Modul 2

(Minggu kedua)

MODUL 2: Preprocessing Data

TUJUAN

1. Mahasiswa Mampu Memahami data pre-processing.
2. Mahasiswa Mampu Mengimplementasikan pre-processing data.
3. Mahasiswa Mampu Menganalisa data pre-processing dari data yang telah disediakan.

PENDAHULUAN

Hal yang harus dilakukan dan acuan yang harus dibaca sebelum praktikum :

1. Menginstal Python pada PC masing-masing praktikan.

DAFTAR PERTANYAAN

1. Apa yang dimaksud dengan data pre-processing?
2. Mengapa perlu adanya tahapan pre-processing?
3. Sebutkan langkah-langkah dari data pre-processing?

**TEORI SINGKAT**

**Data Preparation** atau bisa disebut juga dengan **data preprocessing** adalah suatu proses/langkah yang dilakukan untuk membuat data mentah menjadi data yang berkualitas(input yang baik untuk data mining tools). Tahapan dalam melakukan data mining salah satunya adalah preprosesing data. Pertanyaannya adalah mengapa data perlu di bersihkan sebelum diproses?

Hal ini terjadi karena biasanya data yang akan digunakan belum baik, penyebabnya antara lain :

* **Incomplete** : kekurangan nilai-nilai atribut atau atribut tertentu lainnya.
* **Noisy** : berisi kesalahan atau nilai-nilai outlier yang menyimpang dari yang diharapkan.
* **Inconsisten** : ketidakcocokan dalam penggunaan kode atau nama.

Disini kualitas data yang baik didasarkan oleh keputusan yang baik dan data warehouse memerlukan integrasi kualitas data yang konsisten. Beberapa hal yang perlu diperhatikan untuk mendapatkan data yang baik adalah :

* **Accuracy**
* **Completeness**
* **Consistency**
* **Timeliness**
* **Value added**
* **Interpretability**
* **Accessibility**
* **Contextual**
* **Representational**

**LAB SETUP**

Hal yang harus disiapkan dan dilakukan oleh praktikan untuk menjalankan praktikum modul ini:

1. Menginstal library yang dibutuhkan untuk mengerjakan modul.
2. Menjalankan Visual Studio Code.

**Latihan 1.1**

**Penjelasan Singkat :**