

JAWABAN KUIS 2

Nama Mata Kuliah : Data Mining
Kode Mata Kuliah : MKB22-53501
Semester/Kelas : V / B TRPL
Dosen : Dr. Putu Manik Prihatini, S.T., M.T.
Hari/Tanggal : Jumat / 13 Desember 2024
Nama Lengkap : Farhan Maulana
NIM : 2215354018

Jawaban

1. Jawaban soal nomor 1

Langkah 1: Membentuk Frequent 1-itemset

1. Hitung frekuensi (support count) masing-masing item dalam daftar transaksi.
 - a. **T100**: {M, O, N, K, E, Y}
 - b. **T200**: {D, O, N, K, E, Y}
 - c. **T300**: {M, A, K, E}
 - d. **T400**: {M, U, C, K, Y}
 - e. **T500**: {C, O, K, I, E}
2. Tabel frekuensi item:
 - a. **M**: 3
 - b. **O**: 3
 - c. **N**: 2
 - d. **K**: 5
 - e. **E**: 4
 - f. **Y**: 3
 - g. **D**: 1
 - h. **A**: 1
 - i. **U**: 1
 - j. **C**: 2
 - k. **I**: 1
3. Hapus item dengan support count < 3.
 - a. Frequent 1-itemset: {M, O, K, E, Y}

Langkah 2: Membentuk Frequent 2-itemset

1. Buat pasangan item dari Frequent 1-itemset:
 - a. Pasangan: {M,O}, {M,K}, {M,E}, {M,Y}, {O,K}, {O,E}, {O,Y}, {K,E}, {K,Y}, {E,Y}
2. Hitung support count setiap pasangan:
 - a. {M,O}: 1
 - b. {M,K}: 3
 - c. {M,E}: 3
 - d. {M,Y}: 2
 - e. {O,K}: 3
 - f. {O,E}: 3
 - g. {O,Y}: 2
 - h. {K,E}: 4
 - i. {K,Y}: 3
 - j. {E,Y}: 2
3. Hapus pasangan dengan support count < 3.
 - a. Frequent 2-itemset: {M,K}, {M,E}, {O,K}, {O,E}, {K,E}, {K,Y}

Langkah 3: Membentuk Frequent 3-itemset

1. Buat kombinasi 3-itemset dari Frequent 2-itemset.
 - a. Kombinasi: {M,K,E}, {O,K,E}
2. Hitung support count setiap kombinasi:
 - a. {M,K,E}: 3
 - b. {O,K,E}: 3
3. Frequent 3-itemset: {M,K,E}, {O,K,E}

2. Jawaban soal nomor 2

DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

Keunggulan:

- DBSCAN relatif tahan terhadap noise
- DBSCAN dapat menemukan banyak cluster yang tidak dapat ditemukan dengan menggunakan K-means

Kelemahan:

- DBSCAN memiliki masalah ketika cluster memiliki kepadatan yang sangat bervariasi.
- DBSCAN bermasalah dengan data dimensi tinggi karena densitas lebih sulit ditentukan untuk data dimensi tinggi
- DBSCAN bisa menjadi mahal ketika perhitungan tetangga terdekat membutuhkan komputasi semua kedekatan berpasangan, seperti yang biasanya terjadi pada data dimensi tinggi.

K-Means

Keunggulan:

- sederhana, efisien, dan dapat digunakan untuk berbagai tipe data

Kelemahan:

- tidak dapat menangani cluster non-globular atau cluster dengan ukuran dan kepadatan berbeda
- K-Means memiliki masalah pengelompokan data yang mengandung outlier.
- K-Means dibatasi untuk data yang memiliki konsep pusat (centroid)

3. Jawaban soal nomor 3

Langkah 1: Inisialisasi Cluster

Setiap data awalnya dianggap sebagai cluster sendiri:

Cluster: {p1}, {p2}, {p3}, {p4}, {p5}

Langkah 2: Temukan Pasangan Cluster dengan Similarity Tertinggi

Dari similarity matrix, cari dua cluster dengan similarity **tertinggi** (selain diagonal matriks, karena diagonal = 1).

Pada tabel, similarity tertinggi adalah **0.98** antara **p2** dan **p5**.

Langkah 3: Gabungkan Cluster

Gabungkan cluster {p2} dan {p5} menjadi satu cluster baru:

Cluster: {p1}, {p2, p5}, {p3}, {p4}

Update matriks similarity untuk cluster baru:

Complete Link memilih nilai similarity **terendah** antara semua pasangan data di cluster baru dengan cluster lainnya.

Contoh:

- Similarity antara {p2, p5} dan {p1} = **min(0.10, 0.35)** = 0.10
- Similarity antara {p2, p5} dan {p3} = **min(0.64, 0.85)** = 0.64
- Similarity antara {p2, p5} dan {p4} = **min(0.47, 0.76)** = 0.47

Langkah 4: Ulangi Proses Penggabungan

1. Dari matriks yang diperbarui, cari similarity tertinggi lagi.
2. Gabungkan cluster yang memiliki similarity tertinggi.
3. Perbarui matriks similarity menggunakan Complete Link (ambil nilai **terendah** dari pasangan).

Langkah Akhir: Hentikan Ketika Semua Data Berada dalam Satu Cluster

Lanjutkan proses penggabungan hingga semua data tergabung dalam satu cluster, sambil mencatat langkah-langkah penggabungan.