

# Panduan Praktikum

## Sequential Pattern Analysis

Durasi: 3 × 110 menit

### Capaian Pembelajaran Mata Kuliah

Mahasiswa mampu menjelaskan teknik association rule, sequential pattern analysis, serta penggunaannya pada suatu permasalahan.

### Tools

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

### Materi Praktikum

- Pengertian Sequential Pattern Analysis

Sequential Pattern Analysis adalah teknik dalam data mining yang bertujuan untuk menemukan pola berurutan dalam dataset yang besar. Pola ini menggambarkan urutan item atau kejadian yang sering muncul bersama dalam suatu urutan waktu atau kejadian. Metode ini banyak digunakan dalam analisis perilaku pelanggan, prediksi, atau sistem rekomendasi.

- Konsep Dasar Sequential Pattern Analysis

1. **Sequence**: Urutan dari item atau kejadian yang berhubungan. Contoh:

- Urutan pembelian pelanggan: [roti → susu → telur].
- Log aktivitas pengguna di sebuah aplikasi.

2. **Itemset**: Kumpulan item yang terjadi dalam waktu bersamaan atau dalam suatu kejadian tunggal. Contoh: [roti, mentega].

3. **Support**: Proporsi urutan dalam dataset yang mengandung pola tertentu.

## Formula:

$$\text{Support(Pola)} = (\text{Jumlah Sequence yang Mengandung Pola}) / (\text{Total Sequence})$$

- Metrik Evaluasi Pola Berurutan

1. **Support**: Mengukur seberapa sering pola muncul dalam dataset.
2. **Confidence**: Mengukur kekuatan hubungan antar pola dalam urutan.
3. **Length**: Panjang pola yang dianalisis, biasanya ditentukan oleh jumlah item dalam urutan.

- Algoritma Sequential Pattern Analysis

1. Apriori for Sequential Pattern Mining

- Memanfaatkan pendekatan iteratif dengan properti downward closure (pola dengan support rendah tidak akan menghasilkan pola yang valid).

- Langkah:

1. Identifikasi item tunggal dengan support tinggi.
2. Bentuk kombinasi pola lebih panjang dengan item tersebut.
3. Ulangi hingga tidak ada lagi pola yang memenuhi kriteria.

2. GSP (Generalized Sequential Pattern)

- Algoritma untuk menemukan pola berurutan dengan menggunakan pruning berdasarkan support minimum.

- Memanfaatkan waktu atau urutan kejadian untuk membentuk pola lebih kompleks.

3. PrefixSpan

- Menggunakan pendekatan *divide and conquer* dengan fokus pada prefix (bagian awal dari urutan).

- Lebih efisien dibandingkan GSP karena tidak perlu menghasilkan kandidat pola secara eksplisit.

- Contoh Aplikasi

1. **Analisis Perilaku Pelanggan**:

- Identifikasi urutan pembelian yang sering dilakukan pelanggan untuk meningkatkan strategi pemasaran.

- Contoh: [buku → pulpen → tinta].

## 2. **\*\*Sistem Rekomendasi\*\***:

- Memprediksi barang yang kemungkinan besar akan dibeli pelanggan berdasarkan urutan pembelian sebelumnya.

- Contoh: Jika pelanggan membeli [televisi → kabel HDMI], maka mereka cenderung membeli [braket dinding].

## 3. **\*\*Prediksi Pola Medis\*\***:

- Menemukan pola kronologis dari diagnosis atau pengobatan pasien untuk meningkatkan prediksi hasil kesehatan.

## 4. **\*\*Analisis Log Aktivitas\*\***:

- Menemukan pola penggunaan aplikasi untuk memahami perilaku pengguna atau mendeteksi anomali.

# Panduan Praktikum

Download lalu pelajari dan running code pada link berikut:

<https://github.com/aldinata/Modul-Praktikum-Data-Mining/blob/main/Materi/5%20-%20Sequential%20Pattern%20Analysis.ipynb>