

# **PERLOMBONGAN ATURAN SEKUTUAN**

**STQD6414 PERLOMBONGAN DATA**



**Prof. Madya Dr. Nurulkamal Masseran**

**Jabatan Sains Matematik**

**Universiti Kebangsaan Malaysia**

# PENGENALAN:

- Perlombongan Aturan Sekutuan (*Mining Association Rule*) bertujuan mencari hubungan yang 'menarik' antara set item.
- Ia biasanya digunakan untuk membuat cadangan produk dengan mengenal pasti produk yang sering dibeli bersama.
- Perlombongan Aturan Sekutuan biasanya dilakukan terhadap data urus niaga dari pasaran runcit atau dari perniagaan e-dagang dalam talian.
- Al-khwarizmi *apriori* dan *eclat* digunakan untuk melihat pola dan aturan dalam data ini.
- Apakah itu aturan (*rule*)?
- Aturan adalah tatatanda yang mewakili item yang sering dibeli bersama dengan item-item lain.
- Ia mempunyai LHS dan bahagian RHS yang diwakili seperti berikut:

$$X \Rightarrow Y$$

- Ini bermaksud, item di sebelah kanan sering dibeli bersama-sama dengan item di sebelah kiri.



# DATA TRANSAKSI:

- Contoh Data Transaksi Pembelian Barang:

transaction ID	items
1	milk, bread
2	bread, butter
3	beer
4	milk, bread, butter
5	bread, butter

- Al-khwarizmi membaca data transaksi dalam bentuk pemboleh ubah binari/dedua.
- Iaitu  $I = \{i_1, i_2, \dots, i_n\}$  ialah set  $n$  atribut binari yang disebut sebagai item. Dan  $D = \{t_1, t_2, \dots, t_m\}$  ialah set  $m$  transaksi.

		items			
		milk	bread	butter	beer
transactions	1	1	1	0	0
	2	0	1	1	0
	3	0	0	0	1
	4	1	1	1	0
	5	0	1	1	0



# ASAS ATURAN SEKUTUAN:

- Aturan sekutuan mengambil bentuk  $X \Rightarrow Y$ .
- Dengan syarat:
  - i)  $X, Y \subseteq I$
  - ii)  $X \cap Y = \emptyset$  (X dan Y bukan set item yang sama).
  - iii) X ialah aturan anteseden (peristiwa yang berlaku terlebih dahulu).
  - iv) Y ialah aturan konsekuen (peristiwa yang berlaku disebabkan sesuatu).

## Contoh:

$$\{susu, mentega, roti\} \rightarrow \{telur\}$$



# ASAS ATURAN SEKUTUAN:

- Set Item yang kerap (*Frequent itemsets*) digunakan untuk mendapatkan aturan sekutuan dalam bentuk  $X \Rightarrow Y$ .
- Contoh Aturan Sekutuan ialah:  $\{Telur, Susu\} \Rightarrow \{Yogurt\}$ .
- Berdasarkan Aturan Sekutuan tersebut, pemilik pasar raya mendapati pelanggan yang membeli telur dan susu juga akan membeli Yogurt.
- Maka, pasar raya tersebut boleh mempromosikan yogurt kepada pelanggan yang sering membeli telur dan susu.
- Walaupun, pasar raya boleh mengatur susunan rak bagi jualan yogurt berdekatan dengan rak jualan telur dan susu.



# ASAS ATURAN SEKUTUAN:

- Aturan Sekutuan menerangkan berkenaan hubungan atau korelasi antara set item.
- Tiga ukuran asas dalam memilih aturan sekutuan ialah :
  - i) Sokongan (*support*)
  - ii) Keyakinan (*confidence*)
  - iii) Lif (*lift*)
- Sokongan ialah perkadaran transaksi dalam data yang mengandungi kedua-dua set item X dan Y:

$$\text{support}(X \Rightarrow Y) = P(X \cap Y) = \frac{n_{XY}}{N}$$



- *Keyakinan* ialah perkadaran transaksi akan mengandungi item Y jika item X dibeli:

$$confidence(X \Rightarrow Y) = P(Y | X) = \frac{P(X \cap Y)}{P(X)} = \frac{\binom{n_{XY}}{N}}{\binom{n_X}{N}}$$

- *Lif* ialah nisbah *Keyakinan* kepada perkadaran transaksi yang mengandungi Y:

$$lift(X \Rightarrow Y) = \frac{confidence(X \Rightarrow Y)}{P(Y)} = \frac{P(X \cap Y)}{P(Y)P(X)} = \frac{\binom{n_{XY}}{N}}{\binom{n_X}{N}\binom{n_Y}{N}}$$

- Nilai lif yang diperolehi ditakrif sebagai:

- $lift(X \Rightarrow Y) = 1$  , X dan Y adalah tak bersandar.
- $lift(X \Rightarrow Y) > 1$  , X dan Y mempunyai kesan pelengkap.
- $lift(X \Rightarrow Y) < 1$  , X dan Y mempunyai kesan pengganti



# CONTOH:


- Diberi data bagi transaksi berikut:

Transaksi	Set Item
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

- Contoh Aturan Sekutuan: {Milk, Diaper}  $\Rightarrow$  Beer

i)  $\text{support}(\{Milk, Diaper\} \Rightarrow Beer) = P(\{Milk, Diaper\} \cap Beer) = \frac{2}{5} = 0.4$

ii)  $\text{confidence}(\{Milk, Diaper\} \Rightarrow Beer) = P(Beer | \{Milk, Diaper\})$

$$= \frac{P(\{Milk, Diaper\} \cap Beer)}{P(\{Milk, Diaper\})} = \frac{\left(\frac{2}{5}\right)}{\left(\frac{3}{5}\right)} = \frac{2}{3} = 0.67$$




# CONTOH:

- Diberi data bagi transaksi berikut:

Transaksi	Set Item
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

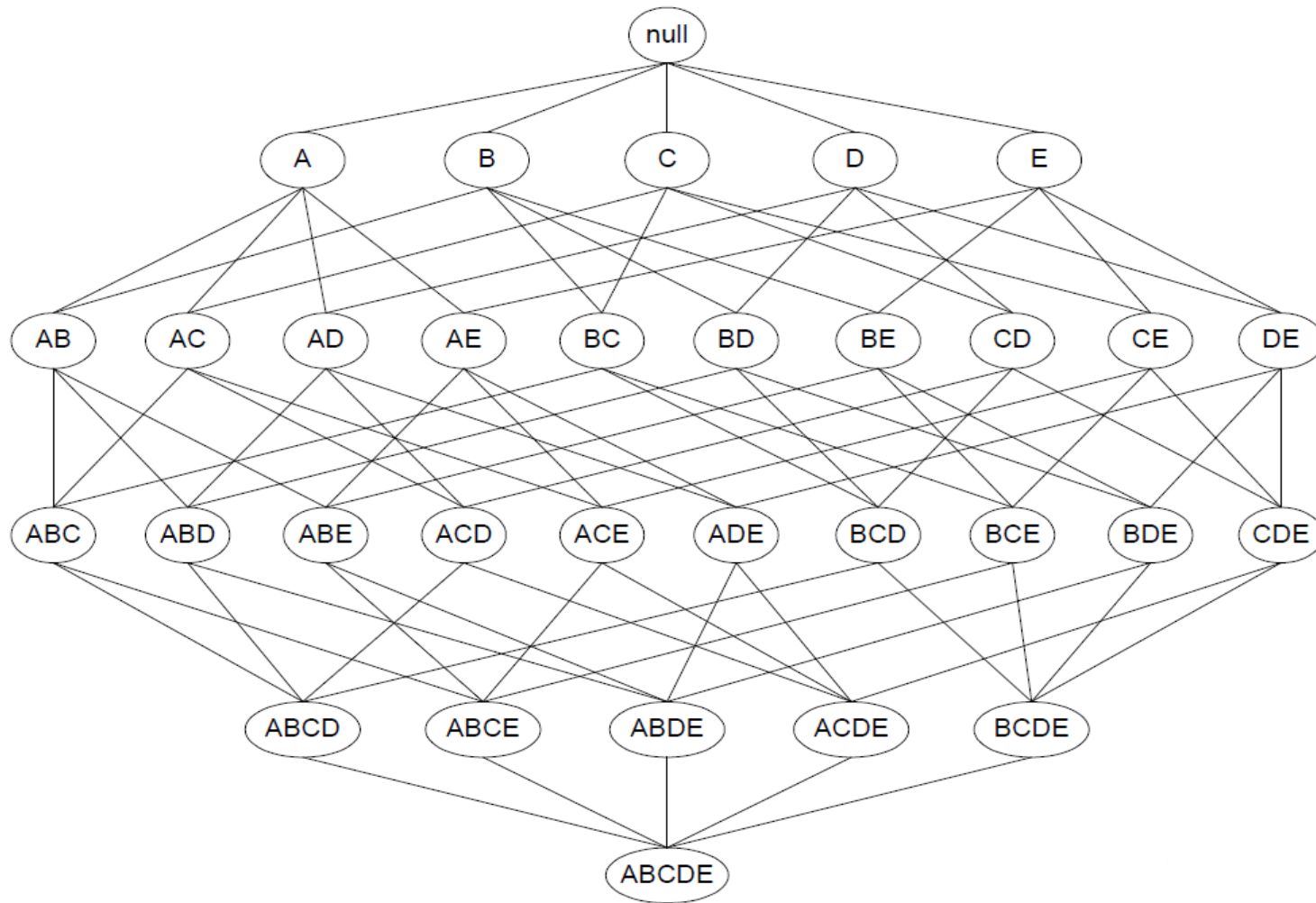
- Contoh Aturan Sekutuan: {Milk, Diaper}⇒Beer

$$\text{iii) } lift(\{Milk, Diaper\} \Rightarrow Beer) = \frac{confidence(\{Milk, Diaper\} \Rightarrow Beer)}{P(Beer)} = \frac{0.67}{\left(\frac{3}{5}\right)} = 1.12$$



# KEMUNGKINAN SEMUA SET ITEM:

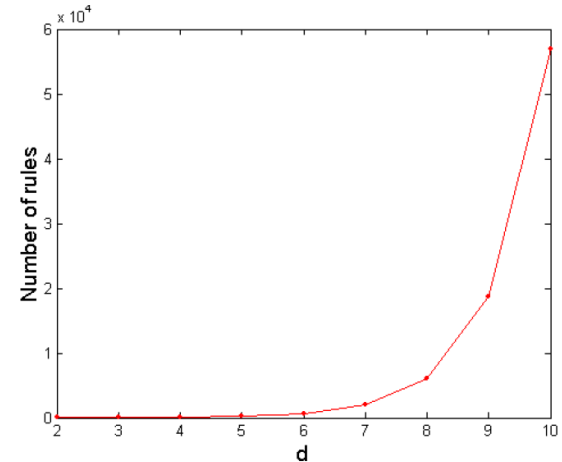
- Jika pasar raya ada  $d=5$  item, maka akan terdapat  $2^d = 2^5 = 32$  kemungkinan set item yang boleh terbentuk iaitu:



# KEMUNGKINAN SEMUA ATURAN SEKUTUAN:

- Jumlah semua kemungkinan set item =  $2^d$
- Jumlah semua kemungkinan Aturan Sekutuan:

$$R = \sum_{k=1}^{d-1} \left[ \binom{d}{k} \times \sum_{j=1}^{d-k} \binom{d-k}{j} \right]$$
$$= 3^d - 2^{d+1} + 1$$



- Umumnya  $d$  adalah sangat besar.
- Maka terdapat terlalu banyak aturan sekutuan yang mungkin berlaku.
- Adalah mustahil untuk mendapatkan aturan sekutuan secara manual.
- Al-khwarizmi *apriori* digunakan untuk mendapatkan set aturan sekutuan yang signifikan.



# ASAS ATURAN SEKUTUAN:

- Setiap aturan sekutuan  $X \rightarrow Y$  perlu memenuhi rangka kerja (*framework*) berikut:

$$\text{support}(X \Rightarrow Y) \geq \sigma$$

$$\text{confidence}(X \Rightarrow Y) \geq \gamma$$

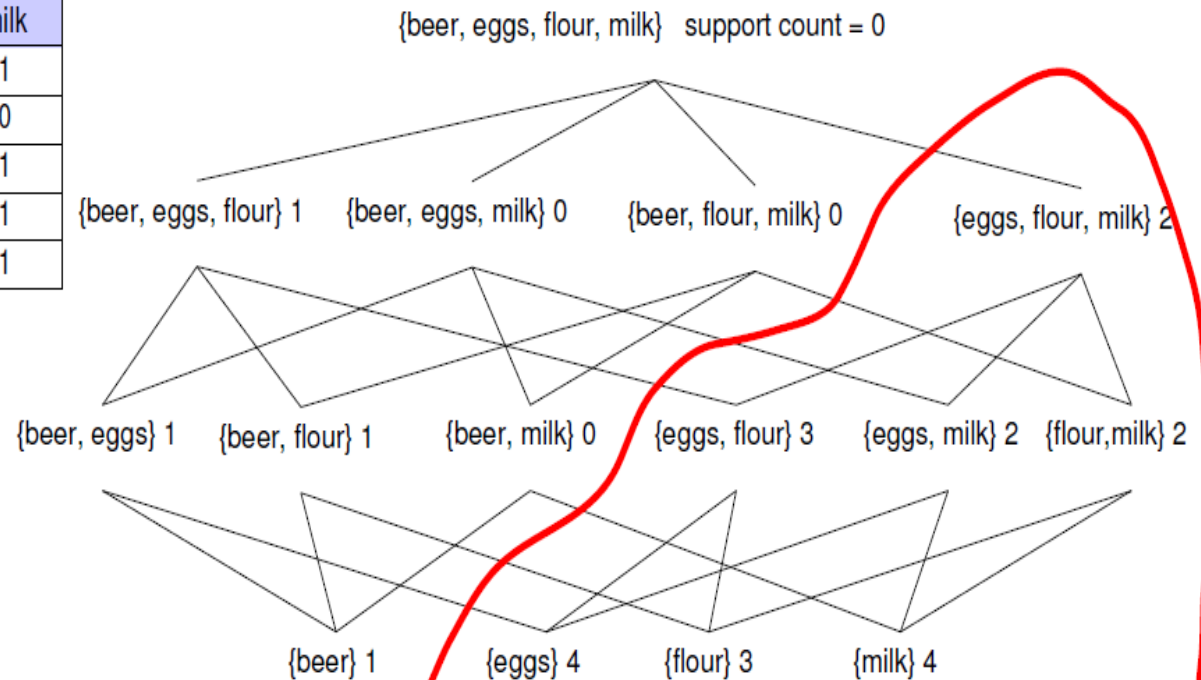
- $\sigma$  ialah nilai ambang (*threshold*) minimum sokongan.
- $\gamma$  ialah nilai ambang minimum keyakinan.
- Prasyarat nilai ambang ini perlu ditentukan oleh penganalisis.
- Nilai ambang ini perlu ditentukan untuk memilih aturan sekutuan yang bermakna.



# MINIMUM SOKONGAN:

- Berdasarkan nilai ambang sokongan yang ditetapkan, kombinasi item set yang paling kerap berlaku saja akan dikekalkan.
- **Contoh:** Bagi data transaksi berikut, dengan minimum sokongan,  $\sigma=0.5$ :

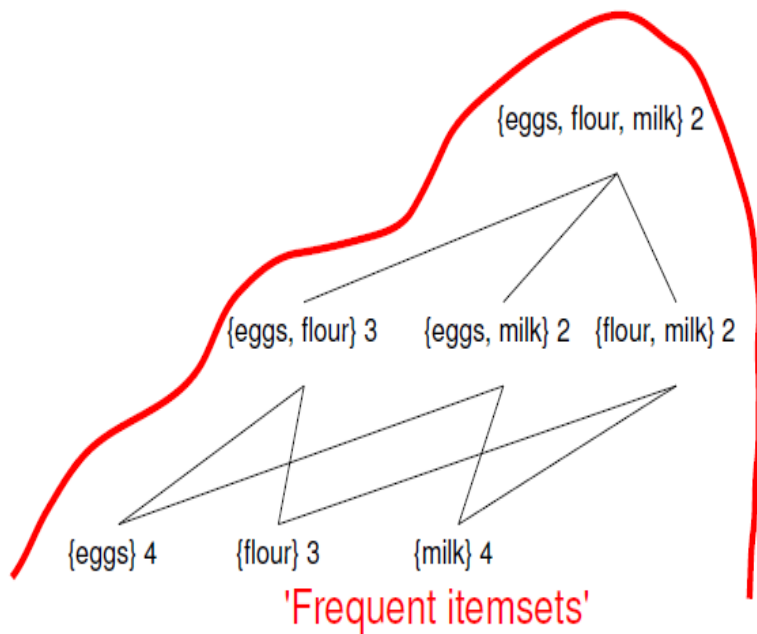
Transaction ID	beer	eggs	flour	milk
1	0	1	1	1
2	1	1	1	0
3	0	1	0	1
4	0	1	1	1
5	0	0	0	1



'Frequent Itemsets'

# MINIMUM KEYAKINAN:

- Berdasarkan set aturan yang terpilih dalam minimum sokongan,  $\sigma=0.5$ , kombinasi bagi set aturan akan dihitung.



			Confidence
{eggs}	→	{flour}	$3/4 = 0.75$
{flour}	→	{eggs}	$3/3 = 1$
{eggs}	→	{milk}	$2/4 = 0.5$
{milk}	→	{eggs}	$2/4 = 0.5$
{flour}	→	{milk}	$2/3 = 0.67$
{milk}	→	{flour}	$2/4 = 0.5$
{eggs, flour}	→	{milk}	$2/3 = 0.67$
{eggs, milk}	→	{flour}	$2/2 = 1$
{flour, milk}	→	{eggs}	$2/2 = 1$
{eggs}	→	{flour, milk}	$2/4 = 0.5$
{flour}	→	{eggs, milk}	$2/3 = 0.67$
{milk}	→	{eggs, flour}	$2/4 = 0.5$



# PENENTUAN ATURAN SEKUTUAN: ALGORITHMMA APRIORI & ECLAT

- Algorithma ini dijalankan berdasarkan.
  - i) Dapatkan semua set item yang kerap: Dapatkan semua kombinasi set item kerap berlaku dalam data.
  - ii) Tentukan aturan sekutuan daripada (i): Kekalkan hanya set item yang memenuhi sifat ambang sokongan minimum dan/atau sifat ambang keyakinan minimum.
- Kedua-dua nilai ambang ini ditetapkan oleh penganalisis ataupun pakar bidang.



# PERLOMBONGAN ATURAN SEKUTUAN MENERUSI R:

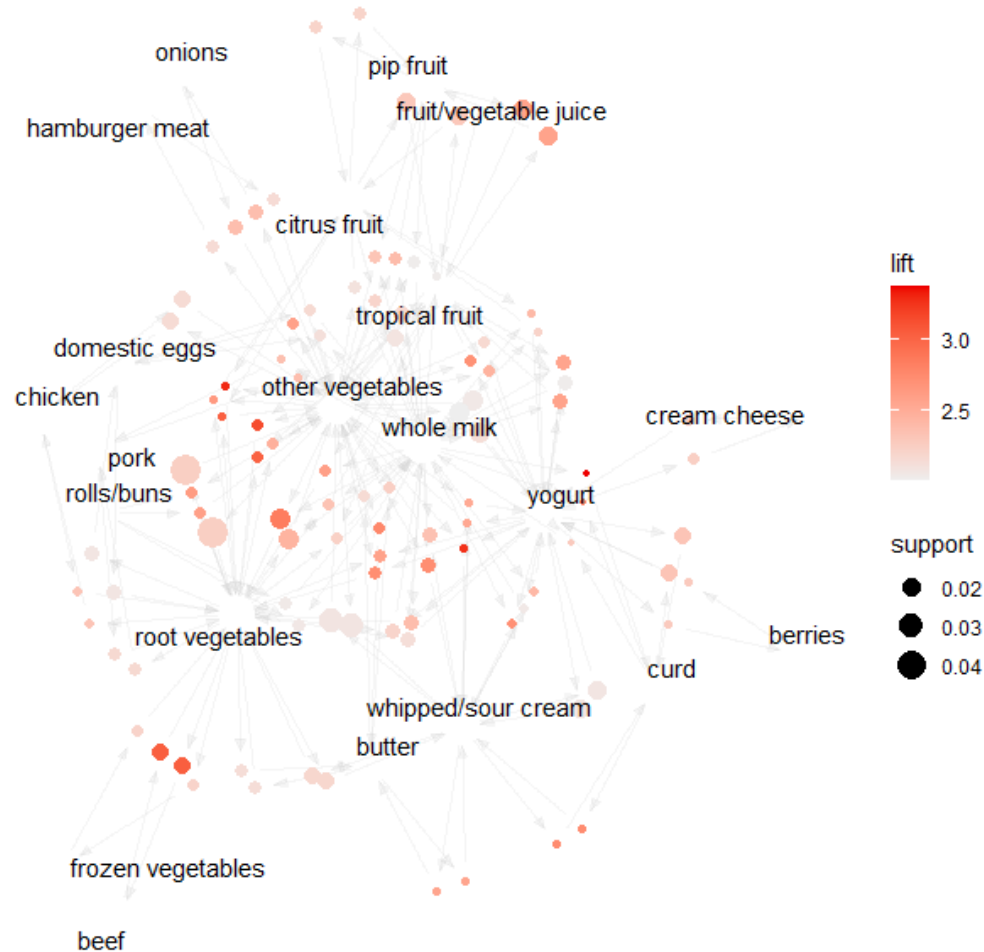
- Menggunakan algoritma apriori atau eclat, antara perkara penting yang perlu ditentukan dalam aturan sekutuan ialah:
  - i) Bagaimana untuk melihat item yang paling kerap dibeli?
  - ii) Bagaimana untuk mendapatkan Aturan Sekutuan yang signifikan bagi cadangan produk?
  - iii) Bagaimana untuk mendapatkan Aturan Sekutuan yang berkait dengan item tertentu?
  - iv) Bagaimana untuk mengvisualkan corak dan pola bagi aturan sekutuan yang signifikan?.





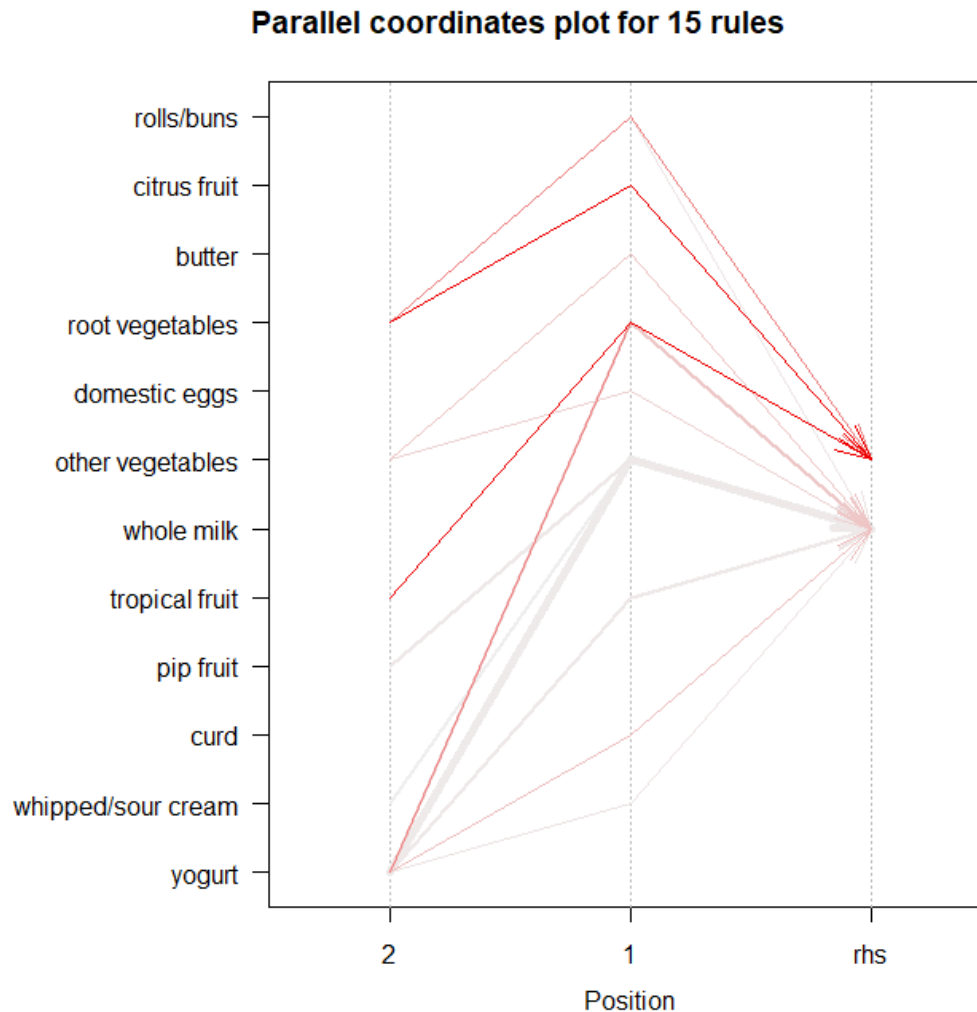
# PENGVISUALAN: PLOT RANGKAIAN

- Dalam plot ini, aturan sekutuan diwakili dalam bentuk rangkaian seperti berikut:



# PENGVISUALAN: PLOT KOORDINAT SELARI

- Plot ini digunakan untuk memahami hubungan dan corak antara atribut atau item yang berbeza dalam aturan sekutuan:



# RUJUKAN:

- Adamo, J-M. (2001). *Data Mining for Association Rules and Sequential Patterns: Sequential and Parallel Algorithms*. Springer.
- Aggarwal, C.C. (2015). *Data Mining: The Textbook*. New York: Springer.
- Hahsler, M., Grün, B., Hornik, K. (2005). arules - A Computational Environment for Mining Association Rules and Frequent Item Sets. *Journal of Statistical Software*, 14(15), 1–25.
- Makhabel, B. (2015). *Learning Data Mining with R: Develop key skills and techniques with R to create and customize data mining algorithms*. Birmingham: Packt Publishing.
- Olson, D. L., Lauhoff, G. (2019). *Descriptive Data Mining*. 2nd edition. Springer.
- Subramanian, G. (2017). *R Data Analysis Projects: Build end to end analytics systems to get deeper insights from your data*. Birmingham: Packt Publishing.
- Tan, P-G., Steinbach, M., Karpatne, A., Kumar, V. 2019. *Introduction to data mining*. 2nd edition. Pearson Education, Boston.



**TOPIK SETERUSNYA:**

# **Perlombongan Data Siri Masa**

