

DATA MINING

3 SKS | Semester 6 | S1 Sistem Informasi

TEKNIK ASOSIASI – ALGORITMA APRIORI

Nizar Rabbi Radliya
nizar@email.unikom.ac.id



TEKNIK ASOSIASI

Teknik Asosiasi (association) atau *association rule mining* adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item. Teknik ini digunakan untuk menemukan pola yang mendeteksi kumpulan atribut-atribut yang muncul bersamaan dalam frekuensi yang sering.

Biasa disebut dengan *affinity analysis* atau *market basket analysis*.

Algoritma:

A Priori

FP-Growth

GRI

dll

*Customers who bought this item ...
also bought ...*



TEKNIK ASOSIASI

Association rule biasanya dinyatakan dalam bentuk :

$\{\text{Teh}\} \rightarrow \{\text{Gula}\}$ (*support* = 40%, *confidence* = 50%)

Yang artinya : "50% dari transaksi di database yang memuat item **Teh** juga memuat item **Gula**. Sedangkan 40% dari seluruh transaksi yang ada di database memuat kedua item tersebut."

Dapat juga diartikan : "Seorang konsumen yang membeli **Teh** punya kemungkinan 50% untuk juga membeli **Gula**. Aturan ini cukup signifikan karena mewakili 40% dari catatan transaksi selama ini."

METODOLOGI DASAR

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap:

1. Analisis pola frekuensi tinggi
2. Pembentukan aturan asosiasi

ANALISIS POLA FREKUENSI TINGGI

Tahap ini mencari kombinasi item yang memenuhi syarat minimum dari nilai *support* dalam *database*. Nilai *support* sebuah item diperoleh dengan rumus berikut:

$$\text{Support (A)} = \frac{\text{Jumlah Transaksi mengandung A}}{\text{Total Transaksi}}$$

Sementara nilai *support* dari 2 item diperoleh dengan rumus berikut:

$$\text{Support (A} \cap \text{B)} = \frac{\text{Jumlah Transaksi mengandung A dan B}}{\text{Total Transaksi}}$$

PEMBENTUKAN ATURAN ASOSIASI

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan assosiatif yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan assosiatif $A \rightarrow B$

Nilai *confidence* dari aturan $A \rightarrow B$ diperoleh dari rumus berikut:

$$\text{Confidence } P(B | A) = \frac{\text{Jumlah Transaksi mengandung } A \text{ dan } B}{\text{Jumlah Transaksi mengandung } A}$$

CONTOH DATA TRANSAKSI

Data Transaksi

Transaksi	Item yang dibeli
1	Susu, Teh, Gula
2	Teh, Gula, Roti
3	Teh, Gula
4	Susu, Roti
5	Susu, Gula, Roti
6	Teh, Gula
7	Gula, Kopi, Susu
8	Gula, Kopi, Susu
9	Susu, Roti, Kopi
10	Gula Teh, Kopi

Daftar Item yang dibeli: **Teh, Gula, Kopi, Susu, Roti**

DAFTAR ISTILAH

I adalah himpunan yang tengah dibicarakan.

Contoh: $I = \{\text{Teh, Gula, Kopi, Susu, Roti}\}$

D adalah Himpunan seluruh transaksi yang tengah dibicarakan

Contoh: $D = \{\text{Transaksi 1, Transaksi 2, ..., Transaksi 10}\}$

DAFTAR ISTILAH

Proper Subset adalah himpunan bagian murni

Contoh:

Ada suatu himpunan $I = \{a, b, c\}$

Himpunan Bagian dari I adalah

Himpunan Kosong = $\{\}$

Himpunan 1 Unsur = $\{a\}, \{b\}, \{c\}$

Himpunan 2 Unsur = $\{a, b\}, \{a, c\}, \{b, c\}$

Himpunan 3 Unsur = $\{a, b, c\}$

Proper subset nya adalah Himpunan 1 Unsur dan Himpunan 2 Unsur

DAFTAR ISTILAH

Item Set adalah himpunan item atau item-item di **I**

Contoh:

Ada suatu himpunan $I = \{a,b,c\}$

Item set nya adalah

$\{a\}$; $\{b\}$; $\{c\}$; $\{a,b\}$; $\{a,c\}$; $\{b,c\}$

K-item set adalah Item set yang terdiri dari **K** buah item yang ada pada **I**. Intinya **K** itu adalah jumlah unsur yang terdapat pada suatu himpunan.

Contoh:

2-item set adalah yang bersifat 2 unsur.

DAFTAR ISTILAH

Item Set Frekuensi adalah Jumlah transaksi di **I** yang mengandung jumlah item set tertentu.

Intinya jumlah transaksi yang membeli suatu item set.

Contoh:

Kita gunakan tabel transaksi di atas:

Frekuensi Item set yang sekaligus membeli Teh dan Gula adalah 4

DAFTAR ISTILAH

Frekuen Item Set adalah item set yang muncul sekurang-kurangnya “sekian” kali di D. Kata “sekian” biasanya di simbolkan dengan Φ . Φ merupakan batas minimum dalam suatu transaksi.

Contoh:

Pertama kita tentukan $\Phi = 3$, karena jika tidak di tentukan maka maka frekuen item set tidak dapat di hitung. Jika $\Phi=3$ untuk {Teh, Gula} apakah frekuen Item set?

Jika kita hitung maka jumlah transaksi yang membeli Teh sekaligus membeli Gula adalah 4.

Karena $4 \geq 3$ maka {Teh, Gula} merupakan Frekuen Item set.

Fk adalah Himpunan semua frekuen Item Set yang terdiri dari K item.

TAHAPAN ALGORITMA

1. Tentukan Φ
2. Tentukan semua Frekuen Item Set
3. Untuk setiap Frekuen Item Set lakukan hal sbb:
 - i. Ambil sebuah unsur, namakanlah s
 - ii. Untuk sisanya namakanlah $ss-s$
 - iii. Masukkan unsur-unsur yang telah di umpamakan ke dalam *rule If (ss-s) then s*
(untuk langkah ke 3 lakukan untuk semua unsur)

CONTOH DATA TRANSAKSI

Transaksi	Item yang dibeli
1	A, B, C
2	D, A, B
3	A, B
4	C, D
5	C, B, D
6	C, B, A
7	D, A
8	B, D
9	B, C, D
10	A, C, D

Nilai minimum frekuensi kemunculan $\Phi = 4$ dan nilai minimal confidence = 65%

FORMAT TABULAR DATA TRANSAKSI (1 ITEM SET / 1 UNSUR / F1)

Transaksi	A	B	C	D
1	1	1	1	0
2	1	1	0	1
3	1	1	0	0
4	0	0	1	1
5	0	1	1	1
6	1	1	1	0
7	1	0	0	1
8	0	1	0	1
9	0	1	1	1
10	1	0	1	1
	6	7	6	7

2-ITEM SET / 2 UNSUR / F2

Transaksi	A	B	f
1	1	1	P
2	1	1	P
3	1	1	P
4	0	0	S
5	0	1	S
6	1	1	P
7	1	0	S
8	0	1	S
9	0	1	S
10	1	0	S
	6	7	4

Transaksi	A	C	f
1	1	1	P
2	1	0	S
3	1	0	S
4	0	1	S
5	0	1	S
6	1	1	P
7	1	0	S
8	0	0	S
9	0	1	S
10	1	1	P
	6	6	3

Transaksi	A	D	f
1	1	0	S
2	1	1	P
3	1	0	S
4	0	1	S
5	0	1	S
6	1	0	S
7	1	1	P
8	0	1	S
9	0	1	S
10	1	1	P
	6	7	3

Transaksi	B	C	f
1	1	1	P
2	1	0	S
3	1	0	S
4	0	1	S
5	1	1	P
6	1	1	P
7	0	0	S
8	1	0	S
9	1	1	P
10	0	1	S
	7	6	4

Transaksi	B	D	f
1	1	0	S
2	1	1	P
3	1	0	S
4	0	1	S
5	1	1	P
6	1	0	S
7	0	1	S
8	1	1	P
9	1	1	P
10	0	1	S
	7	7	4

Transaksi	C	D	f
1	1	0	S
2	0	1	S
3	0	0	S
4	1	1	P
5	1	1	P
6	1	0	S
7	0	1	S
8	0	1	S
9	1	1	P
10	1	1	P
	6	7	4

3-ITEM SET / 3 UNSUR / F3

Transaksi	A	B	C	f
1	1	1	1	P
2	1	1	0	S
3	1	1	0	S
4	0	0	1	S
5	0	1	1	S
6	1	1	1	P
7	1	0	0	S
8	0	1	0	S
9	0	1	1	S
10	1	0	1	S
	6	7	6	2

Transaksi	B	C	D	f
1	1	1	0	S
2	1	0	1	S
3	1	0	0	S
4	0	1	1	S
5	1	1	1	P
6	1	1	0	S
7	0	0	1	S
8	1	0	1	S
9	1	1	1	P
10	0	1	1	S
	7	6	7	2

Transaksi	A	B	D	f
1	1	1	0	S
2	1	1	1	P
3	1	1	0	S
4	0	0	1	S
5	0	1	1	S
6	1	1	0	S
7	1	0	1	S
8	0	1	1	S
9	0	1	1	S
10	1	0	1	S
	6	7	7	1

RULE IF X THEN Y

- Rule yang dipakai adalah *if x then y*, dimana *x* adalah *antecedent* dan *y* adalah *consequent*.
- Berdasarkan *rule* tersebut, maka dibutuhkan 2 buah item yang mana salah satunya sebagai *antecedent* dan sisanya sebagai *consequent*.
- F1 tidak disertakan karena hanya terdiri dari 1 item saja.
- Untuk *antecedent* boleh lebih dari 1 unsur, sedangkan untuk *consequent* terdiri dari 1 unsur.

RULE IF X THEN Y

If x then y	Support	Confidence
A - B	40	66.67
B - A	40	57.14
B - C	40	57.14
C - B	40	66.67
B - D	40	57.14
D - B	40	57.14
C - D	40	66.67
D - C	40	57.14

LATIHAN 1

Nilai minimum frekuensi kemunculan $\Phi = 2$ dan nilai minimal confidence = 50%

Data Transaksi

Transaksi	Item yang dibeli
1	Susu, Teh, Gula
2	Teh, Gula, Roti
3	Teh, Gula
4	Susu, Roti
5	Susu, Gula, Roti
6	Teh, Gula
7	Gula, Kopi, Susu
8	Gula, Kopi, Susu
9	Susu, Roti, Kopi
10	Gula Teh, Kopi

Daftar Item yang dibeli: **Teh, Gula, Kopi, Susu, Roti**

LATIHAN 2

Nilai minimum frekuensi kemunculan $\Phi = 3$ dan nilai minimal confidence = 60%

Data Transaksi

Transaksi	Item yang dibeli
1	L, X
2	L, M, X
3	M, X
4	M, L, X
5	S, L, X
6	S, M, L
7	X, S, M
8	S, M
9	L, M, S
10	X, S

NEXT

TEKNIK KLASIFIKASI

