

DATA MINING

3 SKS | Semester 6 | S1 Sistem Informasi

Pertemuan 2



Nizar Rabbi Radliya
nizar.radliya@yahoo.com



Definisi Data Mining

Tan (2006) mendefinisikan data mining sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Data mining juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan.

Darly Pregibon (2011) menyatakan bahwa data mining adalah campuran dari statistik, kecerdasan buatan, dan riset basis data.

Pramudiono (2006) mengartikan data mining sebagai serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual.

Definisi Data Mining

Proses mendapatkan pola atau informasi (pengetahuan) ...

... dengan beberapa teknik ...

... dari kumpulan data besar.

Contoh Data Mining atau bukan?

Bukan: Pencarian informasi tertentu di internet.

Termasuk: Pengelompokan informasi yang mirip dalam konteks tertentu pada hasil pencarian.

.....
Bukan: Petugas medis mencari data medis untuk menganalisis catatan pasien dengan penyakit tertentu.

Termasuk: Peneliti medis mencari cara pengelompokan data penyakit pasien berdasarkan data diagnosis, umur, alamat.

.....
Bukan: Pembuatan laporan tahunan penjualan perusahaan dengan merekap semua data selama setahun.

Termasuk: Pemanfaatan data penjualan perusahaan untuk mendapatkan pola prediksi stok yang sebaiknya disediakan pada tahun berikutnya.

Proses Data Mining

Tiga langkah utama dalam proses data mining (Gonunescu, 2011)

1. Ekplorasi/pemrosesan awal data
2. Membangun model dan melakukan validasi terhadapnya
3. Penerapan

Proses Data Mining

Kasifikasi (classification) digunakan untuk pembuatan model yang dapat melakukan pemetaan dari setiap himpunan variabel ke setiap targetnya, kemudian menggunakan model tersebut untuk memberikan nilai target pada himpunan variabel baru yang didapat.

.....

Contoh:

1. Menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau bukan.
2. Melakukan deteksi penyakit pasien berdasarkan sejumlah nilai parameter penyakit yang diderita.
3. Memprediksi pelanggan mana yang akan berpindah ke kompetitor kita atau tetap setia.

Proses Data Mining

Penklusteran (clustering) digunakan untuk melakukan pengelompokan data-data ke dalam sejumlah kelompok (cluster) berdasarkan karakteristik masing-masing data pada kelompok-kelompok yang ada.

.....

Contoh:

1. Mengetahui pola pembelian barang oleh para konsumen pada waktu-waktu tertentu.
2. Mendapatkan kelompok-kelompok konsumen untuk target pemasaran.

Proses Data Mining

Asosiasi (association) digunakan untuk menemukan pola yang mendeteksi kumpulan atribut-atribut yang muncul bersamaan dalam frekuensi yang sering, dan membentuk sejumlah kaidah dari kumpulan-kumpulan tersebut.

.....

Contoh:

1. Untuk mengenali item-item yang dibeli bersama-sama oleh cukup banyak pelanggan.
2. Untuk mengetahui suku cadang yang membutuhkan perbaikan setelah suku cadang yang lainnya mengalami kerusakan.

Proses Data Mining

Anomali (anomaly) berkaitan dengan pengamatan sebuah data dari sejumlah data yang secara signifikan mempunyai karakteristik yang berbeda dari sisa data yang lain.

.....

Contoh:

1. Untuk mengetahui pola data yang memasuki jaringan sehingga penyusupan bisa ditemukan jika pola kerja data yang datang berbeda.
2. Diteksi perilaku kondisi cuaca yang mengalami anomali.

Materi Minggu Ke 3

Set Data

1. Definisi dan tipe data
2. Kualitas data
3. Similaritas dan dissimilaritas



PREPARE YOURSELF

