

Pertemuan 3

KONSEP DATA MINNING

Data Science
Pingky Dezar Zulkarnain, SE., M.Sc., Ph.D
Putri Cholidhazia, ST., Mkom

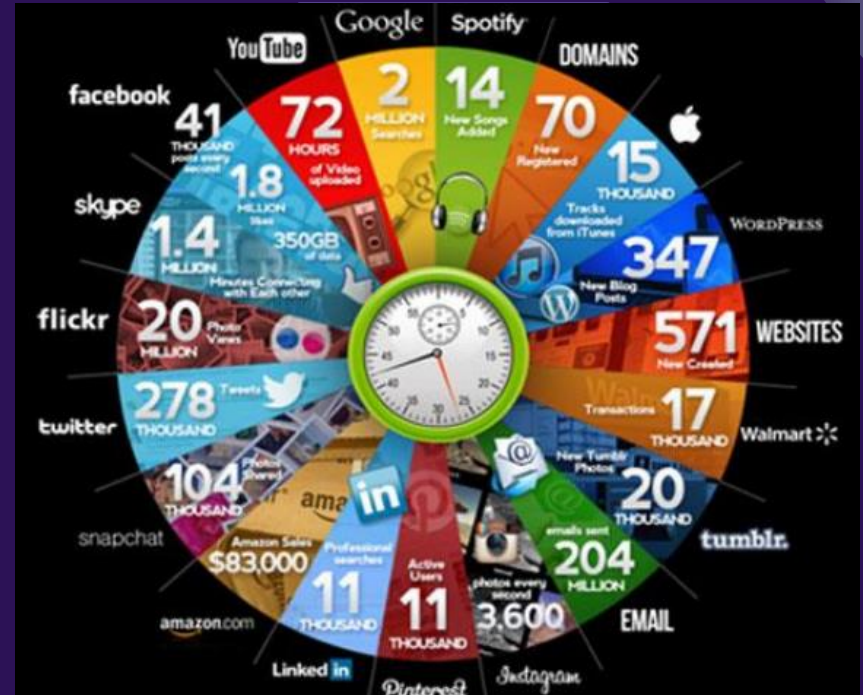
Latar Belakang

- **Banyak data telah terekam dan disimpan (Ledakan Data):**
 - **Transaksi penjualan supermarket**
 - **Transaksi perbankan dan kartu kredit**
 - **Log kunjungan Web**
- **Data terkumpul dan tersimpan pada kecepatan yang luar biasa (GB/hour)**
- **Kebutuhan untuk berkompetisi dengan strategi yang tepat menjadi lebih tinggi**

Data Media Sosial

“ Manusia kebanjiran data tetapi miskin pengetahuan.

- untuk itu diperlukan bagaimana manusia bisa mengubah data menjadi pengetahuan”



Data Mining



Data:

Fakta-fakta yang menggambarkan suatu kejadian yang sebenarnya pada waktu tertentu

Minning:

Proses Penambangan



DEFINISI

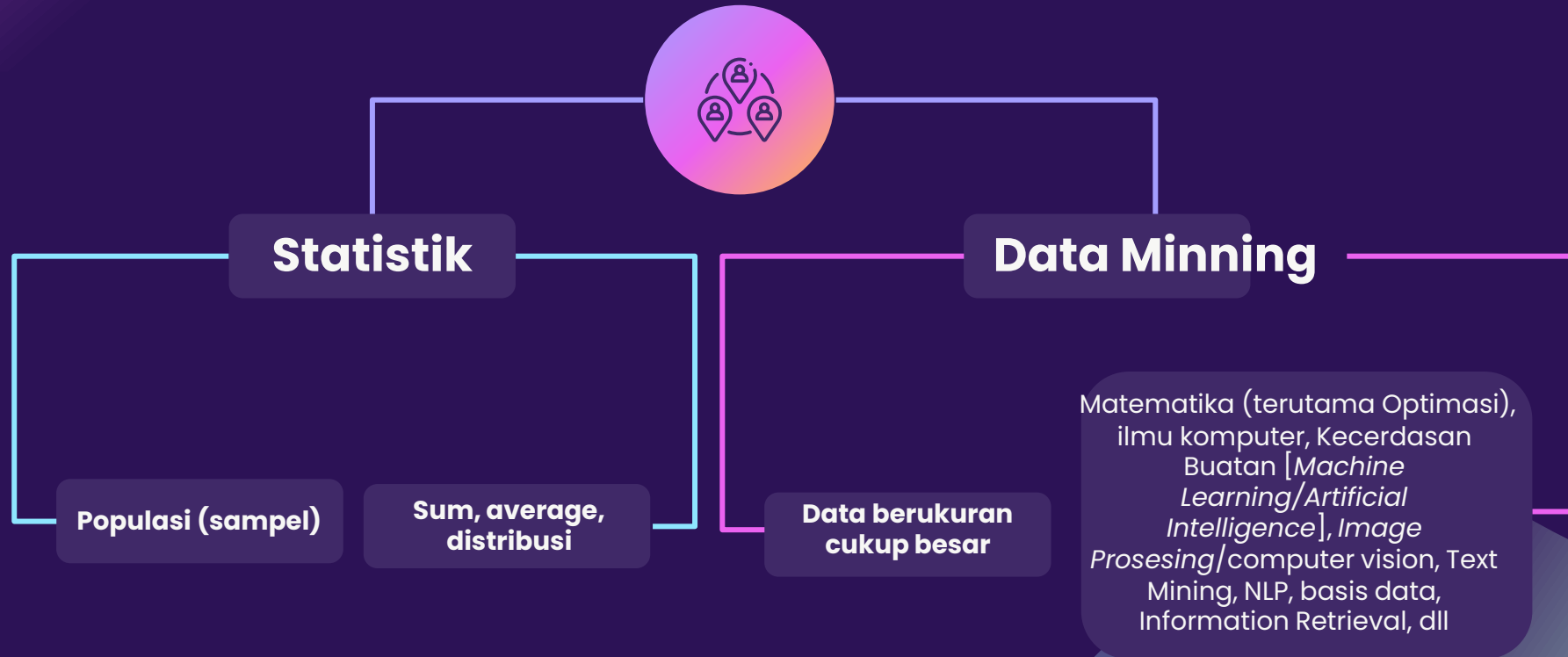
"Data mining merupakan teknik analisa data dengan tujuan untuk **menemukan pola** (kecenderungan yang penting yang biasanya tidak disadari keberadaanya) guna membantu dalam menetapkan keputusan."

« **Pola, rumus, aturan atau model yang muncul dari data** »

Menurut (Witten et al., 2011)

“Melakukan ekstraksi untuk mendapatkan informasi penting yang sifatnya implisit dan sebelumnya tidak diketahui, dari suatu data”

Perbedaan Statistik



Data Science vs Data Mining

S.No.	Ilmu Data	Penambangan Data
1	Ilmu data adalah sebuah area.	Data Mining adalah sebuah teknik.
2	Ini tentang pengumpulan, pemrosesan, analisis, dan pemanfaatan data ke dalam berbagai operasi. Ini lebih konseptual.	Ini tentang mengekstraksi informasi penting dan berharga dari data.
3	Ini adalah bidang studi seperti Ilmu Komputer, Statistik Terapan atau Matematika Terapan.	Ini adalah teknik yang merupakan bagian dari Knowledge Discovery dalam proses Data Base (KDD).
4	Tujuannya adalah untuk membangun produk yang dominan data untuk sebuah usaha.	Tujuannya adalah untuk membuat data lebih penting dan dapat digunakan yaitu dengan mengekstraksi hanya informasi penting.
5	Ini berkaitan dengan semua jenis data yaitu terstruktur, tidak terstruktur atau semi-terstruktur.	Ini terutama berkaitan dengan bentuk terstruktur dari data.
6	Ini adalah satu set super Data Mining karena ilmu data terdiri dari penghapusan Data, pembersihan, visualisasi, statistik, dan banyak lagi teknik.	Ini adalah sub set Ilmu Data sebagai kegiatan penambangan yang berada dalam pipa ilmu Data.
7	Ini terutama digunakan untuk tujuan ilmiah.	Ini terutama digunakan untuk tujuan bisnis.

APA PENTINGNYA DATA MINING TERHADAP DATA SCIENTIST

Data mining pada dasarnya merupakan bagian dari Data Science. Data mining ini merupakan proses pengumpulan informasi untuk melihat pola dari kumpulan data yang sangat besar dan data discovery. Data mining memiliki fokus untuk mengekstrak pola dari data dengan menggunakan metode statistika agar dapat dianalisa.

Data mining akan membantu data scientist dalam menyusun data mentah, merumuskannya serta mengenali berbagai pola melalui algoritma matematika dan komunikasi untuk membuka berbagai wawasan yang bermanfaat

Manfaat Data Mining

1. Memudahkan pengambilan keputusan
2. Membuat prediksi akurat untuk perencanaan.
Pengurangan biaya
3. Mendapat wawasan tentang pelanggan



Contoh

Data Kehadiran Karyawan

NIP	TGL	DATANG	PULANG
1103	02/12/2004	07:20	15:40
1142	02/12/2004	07:45	15:33
1156	02/12/2004	07:51	16:00
1173	02/12/2004	08:00	15:15
1180	02/12/2004	07:01	16:31
1183	02/12/2004	07:49	17:00



Informasi Akumulasi Bulanan Kehadiran Karyawan

NIP	Masuk	Alpa	Cuti	Sakit	Telat
1103	22				
1142	18	2		2	
1156	10	1	11		
1173	12	5			5
1180	10			12	

Hasil

Pola Kebiasaan Kehadiran Mingguan Karyawan

	Senin	Selasa	Rabu	Kamis	Jumat
Terlambat	7	0	1	0	5
Pulang Cepat	0	1	1	1	8
Izin	3	0	0	1	4
Alpa	1	0	2	0	2



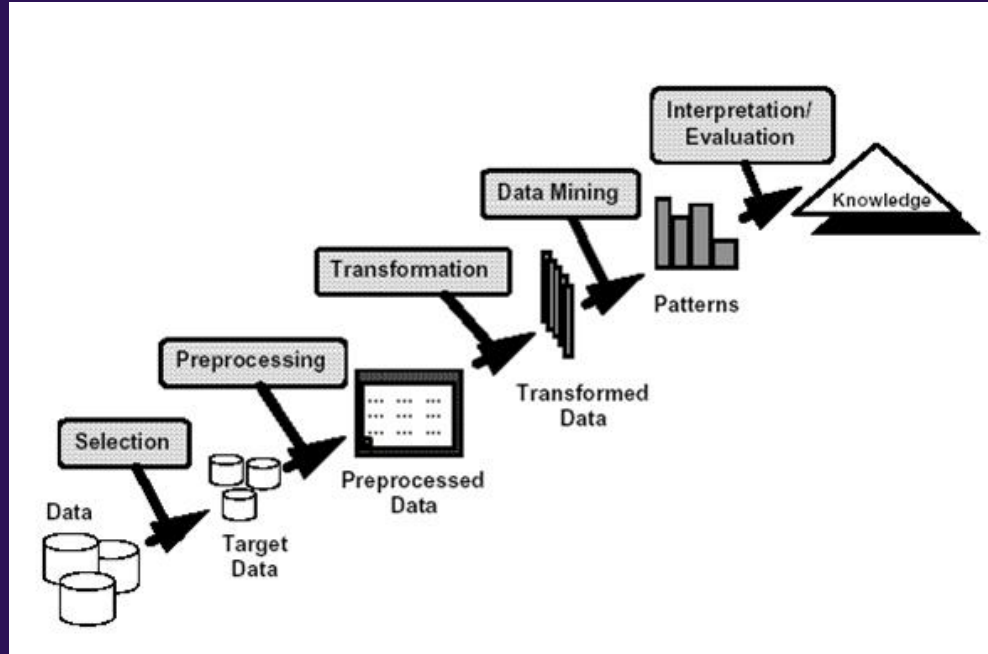
Informasi Akumulasi Bulanan Kehadiran Karyawan

- ❑ Kebijakan penataan jam kerja karyawan khusus untuk hari senin dan jumat
- ❑ Peraturan jam kerja:
 - Hari Senin dimulai jam 10:00
 - Hari Jumat diakhiri jam 14:00
 - Sisa jam kerja dikompensasi ke hari lain

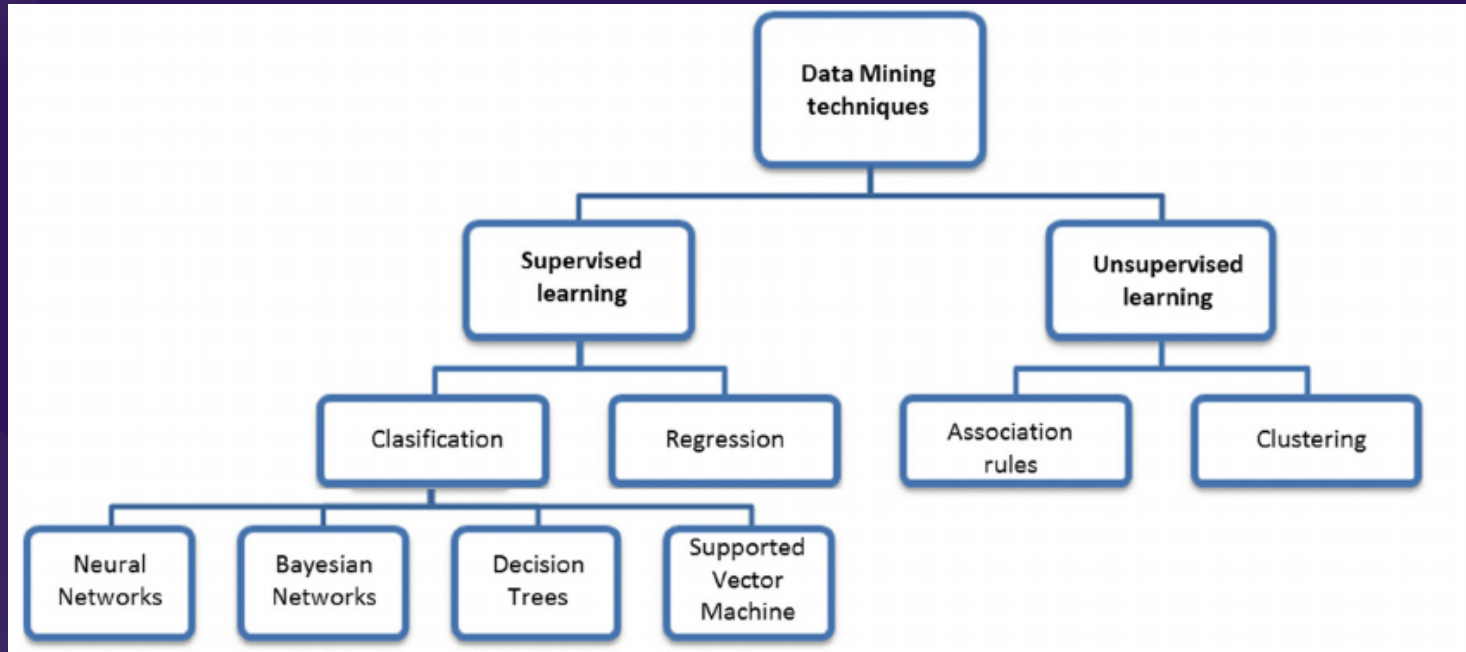
TUGAS DATA MINNING

- ❑ Prediktif. Tujuan dari tugas prediktif adalah untuk memprediksi nilai dari atribut tertentu berdasarkan pada nilai dari atribut-atribut lain.
- ❑ Deskriptif. Tujuan dari tugas deskriptif adalah untuk menurunkan pola-pola (korelasi, trend, cluster, dan anomali) yang meringkas hubungan dalam data

Knowledge Discovery In Database (KDD)



Teknik Data Mining



DATA MINNING DI INDUSTRI:

01 Komunikasi

Memahami banyaknya data pelanggan

02 Asuransi

Mengidentifikasi faktor risiko pada pengajuan klaim

03 Manufaktur

Memprediksi aset produksi

04 Retail

Meningkatkan hubungan pelanggan dan memperkirakan penjualan

05 Pendidikan

Memprediksi tingkat pencapaian mahasiswa

06 Manufaktur

Mendeteksi penipuan

Thanks!

Sampai Ketemu dipertemuan Selanjutnya