

**Nama : Muhamad Naufal Burhanuddin Balit**

**Kelas : Jumatec**

## **Resume Materi Data Processing**

- EDA (Exploration Data Analysis)

Exploration Data Analysis (EDA) bagaikan jiwa bagi semua proses analisis data. Kemampuan untuk melakukan EDA dengan baik adalah syarat dasar utama bagi seluruh profesi yang terkait dengan pengolahan data, baik itu business intelligence, data analyst, data scientist, dan sebagainya.

EDA juga menjadi tahapan awal dari kebanyakan proses analisis data dan menjadi suatu tahapan yang amat menentukan seberapa baik Analisa data selanjutnya akan di akan dihasilkan. Diperkenalkan oleh John Turkey 1961: “Procedures for analyzing data, techniques for interpreting the result of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and result of (mathematical) statistic which apply to analyzing data”.

Komponen EDA meliputi preprocessing, perhitungan berbagai nilai statistic dasar (e.g. ukuran pusat dan penyebaran data), visualisasi, penyusunan hipotesis(dugaan awal), pemeriksaan asumsi, hingga story-telling dan reporting. Di dalamnya juga termasuk proses penanganan missing values, outlier, reduksi dimensi, pengelompokkan, transformasi dan distribusi data.

Beberapa tujuan dari EDA yaitu :

1. menyarankan hipotesis
2. menilai asumsi
3. pemilihan teknik statistik yang tepat
4. pengumpulan data lebih lanjut

- Data(set)

data(set) adalah koleksi beberapa entitas/objek data dan attribute nya. Attribut adalah sifat atau karakteristik dari sebuah objek.

Contoh pada objek dari manusia yaitu :

1. umur
2. berat badan
3. tinggi badan
4. jenis kelamin

setiap attribute memiliki beberapa kemungkinan “state”, sebagai contoh: pria/Wanita. Koleksi attribute mendefinisikan suatu objek

- Preprocessing

Preprocessing adalah kunci utama dalam menadapatkan model data yang valid dan reliable. Preprocessing yang berbeda, akan menghasilkan kesimpulan atau insight yang berbeda pula.

Beberapa proses dasar yaitu :

1. Seleksi variable dan “join”
2. Data cleaning: duplikasi, noise, dan outliers
3. Transformasi data
4. Dimensional reduction

Mengapa perlu adanya preprocessing?

Karena data yang ada di dunia nyata, terkadang tidak lengkap dan sangat banyak.

## AI4Jobs | Kampus Merdeka Batch 3

Langkah Langkah Preprocessing yaitu :

1. Data gathering
  - a. Data warehouse, database, web crawling
  - b. Identifikasi ekstraksi, dan integrasi data
2. Data cleaning
3. Transformasi data (misal encoding var kategorik)
4. Normalisasi/standarisasi
5. Data reduction:
  - a. Variable selection (domain knowledge/automatic)
  - b. Feature engineering
  - c. Variable reduction

Sumber Materi:

[https://www.youtube.com/watch?v=OzxmCTPpbN8&ab\\_channel=taudataAnalytics](https://www.youtube.com/watch?v=OzxmCTPpbN8&ab_channel=taudataAnalytics)