

ANALISA ASOSIASI

DATA MINING

Analisis Asosiasi (Association rule mining)

- Aturan asosiasi merupakan studi tentang hubungan dua item barang atau lebih, Aturan asosiasi ingin memberikan informasi hubungan tersebut dalam bentuk hubungan **IF THEN** dan aturan ini dihitung dari data yang sifatnya *probabilistic* (Santoso, 2007).

Support

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

sedangkan nilai support dari 2 item diperoleh dari rumus berikut:

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

Market Basket Analysis

- Market Basket Analysis* merupakan sebuah teknik datamining untuk melakukan analisis terhadap data pada bidang retail dan pemasaran (vats, 2015).
- Tujuan dari *Market Basket Analysis* adalah untuk memahami kebiasaan pembelian yang dilakukan pelanggan dan untuk menentukan produk apa yang akan dibeli pelanggan secara bersamaan

Aturan Asosiasi

- Aturan Asosiasi adalah pernyataan implikasi bentuk $X \rightarrow Y$, dimana X dan Y adalah itemset yang bersifat lepas/*disjoint*, yang memenuhi persyaratan $X \cap Y = \{\}$.
- Kekuatan aturan asosiasi dapat diukur dengan *support* dan *confidence*. *Support* digunakan untuk menentukan seberapa banyak aturan dapat diterapkan pada set data, sedangkan *confidence* digunakan untuk menentukan seberapa sering item di dalam Y muncul dalam transaksi yang berisi X.

Confidence

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

Contoh

TX	Item
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

TX	Teh	Gula	Kopi	Susu	Roti
1	1	1	0	1	0
2	1	1	0	0	1
3	1	1	0	0	0
4	0	0	0	1	1
5	0	1	0	1	1
6	1	1	0	0	0
7	0	1	1	1	0
8	0	1	1	1	0
9	0	0	1	1	1
10	1	1	1	0	0

Algoritma Apriori

Berusaha untuk menemukan jumlah item-set frekuensi secara efisien, yang menggunakan prinsip jika sebuah item-set itu frekuensi maka semua subset dari itemset tersebut adalah juga frekuensi.

1. Pembangkitan Itemset
2. Pembangkitan Association Rule

1. Pembangkitan Item Set

Frequent 1-Itemset (F1)

TX	Teh	Gula	Kopi	Susu	Roti
1	1	1	0	1	0
2	1	1	0	0	1
3	1	1	0	0	0
4	0	0	0	1	1
5	0	1	0	1	1
6	1	1	0	0	0
7	0	1	1	1	0
8	0	1	1	1	0
9	0	0	1	1	1
10	1	1	1	0	0

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

sedangkan nilai support dari 2 item diperoleh dari rumus berikut:

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

Support:
 Teh : 5/10 = 50%
 Gula : 8/10 = 80%
 Kopi : 4/10 = 40%
 Susu : 6/10 = 60%
 Roti : 4/10 = 40%

Frequent Itemset dengan Nilai minimum ($\phi = 2$)

Frequent 2-Itemset (F2)

- Calon 2-Itemset

Kombinasi	Jumlah
Teh Gula	5
Teh Kopi	1
Teh Susu	1
Teh Roti	1
Gula Kopi	3
Gula Susu	4
Gula Roti	2
Kopi Susu	3
Kopi Roti	1
Susu Roti	3

- Frequent Itemset dengan Nilai minimum ($\phi = 2$) maka terbentuk 2-item set

{(Teh,Gula),(Gula,Kopi),(Gula,Susu),(Gula,Roti),(Kopi,Susu),(susu,Roti)}

Kombinasi	Jumlah
Teh Kopi	1
Teh Susu	1
Teh Roti	1
Gula Kopi	3
Gula Susu	4
Gula Roti	2
Kopi Susu	3
Kopi Roti	1
Susu Roti	3

Frequent 3-Itemset (F3)

- Kombinasi dari F2 dapat digabung menjadi calon 3-itemset
- Itemset dari F2 yang dapat digabung adalah itemset yang memiliki kesamaan dalam k-1 item pertama.

Kombinasi	Jumlah
Gula Teh Kopi	1
Gula Teh Susu	1
Gula Kopi Susu	2
Susu Kopi Roti	1
Susu Roti Gula	1
Gula Teh Roti	1
Gula Kopi Roti	0
Gula Susu Roti	3

- Sehingga dengan ($\phi = 2$) F3 adalah {gula,kopi,susu}

•Pembangkitan Asosiasi Rule

Rule - Asosiasi F3

Kombinasi	Jumlah	Kombinasi	Jumlah
Gula Teh Kopi	1	Teh Gula Susu	5
Gula Teh Susu	1	Teh Kopi Roti	1
Gula Kopi Susu	2	Teh Susu Roti	1
Susu Kopi Roti	1	Gula Kopi Roti	1
Susu Roti Gula	1	Gula Susu Roti	4
Gula Teh Roti	1	Kopi Susu Roti	3
Gula Kopi Roti	0	Kopi Roti Susu	1
Gula Susu Roti	1	Susu Roti	3

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

Aturan	Support	Confidence
IF Gula,Kopi THEN Susu	2/3	67%
IF Gula,Susu THEN Kopi	2/4	50%
IF Kopi,Susu THEN Gula	2/3	67%

Jika ditetapkan minimum confidence sebesar 60% maka RuleF3 yang terbentuk adalah
IF Gula,Susu THEN Kopi
IF Kopi,Susu THEN Gula

Rule-F2

Kombinasi	Jumlah	Tx	Teh	Gula	Kopi	Susu	Roti
Teh Gula	5	1	1	1	0	1	0
Teh Kopi	1	2	1	1	0	0	1
Teh Susu	1	3	1	1	0	0	0
Teh Roti	1	4	0	0	0	1	1
Gula Kopi	3	5	0	1	0	1	1
Gula Susu	4	6	1	1	0	1	1
Gula Roti	2	7	0	1	1	1	0
Kopi Susu	3	8	0	1	1	1	0
Kopi Roti	1	9	0	0	1	1	1
Susu Roti	3	10	1	1	1	0	0
		5	8	4	0	4	4

Aturan	Support	Confidence
IF Teh THEN Gula	5/5	100%
IF Gula THEN Teh	5/8	62.5%
IF Gula THEN Kopi	3/8	37.5%
IF Kopi THEN Gula	3/4	75%
IF Gula THEN Susu	4/8	50%
IF Susu THEN Gula	4/6	67%
IF Kopi THEN Susu	3/4	75%
IF Susu THEN Kopi	3/6	50%
IF Susu THEN Roti	3/6	50%
IF Roti THEN Susu	3/4	75%

Aturan Asosiasi Final F2

Kombinasi	Jumlah	Tx	Teh	Gula	Kopi	Susu	Roti
Teh Gula	5	1	1	1	0	1	0
Teh Kopi	1	2	1	1	0	0	1
Teh Susu	1	3	1	1	0	0	0
Teh Roti	1	4	0	0	0	1	1
Gula Kopi	3	5	0	1	0	1	1
Gula Susu	4	6	1	1	0	1	1
Gula Roti	2	7	0	1	1	1	0
Kopi Susu	3	8	0	1	1	1	0
Kopi Roti	1	9	0	0	1	1	1
Susu Roti	3	10	1	1	1	0	0
		4	8	4	4	4	4

Aturan	Support	Confidence	Total
IF Teh THEN Gula	5/10 50%	5/5 100%	50%
IF Gula THEN Teh	5/10 50%	5/8 62.5%	31%
IF Gula THEN Kopi	3/10 30%	3/8 37.5%	11.25%
IF Kopi THEN Gula	3/10 30%	3/4 75%	22.5%
IF Gula THEN Susu	4/10 40%	4/8 50%	20%
IF Susu THEN Gula	4/10 40%	4/6 67%	26.8%
IF Kopi THEN Susu	3/10 30%	3/4 75%	22.5%
IF Susu THEN Kopi	3/10 30%	3/6 50%	15%
IF Susu THEN Roti	3/10 30%	3/6 50%	15%
IF Roti THEN Susu	3/10 30%	3/4 75%	22.5%

Algoritma apriori

Langkah –langkah dalam Algoritma apriori :

- Membaca semua transaksi sebanyak satu kali untuk menentukan support setiap item. Selanjutnya dipilih 1- itemset yang *frequent*.
- Membangkitkan kandidat *k-itemset* menggunakan $(k-1)$ -*itemset* yang *frequent*.
- Menghitung support count dari setiap kandidat
- Membuang kandidat *itemset* yang memiliki nilai *support-count* kurang dari *min-support*.
- Jika tidak ada lagi *itemset* yang *frequent* maka selesai.

Algoritma Apriori dlm pseudocode

```

//pembangkitan itemset
begin
K=1
Fk={|I|e^alpha(i)}>=Nxminsups //mencari F1
Do while Fk!=empty
k=k+1
Ck=apriori_gen(Fk) // membangkitkan kandidat itemset
For Each t in transaksi T
Ct=subset(Ck,t) // identifikasi semua kandidat yg dimiliki t
for each c in Kandidat
alpha(c)=alpha(c)+1 // naikkan support count
endifor
endifor
Fk={c|cCeK^alpha(c)}>=Nxminsups //pilih k-itemset yang frequent
endWhile
Hasil=Ufk
end
    
```

Procedure

```

Procedure ap_genrule(Fk,Hm)
begin
K=|Fk| ukuran itemset frequent
M=|Hm| ukuran consequent aturan
if k>m+1 then
Hm+1=apriori_gen(Hm)
For Each hm+1 in Hm+1 do
conf=σ(Fk)/σ(Fk-hm+1)
if conf>=minconf then
output aturan(Fk-hm+1) → hm+1
else
hapus hm+1 dari Hm+1
endif
endfor
call ap_genrule(Fk,Hm+1)
Endif
end

apriori_gen(Fk)
begin
For each k-itemset k>=2 do
Hi={i|ieFk} //aturan consequent 1-
item
call ap_genrule(Fk,Hi)
Endfor
end

```

Tugas

- Buatlah aturan asosiasi dengan Nilai minimum ($\phi = 2$) dan nilai minimum confidence 60%

No	Transaksi			
1	Telur	Beras		
2	Telur	Beras	Minyak	Topi
3	Telur	Beras	Minyak	Buku
4	Buku	Minyak	Topi	
5	Telur	Beras	Minyak	Topi
6	Telur	Beras	Minyak	
7	Telur	Minyak	Buku	
8	Beras	Buku		