

Data Mining

Pertemuan 1 Pengenalalan

Pengertian

- Data Mining adalah proses menemukan **pola** dalam kumpulan **data besar** yang melibatkan metode di persimpangan antara **machine learning, statistik, dan database system**.

Persilangan

- Data Mining adalah subbidang ilmu komputer dan statistik interdisipliner dengan tujuan keseluruhan untuk mengekstrak informasi (dengan metode cerdas) dari kumpulan data dan mengubah informasi menjadi struktur yang dapat dipahami untuk digunakan lebih lanjut.

Tugas Data Mining

- Tugas data mining adalah analisis semi-otomatis atau otomatis dari sejumlah besar data untuk mengekstrak pola yang sebelumnya tidak diketahui dan menarik seperti pengelompokan data (Clustering), data yang tidak biasa (deteksi anomali), dan ketergantungan (Association Rule Mining, Sequential Mining)

Indeks Spasial

- Melibatkan penggunaan teknik database seperti **indeks spasial**.
- Pola ini kemudian dapat dilihat sebagai semacam ringkasan data input, dan dapat digunakan dalam analisis lebih lanjut atau, misalnya, dalam pembelajaran mesin dan analitik prediktif.

Contoh Sederhana 1

Suhu	Kelembaban	Cuaca
26	70	mendung
28	72	mendung
32	50	cerah
33	49	cerah
24	80	hujan
22	90	hujan

Apa yang bisa disimpulkan?

- Bagaimana mengklasifikasi kan kondisi mendung, cerah dan hujan?
- Adakah perbedaan numerik pada suhu dan kelembaban?

Yang bisa disimpulkan dari Data

- Mendung :
 - $26 \leq \text{Suhu} \leq 28$
 - $70 \leq \text{Kelembaban} \leq 72$
- Cerah
 - $32 \leq \text{Suhu} \leq 33$
 - $49 \leq \text{Kelembaban} \leq 50$
- Hujan
 - $22 \leq \text{Suhu} \leq 24$
 - $80 \leq \text{Kelembaban} \leq 90$

Lalu...

- Suhu 25 dan kelembaban 69 bagaimana?
- Kalau ada testimoni orang mengatakan pernah kepanasan ditengah hujan bagaimana?
- Kalau ada testimoni orang mengatakan pernah kedinginan waktu cerah bagaimana?

Close world problem

- Dunia tertutup di lingkup data yang ada. Variabel yang sama bisa berbeda di kondisi yang berbeda.
- Kita bisa menjawab sebatas data :
 - Terbatas pada tempat dan waktu

Kok jadi kayak filosofi?

Contoh Sederhana 2

Suhu	Kelembaban	Cuaca
26	70	mendung
33	52	mendung
32	69	cerah
33	49	cerah
33	72	hujan
32	71	hujan

Apa yang bisa disimpulkan

- Bagaimana menentukan cuaca kalau begini???

Bukan salah kalian kok...

- Kalau datanya ga jelas. Algoritma dan komputer terancangipun tidak akan bisa melakukan data mining!!!

Solusinya?

- Cari / ambil data yang lebih banyak
- Jika hasilnya sama, mungkin variabel tersebut tidak berpengaruh terhadap kondisi.
- Tambahkan variabel lainnya, hingga menemukan **korelasi**.

Diperlukan Pre Processing

- Feature Selection
- Feature Extraction
- Statistik
 - Correlation Matrix
 - Standard Deviation
 - Median, Sub Median
 - Etc

Data Interpretation

- Mengacu pada implementasi proses di mana data ditinjau untuk tujuan sampai pada kesimpulan yang diinformasikan.
- Interpretasi data memberi makna pada informasi yang dianalisis dan menentukan signifikansi dan implikasinya.

Data Mining yang akan dibahas

- Classification
- Regression
- Clustering
- Association Rule Mining
- Sequential Mining

Algoritma yang akan dibahas

- Decision Tree
- Genetic Programming
- Fuzzy Association Rule Mining
- Apriori
- etc

Bahasa Pemrograman

- Python (Anaconda 3) **Wajib**
 - Pandas
 - Sci-kit Learn
- Jupyter notebook **Wajib**
- Pyspark **Tidak Wajib**

Tugas dan Ujian

- Tidak ada ujian
- Hanya tugas
 - Review Jurnal 30%
 - Tugas Programming 40%
 - Tugas Laporan Analisis 30%