

AI4Jobs | Kampus Merdeka Batch 3

Nama : Abdullah Latif
Kelas : JumaTec Class
Tanggal : 4 Oktober 2022

Reference : <https://youtu.be/OzxmCTPpbN8>

HASIL RESUME

EDA (Exploratory Data Analysis) adalah bagian awal sebelum seorang data science atau analysis kemudian membuat model yang lebih kompleks, dan digunakan untuk memperoleh insight lewat visualisasi dan statistik dasar. EDA diperkenalkan oleh John Tukey dari tahun 1961 dan kemudian semakin berkembang. Seiring berjalannya berkembangnya ilmu dan teknologi terkait data apa saja yang ada diantaranya adalah data preparation seperti menangani mentransformasi data lalu apa itu tipe data menyesuaikan tipe data. Apabila ada yang kurang tepat noise menangani noise misalnya lalu missing value, outlier dan lain sebagainya.

- Noise dapat terjadi karena beberapa hal, meliputi :
 1. Kesalahan instrumen pengukuran
 2. Kesalahan input/entry
 3. Transmisi yang tidak sempurna
 4. Inkonsistensi penamaan
- Sedangkan tipe dari missing value adalah sebagai berikut :
 1. Missing Completely At Random (MCAR)
Data hilang secara acak dan tidak berkaitan dengan variabel tertentu
 2. Missing At Random (MAR)
Data di suatu variabel hilang hanya berkaitan dengan variabel respon/pengamatan
 3. Missing Not At Random (MNAR)
Data di suatu variabel berkaitan dengan variabel itu sendiri
- Pentingnya preprocessing adalah sebagai berikut :
 1. Data di dunia nyata biasanya tidak sebersih data di buku akademik
 2. Data tidak lengkap, hanya agregat, kurang variabel penting, dsb
 3. Analisa pada data yang tidak di preprocess biasanya menghasilkan insight yang kurang tepat.
- Beberapa langkah utama setelah melakukan data understanding sebagai berikut :
 1. Data gathering
 2. Data Cleaning

AI4Jobs | Kampus Merdeka Batch 3

- 3. Transformasi data
- 4. Normalisasi / Standarisasi
- 5. Data reduction