

NAMA : MUHAMMAD ILHAM AKBAR

NIM : 19.01.013.076

KELAS : SAINS DATA C

1. DATA CULTURE

Tanpa kita sadari bahwa ternyata perkembangan data ini terjadi terlalu pesat tanpa kita sadari. Suka tidak suka perkembangan data ini sudah ada disekitar kita. Pada tahun 2020 ada project pembuatan bigdata yg bertujuan untuk menemukan jawaban dari keruhan Bu menteri mengenai pendataan pejabat yg mengatakan bahwa masuknya 45% persen perempuan 50% laki-laki Tapi yg menjabat 85% laki-laki dan 15% perempuan. Yang menjadi pertanyaan perempuan-perempuan yg sendir ini kenapa? Kenapa cuman menjabat 15%. permasalahannya seperti ini nanti ditemukan jawabannya jika tidak bisa secara statistik informasi maka akan dimanakan dipola. Jika dulu kita berbicara mengenai 140 terabyte dalam 10 tahun ditumpuk ditahun 2002 itu sudah banyak tapi ternyata sekarang sudah tidak logan. bahkan 140 terabyte dalam 5 hari jadi 140 terabyte dalam 10 tahun. 140 terabyte dalam 5 hari ternyata tidak sebesar di era sekarang karena dilihat dari sosial media sekarang melihat dari data statistik dalam 60 detik apa yg terjadi? ada 4 lagu terupload di spotify, kemudian 72 jam video terupload di youtube dan ada \$1,5 milyar transaksi di amazon. Jika dulu kita berbicara 140 terabyte dalam 5 hari atau 140 terabyte dalam 10 tahun maka sekarang bahasanya yaitu 100 tablet perhari. Google itu memproses 100 peta B perhari di tiga juta servernya. Google itu menguasai hampir 20-25 persen Infrastruktur internet di dunia, makanya kalau mencari tempat paling relatif dan aman untuk urusan hosting ya di Google dan tidak ada tempat lain.

2. DATA MINING

Mengubah data menjadi pengetahuan, yg dimana data harus diubah menjadi pengetahuan agar bermanfaat bagi manusia. Sehingga dengan pengetahuan tersebut manusia dapat melakukan estimasi dan prediksi apa yang terjadi di depan, melakukan analisis tentang asosiasi, korelasi dan pengelompokan antar data dan atribut, serta membantu dalam pengambilan keputusan dan pembuatan kebijakan. Metode mengolah data yg banyak dan beraneka ragam itu menjadi pengetahuan disebut data ~~mining~~ mining. Terminology data mining ini sudah dipakai orang sejak 1980 dengan menggunakan kurikulum di komputer science. bidang ilmu komputer yg ada pada saat itu secara umumnya dibagi mengikuti data apa yg diproses. Jika memproses data terstruktur data mining

dia memproses image - image processing, memproses video disebut computer vision, dan memproses fat disebut software engineering. Untuk data tidak terstruktur biasanya disebut information ~~retrieval~~ retrieval meskipun akhir-akhir ini sering disebut text mining atau text processing. Disiplin ilmu yg mempelajari metode untuk menemukan pola, pengetahuan, formula, aturan, ataupun insight dari suatu data yang volumenya besar dan jenisnya beraneka ragam.

TERMINOLOGI & NAMA LAIN DATA MINING.

* Ekstraksi dari data ke pengetahuan:

1. Data : fakta yg terekam dan tidak membawa arti.
2. Informasi : Rekap, rangkuman, informasi, penjelasan, visualisasi, dan statistik dari data.
3. Pengetahuan : Pola, rumus, aturan atau model yg muncul dari data.

* Nama lain data mining:

- * Data science
- * Knowledge discovery in database (KDD).
- * Big data
- * Business Intelligence
- * Knowledge Extraction
- * Predictive Analytics
- * Pattern analysis
- * Information Harvesting.

3. DATA METHODS

Empat Dimensi masalah Big data

- Volume scale of data
- Variety Different Forms of data
- Velocity Analysis of Streaming Data
- Variability uncertainty of data.

Volume itu jumlah datanya gede, volume yg gede ini nanti diprocesing jadi data cleaning dan data cleaning dibagi menjadi dua. yaitu data yg kosong atau data noisy. data noisy bisa dihilangkan atau diperbaiki, data yg missing bisa dihilangkan juga bisa diisi dengan nilai.

4. DATA USE CASES

Bisa digunakan untuk analiti perbankan, healthcare untuk asuransi kediselamatannya, mengknistal mana yg casenya rawan.

5. DATA RESEARCH

Terbagi menjadi dua *computing/non-compute*. Kalau *computing* pasti mengarah ke pengukuran algoritma tapi kalau *non komputer* yaitu menganalisis pola yg dihasilkan. orang komputer gak peduli dengan objek tapi orang *non komputer* peduli pada objek.