

TUGAS 4: FI, AR, dan SP Mining

Tugas ini dikerjakan kelompok. Kerjakan soal nomor 1 dan 2 secara manual dan tuliskan jawabannya pada MS Word/Open Office. Soal nomor 3 dan 4 kerjakan menggunakan tool SPMF. Tuliskan jawaban singkat pada MS Word, dan sertakan screenshot dari luaran SPMF. Konversikan semua jawaban ke PDF dan unggah ke Elok, paling lambat hari Senin, 23 Nopember 2020, pukul 24.00.

1. Tabel 1 adalah database berisikan 5 transaksi dan 5 items. Selesaikan pertanyaan di bawah ini menggunakan algoritma Apriori dengan perhitungan manual.

Table 1. Database memuat 6 transaksi

Transaction id	Items
t1	{1, 2, 4, 5}
t2	{1, 3}
t3	{1, 2, 3, 5}
t4	{2, 3, 5}
t5	{1, 2, 3, 5}

- a. Tentukan **F1** dan **F2**, yaitu himpunan semua frequent itemset yang panjangnya 1 dan 2, jika nilai *minimum support* 50%.
 - b. Tentukan **C3** (candidate frequent 3-itemset) setelah tahap Join dan Pruning.
 - c. Tentukan **F3** (nilai *minimum support* 50%)
 - d. Pilih salah satu frequent 3-itemset pada F3, dan tentukan association rule yang memenuhi *minimum confidence* 80%.
 - e. Tentukan Frequent Closed Itemset dan Frequent Maximal Itemset (jika ada).
2. Diberikan 4 customer sequence seperti pada Tabel 2. Jawablah pertanyaan berikut ini menggunakan algoritma GSP dengan perhitungan manual.
- a. Dari customer sequences pada Table 2, tentukan himpunan semua frequent 1-sequences (F1) dan semua frequent 2-sequences (F2, jika nilai minimum support 60%.
 - b. Dari F2 di (a), tentukan C3 setelah JOIN dan C3 setelah PRUNING.
 - c. Tentukan F3.
 - d. Tentukan Frequent Closed Sequential Pattern dan Frequent Maximal Sequential Pattern (jika ada).

Tabel 2. Customer sequence database

ID	Sequences
S1	(1), (1 2 3), (1 3), (4), (3 6)
S2	(1 4), (3), (2 3), (1 5)
S3	(5 6), (1 2), (4 6), (3), (2)
S4	(5), (7), (1 6), (3), (2), (3)

3. Menggunakan file **mushroom.arff** dan SPMF tool, jawablah pertanyaan di bawah ini.
- Tuliskan jumlah *frequent itemset* yang dihasilkan, jika nilai minimum support 70%. Tuliskan nama algoritma yang anda gunakan.
 - Tuliskan jumlah *frequent closed itemset*, jika nilai minimum support 70%. Tuliskan nama algoritma yang anda gunakan.
 - Tuliskan jumlah *frequent maximum itemset*, jika nilai minimum support 70%. Tuliskan nama algoritma yang anda gunakan.
 - Tuliskan jumlah association rule, jika nilai minimum support 70% dan nilai minimum confidence 95%. Tuliskan nama algoritma yang anda gunakan.
 - Tuliskan jumlah *closed association rule* yang dihasilkan, jika nilai minimum support 70% dan nilai minimum confidence 95%. Tuliskan nama algoritma yang anda gunakan.
4. Menggunakan file **FIFA.txt** yang berisi sequence click stream dari website FIFA World Cup 98 (nomer dalam file menyatakan halaman web (page number)), jawab pertanyaan di bawah ini menggunakan tool SPMF
- Tuliskan jumlah sequential pattern yang dihasilkan, jika digunakan nilai minimum support 20%. Tuliskan nama algoritma yang anda gunakan.
 - Berapakah jumlah sequential pattern di (a) yang nilai *suppot count* nya lebih dari 5000?
 - Berapakah jumlah sequential pattern di (b) yang memuat *page no* 135?
 - Tuliskan jumlah *closed sequential pattern*, jika digunakan nilai minimum support 25%. Tuliskan nama algoritma yang anda gunakan.
 - Tuliskan jumlah *maximal sequential pattern*, jika digunakan nilai minimum support 25%. Tuliskan nama algoritma yang anda gunakan.

