

NAMA : Muhammad Rifkaul

NIM : 19.01.013.111

KELAS : Sains Data C

1. Data CULTURE

Perkembangan data ini terjadi tanpa kita sadari. Perkembangan data ini sudah ada di sekitar kita. Di tahun 2020 pembuat bigdata yang bertujuan untuk menemukan jawaban dari keluhan menteri mengenai Pendataan pejabat yang mengatakan bahwa masuknya 45% Perempuan 50% laki-laki Tapi yang menjabat 85% laki-laki dan 15% Perempuan. Jika dulu kita berbicara mengenai 140 terabyte dalam 10 tahun di kumpul di tahun 2002. dilihat dari sosial media data statistik dalam 60 detik apa yang terjadi? ada 4 lagu terupload di spotify. Jika dulu kita berbicara 140 terabyte dalam 8 hari atau 140 terabyte dalam 10 tahun maka sekarang bahasanya yaitu 1000 tablet perhari.

2. Data Mining

Data ini menjadi pengetahuan yang dimana dapat diubah menjadi pengetahuan yang bermanfaat bagi manusia sehingga dengan pengetahuan tersebut manusia dapat melakukan estimasi dan prediksi yang akan terjadi kedepannya serta dapat membantu dalam pengambilan keputusan dan kebijakan. Data mining ini sudah dipelajari sejak tahun 1980 dengan memasukkan kurikulum komputer sains. Jika memproses data terstruktur data mining melalui proses image processing, Computer vision dan software engineering. Untuk data tidak terstruktur disebut information retrieval atau disebut text mining.

Terminologi dan nama lain Data Mining

- Ekstraksi dari data kepengetahuan

1. Data : Fakta yang teretam dan tidak membawa arti
2. Informasi : Rethap, rangkuman, informasi, dan statistik dari data
3. Pengetahuan : pola, rumus, aturan, dan model yang muncul dari data

- Nama lain dari data mining

1. Data science
2. Big data
3. Business intelligence
4. ~~Knowledge~~ Knowledge extraction.
5. predictive analytics
6. pattern analyse
7. information harvesting

3. Data Methods

Ada 4 masalah big data

- Volume scale of data
- Variety different forms of data
- Velocity Analysis of streaming data
- Variaty uncertainty of data

Jadi data cleaning dibagi menjadi 2 yaitu, data yang kosong atau data noisy, data noisy bisa di perbaiki atau dihilangkan, sedangkan data yang kosong bisa dihilangkan atau diisi dengan nilai.

4. Data Use Cases

Data ini biasa digunakan untuk analisis perbankan dan health care termasuk untuk asuransi kesehatan dapat mengetahui mana lasernya rawan.

5. Data Describer

Data ini terbagi menjadi 2 yaitu, Computing / not-computer.

Computing biasanya merujuk ke pengukuran algoritma sedangkan non-computer

biasanya merujuk pada yang dihasilkan.