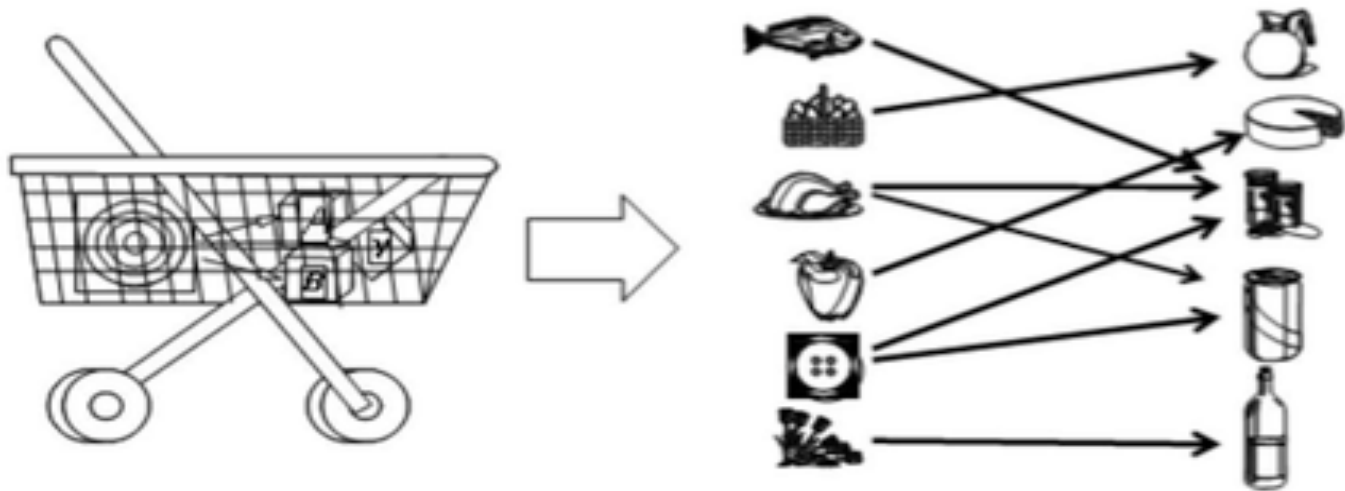


ASSOCIATION RULES



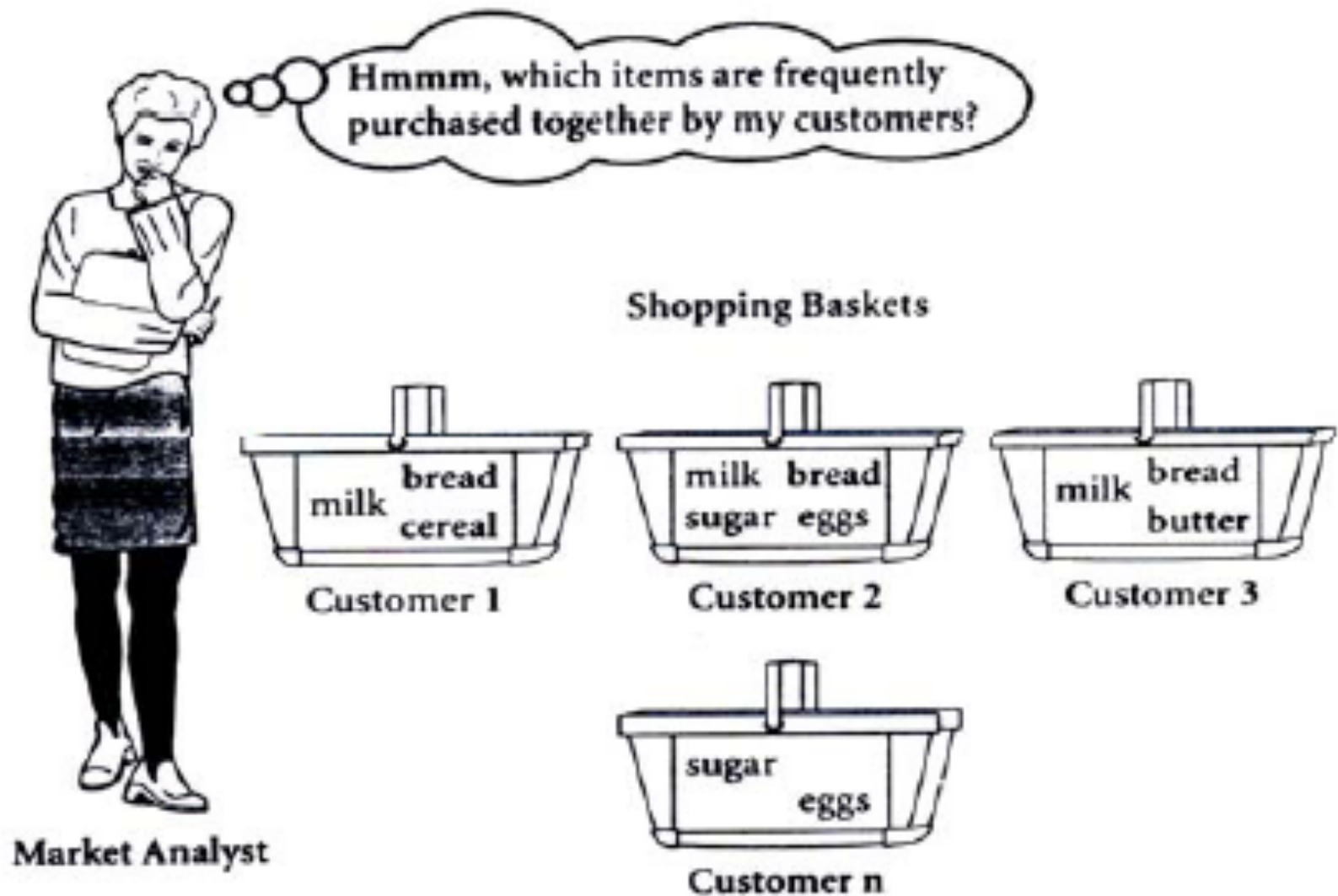
Pendahuluan

- Analisis aturan asosiasi merupakan tugas dasar pada data mining.
- Tujuannya:
 - Menemukan hubungan kemunculan bersamaan (asosiasi) diantara item-item data.
- Aplikasi klasik yang menggunakan metode ini adalah *market basket data analysis*.
 - Tujuannya: menemukan bagaimana item-item barang yang dibeli oleh pelanggan

Ide Awal

- “ barang A biasanya dibeli secara bersamaan dengan barang B oleh seorang konsumen pada suatu waktu tertentu dengan sejumlah pembelian bersama “

Konsep Market Basket Analysis



Contoh:

- Sabun Mandi → Pasta Gigi
 - [support: 40%, confidence = 80%]
 - 40% pelanggan membeli Sabun Mandi dan Pasta Gigi bersamaan
 - 80% pelanggan membeli Sabun Mandi pasti membeli Pasta Gigi.

Ukuran

- Support
 - Seberapa sering aturan yang dihasilkan berlaku pada himpunan transaksi T .
- Confidence
 - Dilihat sebagai probabilitas kondisional terhadap aturan.
- Aturan yang terpilih adalah aturan yang memenuhi minimum support dan minimum confidence

$$\text{support}(A \Rightarrow B) = P(A \cup B)$$

$$\text{confidence}(A \Rightarrow B) = P(B|A).$$

Konsep Dasar Association Rules

$I = \{i_1, i_2, \dots, i_m\}$ Himpunan item

$T = \{t_1, t_2, \dots, t_n\}$ Himpunan transaksi

T_i adalah himpunan item dimana $t_i \subseteq I$

Example case - Swalayan

Transaction	Items Purchased
1	Broccoli, green peppers, corn
2	Asparagus, squash, corn
3	Corn, tomatoes, beans, squash
4	Green peppers, corn, 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, beans, squash
13	Squash, corn, asparagus, beans
14	Corn, green peppers, tomatoes, beans, broccoli

Istilah-istilah

- Aturan asosiasi yang berbentuk “*if....then....*” atau “*jika....maka*”, merupakan pengetahuan yang dihasilkan dari fungsi aturan asosiasi.
- *Item* → barang yang dibeli atau barang yang menjadi objek kegiatan belanja.
- Pada swalayan unsada terdapat 7 jenis item yaitu (urut abjad) **asparagus, beans, brocolli, corn, green peppers, squash** dan **tomatoes**.

Istilah-istilah (lanj)

- Himpunan item dilambangkan dengan I merupakan himpunan dari semua jenis item yang akan dibahas.
- Persamaan himpunan item \rightarrow
- **Persamaan 1:**
- $I = \{\text{asparagus, beans, brocolli, corn, green peppers, squash, tomatoes}\}$
 - Himpunan item yang dibeli pengunjung ke- i disebut transaksi ke $- I$
 - Dilambangkan T_i

Istilah-istilah (lanj)

Persamaan 2:

- $T1 = \{\text{broccoli, green, peppers, corn}\}$
- $T2 = \{\text{Asparagus, squash, corn}\}$
-
- $T14 = \{\text{corn, green, peppers, tomatoes, beans, broccoli}\}$

Persamaan 3:

- Himpunan seluruh transaksi dilambangkan dengan D sehingga persamaan 3 ini menjadi:
- $D = \{T1, T2, \dots, T14\}$

Istilah-istilah

Persamaan $4 \rightarrow$ implikasi

“*jika A, maka B*” atau “ $A \rightarrow B$ ”

- A disebut *anteseden* atau *pendahulu*
- B disebut *konsekuen* atau *pengikut*

Istilah-istilah

1. A maupun B adalah himpunan bagian murni dari I

→ **Persamaan 5** yaitu $A, B \subset I$

2. A dan B adalah dua himpunan yang saling lepas. Sehingga disimbolkan pada

→ **Persamaan 6:** yaitu

$$A \cap B = \emptyset$$

Istilah-istilah

- Salah satu ukuran kinerja bagi aturan asosiasi ” $A \rightarrow B$ ” adalah besar support(dukungan) yang dilambangkan dengan $s(A \rightarrow B)$. Dan didefinisikan sebagaimana di persamaan 7.

Persamaan 7

$$s(A \Rightarrow B) = P(A \cap B) = \frac{\text{Jumlah transaksi yang mengandung item-item pada } A \cap B}{\text{Jumlah total transaksi pada } D}$$

Istilah-istilah (lanj)

- **Persamaan 8**

$$\text{conf}(A \Rightarrow B) = P(A \cap B) = \frac{\text{Jumlah transaksi yang mengandung item-item pada } A \cap B}{\text{Jumlah transaksi yang mengandung } A}$$

$$\text{conf}(B \Rightarrow A) = P(A \cap B) = \frac{\text{Jumlah transaksi yang mengandung item-item pada } A \cap B}{\text{Jumlah transaksi yang mengandung } B}$$

- *Itemset* \rightarrow suatu himpunan yang beranggotakan sebagian atau seluruh item yang menjadi anggota I.

Istilah-istilah (lanj)

- Contoh dari itemset adalah $\{Asparagus\}$ atau $\{Asparagus, Bean\}$, atau $\{Asparagus, Beans, Squash\}$
- Itemset yang beranggotakan k buah item disebut k -itemset.

Istilah-istilah (lanj)

- Himpunan $\{Asparagus\}$ adalah suatu itemset. Lebih spesifik lagi 1-itemset karena hanya beranggotakan satu buah item saja
- Himpunan $\{Asparagus, Beans\}$ adalah suatu itemset. Lebih spesifik lagi 2-itemset karena hanya beranggotakan dua buah item saja
- Himpunan $\{Asparagus, beans, squash\}$ adalah suatu itemset. Lebih spesifik lagi 3-itemset karena beranggotakan tiga buah item saja

Istilah-istilah (lanj)

- Besaran *frekuensi itemset* mengukur berapa kali sebuah itemset muncul sebagai bagian atau keseluruhan transaksi yang menjadi anggota daftar transaksi D .
- Contoh:
- Frekuensi itemset {asparagus} adalah 6 karena himpunan ini menjadi bagian dari enam transaksi (lihat data transaksi slide 3), yaitu T2, T5, T6, T9, T12 dan T13

Istilah-istilah (lanj)

- Frekuensi itemset {asparagus, beans} adalah 5 karena himpunan ini menjadi bagian dari lima transaksi , yaitu T5, T6, T9, T12 dan T13
- Frekuensi itemset {asparagus, beans, squash} adalah 4 karena himpunan ini menjadi bagian dari empat transaksi (slide 3), yaitu T6, T9, T12 dan T13

Istilah-istilah (lanj)

- *Itemset sering/frequent itemset* suatu itemset yang memiliki frekuensi itemset minimal sebesar bilangan F yang ditetapkan.
- Contoh bila kita tetapkan $\phi = 4$, maka:
 1. Itemset {asparagus, beans, squash} termasuk itemset yang sering karena memiliki frekuensi itemset yang telah melebihi atau minimal sebesar $F \ \phi = 4$.

Istilah-istilah (lanj)

2. Itemset {squash, tomatoes} tidak termasuk itemset sering karena memiliki frekuensi itemset sebesar 3, artinya masih di bawah nilai ω yang ditetapkan

Himpunan dari seluruh k-itemset dilambangkan dengan **F_k**.

Istilah-istilah (lanj)

Aturan asosiasi secara ringkas digambarkan sbb:

- Berawal dari data latihan yang tersedia (lihat slide 8)
- Data latihan diolah dengan menggunakan algoritma aturan asosiasi.
- Masalah aturan asosiasi berakhir dengan dihasilkannya pengetahuan yang direpresentasikan dalam bentuk diagram yang disebut aturan asosiasi.

Prototip masalah aturan asosiasi dan pengetahuan yang dihasilkan

“ jika membeli asparagus, maka membeli beans ”

Dapat diartikan:

- Item asparagus mempunyai kecenderungan untuk dibeli bersama-sama dengan item beans, atau
- Pengunjung toko unsada yang membeli asparagus mempunyai kecenderungan untuk juga membeli beans

Algoritma aturan asosiasi

Market Basket Analysis (MBA)

Hasil pembahasan sebelumnya dapat disimpulkan menjadi:

- Data historis merupakan data penting sebagai data latihan/training data
- Data tersebut akan dijadikan input bagi suatu algoritma yang saat ini belum kita ketahui algoritmanya
- Sebagai keluaran algoritma yang saat ini belum kita ketahui jenisnya , kita akan memperoleh pengetahuan yang secara sederhana dapat direpresentasikan dalam bentuk *“jika...., maka....”*

Kahoot

Total transaksi : 1000

- Hammer : 50
- Nails : 80
- Lumber : 20
- Hammer + nails : 15
- Nails + lumber : 10
- Hammer + lumber : 10
- Hammer + nails + lumber : 5

Tabel 1. Data Transaksi

ID Transaksi	Item Terjual
1	pena,roti,mentega
2	roti,mentega,telur,susu
3	buncis,telur,susu
4	roti,mentega
5	roti,mentega,kecap,telur,susu

1. Analisa pola frekuensi tinggi

- Tahap ini mencari kombinasi item yang memenuhi syarat minimum dari nilai *support* dalam database, yang dapat dirumuskan sebagai berikut:
- Rumus

$$\text{Support} = p(A \cap B) = \frac{\text{jumlah transaksi yang memuat A dan B}}{\text{total jumlah transaksi}}$$

Analisa pola frekuensi tinggi

- transaksi yang memuat {roti,mentega} ada 4 (support 80%)
- transaksi yang memuat {roti,mentega,susu} ada 2 (support 40%)
- Transaksi yang memuat {buncis} hanya 1 (support 20%)
- dan sebagainya...

Analisa pola frekuensi tinggi

- Bila ditetapkan syarat minimum dari nilai *support* untuk pola frekuensi tinggi dalam contoh ini adalah 30%, diperoleh pola frekuensi tinggi yang memenuhi syarat minimum nilai *support* adalah :

Kombinasi Item	Support	Kombinasi Item	Support
{roti}	80%	{roti,susu}	40%
{mentega}	80%	{mentega,susu}	40%
{telur}	60%	{roti,telur}	40%
{susu}	60%	{mentega,telur}	40%
{roti,mentega}	80%	{roti,mentega,susu}	40%
{telur,susu}	60%	{roti,mentega,telur,susu}	40%

2. Pembentukan aturan assosiatif

- Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan assosiatif yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan assosiatif
- $A \rightarrow B$ dengan menggunakan rumus berikut :

$$\text{Confidence} = P(B / A) = \frac{\text{Support } (A \cap B)}{P(A)}$$

- salah satu contoh aturan assosiatif yang dapat ditemukan pada Tabel.1
- $\{\text{telur}, \text{susu}\} \rightarrow \{\text{roti}, \text{mentega}\}$ dengan nilai
- *confidence*:

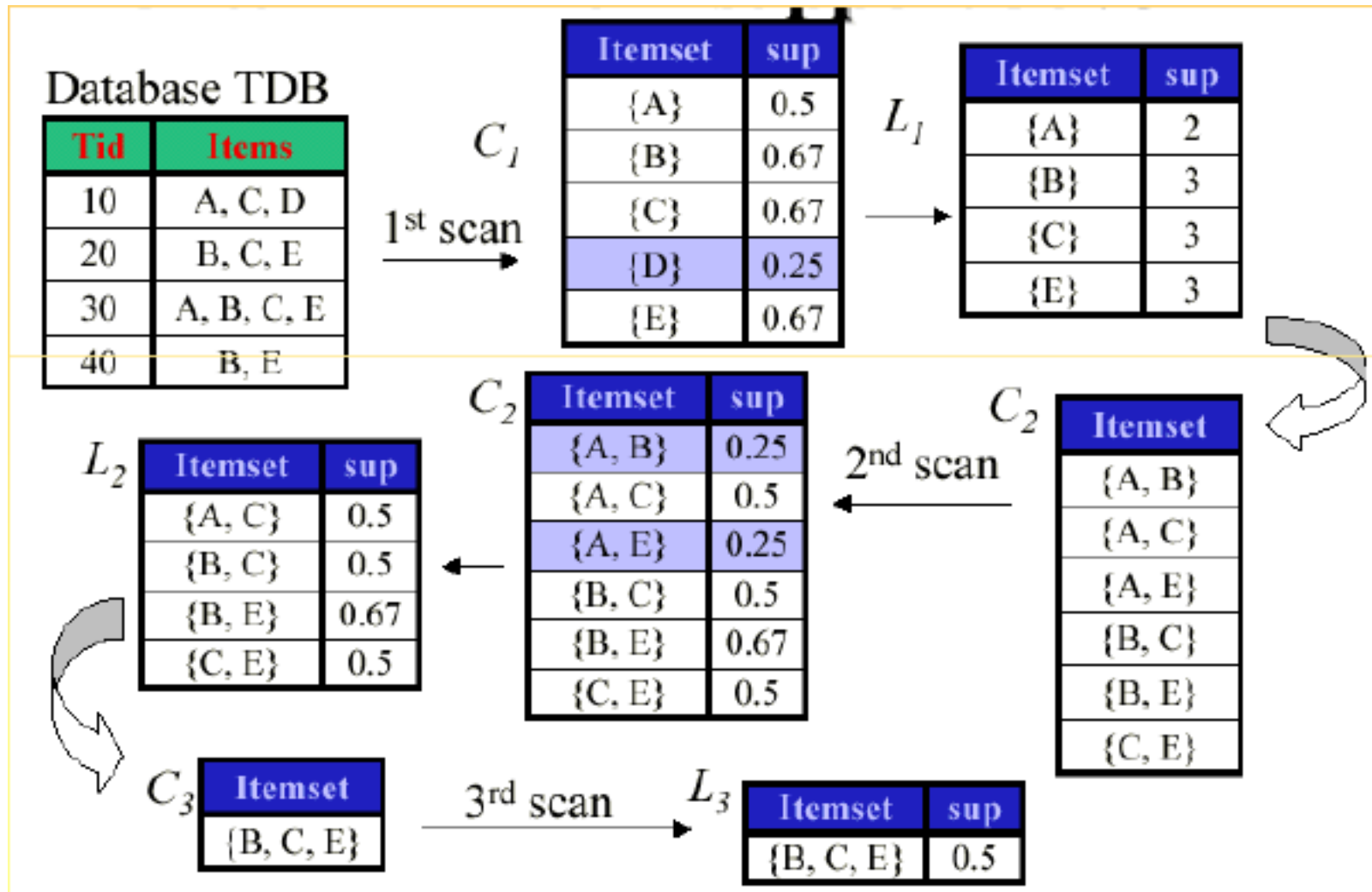
$$= \frac{\text{Support } (A \cap B)}{P(A)} = 40\% / 60\% = 66.67\%$$

Analisa aturan assosiatif

- Bila syarat minimum untuk *confidence* dari contoh diatas adalah 50%
- Aturan assosiatif yang dapat ditemukan dari tabel.1 diantaranya adalah :

Associations Rule	Support	Confidence
$\{\text{telur}, \text{susu}\} \rightarrow \{\text{roti}, \text{mentega}\}$	40%	66.6%
$\{\text{roti}, \text{mentega}\} \rightarrow \{\text{susu}\}$	40%	50%
$\{\text{mentega}, \text{susu}\} \rightarrow \{\text{roti}\}$	40%	100%

Contoh apriori dengan minimum support 50%



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

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

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