



---

# *ASSOCIATION RULE*



# Association rule?

---

- Mencari suatu kaidah keterhubungan dari data
- Diusulkan oleh Agrawal, Imielinski, and Swami (1993)



# Contoh

---

Dalam suatu supermarket kita ingin mengetahui seberapa jauh orang yang membeli celana juga membeli sabuk?



# ASSOCIATION RULE

---

- *Association rule* atau analisis asosiasi adalah teknik data mining untuk menemukan aturan asosiasi antara suatu kombinasi item.
- Contoh aplikasi dari analisis asosiasi adalah menganalisa isi keranjang belanja di pasar swalayan. Analisis asosiasi juga sering disebut dengan istilah *market basket analysis*.
- Terdapat dua parameter yang digunakan untuk mengukur aturan asosiatif yaitu:
  - *Support* (nilai penunjang) yaitu persentase kombinasi item tersebut dalam database
  - *Confidance* (nilai kepastian), yaitu kuatnya hubungan antar item dalam aturan asosiatif.



# Input & problema

---

- Input
  - Adanya sejumlah transaksi
  - Setiap transaksi memuat kumpulan item
- Problema
  - Bagaimana caranya menemukan association rule yang memenuhi minimum support dan minimum confidence yang kita berikan

# Manfaat

- Dapat digunakan untuk Market Basket Analysis (menganalisa kebiasaan customer dengan mencari asosiasi dan korelasi dari data transaksi)
  - Sebagai saran penempatan barang dalam supermarket
  - Sebagai saran produk apa yang dipakai dalam promosi





# Defenisi Frequent Item Set

- **Itemset**
  - A set of one or more items
    - E.g.: {Milk, Bread, Diaper}
  - *k*-itemset
    - An itemset that contains *k* items
- **Support count ( $\sigma$ )**
  - Frequency of occurrence of an itemset (number of transactions it appears)
  - E.g.  $\sigma(\{\text{Milk, Bread, Diaper}\}) = 2$
- **Support**
  - Fraction of the transactions in which an itemset appears
  - E.g.  $s(\{\text{Milk, Bread, Diaper}\}) = 2/5$
- **Frequent Itemset**
  - An itemset whose support is greater than or equal to a *minsup* threshold

<i>TID</i>	<i>Items</i>
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke



# ANALISIS ATURAN ASOSIASI

## Market-Basket transactions

<i>TID</i>	<i>Items</i>
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

## Examples of association rules

$\{\text{Diaper}\} \rightarrow \{\text{Beer}\},$   
 $\{\text{Milk, Bread}\} \rightarrow \{\text{Diaper, Coke}\},$   
 $\{\text{Beer, Bread}\} \rightarrow \{\text{Milk}\},$





# PARAMETER ATURAN ASOSIASI

- *Support*(s) adalah pembagian dari transaksi yang mengandung nilai A dengan total transaksi.

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

*Support* dari dua item diperoleh dari rumus berikut:

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



# PARAMETER ATURAN ASOSIASI

- *Confidance (c)* merupakan ukuran seberapa sering item A muncul di transaksi yang mengandung item A'
- *Confidance* dihitung setelah support ditentukan.
- *Nilai confidance* dari aturan A-> B diperoleh dari rumus berikut:

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

# CONTOH

<i>TID</i>	<i>Items</i>
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

Contoh:

{ Milk, Diaper } -> Beer

$$s = \frac{\sigma(\{\text{Milk, Diaper, Beer}\})}{|T|} = \frac{2}{5} = 0.4$$

$$c = \frac{\sigma(\{\text{Milk, Diaper, Beer}\})}{\sigma(\{\text{Milk, Diaper}\})} = \frac{2}{3} = 0.67$$

Artinya :67% dari transaksi di database yang memuat item milk dan diaper juga memuat beer, sedangkan 40% dari seluruh transaksi memuat ketiga item tersebut

# Contoh

Transaks i	A	B	C	D
T1	1	0	1	14
T2	0	0	6	0
T3	1	0	2	4
T4	0	0	4	0
T5	0	0	3	1
T6	0	0	1	13
T7	0	0	8	0
T8	4	0	0	7
T9	0	1	1	10
T10	0	0	0	18

Jumlah transaksi  $|D| = 10$

Kemunculan item A pada transaksi ( $|T_a|$ ) sebanyak 3 kali yaitu pada T1, T3, T8.

$Supp(A) = |T_a| / |D| = 3/10 = 0.3$ .

$|T_{cd}|$  sebanyak 5 kali, yaitu pada T1, T3, T5, T6, T9.

$Supp(CD) = |T_{cd}| / |D| = 5/10 = 0.5$ .

Frequent itemset adalah itemset yang mempunyai **support  $\geq$  minimum support** yang diberikan oleh user.



# TEKNIK ANALISIS ASOSIASI

---

- **Algoritme Apriori**
- Algoritme FP-Growth



# ALGORITME APRIORI

---

- Algoritme apriori adalah suatu metode untuk mencari pola hubungan antara satu atau lebih item dalam suatu dataset.
- Algoritme ini banyak digunakan pada data transaksi atau biasa disebut *market basket*.
- Dengan adanya algoritma apriori, pemilik swalayan dapat mengetahui pola pembelian seorang konsumen.



# CARA KERJA ALGORITME APRIORI

1. 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,
2. Tentukan minimum support
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.



# Contoh Soal

Sebuah supermarket memiliki data transaksi sebagai berikut:

<u>Id Transaksi (Support)</u>	<u>Items</u>
1	<u>Pena,roti,mentega</u>
2	<u>Roti,mentega,telur,susu</u>
3	<u>Buncis,telur,susu</u>
4	<u>Roti,mentega</u>
5	<u>Roti,mentega,kecap,telur,su su</u>

Gambar 1 : Data Transaksi

minimum dari nilai support pola frekuensi tinggi adalah 2



- Iterasi 1

Untuk 1-itemset hitung dan scan database untuk mendapatkan pola frequent dari support

C1

<u>Itemset</u>	Support Count	Support
<u>Roti</u>	4	80%
<u>Mentega</u>	4	80%
<u>Telur</u>	3	60%
<u>Susu</u>	3	60%
<u>Buncis</u>	1	20%
<u>Kecap</u>	1	20%
<u>Pena</u>	1	20%

Gambar 2 : 1-itemset

F<sub>i</sub>

<u>Itemset</u>	Support Count	Support
<u>Roti</u>	80%	4
<u>Mentega</u>	80%	4
<u>Telur</u>	60%	3
<u>Susu</u>	60%	3

Gambar 3 : Pola Frequent

- Iterasi 2

Pada iterasi sebelumnya pola frequent dari support telah didapatkan dari 1-itemset, untuk 2-itemset, generate k-itemset dari k-itemset iterasi sebelumnya, dengan melakukan kombinasi dari k-itemset tersebut.

$C_2$

<u>Itemset</u>
<u>Roti, mentega</u>
<u>Roti, telur</u>
<u>Roti, susu</u>
<u>Mentega, telur</u>
<u>Mentega, susu</u>
<u>Telur, susu</u>

Gambar 4: kombinasi dari k-itemset

$C_2$

<u>Itemset</u>	Support Count	Support
<u>Roti, mentega</u>	4	80%
<u>Roti, telur</u>	2	40%
<u>Roti, susu</u>	2	40%
<u>Mentega, telur</u>	2	40%
<u>Mentega, susu</u>	2	40%
<u>Telur, susu</u>	3	60%

Gambar 5: 2-itemset



- Iterasi 2

Pada iterasi sebelumnya pola frequent dari support telah didapatkan dari 1-itemset, untuk 2-itemset, generate k-itemset dari k-itemset iterasi sebelumnya, dengan melakukan kombinasi dari k-itemset tersebut.

$F_2$

<u>Itemset</u>	Support Count	Support
<u>Roti,mentega</u>	4	80%
<u>Roti,telur</u>	2	40%
<u>Roti,susu</u>	2	40%
<u>Mentega,telur</u>	2	40%
<u>Mentega,susu</u>	2	40%
<u>Telur,susu</u>	3	60%

Gambar 6: Pola frequent tinggi



- Iterasi 3

$C_3$

<u>Itemset</u>
<u>Roti,mentega,telur</u>
<u>Roti,mentega,susu</u>
<u>Mentega,telur,susu</u>

Gambar 7: Kombinasi dari k-itemset

$C_3$

<u>Itemset</u>	Support Count	Support
<u>Roti,mentega,telur</u>	40%	2
<u>Roti,mentega,susu</u>	40%	2
<u>Mentega,telur,susu</u>	40%	2

Gambar 8: 3-itemset dari scan database



- Iterasi 3

$C_3$

<u>Itemset</u>
<u>Roti,mentega,telur</u>
<u>Roti,mentega,susu</u>
<u>Mentega,telur,susu</u>

Gambar 7: Kombinasi dari k-itemset

$C_3$

<u>Itemset</u>	<b>Support Count</b>	<b>Support</b>
<u>Roti,mentega,telur</u>	40%	2
<u>Roti,mentega,susu</u>	40%	2
<u>Mentega,telur,susu</u>	40%	2

Gambar 8: 3-itemset dari scan database

- 
- Iterasi 3

$F_3$

<u>Itemset</u>	<u>Support Count</u>	<u>Support</u>
<u>Roti, mentega, telur</u>	40%	2
<u>Roti, mentega, susu</u>	40%	2
<u>Mentega, telur, susu</u>	40%	2

Gambar 9: 3-itemset untuk pola frequent tinggi

- Iterasi 4

$C_4$

<u>Itemset</u>
<u>Roti,mentega,telur,susu</u>

Gambar 10: kombinasi 3-itemset untuk k-itemset

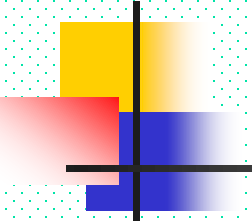
Scan database untuk mendapatkan itemset dari support, itemset yang memenuhi minimum support dipilih sebagai pola frequent tinggi

$F_4$

<u>Itemset</u>	<u>Support Count</u>	<u>Support</u>
<u>Roti,mentega,telur,susu</u>	40%	2

Gambar 11: pola frequent tinggi

Scan database untuk mendapatkan itemset dari support, itemset yang memenuhi minimum support dipilih sebagai pola frequent tinggi

- 
- 
- Iterasi 4
  - Tidak ada lagi kombinasi yang bisa dibentuk untuk k-itemset berikutnya, proses berhenti.
  - Pola frequent tinggi yang ditemukan adalah **“roti, mentega, telur,susu”**.





## Contoh Soal-2

---

- Bentuk association rules yang memenuhi syarat minimum dengan menghitung confidence association rules  $A \rightarrow B$ .

# Contoh Soal

<u>Aturan Asosiatif</u>	<u>Support {AUB}</u>	<u>Support {A}</u>	<u>Confidence</u>
{ <u>Roti,Mentega,Telur</u> } → { <u>susu</u> }	40%	60%	66%
{ <u>Susu,Mentega,Telur</u> } → { <u>Roti</u> }	40%	80%	50%
{ <u>Roti,Susu,Telur</u> } → { <u>Mentega</u> }	40%	80%	50%
{ <u>Susu,mentega,Roti</u> } → { <u>telur</u> }	40%	60%	66%
{ <u>susu,mentega</u> } → { <u>telur,roti</u> }	40%	40%	100%
{ <u>susu,telur</u> } → { <u>mentega,roti</u> }	60%	80%	75%
{ <u>susu,roti</u> } → { <u>mentega,susu</u> }	40%	40%	100%
{ <u>mentega,telur</u> } → { <u>susu,roti</u> }	40%	40%	100%
{ <u>mentega,roti</u> } → { <u>susu,telur</u> }	80%	60%	75%
{ <u>telur</u> } → { <u>mentega,susu</u> }	40%	40%	100%
{ <u>telur</u> } → { <u>mentega,roti</u> }	40%	80%	50%
{ <u>susu</u> } → { <u>mentega,roti</u> }	60%	80%	75%
{ <u>susu</u> } → { <u>mentega,telur</u> }	60%	40%	66%
{ <u>mentega</u> } → { <u>susu,telur</u> }	80%	60%	75%
{ <u>mentega</u> } → { <u>susu,roti</u> }	80%	40%	50%
{ <u>mentega</u> } → { <u>susu</u> }	80%	60%	75%
{ <u>mentega</u> } → { <u>roti</u> }	80%	80%	100%
{ <u>mentega</u> } → { <u>telur</u> }	80%	60%	75%
{ <u>susu</u> } → { <u>roti</u> }	60%	80%	75%
{ <u>susu</u> } → { <u>telur</u> }	60%	60%	100%
{ <u>telur</u> } → { <u>roti</u> }	60%	80%	75%

Gambar 11 : Aturan Asosiatif

# Contoh apriori dengan minimum support 50%

Database TDB

Tid	Items
10	A, C, D
20	B, C, E
30	A, B, C, E
40	B, E

1<sup>st</sup> scan

$C_1$

Itemset	sup
{A}	0.5
{B}	0.67
{C}	0.67
{D}	0.25
{E}	0.67

$L_1$

Itemset	sup
{A}	2
{B}	3
{C}	3
{E}	3

$L_2$

Itemset	sup
{A, C}	0.5
{B, C}	0.5
{B, E}	0.67
{C, E}	0.5

$C_2$

Itemset	sup
{A, B}	0.25
{A, C}	0.5
{A, E}	0.25
{B, C}	0.5
{B, E}	0.67
{C, E}	0.5

2<sup>nd</sup> scan

$C_2$

Itemset
{A, B}
{A, C}
{A, E}
{B, C}
{B, E}
{C, E}

$C_3$

Itemset
{B, C, E}

3<sup>rd</sup> scan

$L_3$

Itemset	sup
{B, C, E}	0.5



# Tahapan Algoritma FP Growth

---

1. Penyiapan Dataset
2. Pencarian Frequent Itemset (Item yang sering muncul)
3. Dataset diurutkan Berdasarkan Priority
4. Pembuatan FP-Tree Berdasarkan Item yang sudah diurutkan
5. Pembangkitan Conditional Pattern Base
6. Pembangkitan Conditional FP-tree
7. Pembangkitan Frequent Pattern
8. Mencari Support
9. Mencari Confidence

# 1. Penyiapan Dataset

Customer	Gula	Kopi	Aqua	Popok	Sprei	Sabun	Sampo	Kemeja	Celana	Boneka	
1	1	1	1	0	0	0	1	1	0	0	0
2	1	1	1	0	1	1	0	0	1	1	1
3	0	0	0	0	1	1	0	0	0	0	1
4	1	0	0	1	0	0	0	0	0	0	0
5	0	0	0	1	1	0	0	1	0	0	0
6	1	1	1	0	0	0	1	0	0	0	0
7	0	0	0	0	0	1	0	0	0	1	1
8	0	0	0	1	1	1	1	1	1	0	0
9	1	1	1	0	0	0	0	0	0	1	0
10	0	0	0	1	0	0	0	1	0	0	0
11	1	1	1	1	0	0	0	0	0	0	0
12	0	0	0	0	0	1	1	1	0	0	0
Frequent	6	5	5	4	5	4	5	2	3	3	
Priority	1	2	3	6	4	7	5	10	8	9	



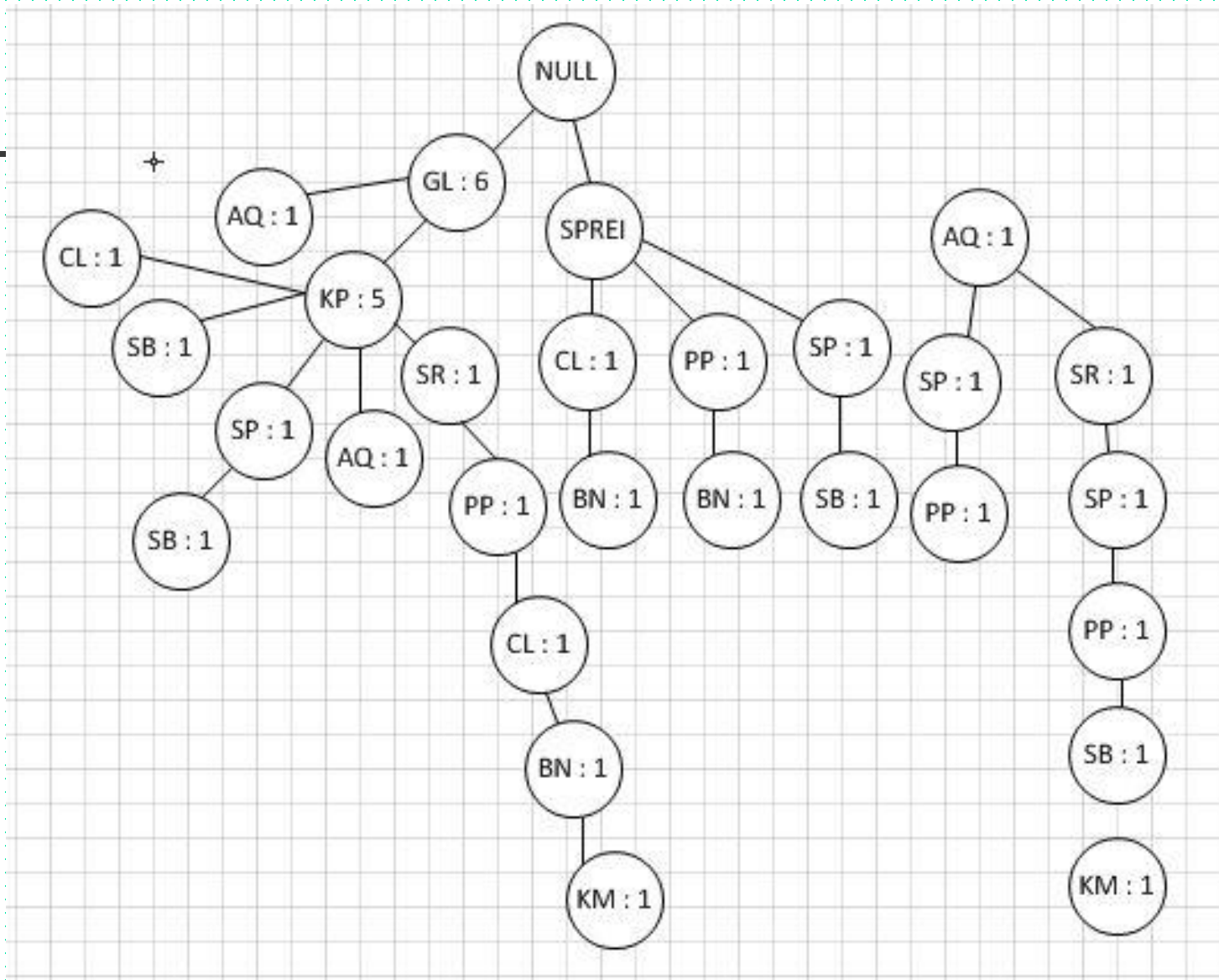
## 2. Pencarian Frequent Itemset

Frequent	
Gula	6
Kopi	5
Aqua	5
Popok	4
Sprei	5
Sabun	4
Sampo	5
Kemeja	2
Celana	3
Boneka	3

### 3. Dataset diurutkan Berdasarkan Priority

Customer	Item	Urutan
1	GL, KP, SB, SP	GL, KP, SP, SB
2	GL, KP, PP, SR, KM, CL, BN	GL, KP, SR, PP, CL, BN, KM
3	PP, SR, BN	SR, PP, BN
4	GL, AQ	GL, AQ
5	AQ, PP, SP	AQ, SP, PP
6	GL, KP, SB	GL, KP, SB
7	SR, CL, BN	SR, CL, BN
8	AQ, PP, SR, SB, SP, KM	AQ, SR, SP, PP, SB, KM
9	GL, KP, CL	GL, KP, CL
10	AQ, SP	AQ, SP
11	GL, KP, AQ	GL, KP, AQ
13	SR, SB, SP	SR, SP, SB

# 4. Pembuatan FP-Tree





## 5. Pembangkitan Conditional Pattern Base

Conditional Pattern Base
Kemeja : {{GL, KP, SR, PP, CL, BN :1}, {AQ, SR, SP, PP, SB :1}}
Boneka : {{GL, KP, SR, PP, CL :1}, {SR, CL:1}, {SR,PP}}
Celana : {{GL, KP, SR, PP:1}, {GL, KP :1}, {SR:1}}
Sabun : {{GL, KP, SP :1}, {GL, KP :1}, {AQ, SR, SP, PP :1}, {SR, SP :1}}
Popok : {{GL, KP, SR :1}, {SR :1}, {AQ, SP :1}, {AQ, SR, SP :1}}
Sampo : {{GL, KP :1}, {AQ :2}, {AQ, SR :1}, {SR :1}}
Sprei : {{GL, KP :1}, {AQ :1}}
Aqua : {{GL :1}, {GL, KP :1}}
Kopi : {GL : 5}

## 6. Pembangkitan Conditional FP-tree

Conditional FP Tree
Kemeja : (GL :1, KP : 1), (AQ : 1, SR :1)
Boneka : (GL :1, KP : 1), (SR :2)
Celana : (GL : 2, KP :2), (SR :1)
Sabun : (GL :2, KP :2), (SR : 1), (AQ :1)
Popok : (SR :1, AQ :2)
Sampo : (GL :1, KP :1),
Sprei : (GL : 1, KP :2), (SR :1)
Aqua : (GL :2, KP :2)
Kopi : (GL :1)



## 7. Pembangkitan Frequent Pattern

Pattern Generated
Kemeja : {GL, KM : 1}, {KP, KM : 1}, {AQ, KM : 1}
Boneka : {GL, BN : 1}, {SR, BN : 1}
Celana : {GL, CL : 2}, {KP, CL : 2}
Sabun : {GL, SB : 2}, {KP, SB : 2}, {SR, SB : 1}, {AQ, SB : 1}
Popok : {GL, PP : 1}, {KP, PP : 1}, {SR, PP : 1}, {AQ, PP : 1}
Sampo : {GL, SP : 1}, {KP, SP : 1}, {AQ, SP : 2}, {SR, SP : 1}
Sprei : {GL, SR : 1}, {KP, SR : 1}, {AQ, SR : 1}
Aqua : {GL, AQ : 2}, {KP, AQ : 2}
Kopi : {GL, KP : 5}

# Frequent 2 Itemset

Frequent 2 Item	Gula	Kopi	Aqua	Popok	Sprei	Sabun	Sampo	Kemeja	Celana	Boneka
Gula	6	5	2	1	1	2	1	1	2	1
Kopi	5	5	1	1	1	2	1	1	2	1
Aqua	2	1	5	2	1	1	3	1	0	0
Popok	1	1	2	4	3	1	2	2	1	2
Sprei	1	1	1	3	5	2	2	2	2	3
Sabun	2	2	1	1	2	4	3	1	0	0
Sampo	1	1	3	2	2	3	5	1	0	0
Kemeja	1	1	1	2	2	1	1	2	1	1
Celana	2	2	0	1	2	0	0	1	3	2
Boneka	1	1	0	2	3	0	0	1	2	3

## 8. Mencari Support 2 Itemset

Support (2 Items)	Gula	Kopi	Aqua	Popok	Sprei	Sabun	Sampo	Kemeja	Celana	Boneka
Gula	0,5	0,4167	0,1667	0,0833	0,0833	0,16667	0,0833	0,08333	0,16667	0,08333
Kopi	0,4167	0,4167	0,0833	0,0833	0,0833	0,16667	0,0833	0,08333	0,16667	0,08333
Aqua	0,1667	0,0833	0,4167	0,1667	0,0833	0,08333	0,25	0,08333	0	0
Popok	0,0833	0,0833	0,1667	0,3333	0,25	0,08333	0,1667	0,16667	0,08333	0,16667
Sprei	0,0833	0,0833	0,0833	0,25	0,4167	0,16667	0,1667	0,16667	0,16667	0,25
Sabun	0,1667	0,1667	0,0833	0,0833	0,1667	0,33333	0,25	0,08333	0	0
Sampo	0,0833	0,0833	0,25	0,1667	0,1667	0,25	0,4167	0,08333	0	0
Kemeja	0,0833	0,0833	0,0833	0,1667	0,1667	0,08333	0,0833	0,16667	0,08333	0,08333
Celana	0,1667	0,1667	0	0,0833	0,1667	0	0	0,08333	0,25	0,16667
Boneka	0,0833	0,0833	0	0,1667	0,25	0	0	0,08333	0,16667	0,25

## 9. Mencari Confidence 2 Itemset

Confidence	Gula	Kopi	Aqua	Popok	Sprei	Sabun	Sampo	Kemeja	Celana	Boneka
Gula	1	0,833333	0,333333	0,166667	0,16666667	0,333333	0,166667	0,166667	0,333333	0,166667
Kopi	1	1	0,2	0,2	0,2	0,4	0,2	0,2	0,4	0,2
Aqua	0,4	0,2	1	0,4	0,2	0,2	0,6	0,2	0	0
Popok	0,25	0,25	0,5	1	0,75	0,25	0,5	0,5	0,25	0,5
Sprei	0,2	0,2	0,2	0,6	1	0,4	0,4	0,4	0,4	0,6
Sabun	0,5	0,5	0,25	0,25	0,5	1	0,75	0,25	0	0
Sampo	0,2	0,2	0,6	0,4	0,4	0,6	1	0,2	0	0
Kemeja	0,5	0,5	0,5	1	1	0,5	0,5	1	0,5	0,5
Celana	0,666667	0,666667	0	0,333333	0,66666667	0	0	0,333333	1	0,666667
Boneka	0,333333	0,333333	0	0,666667	1	0	0	0,333333	0,666667	1