

DATA MINING

Outline

- Pengenalan Data Mining
- Data Mart
- Teknik-Teknik Data Mining
- Peralatan Data Mining
- Data Mining dan Data Warehouse

Pengenalan Data Mining

- Perangkat lunak yang digunakan untuk menemukan pola-pola tersembunyi maupun hubungan-hubungan yang terdapat dalam basis data yang besar dan menghasilkan aturan-aturan yang digunakan untuk memperkirakan perilaku di masa mendatang
- *Data mining* sering dikatakan berurusan dengan “penemuan pengetahuan” dalam basis data. Suatu aturan yang dihasilkan oleh *data mining* misalnya seperti berikut :
“Kebanyakan pembeli mobil Forsa adalah wanita berusia di atas 30 th”

Data Mart

- Bagian dari *data warehouse* yang mendukung kebutuhan pada tingkat departemen atau fungsi bisnis tertentu dalam perusahaan.
- Karakteristik yang membedakan data mart dan *data warehouse* adalah sbb :
(Connolly, Begg, Strachan 1999).

1. Data mart memfokuskan hanya pada kebutuhan-kebutuhan pemakai yang terkait dalam sebuah departemen atau fungsi bisnis.
2. Data mart biasanya tidak mengandung data operasional yang rinci seperti pada *data warehouse*.
3. Data mart hanya mengandung sedikit informasi dibandingkan dengan *data warehouse*. Data mart lebih mudah dipahami dan dinavigasi.

Teknik-Teknik Data Mining

Operasi-operasi dan teknik-teknik yang berhubungan :

1. Operasi Predictive modeling :
classification, value prediction
2. Database segmentation :
demographic clustering,
neural clustering
3. Link Analysis :
association discovery, sequential
pattern discovery, similar time-
sequence discovery
4. Deviation detection :
statistics, visualization

Peralatan Data Mining

- Karakteristik yang terpenting dari peralatan data mining :
 - Fasilitas persiapan data
 - Skalabilitas produk dan kinerja
 - Fasilitas untuk visualisasi hasil

Data Mining dan Data Warehouse

- Kualitas dan konsistensi data merupakan persyaratan untuk data mining untuk menjamin keakuratan model-model prediksi. Data warehouse dipopulasikan dengan data yang bersih dan konsisten.
- Data mining akan sangat berguna jika mendapatkan data dari banyak sumber untuk mendapatkan sebanyak mungkin keterhubungan antar data. Data warehouse hanya berisikan data dari beberapa sumber

- Pemilihan subset-subset record dan field untuk data mining membutuhkan kemampuan query dari data warehouse
- Hasil pembelajaran data mining berguna jika ada cara untuk menyelidiki lebih jauh lagi mengenai pola-pola yang ditemukan. Data warehouse menyediakan kemampuan untuk melihat sumber data masa lalu.

Referensi

- Connolly, Thomas; Begg, Carolyn; Strachan, Anne; **Database Systems : A Practical Approach to Design, Implementation and Management**, 3rd edition, Addison Wesley, 2003.
- <http://mti.ugm.ac.id/~adji/courses/resources/lectures/InformSystem/AKA/si-6-2.ppt>

Kesimpulan

- Data mining merupakan perangkat lunak yang digunakan untuk menemukan pola tersembunyi maupun hubungan yang terdapat dalam basis data yang besar dan menghasilkan aturan yang digunakan untuk memperkirakan perilaku di masa mendatang
- Teknik-teknik data mining :
classification, value prediction, demographic clustering, neural clustering, association discovery, sequential pattern discovery, similar time-sequence discovery, statistics, visualization

- Karakteristik yang terpenting dari peralatan data mining :
fasilitas persiapan data,
skalabilitas produk dan kinerja,
fasilitas untuk visualisasi hasil
- Data mart merupakan bagian dari *data warehouse* yang mendukung kebutuhan pada tingkat departemen atau fungsi bisnis tertentu dalam perusahaan