

Tugas Praktikum: Sequential Pattern Analysis

Deskripsi Tugas

Mahasiswa diminta untuk menemukan pola berurutan dalam dataset menggunakan algoritma Sequential Pattern Mining seperti PrefixSpan atau GSP. Praktikum ini bertujuan untuk memahami langkah-langkah menemukan pola berurutan dan mengevaluasi pola tersebut berdasarkan support dan confidence.

Tujuan Praktikum

1. Memahami konsep Sequential Pattern Analysis dan aplikasinya dalam data mining.
2. Menerapkan algoritma Sequential Pattern Mining seperti PrefixSpan atau GSP pada dataset transaksi.
3. Mengevaluasi pola berurutan berdasarkan metrik seperti support dan confidence.

Langkah-Langkah Praktikum

1. **Unduh Dataset**

Gunakan dataset Retail Transaction Dataset dari Kaggle. Dataset ini berisi data transaksi pelanggan di sebuah ritel, termasuk urutan pembelian mereka.

Link dataset: <https://www.kaggle.com/datasets/acostasg/random-transaction-dataset>

2. **Eksplorasi Data**

- Tampilkan struktur dataset dan atribut yang tersedia.
- Identifikasi kolom yang relevan untuk analisis pola berurutan (misalnya, `TransactionID` dan `Items`).
- Bersihkan data dengan menghapus transaksi yang tidak relevan atau duplikat.

3. **Persiapan Data untuk Sequential Pattern Mining**

- Format data menjadi urutan transaksi berdasarkan pelanggan atau waktu.
- Gunakan pustaka Python seperti `pyfim`, `mlxtend`, atau SPMF untuk memproses dataset.

4. **Penerapan Algoritma**

- Terapkan algoritma Sequential Pattern Mining (PrefixSpan atau GSP) dengan parameter support minimal (misalnya, 0.01).
- Temukan pola berurutan yang memenuhi kriteria support.

5. **Evaluasi Pola**

- Analisis pola yang ditemukan dan tentukan relevansinya.
- Pilih tiga pola terbaik berdasarkan nilai support tertinggi.

6. ****Laporan Praktikum****

- Sertakan hasil eksplorasi data, pola yang ditemukan, dan analisis pola berurutan.
- Tulis kesimpulan tentang pola yang ditemukan dan aplikasinya.

Hasil yang Diharapkan

1. Dataset yang telah diproses dan diformat untuk Sequential Pattern Mining.
2. Daftar pola berurutan yang ditemukan dengan nilai support dan confidence.
3. Laporan singkat yang mencakup langkah-langkah, hasil, dan analisis pola berurutan.

Penilaian

1. ****Eksplorasi Data (20%)****: Pemahaman tentang struktur dan persiapan data.
2. ****Implementasi Algoritma (40%)****: Kesesuaian penerapan algoritma PrefixSpan atau GSP.
3. ****Evaluasi Pola (30%)****: Relevansi pola yang ditemukan dan interpretasinya.
4. ****Laporan Praktikum (10%)****: Kerapihan, kelengkapan, dan analisis hasil.