

Resume video youtube dengan judul “Konsep, Proses dan Contoh Nyata Data Mining dan Big Data: Ga Galau Lagi Jadi Data Scientist”

<https://www.youtube.com/watch?v=h82NuHDNhKI>

Manusia memproduksi data dalam jumlah yang sangat besar setiap harinya dalam bidang seperti Bisnis, kedokteran, ekonomi, finansial dan lain lain..

Aplikasi web dan sosial mengenerate banyak sekali data seperti facebook dengan 300 PB(peta byte) user data setiap harinya.

Kebanjiran data tapi miskin pengetahuan, data harus diolah menjadi pengetahuan agar bisa bermanfaat bagi manusia. Dengan pengetahuan kita dapat, melakukan estimasi dan prediksi, melakukan analisis tentang asosiasi, korelasi dan pengelompokan, serta membantu mengambil keputusan dan membuat kebijakan.

Data dapat diolah melalui urutan seperti dari data absen, kemudian diperoleh informasi, dari informasi tersebut kita memperoleh pengetahuan dengan pengetahuan itu kita bisa membuat kebijakan seperti yang dapat dilakukan melalui data kehadiran pegawai untuk membuat kebijakan jadwal melalui kebiasaan kehadiran yang kita peroleh setelah mengolah data absensi pegawai.

Data mining merupakan disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data yang besar

Ekstraksi data ke pengetahuan:

Data: fakta yang terekam dan tidak membawa arti

Informasi: rekap, rangkuman, penjelasan dan statistik dari data

Pengetahuan: pola, rumus, aturan atau model yang muncul dari data

Nama lain data mining:

Big data

Business intelligence

Pattern analysis

Contoh data mahasiswa pada kampus

NIM	Gender	Nilai UN	Asal Sekolah	IPS1	IPS2	IPS3	IPS 4	...	Lulus Tepat Waktu
10001	L	28	SMAN 2	3.3	3.6	2.89	2.9		Ya
10002	P	27	SMA DK	4.0	3.2	3.8	3.7		Tidak
10003	P	24	SMAN 1	2.7	3.4	4.0	3.5		Tidak
10004	L	26,4	SMAN 3	3.2	2.7	3.6	3.4		Ya
...									
...									
11000	L	23,4	SMAN 5	3.3	2.8	3.1	3.2		Ya

Dari data mahasiswa diatas kita bisa melakukan prediksi kelulusan mahasiswa dengan metode klasifikasi contoh algoritma seperti Naive Bayes

Contoh perubahan aplikasi setelah menerapkan data mining seperti
Sistem informasi akademik menjadi sistem prediksi kelulusan mahasiswa
Sistem pencatatan pemilu menjadi sistem prediksi hasil pemilu
Sistem laporan kekayaan pejabat menjadi sistem prediksi koruptor

Revolusi industri 4.0

1784 industri 1.0 mechanization, steam power, weaving loom
1870 industri 2.0 mass production, electrical energy, assembly line
1969 industri 3.0 automation, computer, electronics
Today industry 4.0 cyber physical systems, internet of things network

Perusahaan pengolah pengetahuan
Uber, the world largest taxi company
Tidak mempunyai kendaraan taxi
Google world's largest media/advertising company
Tidak menciptakan konten
Alibaba the most valuable retailer
Tidak mempunyai inventaris

Data mining tasks and roles

IT/ DBA

- Data source
- Data preprocessing, data integration, data warehouse

Data Scientist

- Data exploration
- Data mining

Business Analyst

- Data presentation

End User

- Decision Making

Data mining terhubung dalam beberapa bidang yaitu

Pattern recognition, machine learning, statistics, computing algorithms, database technology

Masalah pada data mining

1. Tremendous amount of data
2. High dimensionality data
3. High complexity data
4. New and sophisticated applications

Metode data mining:

1. Estimation (Estimasi): Linear Regression, Neural Network, Support Vector Machine, etc
2. Prediction/Forecasting (Prediksi/Peramalan): Linear Regression, Neural Network, Support Vector Machine, etc

3. Classification (Klasifikasi): Naive Bayes, K-Nearest Neighbor, C4.5, ID3, CART, Linear Discriminant Analysis, Logistic Regression, etc
4. Clustering (Klastering): K-Means, K-Medoids, Self-Organizing Map (SOM), Fuzzy C-Means, etc
5. Association (Asosiasi): FP-Growth, A Priori, Coefficient of Correlation, Chi Square, etc

Metode learning algoritma data mining

Supervised learning, unsupervised learning, semi- supervised learning

Association based learning

Sejarah sains

Sebelum 1600: empirical science

Disebut sains kalau bentuknya kasat mata

1600 - 1950: theoretical science

Disebut sains kalau bisa dibuktikan secara matematis atau eksperimen

1950-1990: computational science

Seluruh disiplin ilmu bergerak ke komputasi

Lahirnya banyak model komputasi

1990-sekarang: data science

Kultur manusia menghasilkan data besar

Kemampuan komputer mengolah data besar

Datangnya data mining sebagai arus utama sains

Business objective are the origin of every data mining solution

The maxim: 'data mining is a business process'

Contoh penerapan data mining:

Penentuan kelayakan kredit rumah

Penentuan pasokan listrik pln

Perkiraan harga saham dan tingkat inflasi