

Modul Manajemen Pengetahuan

Data Mining Business Intelligent

Latar Belakang

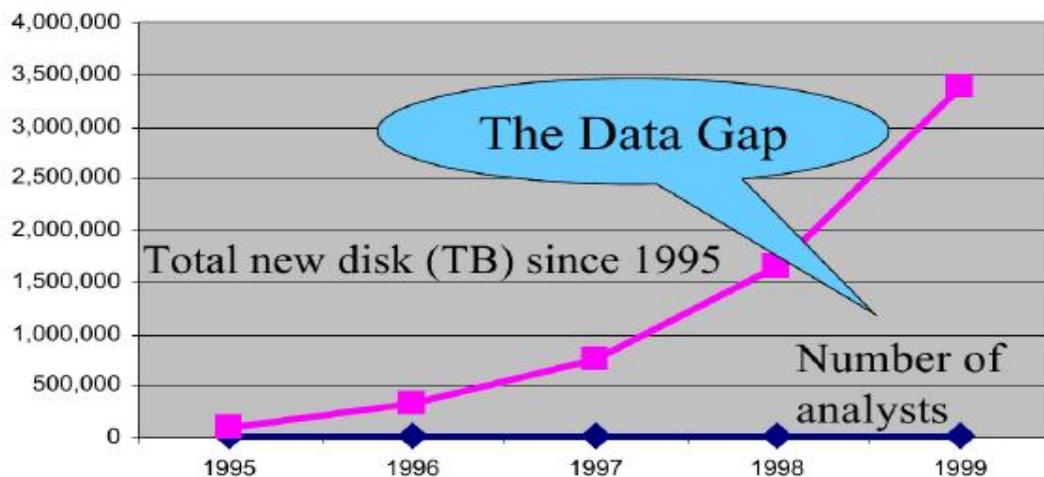
- Banyak data yang telah direkam dan disimpan:
 - Transaksi penjualan supermarket
 - Transaksi perbankan dan kartu kredit
 - Log kunjungan Web (access_log)
 - Akuisisi data dalam penelitian-penelitian seperti astronomi, kesehatan, dll
- Sistem komputer lebih murah dan cepat (Moore's Law)



- Kebutuhan untuk berkompetisi dengan strategi yang tepat menjadi lebih tinggi

Mengapa Harus Data Mining?

- Data yang sedemikian besar kadang memiliki informasi yang tersembunyi
- Kemampuan manusia terbatas untuk “mempelototi” data-data tersebut dalam analisis



Definisi Data dan Data Mining

Data

Rekaman atau catatan terhadap fakta / transaksi / obyek

Definisi

- Ekstraksi informasi yang implisit, tidak diketahui sebelumnya, dan berpotensi berguna
- Eksplorasi dan analisis, secara otomatis atau tidak, data yang sangat besar untuk menemukan pola-pola yang berguna dan mempunyai arti

Data Mining

- Data Mining adalah kegiatan untuk menemukan informasi atau pengetahuan yang berguna secara otomatis dari data yang jumlahnya besar.
- Data Mining merupakan salah satu proses dari keseluruhan proses yang ada pada Knowledge Discovery in Databases (KDD).

Data Mining dan KDD

- *Knowledge Discovery in Databases (KDD)* merupakan sekumpulan proses untuk menemukan pengetahuan yang bermanfaat dari data.
- Kumpulan proses dalam KDD meliputi : pembersihan data (*data cleaning*), integrasi data (*data integration*), pemilihan data (*data selection*), transformasi data (*data transformation*), penambangan data (*data mining*), evaluasi pola (*pattern evaluation*), dan presentasi pengetahuan (*knowledge presentation*).
- Berdasarkan definisi ini terlihat bahwa data mining hanya merupakan salah satu proses dari keseluruhan proses yang ada pada KDD, tetapi merupakan proses yang sangat penting dalam menemukan pola-pola yang berguna dari sejumlah data yang besar (data tersebut bisa disimpan dalam basisdata, Data Warehouse, atau media penyimpanan informasi lainnya).

DATA MINING & KDD



DATA WAREHOUSE

KDD vs Data Mining

- Menurut Cabena, Data Mining = Knowledge Discovery in Database (KDD).
- Menurut Jiawei Han, Data Mining merupakan Subset atau salah satu tahap dari KDD saja. Sehingga, batasan ini yang selanjutnya digunakan.
- Data Mining bertujuan mengeksplorasi basis data untuk menemukan pola-pola pengetahuan yang tersembunyi di dalam data tersebut.

Pengertian yang Salah

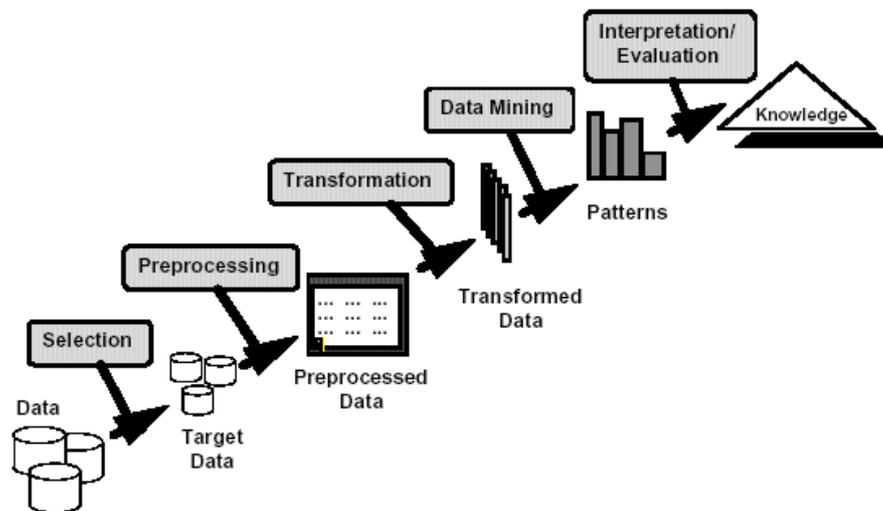
Bukan Data Mining

- Mencari nomor telepon “Bambang Gunawan” di buku telepon Indonesia
- Mencari informasi mengenai “Bunga” di google.com

Data Mining

- Menemukan bahwa banyak orang bernama Bambang di daerah Jawa Timur
- Mengelompokkan dokumen web mengenai “Bunga” sesuai dengan konteks
 - Bunga Bank atau Kredit (Keuangan)
 - Bunga - Tanaman / Pertanian
 - BCL (Artis)

Tahapan Ekstraksi Pengetahuan



Manfaat Data Mining

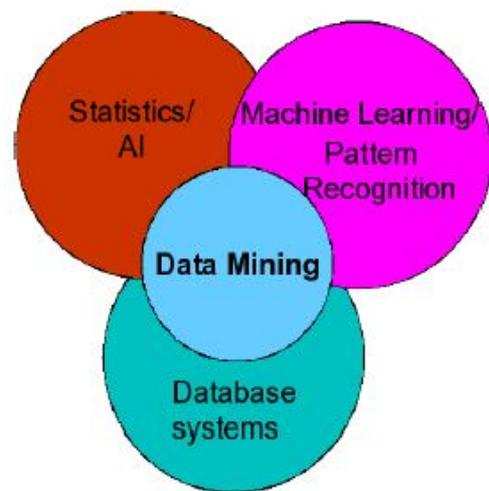
- Pemanfaatan data mining dilihat dari dua sudut pandang, yaitu sudut pandang komersial dan sudut pandang keilmuan.
- Dari sudut pandang komersial, menghasilkan informasi-informasi yang dibutuhkan yang merupakan asset yang dapat meningkatkan daya saing suatu institusi:
 - Bagaimana mengetahui hilangnya pelanggan karena pesaing
 - Bagaimana mengetahui item produk atau konsumen yang memiliki kesamaan karakteristik
 - Bagaimana mengidentifikasi produk-produk yang terjual bersamaan dengan produk lain
 - Bagaimana memprediksi tingkat penjualan
 - Bagaimana menilai tingkat resiko dalam menentukan jumlah produksi suatu item
 - Bagaimana memprediksi perilaku bisnis dimasa yang akan datang

Manfaat Data Mining

- Dari sudut pandang keilmuan, data mining dapat digunakan untuk *capture*, menganalisis serta menyimpan data yang bersifat real time dan sangat besar:
 - Remote sensor yang ditempatkan pada suatu satelit
 - Telescope yang digunakan untuk memindai langit

Ilmu Data Mining

- Gabungan dari beberapa bidang ilmu dalam Matematik dan Ilmu Komputer
- Diperlukan karena:
 - Data yang s(u)angat b(u)esar
 - Dimensi data yang besar
 - Data Heterogen

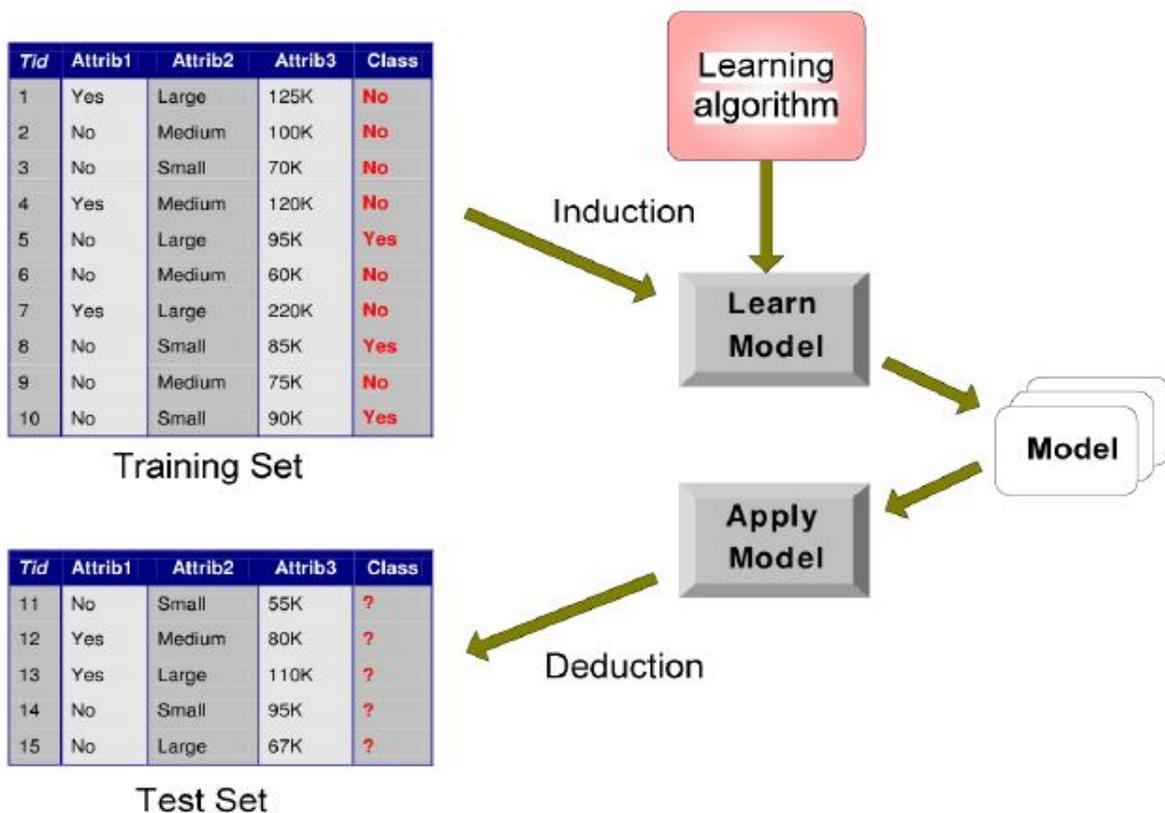


Fungsi-Fungsi dalam Data Mining (Haskett, 2000)

- *Association*, adalah proses untuk menemukan aturan asosiatif antara suatu kombinasi item dalam suatu waktu
- *Clustering*, adalah proses pengelompokan sejumlah data/obyek ke dalam kelompok data sehingga setiap kelompok berisi data yang mirip.
- *Classification*, proses penemuan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui.
- *Forecasting*, adalah proses pengestimasian nilai prediksi berdasarkan pola-pola di dalam sekumpulan data.

Classification Data Mining

- Ekstraksi pola pengelompokan atau pengklasifikasian sebuah himpunan obyek / data (training-set) ke dalam kelas (class) tertentu berdasarkan atribut-atributnya
- Pola pengelompokan yang ditemukan akan menjadi model pengelompokan
- Model digunakan untuk memprediksi kelompok data / obyek baru (test-set)



- Pemakaian Kartu Kredit secara Ilegal

- Tujuan : mendeteksi adanya penggunaan kartu kredit secara ilegal
- Pendekatan :
 - Data transaksi sebelumnya (*lokasi & waktu transaksi, jenis barang yang dibeli, besar transaksi*)
 - Label data-data tersebut dengan **Curang** atau **Aman**
 - DM mencari model klasifikasi **Curang** atau **Aman** berdasarkan atribut transaksi
 - Menerapkan model tersebut jika ada transaksi baru untuk mempercepat / tepat tindakan preventif

..
...

- Deteksi SPAM
 - Tujuan : mendeteksi email yang tidak diharapkan secara dini
- Direct Marketing
 - Tujuan : mencari pengelompokan profil pelanggan agar target marketing sesuai
- Sky Survey Cataloging
 - Tujuan : mengelompokkan obyek langit hasil pemotretan teleskop ke dalam class-nya

Teknik / Metode Classification

- Decision Tree (Pohon Keputusan)
- Rule-based Methods
- Memory based reasoning
- Neural Networks (Jaringan Syaraf Tiruan)
- Naive Bayes dan Bayesian Belief Networks
- Support Vector Machines

- Menemukan pola asosiasi dalam data
 - Adanya kemunculan obyek tertentu berdasarkan kemunculan obyek yang lain

<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}

- Marketing & Sales Promotion

- Misalnya pola yang ditemukan :
 $\{\text{Susu Anak, ...}\} \rightarrow \{\text{Kwaci}\}$
- Kwaci sebagai konsekuen : bagaimana caranya menaikkan penjualan kwaci
- Susu Anak sebagai anteseden : jika tidak lagi menjual susu anak, memprediksi produk lain yang ikut jatuh penjualannya
- Dua-duanya : membuat paket promo Susu Anak, Kwaci, dll

...

- Pengelolaan Rak di Supermarket
 - Tujuan: memudahkan pelanggan berbelanja barang-barang yang sering dibeli bersama
 - Misalnya: ada rak kecil berisi kwaci diletakkan pada bagian susu anak
- Sistem Rekomendasi Pintar
 - Tujuan: memberikan rekomendasi kepada pelanggan toko buku on-line tentang buku-buku lain yang sering dibeli juga oleh pelanggan lainnya jika membeli buku tertentu

Teknik / Metode Association

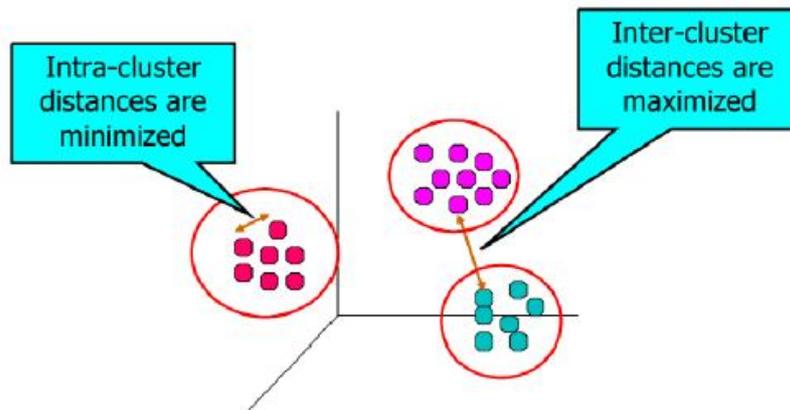
...

- Algoritma Apriori
- Algoritma Hash Based
- Algoritma *Generalized Rule Induction*
- Fuzzy C-Covering

Clustering

Definition

Proses mencari gugus-gugus dari sekumpulan obyek sehingga obyek-obyek di dalam sebuah gugus mirip satu dengan lainnya, dan berbeda dengan obyek di luar gugusnya.



- Segmentasi Pasar:
 - Tujuan: mencari gugus segmentasi pasar berdasarkan data transaksi untuk keperluan marketing
 - Pendekatan:
 - mempersiapkan data beserta atribut data pelanggan berdasarkan geografi dan data pribadi lainnya
 - mencari gugus pelanggan yang mirip berdasarkan atribut2 tsb
 - melakukan observasi perilaku pasar berdasarkan gugus-gugus pelanggan yang ditemukan
 - Hasil: strategi marketing yang tepat sasaran

Teknik / Metode Clustering

- Algoritma K-Means
- Algoritma Fuzzy C Means

TAHAPAN KNOWLEDGE DISCOVERY IN DATABASE

Menurut Peter Cabena, Tahapan – tahapan dalam KDD adalah sebagai berikut :

- Penentuan Sasaran Bisnis (**Business Objective Determination**)
- Persiapan Data (**Preparation Data**)
 - Data Selection
 - Data Preprocessing
 - Data Transformation
- **Data Mining**
- **Anaysis of Result**
- **Assimilation of Knowledge**

Business Objective Determination - 1

- Merupakan sebuah tahapan yang mendefinisikan permasalahan atau tantangan bisnis dengan jelas. Hal ini merupakan aspek yang sangat esensial dalam setiap proyek data mining.
- **Contoh Sasaran Bisnis :**
Mengembangkan suatu strategi marketing untuk mempertahankan loyalitas customer di Jawa Tengah dan Jawa Timur untuk produk soft drink dengan brand dan ukuran tertentu (200ml dalam kotak alumunium) selama bulan April, Mei , Juni yang akan datang. Perusahaan akan menggunakan kombinasi dari berbagai strategi marketing (*mixed marketing*), yang salah satunya adalah *direct mail campaign* kepada customer yang tampaknya "mudah rusak" loyalitasnya.

Business Objective Determination - 2

- Sehingga dampak keseluruhan KDD adalah :
- – **Data Selection:** dipilih customers yang membeli produk soft drink 200 ml dalam kotak aluminium di Jawa Tengah dan Jawa Timur.
- – **Data Transformation:** customers yang membeli produk soft drink 200 ml dalam kotak aluminium disorting dalam 10 kategori, yang masing-masing membedakan tingkat loyalitasnya: membeli produk tersebut 0-10%, 11-20%, , 81-90%, 91-100% sepanjang waktu pembeliannya. **Selanjutnya data inilah yang akan dibawa ke tahap data mining.**

Persiapan Data (Preparation Data) -1

- Merupakan tahapan untuk mempersiapkan data yang diperlukan untuk proses data mining. Tujuannya adalah agar data yang digunakan benar-benar sesuai dengan permasalahan yang akan dipecahkan, dapat dijamin kebenarannya, dan dalam format yang sesuai/tepat.
- Tahap yang paling banyak mengkonsumsi resources (manusia, biaya, waktu) yang tersedia. Biasanya mencapai 60% keseluruhan proyek KDD.

Persiapan Data (Preparation Data) -2

- **Data Selection**
 - Mengidentifikasi semua sumber informasi internal dan eksternal dan memilih sebagian saja dari data yang diperlukan untuk aplikasi data mining.
- **Data Preprocessing**
 - Meyakinkan kualitas data yang telah dipilih pada tahapan sebelumnya. 2 issue yang sering dihadapkan pada tahapan ini adalah Noisy Data dan Missing Value.
- **Data Transformation**
 - Mengubah data ke dalam model analitis serta memodelkan data agar sesuai dengan analisa yang diharapkan dan format data yang diperlukan oleh algoritma data mining.

Data Mining - 1

- Melakukan proses pencarian pengetahuan terhadap data yang ditransformasikan pada tahap sebelumnya.
- Contoh Pengetahuan berbentuk *Association Rule* untuk kasus "Soft Drink":
 - IF soft drink sejenis dengan ukuran yang lebih besar (bukan botol kecil) dibeli lebih dari 58% dalam sejarah pembelian soft drink seorang consumer
 - THEN consumer tersebut diprediksi Loyal.
- Pemilihan tugas data mining : pemilihan goal dari proses KDD misalnya klasifikasi, regresi, clustering, dll.

Data Mining - 2

- *Proses Data mining* yaitu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

Analysis of Result

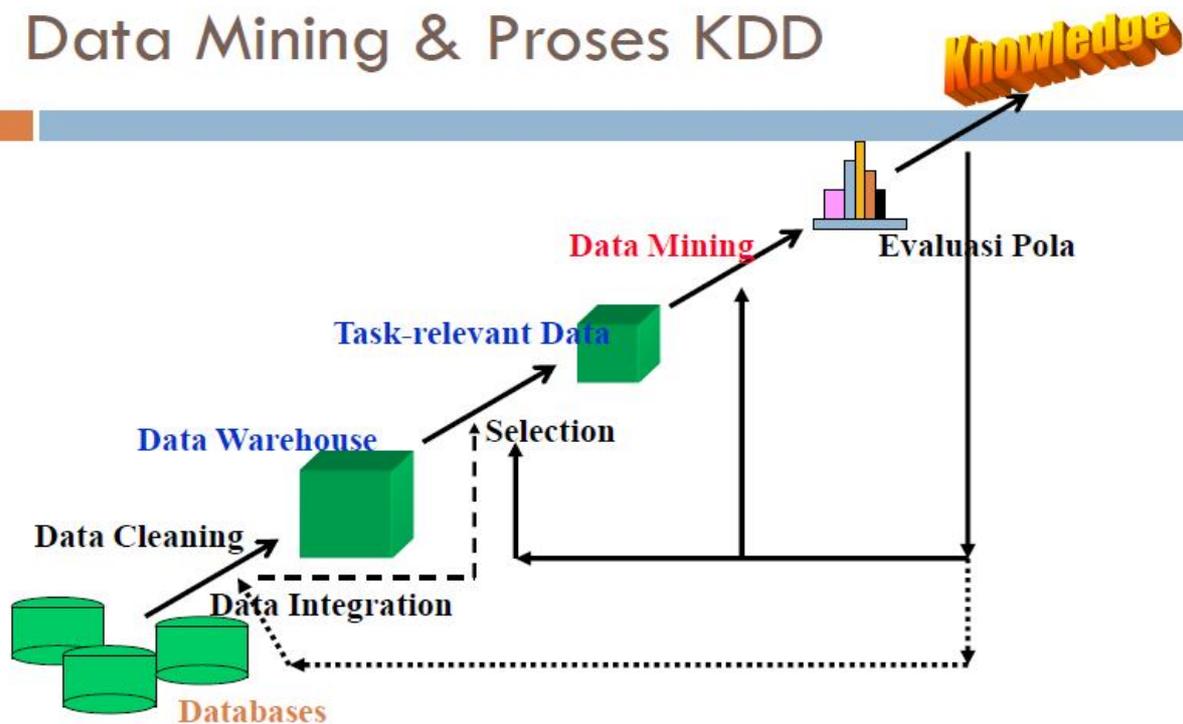
- Menginterpretasikan dan mengevaluasi output dari tahap mining: **patterns**.
- Pendekatan analisa yang digunakan akan bervariasi menurut operasi data mining yang digunakan, tetapi biasanya akan melibatkan teknik visualisasi.

Assimilation of Knowledge

- Menggunakan hasil mining yang telah dievaluasi ke dalam perilaku organisasi dan sistem informasi perusahaan.

DATA MINING dan PROSES KDD – 1

Data Mining & Proses KDD



Source : Han 2004

DATA MINING dan PROSES KDD - 2

- **Pembersihan data (Data Cleaning)**
 - Digunakan untuk membuang data yang tidak konsisten dan noise.
- **Integrasi Data (Data Integration)**
 - Data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Hasil integrasi data sering diwujudkan dalam sebuah data warehouse.
- **Transformasi data**
 - Transformasi dan pemilihan data ini untuk menentukan kualitas dari hasil data mining, sehingga data diubah menjadi bentuk sesuai untuk di-Mining.
- **Aplikasi Teknik Data Mining**

DATA MINING dan PROSES KDD - 2

- Aplikasi teknik data mining sendiri hanya merupakan salah satu bagian dari proses data mining. Ada beberapa teknik data mining yang sudah umum dipakai.

DATA MINING dan PROSES KDD - 2

- **Evaluasi pola yang ditemukan**

- Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai.

- **Presentasi Pengetahuan**

- Presentasi pola yang ditemukan untuk menghasilkan aksi tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat.