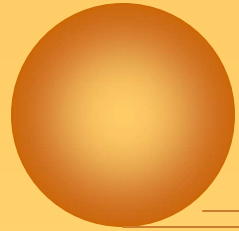


# Association Rule

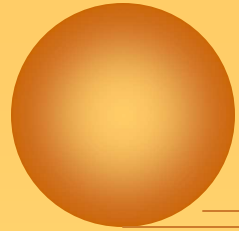
Ali Ridho Barakbah



# Association rule?

---

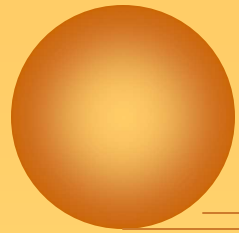
- Mencari suatu kaidah keterhubungan dari data
- Diusulkan oleh Agrawal, Imielinski, and Swami (1993)



# Contoh

---

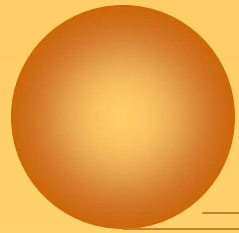
Dalam suatu supermarket kita ingin mengetahui seberapa jauh orang yang membeli celana juga membeli sabuk?



# Input & problema

---

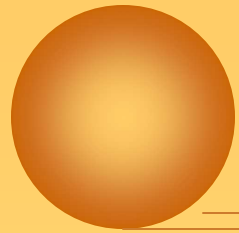
- Input
  - Adanya sejumlah transaksi
  - Setiap transaksi memuat kumpulan item
- Problema
  - Bagaimana caranya menemukan association rule yang memenuhi minimum support dan minimum confidence yang kita berikan



# 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

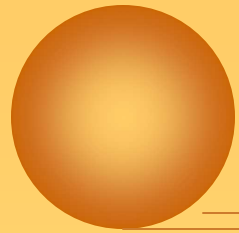




# Definisi umum

---

- Itemset: himpunan dari item-item yang muncul bersama-sama
- Kaidah asosiasi: peluang bahwa item-item tertentu hadir bersama-sama.
- Support dari suatu itemset  $X$  ( $\text{supp}(X)$ ) adalah rasio dari jumlah transaksi dimana itemset muncul dengan total jumlah transaksi



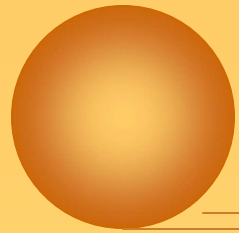
# Definisi umum

---

- Konfidence (keyakinan) dari kaidah  $X \rightarrow Y$ , ditulis  $\text{conf}(X \rightarrow Y)$  adalah
  - $\text{conf}(X \rightarrow Y) = \text{supp}(X \cup Y) / \text{supp}(X)$
  - Konfindence bisa juga didefinisikan dalam terminologi peluang bersyarat

$$\text{conf}(X \rightarrow Y) = P(Y|X) = P(X \cap Y) / P(X)$$

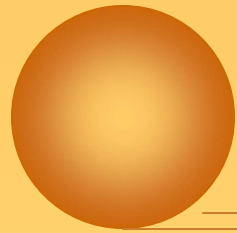
- Database transaksi menyimpan data transaksi. Data transaksi bisa juga disimpan dalam suatu bentuk lain dari suatu database  $m \times n$ .



# Ukuran support

- Misalkan  $I = \{I_1, I_2, \dots, I_m\}$  merupakan suatu himpunan dari literal, yang disebut item-item.
- Misalkan  $D = \{T_1, T_2, \dots, T_n\}$  merupakan suatu himpunan dari  $n$  transaksi, dimana untuk setiap transaksi  $T \in D$ ,  $T \subseteq I$ .
- Suatu himpunan item  $X \subseteq I$  disebut itemset.
- Suatu transaksi  $T$  memuat suatu itemset  $X$  jika  $X \subseteq T$ .
- Setiap itemset  $X$  diasosiasikan dengan suatu himpunan transaksi  $T_X = \{T \in D \mid T \supseteq X\}$  yang merupakan himpunan transaksi yang memuat itemset
- Support dari itemset  $X \rightarrow \text{supp}(X) :$   
$$|T_X|/|D|$$





# Contoh

Transaksi	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.

$\text{Supp}(A) = |T_a|/|D| = 3/10 = 0.3$ .

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

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

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

Itemset	Sp
A	0.3
B	0.1
C	0.8
D	0.7
AB	0
AC	0.2
AD	0.3
BC	0.1
BD	0.1
CD	0.5
ABC	0
ABD	0
ACD	0.2
BCD	0.1
ABCD	0

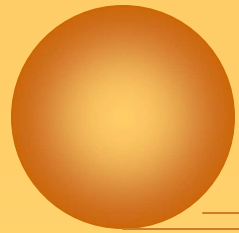
Jika minsupport diberikan oleh user sebagai threshold adalah 0.2, maka frequent itemset adalah semua itemset yang support-nya  $\geq 0.2$ , yakni

A, C, D, AC, AD, CD, ACD

Dari frequent itemset bisa dibangun kaidah asosiasi sbb:

$A \rightarrow C$	$C \rightarrow A$	$A \rightarrow D$
$D \rightarrow A$	$C \rightarrow D$	$D \rightarrow C$
$A, C \rightarrow D$	$A, D \rightarrow C$	$C, D \rightarrow A$

$$\text{Conf}(A \rightarrow C) = \text{supp}(A, C) / \text{supp}(A)$$

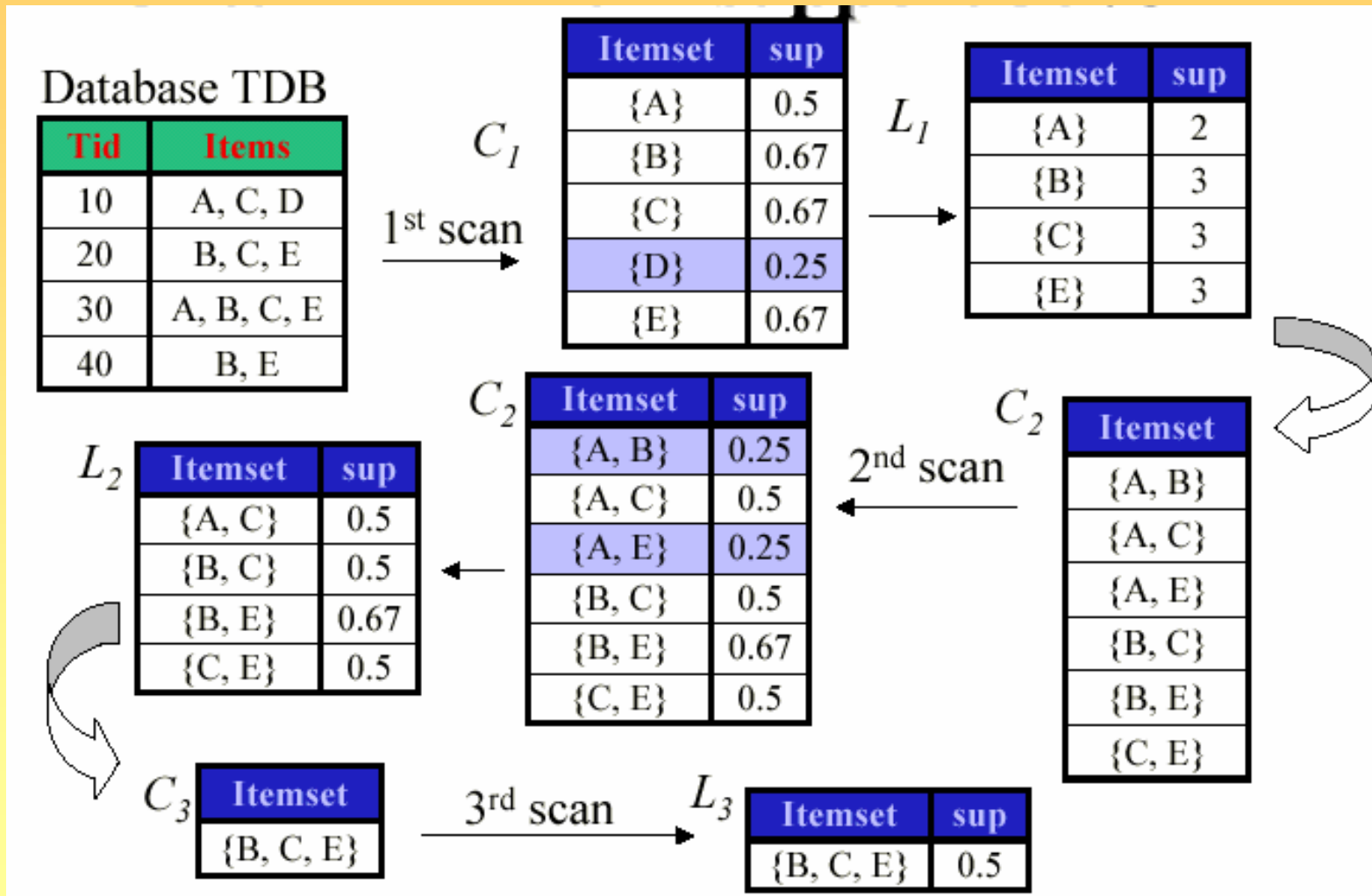


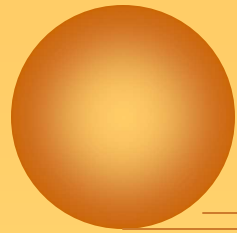
# Apriori

---

- Prinsip apriori :  
Subset apapun dari suatu frequent itemset harus frequent
- $L3 = \{abc, abd, acd, ace, bcd\}$
- Penggabungan sendiri :  $L3 * L * L3$ 
  - abcd dari abc dan abd
  - acde dari acd dan ace
- Pemangkasan Pemangkasan:
  - acde dibuang sebab ade tidak dalam  $L3$
- $C4 = \{abcd\}$

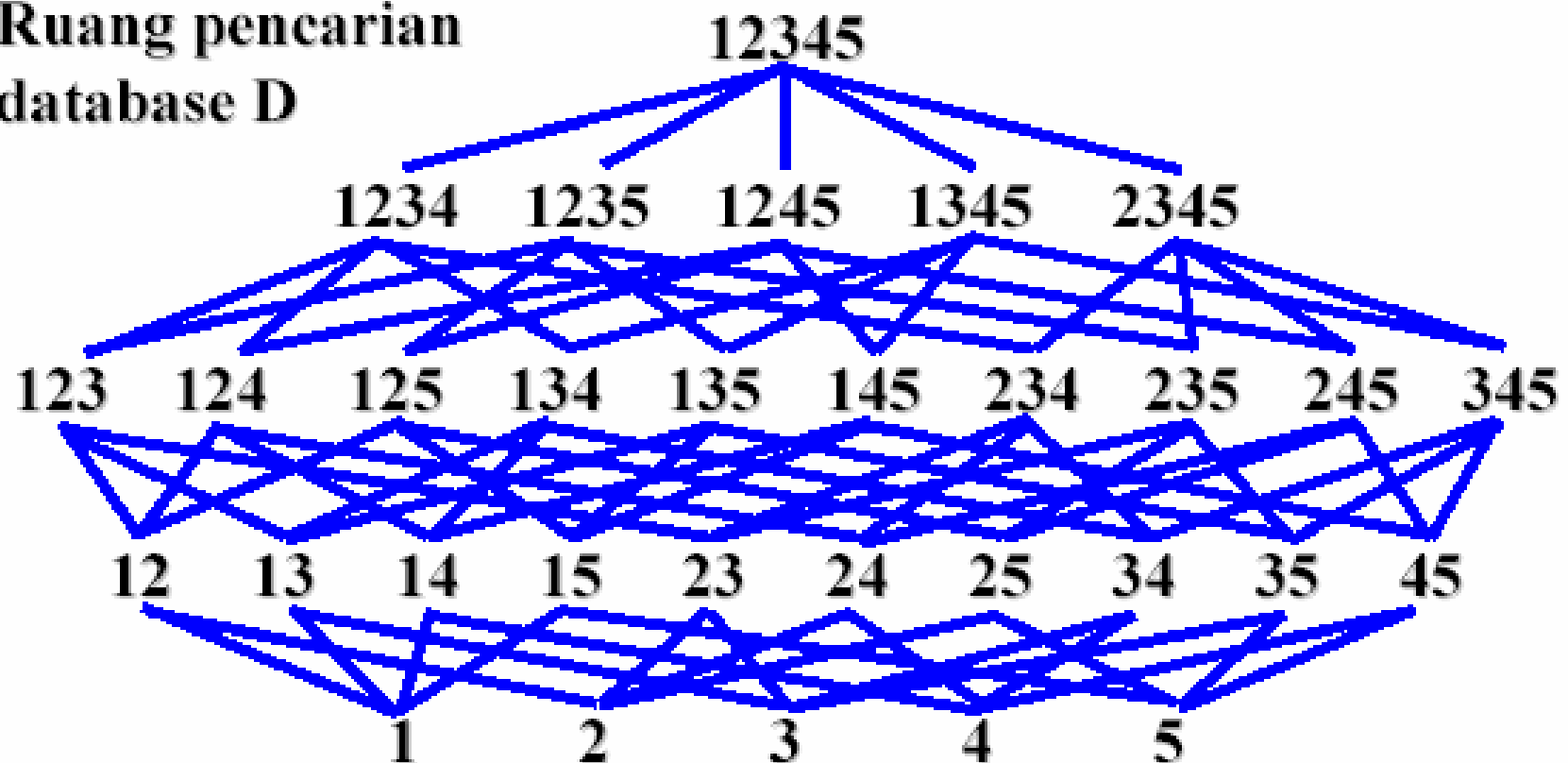
# Contoh apriori dengan minimum support 50%

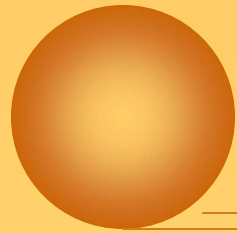




# Search space pada apriori

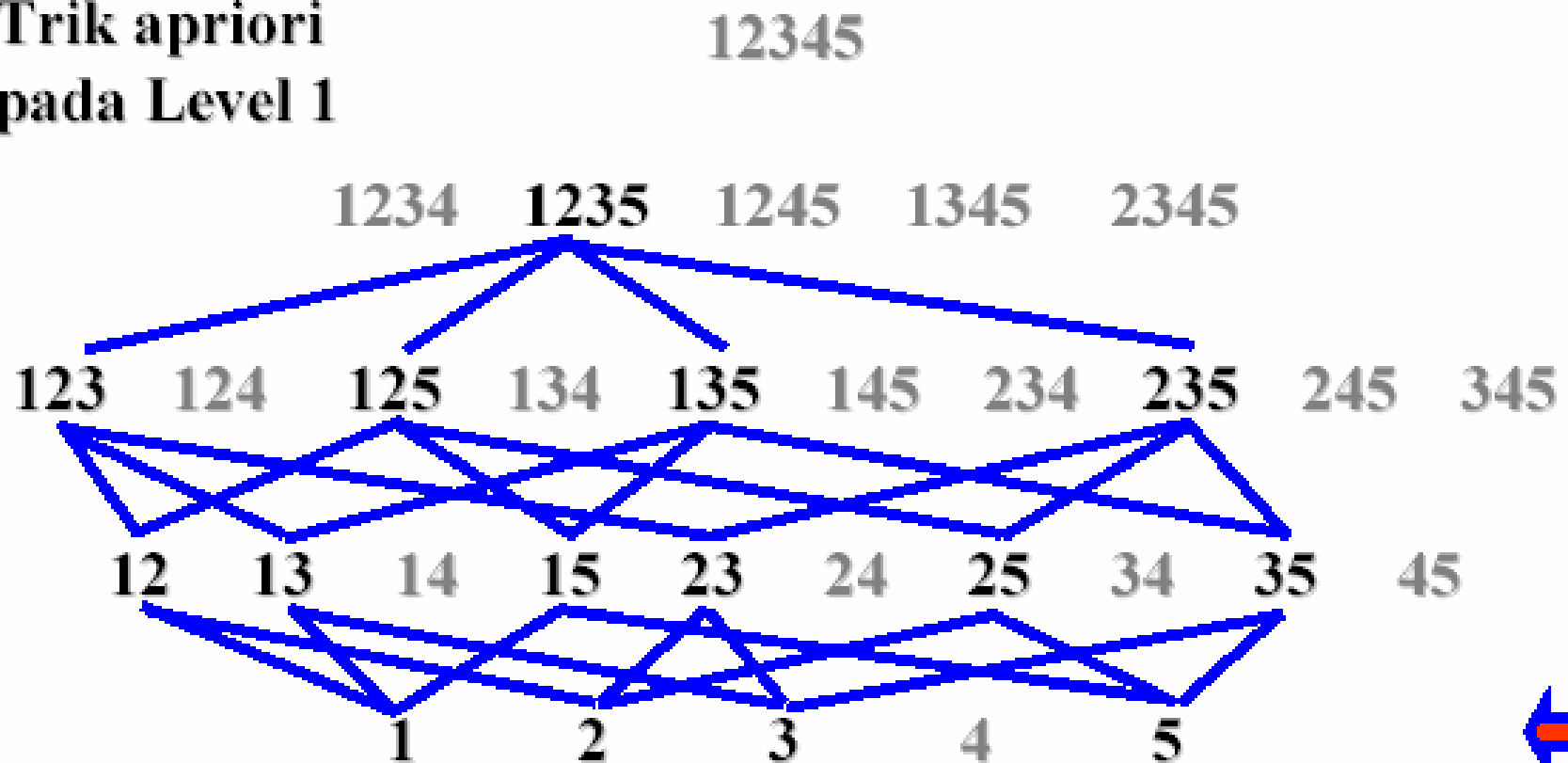
Ruang pencarian  
database D

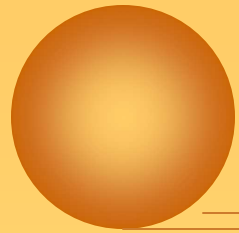




# Search space pada apriori

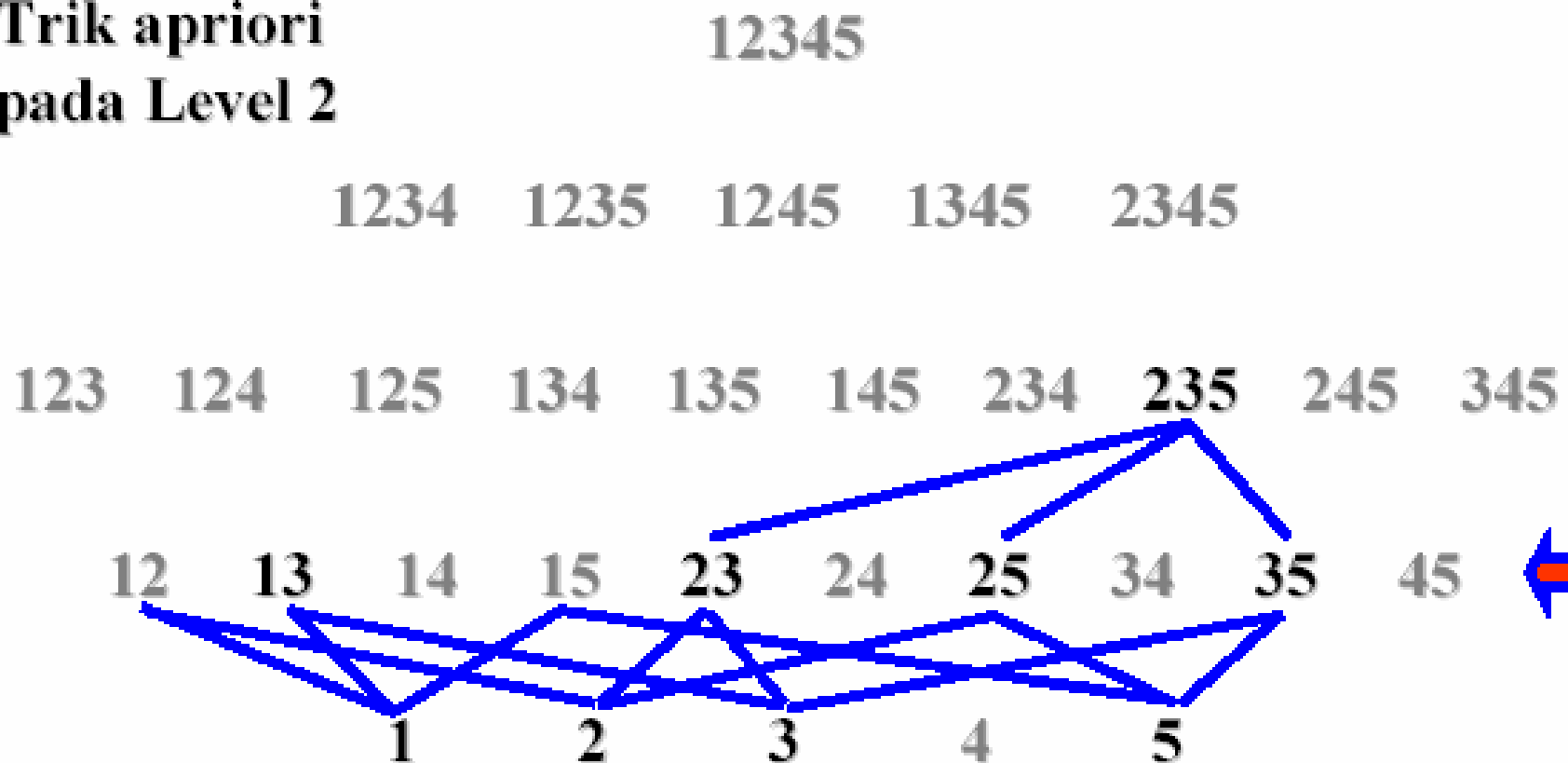
**Trik apriori  
pada Level 1**

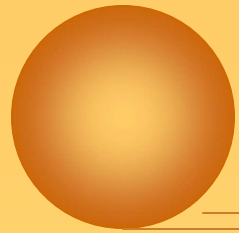




# Search space pada apriori

**Trik apriori  
pada Level 2**



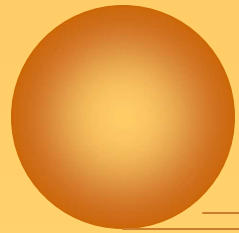


# Latihan

T1	{roti, selai, mentega}
T2	{roti, mentega}
T3	{roti, susu, mentega}
T4	{coklat, roti}
T5	{coklat, susu}

- Suatu supermarket mempunyai sejumlah transaksi seperti dalam tabel
- Buatlah association rule dari data tersebut dengan cara menghitung support dan confidence
- Pakailah metode apriori dengan minimum support=0.3 dan confidence=0.8





# Latihan

T1	{roti, selai, mentega}
T2	{roti, mentega}
T3	{roti, susu, mentega}
T4	{coklat, roti}
T5	{coklat, susu}

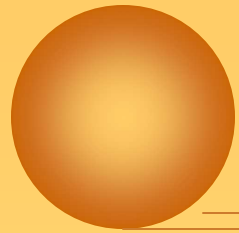
Itemset	Sp
{roti}	0.8
{selai}	0.2
{mentega}	0.6
{susu}	0.4
{coklat}	0.4

Itemset	Sp
{roti,mentega}	0.6
{roti,susu}	0.2
{roti,coklat}	0.2
{mentega,susu}	0.2
{mentega,coklat}	0
{susu,coklat}	0.2



$$\begin{aligned}\text{Conf}(\text{roti} \rightarrow \text{mentega}) &= \text{Supp}(\{\text{roti}, \text{mentega}\}) / \text{Supp}(\{\text{roti}\}) \\ &= 0.6 / 0.8 = 0.75 \rightarrow 75\%\end{aligned}$$

$$\begin{aligned}\text{Conf}(\text{mentega} \rightarrow \text{roti}) &= \text{Supp}(\{\text{mentega}, \text{roti}\}) / \text{Supp}(\{\text{mentega}\}) \\ &= 0.6 / 0.6 = 1 \rightarrow \mathbf{100\%}\end{aligned}$$



# Tugas

T1	{roti, selai, mentega}
T2	{roti, mentega}
T3	{roti, susu, mentega}
T4	{coklat, roti, susu, mentega}
T5	{coklat, susu}

- Suatu supermarket mempunyai sejumlah transaksi seperti dalam tabel
- Buatlah association rule dari data tersebut dengan cara menghitung support dan confidence
- Pakailah metode apriori dengan minimum support=0.3 dan confidence=0.8