

Nama : Muhammad Bagus
Nim : 19.01.013.046
MK : Sains Data C

Tugas Individu II

1. Culture

Tanpa kita sadari bahwa ternyata perkembangan data yang terjadi terlalu pesat. Perkembangan data ini sudah ada di sekitar kita. Pada tahun 2020 ada project pembuatan big data yang bertujuan untuk menemukan jawaban dari keluhan Bu menteri mengenai penjataan pejabat yang mengatakan bahwa masuknya 45% perempuan 50% laki-laki tapi yang menjabat 80% permasalahan laki-laki dan 15% perempuan. yang menjadi pertanyaan perempuan-perempuan yang cerdas ini (winya kemana?). Permasalahan seperti ini ditemukan jawabannya. jika tidak bisa secara statistik informasi maka akan dimainkan dipolda. jika kita membicarakan mengenai 140 terabyte dalam 10 tahun dikumpulkan di tahun 2002 itu sudah banyak, tapi ternyata sekarang sudah tidak logis. 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 sasmed sekarang melihat dari data statistik dalam 60 detik apa yang terjadi? ada 4 lagu terupload di Spotify. Kemudian 72 jam video terupload di youtube dan ada 1,5 miliar transaksi di amazon.

2. Mining

Dari data culture, data harus kita olah menjadi pengetahuan supaya bisa bermanfaat bagi manusia. 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, membantu pengambilan keputusan dan pembuatan kebijakan. Metode mengolah data-data yang banyak beraneka ragam itu menjadi pengetahuan itu disebut data mining minep, terminologi data mining itu sudah dipakai orang sejak 1980. memproses data struktur (data mining), memproses video (computer vision), memproses sat (Software engineering). Disiplin ilmu mempelajari metode mengolah data



yang masuk kemudian menjadi pengetahuan rumus yang namanya data mining. karena konsepnya yang memainkan data, kumpulan data di lalui algoritma jadilah pengetahuan ini konsep building blok dari proses data mining. Terminologi dan nama lain Data mining:

* Ekstraksi dari data ke pengetahuan:

- Data: faktor yang terkumupan dan tidak membawa arti
- Informasi: Lekap, Rangkuman, Penjelasan, Visualisasi dan Statistik dari data
- Pengetahuan: pola, rumus, aturan atau model yang 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. Metode

Disini membicarakan tentang cara mengolah, memining data yang telah ada, termasuk teknologinya. Data Science atau big data berdasarkan Variety, Velocity, dan Variability.

Volume adalah jumlah datanya dan formatnya beraneka ragam ~~dan~~ kecapatannya, dan ketidak jelasannya. Volume yang besar akan diberikan processing yaitu Data Cleaning. Data Cleaning ada dua yaitu data noisy dan missing. Data noisy bisa dihilangkan bisa juga diperbaiki dengan cara Smoothing. Data missing bisa dihilangkan juga bisa diisi dengan nilai rata-rata.

1. Text mining

- mengolah data tidak terstruktur (unstructured) dalam bentuk text, web, Social media, dsb
- Menggunakan metode text processing untuk mengkonversi data tidak terstruktur menjadi terstruktur. kemudian diolah dengan data mining

2. Data mining

- mengolah data terstruktur (structured) dalam bentuk



- tabel yang memiliki atribut dan kelas,
- menggunakan metode data mining (estimasi, forecasting, klasifikasi, klastering atau asosiasi), yang dasar berpikirnya menggunakan konsep matematika statistika atau heuristik di machine learning

4. use cases

bisa digunakan untuk analiti perbankan health care untuk asuransi kelewatfaanya mengidentifikasi yang berisiko rawan

5. Data Research

- Terbagi menjadi two computing/non - compute . kalo komputing pasti mengandung k & pengukuran algoritma tapi kalau non komputer yaitu menganalisis pola yang dihasilkan. orang komputer jadi peduli dengan objek tapi orang non komputer peduli pada objek