

Laporan Seminar Big Data

Author: Asma A. P.

Poster:



Big Data

Big data:

- Kumpulkan data, lalu
- lakukan sesuatu dengan data tersebut.

Contoh:

- Google traffic (data dari handphone pengguna)
- Rekomendasi di [belanja online, video online, dsb]
- Iklan online
- Olahraga
- Ocean health

Tools and Technique

1. Basic data manipulation & analysis

- bikin query
- pakai spreadsheet, DB

2. Data mining

- cari pola

3. Machine learning

- buat prediksi
- algoritma: regression, classification, clustering

4. Data visualisation

- bar chart, pie chart, scatter plot, peta, dsb

5. Mengumpulkan dan mempersiapkan data

- bisa makan 80% pekerjaan

Pitfalls

- correlation belum tentu causation
- underfitting dan overfitting

Data Privacy

Bahas tentang privacy.

Languages, Systems, Platform

- Spreadsheet
- Programming language: R, python
- Database: relational, NoSQL
- MapReduce/Hadoop
- Spark

Latihan

1. Pakai spreadsheet (Google Sheet)

- bikin kolom, tambah formula, dsb
- pivot table
- buat grafik

2. Pakai Tableau

- buat grafik