

Assignment 04

Worksheet Association Rule



Aturan asosiasi (*association rule*) adalah sebuah teknik dalam analisis data yang digunakan untuk mengidentifikasi pola hubungan antara item-item dalam kumpulan data. Teknik ini biasanya diterapkan pada data transaksional, seperti data pembelian pelanggan di toko online atau data penjualan di sebuah toko konvensional. Dalam konsepnya, aturan asosiasi mencoba untuk menemukan korelasi antara item-item tertentu dalam dataset, seperti produk yang sering dibeli bersama atau item yang sering muncul bersama dalam transaksi.

Salah satu metode yang paling umum digunakan untuk menemukan aturan asosiasi adalah Algoritma Apriori. Prinsip apriori adalah sebagai berikut, jika suatu item itu sering muncul atau *frequent* maka semua subsetnya pun harus *frequent* sebagai contoh, $\forall A, B \ A \subseteq B \Leftrightarrow \text{supp}(A) \geq \text{supp}(B)$ Ide utama Apriori: pertama, tentukan frequent 1-itemset, lalu frequent 2-itemset, dan seterusnya. hanya satu pemindaian per panjang k melalui database transaksi.

Algoritma ini bekerja dengan cara mengidentifikasi item-item yang sering muncul bersama dalam transaksi, lalu membangun aturan asosiasi berdasarkan frekuensi kemunculan item-item tersebut. Sebagai contoh, algoritma ini dapat menghasilkan aturan seperti "Jika pelanggan membeli X dan Y, maka mereka juga cenderung membeli Z."

Aturan asosiasi memiliki berbagai aplikasi yang luas dalam berbagai bidang, termasuk e-commerce, manajemen inventaris, pemasaran, dan keamanan data. Dalam e-commerce, aturan asosiasi digunakan untuk menyusun rekomendasi produk kepada pelanggan berdasarkan pola pembelian mereka sebelumnya. Di bidang manajemen inventaris, aturan ini membantu dalam merencanakan stok barang dan mengoptimalkan rantai pasokan dengan memahami hubungan antara produk-produk yang berbeda. Selain itu, aturan asosiasi juga digunakan dalam pemasaran untuk merancang strategi promosi yang lebih efektif berdasarkan pola pembelian pelanggan.

Dengan demikian, aturan asosiasi merupakan alat yang kuat dalam mengungkap wawasan berharga dari data dan membantu organisasi membuat keputusan yang lebih baik berdasarkan pemahaman yang lebih dalam tentang pola-pola dalam data mereka.

Soal:

Diberikan suatu data transaksi penjualan yang berhasil dilakukan pada suatu toko elektronik sebagai berikut:

Transaction 1: {Laptop, Mouse, Keyboard, Headphones}

Transaction 2: {Laptop, Mouse, Keyboard}

Transaction 3: {Mouse, Keyboard, Headphones}

Transaction 4: {Laptop, Keyboard, Headphones}

Transaction 5: {Laptop, Mouse, Keyboard, Webcam}

Pertanyaan:

1. Dengan minimum support = 0.6, jalankan algoritma secara naif untuk mendapatkan frequent itemset dengan memanfaatkan data transaksi diatas! Jelaskan setiap langkah yang perlu dilakukan.
2. Dengan minimum support = 0.6, jalankan algoritma Apriori dengan menuliskan $F < K >$ sebagai *frequent itemsets* pada suatu K dan $C < K+1 >$ sebagai kandidat untuk *frequent itemsets* selanjutnya dengan K adalah jumlah item dalam himpunan berdasarkan data transaksi diatas!
Hint: Anda dapat memanfaatkan Apriori Principle, lihat referensi pada slide berikut [ini](#)
3. Dengan minimum confidence sebesar 0.5, temukan semua aturan asosiasi dari frequent sets yang telah ditemukan sebelumnya!
4. Untuk data transaksi yang lebih besar, pengerjaan dengan tangan akan sangat tidak efisien sehingga diperlukan untuk dapat mengimplementasikan algoritma untuk menemukan frequent itemsets, implementasikan algoritma untuk menghitung frequent itemsets tersebut tanpa menggunakan library module (*make it from scratch*)!

Input : Kumpulan himpunan yang merepresentasikan data transaksi

Output : Himpunan gabungan dari semua *frequent itemsets*

Hint: Anda dapat memanfaatkan contoh implementasi pseudocode dibawah ini (diperbolehkan menggunakan library bantuan untuk melakukan generasi).

$F1 = \{ \text{All frequent 1-itemsets} \}$

Generate $C2 = \text{pairs of items from } F1$

$F2 = \{ \text{All frequent 2-itemsets selected from } C2 \}$

$k = 2$

While FK is not empty do begin

Generate C_{k+1} by joining itemset-pairs in F_k ;

Prune itemsets from C_{k+1} that violate downward closure;

Determine F_{k+1} by support counting on C_{k+1} ;

$k = k + 1$

return union($F1, F2, \dots, F_n$)

5. Menurut Anda, apa kelebihan dan kekurangan dari pemanfaatan aturan asosiasi sebagai sistem rekomendasi jika dibandingkan dengan kebanyakan *machine*

learning based algorithm untuk sistem rekomendasi e.g *matrix factorization algorithm*?

Pengumpulan Tugas

Pengumpulan tugas ini diperbolehkan untuk ditulis tangan (pengumpulan di-scan) atau diketik dan dikumpulkan dalam format berikut:

Assignment4_[NPM]_[NamaLengkap].pdf

Contoh:

Assignment4_1906438834_TimothyOrvinEdwardo.pdf

Apabila Anda mencari referensi dari internet, Anda WAJIB menyertakan sumbernya. Dilarang keras menyontek. Plagiarisme tidak ditoleransi dan akan dikenai penalti atau nilai akhir E.

Pengurangan nilai akibat keterlambatan pengumpulan tugas akan ditentukan berdasarkan jumlah menit keterlambatan Anda dalam mengumpulkan. Misalnya, apabila terlambat 1 menit, nilai akhir akan dikurangi 1 poin, apabila terlambat 10 menit, nilai akhir akan dikurangi 10 poin, dan seterusnya.