



16

Data Warehouse dan Data Mining

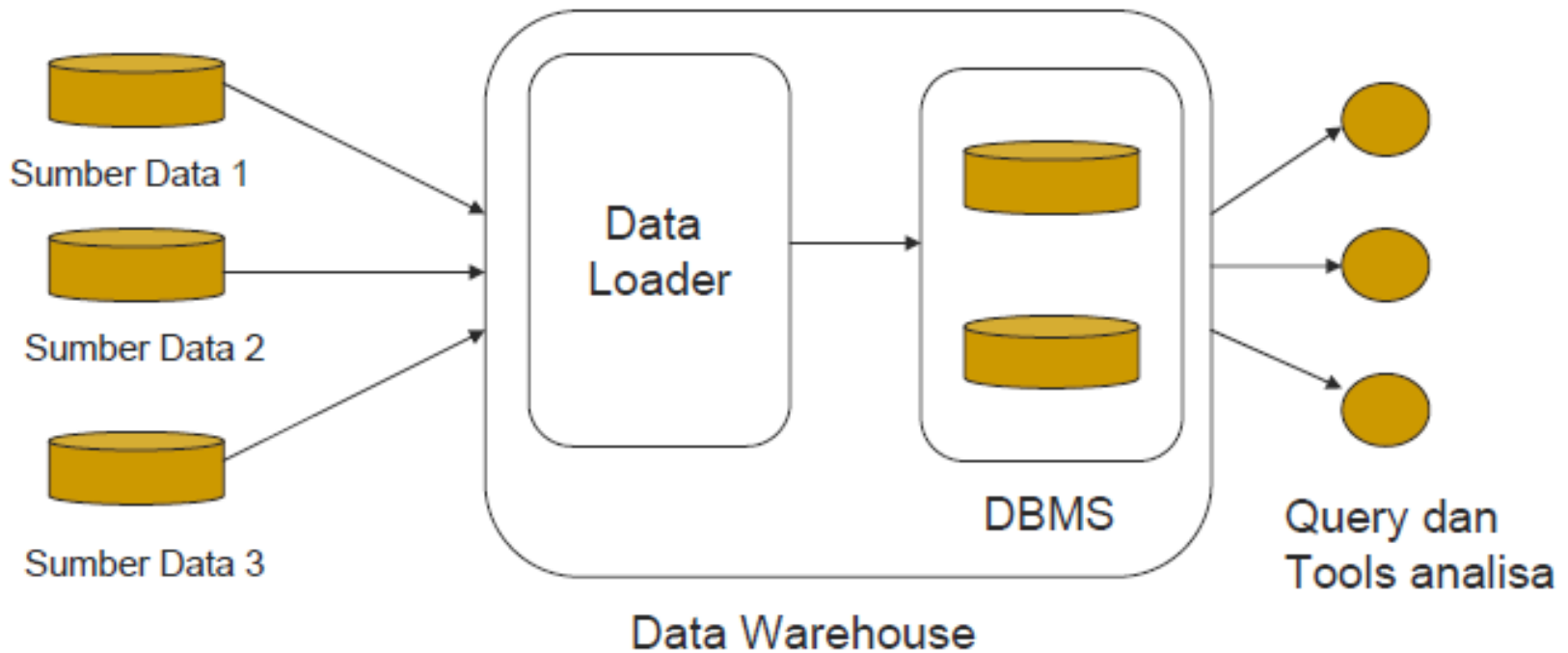
by: Ahmad Syauqi Ahsan

Data Warehouse

- Data warehouse merupakan gudang (atau arsip) dari informasi yang diperoleh dari banyak sumber, disimpan dalam skema basis data yang tunggal, pada satu lokasi saja.
- Begitu diperoleh, data akan disimpan dalam waktu yang lama sehingga memungkinkan pengaksesan data historis, karena itu data warehouse memberikan pemakai sebuah akses ke data yang terkonsolidasi.

Arsitektur Data Warehouse

3



Yang perlu dipertimbangkan

Ada sejumlah hal penting yang harus dipertimbangkan dalam pembentukan data warehouse, diantaranya:

1. **Kapan dan bagaimana mendapatkan data.**

Pengambilan data biasanya diluar jam sibuk, mengingat data transaksi tidak dapat diganggu gugat dalam operasi pengambilan data. Serta bagaimana cara mendapatkannya, bisa sumber data secara aktif mengirimkan data kepada data warehouse atau data warehouse yang melakukan permintaan kepada sumber data. Keduanya bisa dilakukan secara periodik.

2. **Skema konsolidasi yang digunakan**

Biasanya sumber data berasal dari DBMS yang berbeda-beda. Salah satu tugas data warehouse adalah membentuk skema konsolidasi (gabungan) dari berbagai sumber data tersebut menjadi satu kesatuan skema basis data.

Yang perlu dipertimbangkan (2)

5

3. **Bagaimana menjalankan perubahan**

Perubahan data pada sebuah tabel disumber data harus pula dijalankan pada data warehouse

4. **Data apa yang direkapitulasi**

Data mentah yang dihasilkan pada data transaksi bisa jadi menjadi teramat besar bagi data warehouse, sehingga kadangkala hanya diperlukan rekapitulasi (summary).

Definisi Data Mining

6

- Merupakan istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database
- Penggunaan teknik-teknik analisa data, matematika, kecerdasan tiruan dan machine learning untuk mengekstrak informasi yang berkaitan dan pengetahuan terkait dari berbagai database besar
- Merupakan proses dari pola2 rancang bangun aturan, korelasi atau tren atau model prediksi

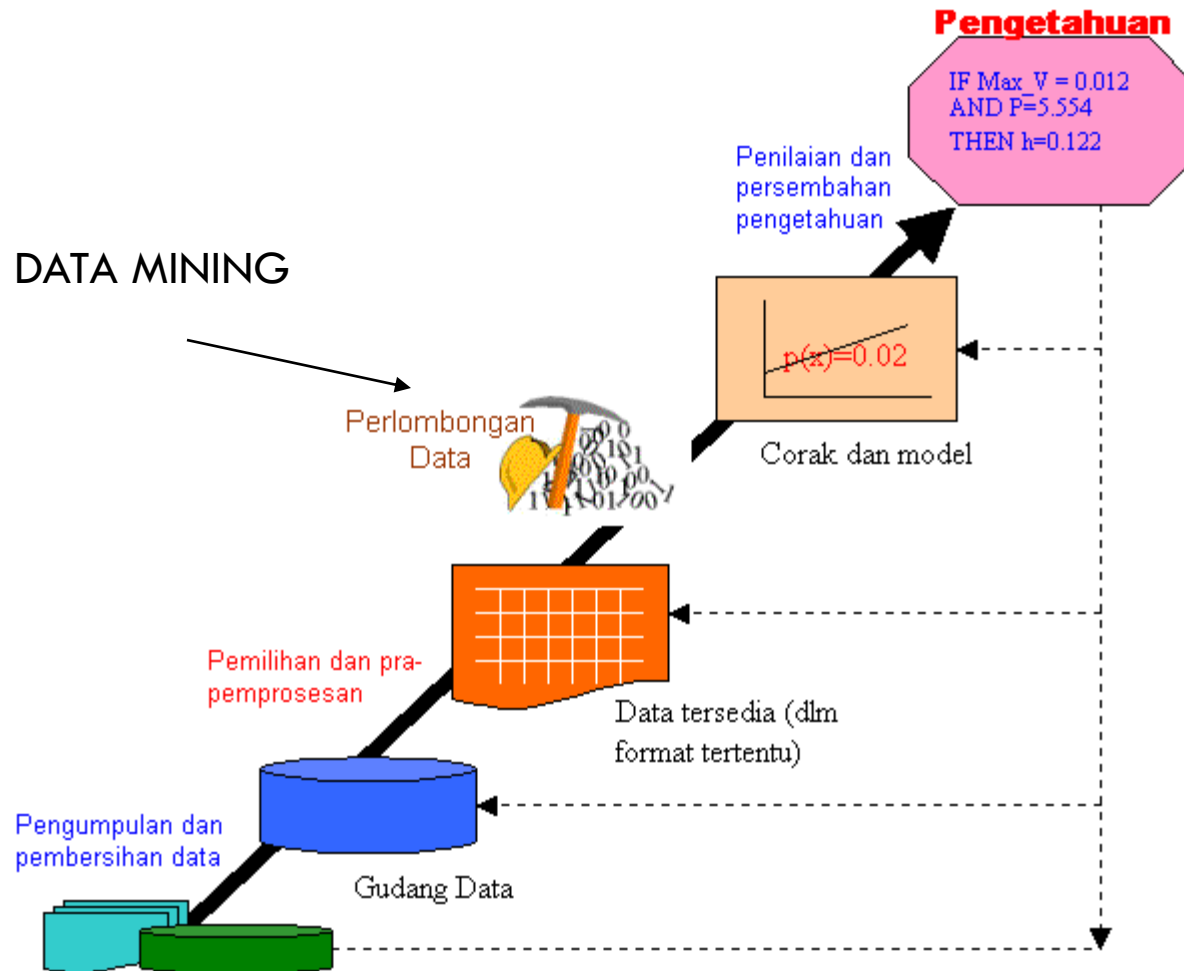
Karakteristik Utama

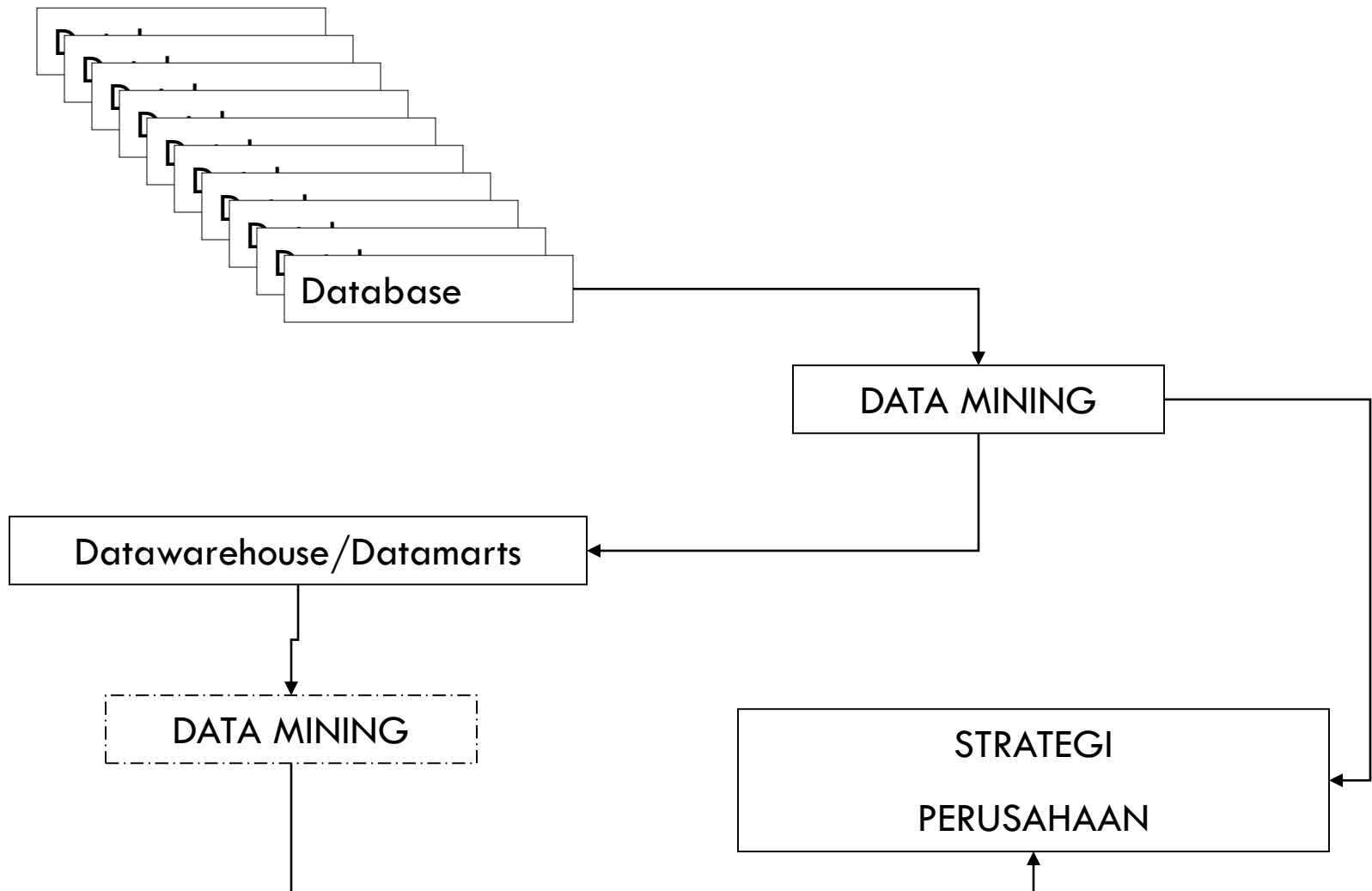
7

- Berasal dari database yang sangat besar, merupakan data beberapa tahun
- Berada dilingkungan arstektur sistem client/server atau arsitektur berbasis web
- Perlu piranti visualisasi data yang canggih
- Kadang-kadang perlu paralel processing
- Pengguna biasanya merupakan pengguna akhir yang tidak mempunyai skill programming

Penemuan Pengetahuan

8





Bagaimana Data Mining Bekerja

10

- Data mining bekerja untuk menemukan pola-pola tertentu dalam data.
- Data mining cerdas menghasilkan informasi dari datawarehouse yang tidak dapat divisualisasikan oleh laporan-laporan dan query.

METODE

11

- Tiga metode yang digunakan untuk mengidentifikasi pola-pola data (Nemati dan Barko 2001)
 - ▣ Model Sederhana (Query berbasis SQL, OLAP, Pertimbangan Manusia)
 - ▣ Model Sedang (Regresi, pohon keputusan, clustering)
 - ▣ Model Kompleks (jaringan saraf, induksi aturan lain)

Pendekatan

12

- Untuk menemukan hubungan yang relevan dalam data berbagai teknik data mining menggunakan pendekatan berikut :
 - ▣ Klasifikasi dan Clustering: pengelompokan berdasar karakteristik tertentu (contoh: pelanggan yang hilang karena pesaing)
 - ▣ Asosiasi: hubungan peristiwa yang terjadi pada satu waktu (produk apa yang terjual bersama produk lain)
 - ▣ Sekuensi: Sama dengan asosiasi tetapi hubungan terjadi pada lebih dari satu periode (kunjungan berulang pada sebuah supermarket)
 - ▣ Regresi: Digunakan untuk memetakan data untuk suatu nilai regresi
 - ▣ Forecasting: Memprediksi masa depan berdasarkan pola-pola data set tertentu

PENDEKATAN->TUGAS

PERAMALAN	Meramal keadaan masa depan berdasarkan data masa lalu	Ramalan nilai jual untuk minggu berikutnya
KLASIFIKASI	Mengatur atau memberikan kelas kepada data berdasarkan kepada atribut	Pengkelasan pelajar berdasarkan kepada hasil ujian.
PENGELOMPOKAN	Mengumpulkan data-data menurut kumpulan mengikut corak dan ciri data.	Mendapatkan kumpulan tertentu dari corak dan jenis belanja
HUBUNGAN	Kajian terhadap keberadaan dan hubungan setiap elemen dalam sumber data yang bersifat transaksi.	Mendapatkan frekuensi item yang selalu dibeli bersama dalam semua transaksi dan mendapatkan bahawa (cth.): Roti selalu dibeli bersama-sama dengan jem

Trigger

14

- Data mining dapat dikendalikan oleh dua hal
 - ▣ Hipotesa
 - Mengasumsikan sesuatu untuk kemudia diteliti dan divalidasi kebenarannya
 - ▣ Penemuan
 - Memeriksa data set yang kemudian melahirkan penemuan akan pola-pola data tertentu

Pengelompokan Piranti Data Mining

15

- Bucks(2000) mengelompokan piranti DM sesuai hubungannya dengan teknologi informasi dan Business Intelligent
 - ▣ Paket analisis statistik dan matematis
 - ▣ Piranti personalisasi untuk pemasaran web
 - ▣ Analitik kedalam platform pemasaran
 - ▣ Peranti CRM tingkat lanjut
 - ▣ Analitik pada platform spesifik industri vertikal
 - ▣ Analitik pada piranti database (OLAP, OLTP)
 - ▣ Piranti data mining standalone

7 Langkah Data Mining

16

1. Menentukan Masalah Bisnis
2. Membangun (menemukan) database DM
3. Menyelidiki data
4. Menyiapkan data untuk pemodelan
5. Membangun (menemukan model)
6. Mengevaluasi model.
7. Bertindak atas hasil yang diperoleh

Mitos DM

17

1. DM menyediakan prediksi sekejap, kenyataannya merupakan proses yang panjang dan proaktif
2. DM belum jelas untuk aplikasi bisnis apa, kenyataannya hampir semua aplikasi bisnis dapat menggunakan DM
3. DM memerlukan database terpisah, kenyataannya oleh perkembangan teknologi informasi DM dapat tidak memiliki database terpisah
4. Hanya seorang ilmuwan yang dapat menerapkan DM, kenyataannya aplikasi DM sekarang memungkinkan seorang manager menengah menggunakan DM
5. DM hanya untuk perusahaan yang mempunyai data yang sangat besar, kenyataannya jika data secara akurat dapat mencerminkan bisnis maka perusahaan kecilpun dapat menggunakan DM

10 Kesalahan DM

18

1. Memilih masalah yang salah
2. Mengabaikan apa yang dipikirkan oleh sponsor anda dalam soal apa yang dapat dilakukan apa yang tidak
3. Tidak memberi cukup waktu untuk persiapan data
4. Hanya melihat pada hasil yang dikumpulkan tidak pernah melihat arsip individu
5. Asal-asalan dalam pengawasan prosedur DM
6. Mengabaikan penemuan yang mencurigakan dan langsung berpindah pada proses selanjutnya\
7. Menjalankan algoritma secara serampangan tanpa memikirkan tahapan-tahapannya
8. Mempercayai segala informasi yang diterima
9. Mempercayai segala informasi mengenai hasil analisis
10. Mengukur hasil DM dengan cara yang berbeda dengan ukuran yang digunakan oleh sponsor anda

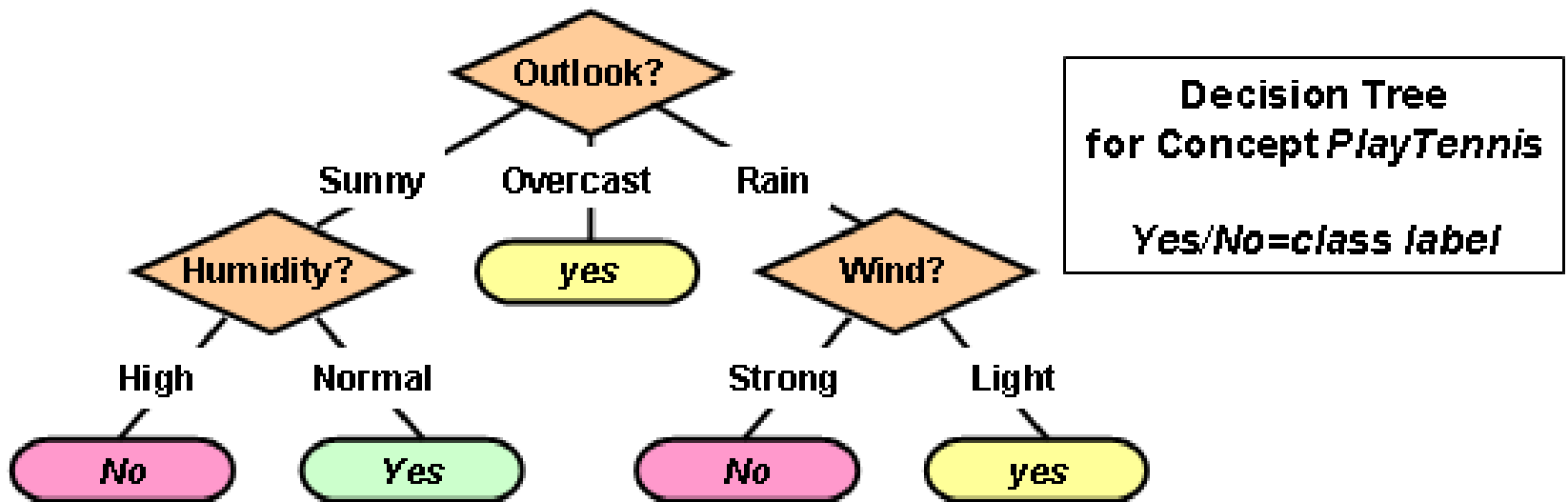
Teknik DM

19

- Metode statistik (regresi, estimasi, probabilitas)
- Pohon keputusan
- Pemikiran berbasis kasus
- Komputasi syaraf (mencari pelanggan prospektif untuk produk baru)
- Agent cerdas
- Algoritma genetika

Contoh Pohon Keputusan

20



Tanya Jawab

Terima Kasih