

* Data adalah fakta terencana tidak membawa makna. Data Science ialah new jargon akan tetapi termasuk ke dalam AI methods. Dahulu penyebutannya yaitu data terstruktur dan data tidak terstruktur. 5 hal yang menyangkut dalam materi data science yaitu:

1. Perubahan culture
2. Me-mining
3. Methods dan teknologinya
4. Contoh kasus (Use Cases)
5. Penelitian (Research)

1. Data Culture

Data sejatinya berada di lingkungan sekitar kita. Manusia memproduksi beragam data yang jumlah dan ukurannya sangat besar. Contohnya sbh:

- a. Astronomi
- Bisnis
- Kedokteran
- Ekonomi
- Olahraga
- Cuaca
- Financial
- dan lain sebagainya

* Pada tahun 2000, data sebanyak 140 tera byte dalam 10 tahun dianggap sudah besar. Akan tetapi pada tahun 2016, data sebanyak 140 TB dalam 5 hari sudah dianggap tidak besar. Hal ini disebabkan pengaruh sosial media. Sebagai contoh: 60 detik waktu kita diskusi, diagram awarane sudah 15M transkrip, 72 jam video sudah terupload, dalam selang waktu 60 detik.

Kata John Naisbitt, Megatrends 1980. Kita ini banjir data dan miskin pengetahuan

2. Data Mining

↳ mengubah data menjadi pengetahuan. Data harus doolah menjadi pengetahuan supaya bisa bermanfaat bagi manusia.

Dengan pengetahuan tsb, manusia dapat:

- Melakukan estimasi dan prediksi apa yang terjadi di masa depan
- Melakukan analisis tentang asosiasi, korelasi, dan pengelompokan antar data dan atribut
- Membantu pengambilan keputusan dan pembuatan kebijakan

Okh karena, data-data tersebut agar bermanfaat harus

diubah menjadi knowledge (pengetahuan) yang volumenya lebih besar. Dahulu, Mining data tujuannya untuk urusan bisnis, ekonomi, dll. Tetapi, data kita ubah jadi informasi dan diolah menjadi

pengetahuan, kemudian digunakan untuk menghemat uang serta dapat mengurangi biaya operasional dengan kinerja yang bagus.
Permasalahannya dalam tabel di bawah ini:

Data Kehadiran Pegawai

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

❌ Pegawai dengan NIP 1156, 1180, dan 1183 ialah pegawai yang rajin sedangkan NIP 1103, 1142, dan 1173 pegawai yang tidak rajin. Akan tetapi kita tidak boleh menjudge pegawai tersebut rajin atau tidaknya karena masih data satu hari. Jika kita melihat data dalam seminggu, sebulan, dan setahun, baru terbentuk sebuah informasi. Jika kita perhatikan data diatas alasan pegawai pulang sebelum jam 4 sore dikarenakan lokasi rumahnya di Depok, Bogor, Tangerang (Jauh dari kantor). Berikut ini informasi dan statistik Akumulasi Bulanan Kehadiran Pegawai.

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

Jadi info grafik yang dibuat itu sudah termasuk ke dalam informasi Statistik dari data itu levelnya informasi karena sudah bermakna.

* Mining

Data Mining ialah disiplin ilmu yang mempelajari metode untuk menemukan pola, pengetahuan, formula, aturan, ataupun insight dari suatu data yang volumenya besar dan jenisnya beraneka ragam. Konsep re-mining data dimulai dari kumpulan data, lalu data mining methods, dan kemudian knowledge (pengetahuan)

* Terminologi dan Nama lain dari Data Mining

Ekstraksi dari Data ke Pengetahuan

1. Data: fakta yang terakam tidak membawa arti
2. Informasi: Rekap, rangkuman, penjelasan, visualisasi, dan statistik dari data
3. Pengetahuan: pola, rumus, aturan, atau model yang muncul dari data

* Nama lain Data mining yaitu sbb:

- Data Science
- Knowledge Discovery in Database (KDD)
- Big Data
- Business Intelligence
- Knowledge Extraction
- Predictive Analytics
- Pattern Analysis
- Information Harvesting

Contoh data di kampus:

- Puluhan ribu data mahasiswa di kampus yang diambil dari sistem informasi akademik
- Seperti apa pengetahuan itu? Rumus, pola, aturan

Contoh data di KPU

- Puluhan ribu data calon anggota legislatif di KPU
- Apakah pernah kita ubah menjadi pengetahuan yang lebih bermanfaat? Tidak!
- dan lain-lain

* Apa saja Stupid Applications dan Smart Applications itu

Stupid Applications	Smart Applications
System informasi akademik	System Prediksi kelulusan Mahasiswa
System pencatatan pemilu	System prediksi hasil pemilu
System Laporan ketayangan pejabat	System prediksi Korrupsion
System Pencatatan kredit	System penentu kelayakan Kredit

* Evolusi Sciences

- Sebelum 1600 : Empirical Science
disebut sains kalau bentuknya kasar mata
 - 1600 - 1950 : Theoretical Science
disebut sains kalau bisa dibuktikan secara matematis atau eksperimen
 - 1950 s - 1990 : Computational science
 - Seluruh disiplin ilmu bergerak ke komputasi
 - lahirnya banyak model komputasi
 - 1990 - sekarang : Data Science
 - kultur manusia menghasilkan data besar
 - kemampuan komputer untuk mengolah data besar
 - Datangnya data mining sebagai arus utama sains
- * Tugas dan tujuan adanya Data Mining yaitu:

- Mengambil data & dijadikan pola pengetahuan (model, rule, formula, tree)
- Pola pengetahuan dimasukkan ke dalam sistem (software)
- Sistem menjadi cerdas dan bermanfaat signifikan dalam meningkatkan ~~value~~ value perusahaan dan bisnisnya

* 4 Dimensi Masalah Big Data yaitu:

- Volume → data preprocessing (Data cleaning, reduction, transformation, dan data integration)
- Variety → Structured vs unstructured data (Text mining) vs (Data mining terstruktur)
- Velocity
- Veracity

* Peran data mining yaitu: estimasi, forecasting, klasifikasi, Klastering, dan asosiasi

* untuk output / pola / Model / Knowledge sendiri yaitu:

- Formula / Function (waktu tunggu)
- Decision Tree (pohon keputusan)
- Korelasi dan Asosiasi
- Rule / Aturan
- Cluster

* Penerapan Data science di Organisasi agar berhasil

- Meningkatkan pemahaman dan kompetensi SDM di bidang data mining, pahami roles & tugas
- Kumpulkan dan analisis berbagai database organisasi yang banyak dan memunculkan etika mining pengetahuan yang bermanfaat

- Gunakan tools dan teknologi yang sederhana dan mudah yang mendukung visi misi organisasi.
- Meningkatkan kompetensi SDM dan pengalaman pengolahan data organisasi agar lebih akurat sesuai kebutuhan.
- * Beberapa platform yang banyak digunakan dalam data mining yaitu : KNIME, Rapidminer, dll
- * Private and Sektor Komersial yakni :
 - Marketing, Finance, Banking and Insurance, Keamanan, Manufacturing, Web dan Internet Software Engineering, dan Telecommunication
- * Adapun public and Government sector yaitu.
 - Finance, taxation, Medicine and Health Care, Education, Insurance, Security, Transportation, Law, and politic
- * Perbedaan Gaya Penelitian Data Science

Computing Researcher	Non Computing Researnder
<ul style="list-style-type: none"> • Perbaikan Algoritma dan metode • Pengukuran performa dari Algoritma 	<ul style="list-style-type: none"> Analisis Pola yang dihasilkan algoritma Multi aspect dan perspective (Finance, Energi, Hukum, dll) Pengukuran ke manfaat pola untuk kehidupan

- * Perbedaan Pola Pikir Akademik dan Industri dunia Data Science

Akademik	Industri
<ul style="list-style-type: none"> • Kontribusi ke pengetahuan • Perbaikan algoritma, model, atau metode • Publikasi paper ke jurnal terindeks 	<ul style="list-style-type: none"> Kontribusi ke masyarakat Produk yg dibutuhkan masyarakat Intellectual Property Right (IPK): paten, hak cipta, merk, dsb.