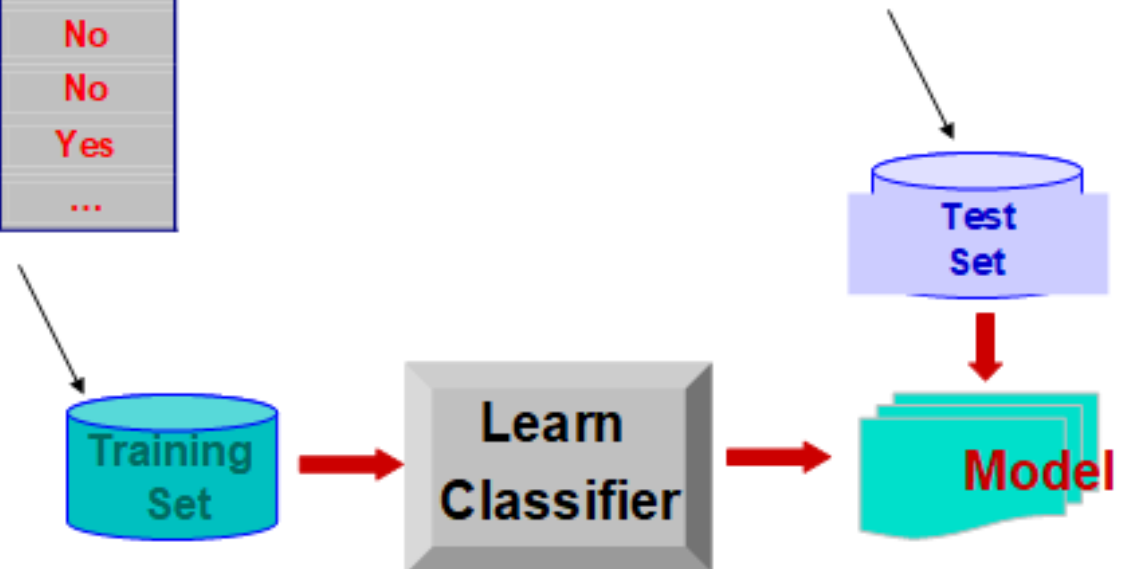
12

Classification Example

categorical *categorical* *quantitative* *class*

<i>Tid</i>	Employed	Level of Education	# years at present address	Credit Worthy
1	Yes	Graduate	5	Yes
2	Yes	High School	2	No
3	No	Undergrad	1	No
4	Yes	High School	10	Yes
...	...	...	...	...

<i>Tid</i>	Employed	Level of Education	# years at present address	Credit Worthy
1	Yes	Undergrad	7	?
2	No	Graduate	3	?
3	Yes	High School	2	?
...	...	...	...	...



Contoh Klasifikasi

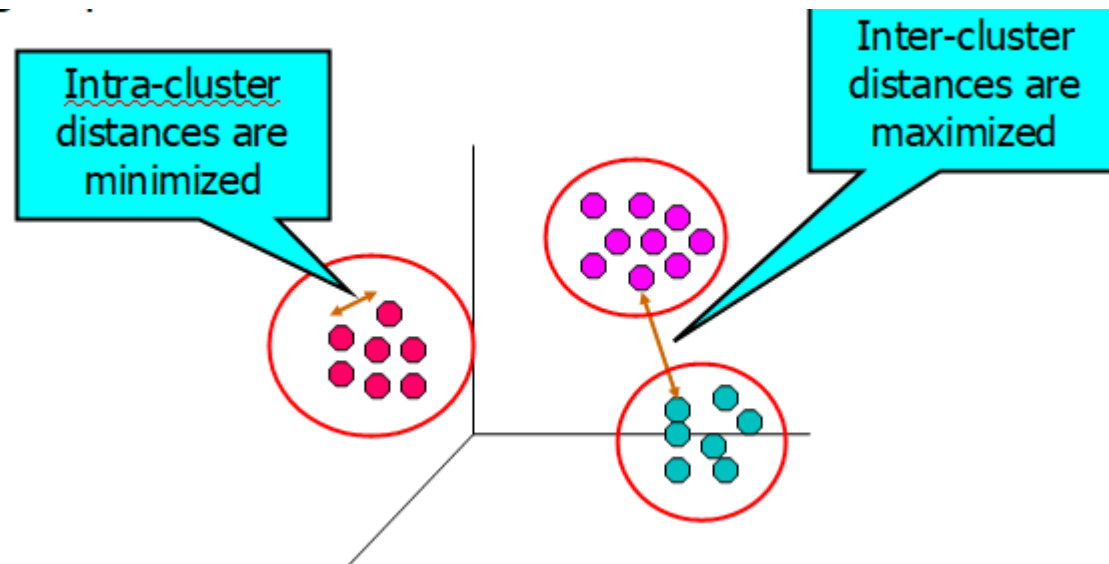
- Klasifikasi transaksi kartu kredit dengan kategori normal atau menipu
- Klasifikasi wilayah (badan air, daerah perkotaan, hutan, dll.) Menggunakan data satelit
- Mengkategorikan berita sebagai berita keuangan, cuaca, hiburan, olahraga, dll

Regression

- Memprediksi nilai variabel bernilai kontinu yang diberikan berdasarkan nilai-nilai variabel lain, dengan asumsi model ketergantungan linear atau nonlinear.
- Diperdalam pada Statistik dan neural network
- Contoh:
 - Memprediksi jumlah penjualan produk baru berdasarkan pembelian yang menguntungkan.
 - Prediksi deret waktu dari indeks pasar saham.

Clustering

Menemukan kelompok objek sedemikian rupa sehingga objek dalam klaster akan serupa (atau terkait) satu sama lain dan berbeda dari (atau tidak terkait dengan) objek dalam grup lain



Contoh Klasterisasi

- Pengelompokkan pelanggan untuk target marketing
- Pengelompokan dokumen berdasarkan kesamaan dokumen yang dicari
- Summarization : Abstrak dan Keyword

Definisi Association Rule

- Berisi sekumpulan transaksi pembelian barang yang berisi item-item barang (susu, keju, indomie)
- Menghasilkan aturan ketergantungan yang akan memprediksi terjadinya pembelian barang berdasarkan kemunculan barang-barang lainnya.

<i>TID</i>	<i>Items</i>
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

Rules Discovered:

{Milk} --> {Coke}

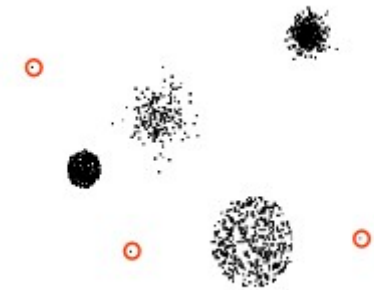
{Diaper, Milk} --> {Beer}

Contoh Association Analysis

- Market Based analysis
- Medical Informatics
- Telecommunication alarm diagnosis
- Keterkaitan Peristiwa Bencana Alam
- Kelulusan Tepat waktu Mahasiswa

Anomaly Detection

- Mendeteksi penyimpangan yang signifikan dari perilaku normal
- Contoh:
 - Deteksi Penipuan Kartu Kredit
 - Network Intrusion
 - Identifikasi perilaku anomali dari jaringan sensor untuk pemantauan dan pengawasan.



Dataset dan Data Preprocessing (Praktikum)

- Aggregation
- Sampling
- Dimensionality Reduction
- Feature subset selection
- Feature creation
- Discretization and Binarization
- Attribute Transformation

**Matur Nuwun
(Terima Kasih)**