

# Algoritma Apriori

# Pengertian Algoritma Apriori

- ▶ Algoritma apriori merupakan salah satu algoritma klasik data mining. Algoritma apriori digunakan agar komputer dapat mempelajari aturan asosiasi, mencari pola hubungan antar satu atau lebih item dalam suatu dataset.
- ▶ Algoritma apriori banyak digunakan pada data transaksi atau biasa disebut market basket, misalnya sebuah swalayan memiliki market basket, dengan adanya algoritma apriori, pemilik swalayan dapat mengetahui pola pembelian seorang konsumen, jika seorang konsumen membeli item A , B, punya kemungkinan 50% dia akan membeli item C, pola ini sangat signifikan dengan adanya data transaksi selama ini.

- ❑ 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 :  $\{roti, mentega\} \rightarrow \{susu\}$  (support = 40%, confidence = 50%).

Yang artinya : “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).

Tetapi di lain pihak Apriori memiliki kelemahan karena harus melakukan scan database setiap kali iterasi, sehingga waktu yang diperlukan bertambah dengan makin banyak iterasi. Masalah ini yang dipecahkan oleh algoritma-algoritma baru seperti FP-growth.

## Contoh Pemakaian Hasil dari Mempelajari Aturan Asosiasi

1. Meletakkan barang-barang yang sering dibeli bersamaan dengan posisi berdekatan atau mudah dijangkau sehingga diharapkan pembeli membeli barang lebih banyak. Cara ini dikembangkan oleh Wal-Mart yang merupakan salah satu pasar swalayan populer di Amerika. Saat itu Wal-Mart menganalisis data yang dimilikinya, dan menemukan bahwa pada hari Jumat sore, pembeli laki-laki yang membeli popok, ternyata cenderung membeli bir. Dari hasil temuan tersebut, Wal-Mart menempatkan bir di dekat tempat penjualan popok, dan alhasil penjualan bir meningkat. Kasus ini menjadi terkenal, karena sebelumnya banyak yang tidak menduga akan ampuhnya data mining.
2. Amazon.com, mengembangkan perekomendasi (recommender), yaitu sebuah program untuk merekomendasikan barang-barang lain kepada pembeli pada saat pembeli melakukan browsing atau membeli suatu barang berdasarkan tingkat keyakinan (confidence).
3. Google mengembangkan fitur auto-complete, yaitu saat pemakai mengetikkan suatu kata, program akan menampilkan daftar kata-kata berikutnya, yang paling banyak memiliki asosiasi pada kata yang diketik

# Definisi-Definisi yang Terdapat Pada Association Rule

1. E adalah himpunan yang tengah dibicarakan. Contoh:  $\{Asparagus, Beans, \dots, Tomatoes\}$
2. D adalah Himpunan seluruh transaksi yang tengah dibicarakan. Contoh:  $\{Transaksi\ 1, transaksi\ 2, \dots, transaksi\ 14\}$
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 E. Contoh:  
Ada suatu himpunan  $E=\{a,b,c,\}$ , 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 E. Intinya K itu adalah jumlah unsur yang terdapat pada suatu Himpunan. Contoh: 2-item set adalah yang bersifat 2 unsur  $\{a,b\},\{a,c\},\{b,c\}$ .
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 (minimum support). Kata “sekian” biasanya di simbolkan dengan  $\Phi$ .  $\Phi$  merupakan batas minimum dalam suatu transaksi.
8. Fk atau K-itemset adalah himpunan semua frekuensi Item Set yang terdiri dari K item, misalnya beras,telur,minyak adalah 3-itemset (Dinotasikan sebagai K-itemset)

# Langkah atau Cara Kerja Apriori

1. Tentukan minimum support.
2. Iterasi 1 : hitung item-item dari support(transaksi yang memuat seluruh item) dengan men-scan database untuk 1-itemset, setelah 1-itemset didapatkan, dari 1-itemset apakah diatas minimum support, apabila telah memenuhi minimum support, 1-itemset tersebut akan menjadi pola frequent tinggi.
3. Iterasi 2 : untuk mendapatkan 2-itemset, harus dilakukan kombinasi dari k-itemset sebelumnya, kemudian scan database lagi untuk hitung item-item yang memuat support. itemset yang memenuhi minimum support akan dipilih sebagai pola frequent tinggi dari kandidat.
4. Tetapkan nilai k-itemset dari support yang telah memenuhi minimum support dari k-itemset.
5. Lakukan proses untuk iterasi selanjutnya hingga tidak ada lagi k-itemset yang memenuhi minimum support.

## Formula Pencarian Nilai Support & Confidence

- 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:

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

- 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:

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

# Latihan Soal



## Latihan Soal 1

ID Transaksi	Barang yang Dibeli
T1	{Mango, Onion, Nintendo, Key-chain, Eggs, Yo-yo}
T2	{Doll, Onion, Nintendo, Key-chain, Eggs, Yo-yo}
T3	{Mango, Apple, Key-chain, Eggs}
T4	{Mango, Umbrella, Corn, Key-chain, Yo-yo}
T5	{Corn, Onion, Onion, Key-chain, Ice-cream, Eggs}

Golden rule (threshold) yang digunakan adalah : 60% atau barang yang dibeli paling sedikit 3. Untuk mempermudah, nama-nama item di Tabel 1, disingkat dengan diambil huruf awalnya saja, sebagai contoh :

*M = Mango*

*O = Onion*

*Dan sebagainya.*

ID Transaksi	Barang yang Dibeli
T1	{M, O, N, K, E, Y }
T2	{D, O, N, K, E, Y }
T3	{M, A, K, E}
T4	{M, U, C, K, Y }
T5	{C, O, O, K, I, E}

## Latihan Soal 1

### Langkah ke-1 :

Hitung banyaknya transaksi untuk setiap item. Hati-hati, untuk item O (Onion) dibeli sebanyak 4 biji, namun pembelian O hanya ada 3 transaksi.

Item	Banyaknya Transaksi
M	3
O	3
N	2
K	5
E	4
Y	3
D	1
A	1
U	1
C	2
I	1

## Latihan Soal 1

### Langkah ke-2 :

Berdasarkan golden rule yang telah disebutkan di atas, saring data pada Tabel 3, hanya memilih item yang memiliki transaksi minimal sebanyak 3 transaksi. Item yang banyaknya transaksi kurang dari 3, dibuang.

Item	Banyaknya Transaksi
M	3
O	3
K	5
E	4
Y	3

### Langkah ke-3 :

Buat pasangan item dimulai dari item pertama, yaitu MO, MK, ME, MY. Kemudian dilanjutkan dengan item kedua. Misalnya OK, OE, OY. Perhatikan, OM tidak dibuat karena OM = MO (pasangan yang dibuat dari item pertama).

Pasangan Item
MO
MK
ME
MY
OK
OE
OY
KE
KY
EY

## Latihan Soal 1

### Langkah ke-4 :

Hitung berapa kali suatu pasangan item dibeli bersamaan. Contohnya pasangan MO dibeli secara bersamaan dalam itemset {M, O, N, K, E, Y}. Pasangan MK dibeli bersamaan sebanyak 3 kali yaitu di dalam {M,O,N,K,E,Y}, {M,A,K,E}, dan {M,U,C, K, Y}.

Pasangan Item	Banyaknya Transaksi
MO	1
MK	3
ME	2
MY	2
OK	3
OE	3
OY	2
KE	4
KY	3
EY	2

### Langkah ke-5 :

Gunakan golden rule, hapus semua pasangan item yang banyaknya transaksi kurang dari tiga.

Pasangan Item	Banyaknya Transaksi
MK	3
OK	3
OE	3
KE	4
KY	3

## Latihan Soal 1

### Langkah ke-6 :

Buat pasangan tiga item dengan aturan menggunakan pasangan item pada Tabel 7 yang memiliki huruf awal yang sama yaitu :

- OK dan OE, menjadi OKE
- KE dan KY, menjadi KEY

Kemudian hitung ada beberapa banyaknya transaksi dari pasangan tiga item berdasarkan Tabel 3.

Pasangan Item	Banyaknya Transaksi
OKE	3
KEY	2

Dalam langkah ini, misalkan ada tiga pasangan item ABC, ABD, ACD, ACE, BCD dan akan dibuatkan pasangan 4 item, carilah 2 huruf awal yang sama. Contoh :

- **ABC** dan **ABD**, menjadi ABCD
- **ACD** dan **ACE**, menjadi ACDE

Dan seterusnya. Secara umum, carilah pasangan item yang huruf (item) terakhirnya berbeda.

## Latihan Soal 1

### Langkah ke-7 :

Gunakan lagi golden rule, dengan membuang pasangan tiga item yang banyaknya transaksi kurang dari 3. Hasilnya tinggal OKE karena KEY hanya dibeli bersamaan dua kali.

**Kesimpulan :** Tiga item yang sering dibeli bersamaan adalah O, K, dan E.

## Latihan Soal 1

Untuk mengetahui tingkat keyakinan (confidence), frequent itemset (yaitu {O, K, E}) dapat digunakan untuk mencari aturan-aturan asosiasi antar item di dalam frequent itemset tersebut. Caranya adalah :

1. Itemset dibuatkan himpunan bagiannya. Hasilnya seperti pada contoh di bawah :

- {O}
- {K}
- {E}
- {O, K}
- {K, E}
- {O, E}

2. Cari asosiasi pada semua himpunan bagian yang telah dibuat, misal : {O} => {K, E} artinya : jika O dibeli, bagaimana kemungkinan K dan E akan dibeli pada transaksi yang sama. O dibeli pada 3 transaksi dan di dalam 3 transaksi tersebut, K dan E juga dibeli. Maka keyakinannya adalah :

{O} => {K, E} : keyakinannya adalah  $\frac{3}{3} \times 100\% = 100\%$ .

{K} => {O,E} : keyakinannya adalah  $\frac{3}{5} \times 100\% = 60\%$

{E} => {O,K} : keyakinannya adalah  $\frac{3}{4} \times 100\% = 75\%$

{K,E} => {O} : keyakinannya adalah  $\frac{3}{3} \times 100\% = 100\%$

{O,E} => {K} : keyakinannya adalah  $\frac{3}{3} \times 100\% = 100\%$

{O,K} => {E} : keyakinannya adalah  $\frac{3}{4} \times 100\% = 100\%$