

Panduan Praktikum Eksplorasi Data

Durasi: 3 × 110 menit

Capaian Pembelajaran Mata Kuliah

Setelah menyelesaikan praktikum ini diharapkan mahasiswa mampu menjelaskan tahapan data mining, karakteristik data, eksplorasi data, dan praproses data serta penerapannya pada suatu permasalahan.

Tools

- Google Colab
- Jupyter Notebook
- PyCharm
- Spyder
- Python IDE yang lain

Materi Praktikum

- Pengertian Eksplorasi Data

Eksplorasi data adalah tahap awal dalam proses analisis data yang bertujuan untuk memahami struktur, pola, dan karakteristik data. Tahap ini merupakan langkah penting sebelum melanjutkan ke proses pembersihan data, transformasi, atau penerapan algoritma data mining.

- Tujuan Eksplorasi Data

1. Memahami struktur dataset (jumlah atribut, jenis data, distribusi).
2. Mengidentifikasi pola, tren, atau hubungan antar variabel.
3. Mendeteksi nilai yang hilang (missing values) atau outlier.
4. Memastikan kualitas data untuk analisis lebih lanjut.
5. Mendukung keputusan tentang teknik analisis data atau algoritma yang akan digunakan.

- Jenis-Jenis Atribut Data

1. Nominal Attributes:

- Variabel kategoris tanpa urutan tertentu.
- Contoh: warna (merah, biru, hijau), jenis kelamin (laki-laki, perempuan).

2. Binary Attributes:

- Atribut kategoris dengan dua nilai.
- Contoh: Ya/Tidak, 1/0, True/False.

3. Ordinal Attributes:

- Variabel kategoris dengan urutan tertentu, tetapi selisih antar kategori tidak seragam.

- Contoh: tingkat pendidikan (SD, SMP, SMA), tingkat kepuasan (rendah, sedang, tinggi).

4. Numeric Attributes:

- Variabel numerik yang dapat dihitung secara matematis.
- Diskrit: Nilai terbatas, seperti jumlah anak.
- Kontinu: Nilai dalam rentang tertentu, seperti tinggi badan.

- Tahapan Eksplorasi Data

1. Memahami Dataset:

- Menentukan jumlah atribut dan data.
- Mengetahui jenis data: numerik atau kategoris.

2. Analisis Statistik Dasar:

- Menghitung rata-rata, median, modus, standar deviasi.
- Mengidentifikasi distribusi data.

3. Deteksi Nilai yang Hilang:

- Menemukan atribut dengan nilai kosong.
- Metode penanganan:

- Menghapus data (jika jumlahnya kecil).
- Mengisi dengan nilai rata-rata, median, atau algoritma prediksi.

4. Deteksi Outlier:

- Outlier adalah data yang jauh berbeda dari nilai lainnya.
- Teknik deteksi:
 - Metode statistik seperti z-score.
 - Visualisasi seperti boxplot.

5. Transformasi Data:

- Normalisasi: Mengubah nilai ke rentang 0-1.
- Standardisasi: Membuat data memiliki rata-rata 0 dan standar deviasi 1.
- Encoding: Mengubah data kategoris menjadi numerik (contoh: one-hot encoding).

6. Visualisasi Data:

- Membuat diagram atau grafik untuk memahami pola data:
 - Histogram untuk distribusi.
 - Scatter plot untuk hubungan antar variabel.
 - Boxplot untuk mendeteksi outlier.

- Pentingnya Eksplorasi Data

1. Menghindari kesalahan analisis akibat data yang tidak dipahami.
2. Menentukan langkah pra-pemrosesan yang diperlukan.
3. Mendapatkan wawasan awal tentang data yang dapat mendukung interpretasi hasil.

- Studi Kasus Sederhana

- Dataset: Informasi pelanggan bank (atribut: usia, pendapatan, status pekerjaan, pinjaman).
- Langkah Eksplorasi:
 1. Memeriksa data hilang pada atribut "pendapatan".

2. Membuat histogram untuk distribusi usia.
3. Mendeteksi outlier pada atribut "pendapatan" menggunakan boxplot.
4. Menggunakan one-hot encoding untuk atribut "status pekerjaan".

Panduan Praktikum

Download lalu pelajari dan running code pada link berikut:

<https://github.com/aldinata/Modul-Praktikum-Data-Mining/blob/main/Materi/1%20-%20Eksplorasi%20Data.ipynb>