

Kelas : Sains Data

1. Culture

2. Mining

3. Methods

Disini membicarakan tentang cara mengolah data yang telah ada, termasuk teknologinya. Data science atau big data berdasarkan variety, velocity, dan veracity. Volume adalah jumlah data dan formatnya, parameter, ragam, kecepatannya, dan ketidakkelengkapan. Volume yang besar



akan diberikan processing yaitu data cleaning. Data cleaning ada dua yaitu data missing dan data noisy. Data noisy bisa dihilangkan bisa juga diperbaiki dengan cara smoothing. Data missing bisa dihilangkan juga bisa diisi dengan nilai rata-rata.

- Text mining

- mengolah data tidak terstruktur dalam bentuk text, web, social media
- menggunakan metode text processing untuk mengkonversi data tidak terstruktur, ~~menjadi~~ terstruktur, kemudian diolah dengan data mining.

- Data mining

- mengolah data terstruktur dalam bentuk tabel yang memiliki atribut dan kelas
- menggunakan metode data mining yang dasar berpikirnya menggunakan konsep matematika statistika atau heuristik atau machine learning.

4. Use Cases

Bisa digunakan untuk analitis perbankan health care untuk asuransi: kedewasaannya, mengkrystal mana yang usernya rawan.

5. Data Research

Terbagi menjadi dua computing / non compute, kalau computing pasti mengarah ke pengukuran algoritma tapi kalau non komputer yaitu menganalisis pola yang dihasilkan orang. komputer tidak peduli dengan objek tapi orang non komputer peduli pada objek.