

## ASSOCIATION RULE (ALGORITMA A PRIORI)

Algoritma A Priori termasuk jenis aturan asosiasi pada data mining. Selain a priori, yang termasuk pada golongan ini adalah metode *generalized rule induction* dan *algoritma hash based*. Aturan yang menyatakan asosiasi antara beberapa atribut sering disebut affinity analysis atau market basket analysis.

Analisis asosiasi atau *association rule mining* adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item. Contoh aturan asosiatif dari analisa pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan susu. Dengan pengetahuan tersebut pemilik pasar swalayan dapat mengatur penempatan barangnya atau merancang kampanye pemasaran dengan memakai kupon diskon untuk kombinasi barang tertentu.

Analisis asosiasi menjadi terkenal karena aplikasinya untuk menganalisa isi keranjang belanja di pasar swalayan. Analisis asosiasi juga sering disebut dengan istilah *market basket analysis*.

Analisis asosiasi dikenal juga sebagai salah satu teknik data mining yang menjadi dasar dari berbagai teknik data mining lainnya. Khususnya salah satu tahap dari analisis asosiasi yang disebut analisis pola frekuensi tinggi (*frequent pattern mining*) menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien.

Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, *support* (nilai penunjang) yaitu persentase kombinasi item tersebut dalam database dan *confidence* (nilai kepastian) yaitu kuatnya hubungan antar item dalam aturan asosiatif.

Aturan asosiatif biasanya dinyatakan dalam bentuk :  
 {roti, mentega} → {susu} (support = 40%, confidence = 50%)

Yang artinya : "50% dari transaksi di database yang memuat item roti dan mentega juga memuat item susu. Sedangkan 40% dari seluruh transaksi yang ada di database memuat ketiga item itu."

Dapat juga diartikan : "Seorang konsumen yang membeli roti dan mentega punya kemungkinan 50% untuk juga membeli susu. Aturan ini cukup signifikan karena mewakili 40% dari catatan transaksi selama ini."

Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan asosiatif yang memenuhi syarat minimum untuk *support* (*minimum support*) dan syarat minimum untuk *confidence* (*minimum confidence*).

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap :

### a. Analisa 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:

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

sedangkan nilai support dari 2 item diperoleh dari rumus berikut:

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

**b. Pembentukan aturan asosiatif**

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

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

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

Sebagai contoh ambil suatu data transaksi yang didapat dari penjualan sayur dengan data transaksi sebagai berikut :

Transaksi	Item yang di beli
1	Broccoli, Green Peppers, Corn
2	Asparagus, Squash, Corn
3	Corn, Tomatoes, Beans, Squash
4	Green Peppers, Corns, Tomatoes, Beans
5	Beans, Asparagus, Broccoli
6	Squash, Asparagus, Beans, Tomatoes
7	Tomatoes, corn
8	Broccoli, Tomatoes, Green Peppers
9	Squash, Asparagus, Beans
10	Beans, Corn
11	Green Peppers, Broccoli, Beans, Squash
12	Asparagus, Bean, Squash
13	Squash, Corn, Asparagus, Beans
14	Corn, Green Peppers, Tomatoes, Beans, Broccoli

I. Definisi-definisi yang terdapat pada *Association Rule*

1. I adalah himpunan yang tengah dibicarakan.  
Contoh:  
{Asparagus, Beans, ..., Tomatoes}
2. D adalah Himpunan seluruh transaksi yang tengah dibicarakan  
Contoh:  
{Transaksi 1, transaksi 2, ..., transaksi 14}
3. Proper Subset adalah Himpunan Bagian murni  
Contoh:  
Ada suatu himpunan  $A = \{a, b, c\}$   
Himpunan Bagian dari A 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
4. Item set adalah Himpunan item atau item-item di I  
Contoh:  
Ada suatu himpunan  $A = \{a, b, c\}$   
Item set nya adalah  
{a}; {b}; {c}; {a, b}; {a, c}; {b, c}

5. 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:  
 3-item set adalah yang bersifat 3 unsur
6. 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 penjualan sayur di atas
  - frekuensi Item set yang sekaligus membeli Beans dan Broccoli adalah 3
  - frekuensi item set yang sekaligus membeli Beans, Squash dan Tomatoes adalah 2
7. Frekuensi Item Set adalah item set yang muncul sekurang-kurangnya “sekian” kali di D. Kata “sekian” biasanya di simbolkan dengan  $\Phi$ .  $\Phi$  merupakan batas minimum dalam suatu transaksi  
 Contoh:  
 Pertama kita tentukan  $\Phi = 3$ , karena jika tidak di tentukan maka frekuensi item set tidak dapat di hitung.  
 Jika  $\Phi=3$  untuk {Asparagus, Beans} apakah frekuensi Item set?  
 Jika kita hitung maka jumlah transaksi yang membeli asparagus sekaligus membeli beans adalah 5.  
 Karena  $5 \geq 3$  maka {Asparagus, Beans} merupakan Frekuensi Item set.
8.  $F_k$  adalah Himpunan semua frekuensi Item Set yang terdiri dari K item.

II. Langkah-langkah algoritma pada **Association Rule**

1. Tentukan  $\Phi$
2. Tentukan semua Frekuensi Item set
3. Untuk setiap Frekuensi 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.

**Latihan.**

Ada transaksi seperti yang terlihat dibawah ini.

Transaksi	Item yang di beli
1	C, E, D
2	A, F, D
3	D, G, B, F
4	E, D, G, B
5	B, A, C
6	F, A, B, G
7	G, D
8	C, G, E
9	F, A, B
10	B, D

1. Pisahkan masing-masing item yang dibeli

Item yang di beli
A
B
C
D
E
F
G

2. Kemudian Buat Tabel seperti dibawah ini:

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

3. Kemudian hitung jumlah banyaknya pembelian untuk setiap item.

Transaksi	A	B	C	D	E	F	G
1	0	0	1	1	1	0	0
2	1	0	0	1	0	1	0
3	0	1	0	1	0	1	1
4	0	1	0	1	1	0	1
5	1	1	1	0	0	0	0
6	1	1	0	0	0	1	1
7	0	0	0	1	0	0	1
8	0	0	1	0	1	0	1
9	1	1	0	0	0	1	0
10	0	1	0	1	0	0	0
$\Sigma$	4	6	3	6	3	4	5

4. Tentukan  $\Phi$ .

Misalkan kita tentukan  $\Phi = 3$ , maka kita dapat menentukan frekuensi itemset. Dari tabel di atas diketahui total  $\Phi$  untuk transaksi  $k = 1$ , semuanya lebih besar dari  $\Phi$ . Maka:

$$F_1 = \{\{A\}, \{B\}, \{C\}, \{D\}, \{E\}, \{F\}, \{G\}\}$$

Untuk  $k = 2$  (2 unsur), diperlukan tabel untuk tiap-tiap pasang item. Himpunan yang mungkin terbentuk adalah:  $\{A,B\}, \{A,C\}, \{A,D\}, \{A,E\}, \{A,F\}, \{A,G\}, \{B,C\}, \{B,D\}, \{B,E\}, \{B,F\}, \{B,G\}, \{C,D\}, \{C,E\}, \{C,F\}, \{C,G\}, \{D,E\}, \{D,F\}, \{D,G\}, \{E,F\}, \{E,G\}, \{F,G\}$ .

Tabel-tabel untuk calon 2 item set:

T	A	B	f
1	0	0	S
2	1	0	S
3	0	1	S
4	0	1	S
5	1	1	P
6	1	1	P
7	0	0	S
8	0	0	S
9	1	1	P
10	0	1	S
		$\Sigma$	3

T	A	C	f
1	0	1	S
2	1	0	S
3	0	0	S
4	0	0	S
5	1	1	P
6	1	0	S
7	0	0	S
8	0	1	S
9	1	0	S
10	0	0	S
		$\Sigma$	1

T	A	D	f
1	0	1	S
2	1	1	P
3	0	1	S
4	0	1	S
5	1	0	S
6	1	0	S
7	0	1	S
8	0	0	S
9	1	0	S
10	0	1	S
		$\Sigma$	1

T	A	E	f
1	0	1	S
2	1	0	S
3	0	0	S
4	0	1	S
5	1	0	S
6	1	0	S
7	0	0	S
8	0	1	S
9	1	0	S
10	0	0	S
		$\Sigma$	0

T	A	F	f
1	0	0	S
2	1	1	P
3	0	1	S
4	0	0	S
5	1	0	S
6	1	1	P
7	0	0	S
8	0	0	S
9	1	1	P
10	0	0	S
		$\Sigma$	3

T	A	G	f
1	0	0	S
2	1	0	S
3	0	1	S
4	0	1	S
5	1	0	S
6	1	1	P
7	0	1	S
8	0	1	S
9	1	0	S
10	0	0	S
		$\Sigma$	1

T	B	C	f
1	0	1	S
2	0	0	S
3	1	0	S
4	1	0	S
5	1	1	P
6	1	0	S
7	0	0	S
8	0	1	S
9	1	0	S
10	1	0	S
		$\Sigma$	1

T	B	D	f
1	0	1	S
2	0	1	S
3	1	1	P
4	1	1	P
5	1	0	S
6	1	0	S
7	0	1	S
8	0	0	S
9	1	0	S
10	1	1	P
		$\Sigma$	3

T	B	E	f
1	0	1	S
2	0	0	S
3	1	0	S
4	1	1	P
5	1	0	S
6	1	0	S
7	0	0	S
8	0	1	S
9	1	0	S
10	1	0	S
		$\Sigma$	1

T	B	F	f
1	0	0	S
2	0	1	S
3	1	1	P
4	1	0	S
5	1	0	S
6	1	1	P
7	0	0	S
8	0	0	S
9	1	1	P
10	1	0	S
		$\Sigma$	3

T	B	G	f
1	0	0	S
2	0	0	S
3	1	1	P
4	1	1	P
5	1	0	S
6	1	1	P
7	0	1	S
8	0	1	S
9	1	0	S
10	1	0	S
		$\Sigma$	3

T	C	D	f
1	1	1	P
2	0	1	S
3	0	1	S
4	0	1	S
5	1	0	S
6	0	0	S
7	0	1	S
8	1	0	S
9	0	0	S
10	0	1	S
		$\Sigma$	1

T	C	E	f
1	1	1	P
2	0	0	S
3	0	0	S
4	0	1	S
5	1	0	S
6	0	0	S
7	0	0	S
8	1	1	P
9	0	0	S
10	0	0	S
		Σ	2

T	C	F	f
1	1	0	S
2	0	1	S
3	0	1	S
4	0	0	S
5	1	0	S
6	0	1	S
7	0	0	S
8	1	0	S
9	0	1	S
10	0	0	S
		Σ	0

T	C	G	f
1	1	0	S
2	0	0	S
3	0	1	S
4	0	1	S
5	1	0	S
6	0	1	S
7	0	1	S
8	1	1	P
9	0	0	S
10	0	0	S
		Σ	1

T	D	E	f
1	1	1	P
2	1	0	S
3	1	0	S
4	1	1	P
5	0	0	S
6	0	0	S
7	1	0	S
8	0	1	S
9	0	0	S
10	1	0	S
		Σ	2

T	D	F	f
1	1	0	S
2	1	1	P
3	1	1	P
4	1	0	S
5	0	0	S
6	0	1	S
7	1	0	S
8	0	0	S
9	0	1	S
10	1	0	S
		Σ	2

T	D	G	f
1	1	0	S
2	1	0	S
3	1	1	P
4	1	1	P
5	0	0	S
6	0	1	S
7	1	1	P
8	0	1	S
9	0	0	S
10	1	0	S
		Σ	3

T	E	F	f
1	1	0	S
2	0	1	S
3	0	1	S
4	1	0	S
5	0	0	S
6	0	1	S
7	0	0	S
8	1	0	S
9	0	1	S
10	0	0	S
		Σ	0

T	E	G	f
1	1	0	S
2	0	0	S
3	0	1	S
4	1	1	P
5	0	0	S
6	0	1	S
7	0	1	S
8	1	1	P
9	0	0	S
10	0	0	S
		Σ	2

T	F	G	f
1	0	0	S
2	1	0	S
3	1	1	P
4	0	1	S
5	0	0	S
6	1	1	P
7	0	1	S
8	0	1	S
9	1	0	S
10	0	0	S
		Σ	2

Dari tabel-tabel 2 unsur di atas, P artinya item-item yang dijual bersamaan, sedangkan S berarti tidak ada item yang dijual bersamaan atau tidak terjadi transaksi. Σ melambangkan jumlah Frekuensi item set. Jumlah frekuensi item set harus lebih besar atau sama dengan jumlah Frekuensi item set ( $\Sigma \geq \Phi$ ). Dari tabel diatas, maka didapat:

$$F_2 = \{\{A,B\}, \{A,F\}, \{B,D\}, \{B,F\}, \{B,G\}, \{D,G\}\}$$

Kombinasi dari itemset dalam F2, dapat kita gabungkan menjadi calon 3-itemset. Itemset-itemset yang dapat digabungkan adalah itemset-itemset yang memiliki kesamaan dalam k-1 item pertama. Misalnya {A,B} dan {A,F} memiliki itemset k-1 pertama yg sama, yaitu A, maka dapat digabungkan menjadi 3-itemset baru yaitu {A, B, F}.

Untuk k = 3 (3 unsur), himpunan yang mungkin terbentuk adalah: {A, B, F}, {B, D, F}, {B, D,G}, {B, F, G}

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

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

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

T	B	F	G	f
1	0	0	0	S
2	0	1	0	S
3	1	1	1	P
4	1	0	1	S
5	1	0	0	S
6	1	1	1	P
7	0	0	1	S
8	0	0	1	S
9	1	1	0	S
10	1	0	0	S
			Σ	2

Dari tabel-tabel di atas, didapat  $F_3 = \{ \}$ , karena tidak ada  $\Sigma \geq \Phi$  sehingga  $F_4, F_5, F_6$  dan  $F_7$  juga merupakan himpunan kosong.

- 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. Dari langkah 4 didapat 1 buah  $F_k$  yaitu  $F_2$ .  $F_1$  tidak disertakan karena hanya terdiri dari 1 item saja. Untuk antecedent boleh lebih dari 1 unsur, sedangkan untuk consequent terdiri dari 1 unsur.
- Tentukan (ss-s) sebagai antecedent dan s sebagai consequent dari  $F_k$  yang telah di dapat berdasarkan rule pada langkah 5.

Pada  $F_2$  didapat himpunan  $F_2 = \{ \{A,B\}, \{A,F\}, \{B,D\}, \{B,F\}, \{B,G\}, \{D,G\} \}$

Maka dapat disusun:

- Untuk  $\{A,B\}$ :
  - Jika (ss-s) = A, Jika s = B, Maka  $\rightarrow$  If buy A then buy B
  - Jika (ss-s) = B, Jika s = A, Maka  $\rightarrow$  If buy B then buy A
- Untuk  $\{A,F\}$ :
  - Jika (ss-s) = A, Jika s = F, Maka  $\rightarrow$  If buy A then buy F
  - Jika (ss-s) = F, Jika s = A, Maka  $\rightarrow$  If buy F then buy A
- Untuk  $\{B,D\}$ :
  - Jika (ss-s) = B, Jika s = D, Maka  $\rightarrow$  If buy B then buy D
  - Jika (ss-s) = D, Jika s = B, Maka  $\rightarrow$  If buy D then buy B
- Untuk  $\{B,F\}$ :
  - Jika (ss-s) = B, Jika s = F, Maka  $\rightarrow$  If buy B then buy F
  - Jika (ss-s) = F, Jika s = B, Maka  $\rightarrow$  If buy F then buy B
- Untuk  $\{B,G\}$ :
  - Jika (ss-s) = B, Jika s = G, Maka  $\rightarrow$  If buy B then buy G
  - Jika (ss-s) = G, Jika s = B, Maka  $\rightarrow$  If buy G then buy B

- Untuk  $\{D,G\}$ :
  - Jika  $(ss-s) = D$ , Jika  $s = G$ , Maka  $\rightarrow$  If buy D then buy G
  - Jika  $(ss-s) = G$ , Jika  $s = D$ , Maka  $\rightarrow$  If buy G then buy D

7. Dari langkah 6, kita mendapatkan 12 rule yang dapat digunakan yaitu

- If buy A then buy B
- If buy B then buy A
- If buy A then buy F
- If buy F then buy A
- If buy B then buy D
- If buy D then buy B
- If buy B then buy F
- If buy F then buy B
- If buy B then buy G
- If buy G then buy B
- If buy D then buy G
- If buy G then buy D

8. Dari langkah 7 dibuat tabel kandidat asosiasi rule untuk 1 antecedent

If antecedent then consequent	Support	Confidence
If buy A then buy B		
If buy B then buy A		
If buy A then buy F		
If buy F then buy A		
If buy B then buy D		
If buy D then buy B		
If buy B then buy F		
If buy F then buy B		
If buy B then buy G		
If buy G then buy B		
If buy D then buy G		
If buy G then buy D		

9. Hitung support dan confidence.

$$SUPPORT = \frac{\Sigma \text{item.yang.dibeli.sekaligus}}{\Sigma \text{jumlah.seluruh.transaksi}} \times 100\%$$

$$CONFIDENCE = \frac{\Sigma \text{item.yang.dibeli.sekaligus}}{\Sigma \text{jumlah.transaksi.pada.bagian.antecedent}} \times 100\%$$

Untuk  $\Sigma$  item yang dibeli sekaligus pada If buy A then buy B, ada 3 transaksi. Jumlah seluruh transaksi adalah 10 transaksi, sehingga support-nya adalah:

$$SUPPORT = \frac{3}{10} \times 100\% = 33,33\%$$

Untuk  $\Sigma$  item yang dibeli sekaligus pada If buy A then buy B, ada 3 transaksi, sedangkan jumlah transaksi yang membeli A adalah 4 transaksi, sehingga confidence-nya adalah:

$$CONFIDENCE = \frac{3}{4} \times 100\% = 75\%$$

Sehingga di dapat tabel sbb:

If antecedent then consequent	Support	Confidence
If buy A then buy B	$(3/10) \times 100\% = 33,33\%$	$(3/4) \times 100\% = 75\%$
If buy B then buy A	$(3/10) \times 100\% = 33,33\%$	$(3/6) \times 100\% = 50\%$
If buy A then buy F	$(3/10) \times 100\% = 33,33\%$	$(3/4) \times 100\% = 75\%$
If buy F then buy A	$(3/10) \times 100\% = 33,33\%$	$(3/4) \times 100\% = 75\%$
If buy B then buy D	$(3/10) \times 100\% = 33,33\%$	$(3/6) \times 100\% = 50\%$
If buy D then buy B	$(3/10) \times 100\% = 33,33\%$	$(3/6) \times 100\% = 50\%$
If buy B then buy F	$(3/10) \times 100\% = 33,33\%$	$(3/6) \times 100\% = 50\%$
If buy F then buy B	$(3/10) \times 100\% = 33,33\%$	$(3/4) \times 100\% = 75\%$
If buy B then buy G	$(3/10) \times 100\% = 33,33\%$	$(3/6) \times 100\% = 50\%$
If buy G then buy B	$(3/10) \times 100\% = 33,33\%$	$(3/5) \times 100\% = 60\%$
If buy D then buy G	$(3/10) \times 100\% = 33,33\%$	$(3/6) \times 100\% = 50\%$
If buy G then buy D	$(3/10) \times 100\% = 33,33\%$	$(3/5) \times 100\% = 60\%$

10. Setelah di dapat support dan confidence untuk masing-masing kandidat, lakukan perkalian antara support dan confidence, dimana confidence-nya diambil 70% ke atas, sehingga di dapat tabel sbb:

If antecedent then consequent	Support	Confidence	Support x confidence
If buy A then buy B	33,33%	75%	0.249975
If buy A then buy F	33,33%	75%	0.249975
If buy F then buy A	33,33%	75%	0.249975
If buy F then buy B	33,33%	75%	0.249975

11. Setelah didapat hasil perkalian antara support dan confidence, pilihlah yang hasil perkaliannya paling besar. Hasil paling besar dari perkalian perkalian tersebut merupakan rule yang dipakai pada saat menjual. Karena hasil perkalian dari ke-4 penjualan diatas bernilai sama, maka semuanya bisa dijadikan rule.

- Jika membeli A maka akan membeli B dengan support 33,33% dan confidence 75%
- Jika membeli A maka akan membeli F dengan support 33,33% dan confidence 75%
- Jika membeli F maka akan membeli A dengan support 33,33% dan confidence 75%
- Jika membeli F maka akan membeli B dengan support 33,33% dan confidence 75%