

LOB2203

**ระบบเทคโนโลยีสารสนเทศด้านโลจิสติกส์
สำหรับธุรกิจออนไลน์**

CH 5 - CH 6



rapidminer



**สาขาการจัดการโลจิสติกส์สำหรับธุรกิจออนไลน์
วิทยาลัยโลจิสติกส์และซัพพลายเชน มหาวิทยาลัยราชภัฏสุรินทร์**



Chapter 5

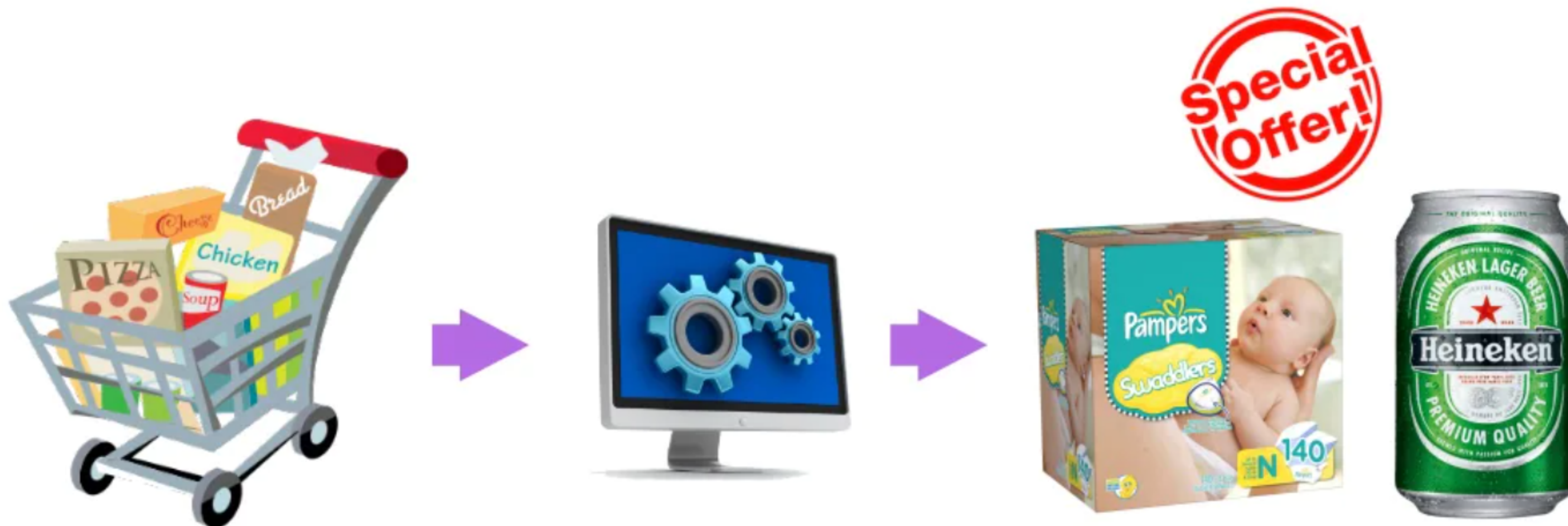
Association Rules





Market Basket Analysis

- เริ่มจากการสังเกตการซื้อสินค้าของลูกค้าแต่ละราย
- วิเคราะห์ข้อมูลเพื่อดูว่า สินค้าอะไรบ้างที่ลูกค้ามักจะซื้อพร้อมกันบ่อยๆ
- ใช้ในการออกโปรโมชั่นเพื่อเพิ่มยอดขาย





Market Basket Analysis

- Amazon.com example

Customers Who Bought This Item Also Bought

Page 1 of 20

The screenshot displays five book recommendations from Amazon. Each recommendation includes a book cover, title, author, star rating, number of reviews, and price. The books are:

- Data Mining: (The Morgan Kaufmann Series in Data Management Systems)** by Ian H. Witten. Kindle Edition, \$38.00, 61 reviews.
- Data Science for Business: What you need to know about data mining and...** by Foster Provost. Kindle Edition, \$21.99, 120 reviews. Labeled as "#1 Best Seller in Statistics".
- Data Science from Scratch: First Principles with Python** by Joel Grus. Kindle Edition, \$21.99, 33 reviews.
- R for Everyone: Advanced Analytics and Graphics (Addison-Wesley Data &...** by Jared Lander. Kindle Edition, \$21.79, 69 reviews.
- Web Scraping with Python: Collecting Data from the Modern Web** by Ryan Mitchell. Kindle Edition, \$17.49, 16 reviews.



Market Basket Analysis

- แปลงข้อมูลจาก POS database เป็น transaction database โดย group by ตามเวลาที่ซื้อสินค้า

TID	Transaction time	Product
1	01-13-2014 20:04	Apple
1	01-13-2014 20:04	Cereal
1	01-13-2014 20:04	Diapers
2	01-14-2014 11:30	Beer
2	01-14-2014 11:30	Cereal
2	01-14-2014 11:30	Eggs
3	01-15-2014 14:30	Apple
3	01-15-2014 14:30	Beer
3	01-15-2014 14:30	Cereal
3	01-15-2014 14:30	Eggs
4	01-16-2014 14:15	Beer
4	01-16-2014 14:15	Eggs

Transaction ID	Items
1	Apple, Cereal, Diapers
2	Beer, Cereal, Eggs
3	Apple, Beer, Cereal, Eggs
4	Beer, Eggs



Market Basket Analysis

- Support
 - นับจำนวนครั้งการซื้อสินค้าแต่ละชนิดคิดเป็น % ของการซื้อสินค้า
 - **Apple** พบว่ามีการซื้อในครั้งที่ 1 และ 3 support(**Apple**) คือ $2/4 = 50\%$
 - **Beer** พบว่ามีการซื้อในครั้งที่ 2, 3 และ 4 support(**Beer**) คือ $3/4 = 75\%$

Transaction ID	Items
1	Apple , Cereal, Diapers
2	Beer, Cereal, Eggs
3	Apple , Beer, Cereal, Eggs
4	Beer, Eggs



Items	Transaction ID				Support
	1	2	3	4	
Apple	1	0	1	0	$2/4 = 50\%$
Beer	0	1	1	1	$3/4 = 75\%$
Cereal	1	1	1	0	$3/4 = 75\%$
Diapers	1	0	0	0	$1/4 = 25\%$
Eggs	0	1	1	1	$3/4 = 75\%$



Market Basket Analysis

- frequent itemset
 - รูปแบบของการซื้อสินค้า (itemset) ที่มีค่า support \geq minimum support
 - กำหนดค่า minimum support = 50% ตัวอย่างของ Frequent Itemset คือ {Apple}, {Beer}, {Cereal}, {Eggs}
- infrequent itemset
 - รูปแบบของการซื้อสินค้า (itemset) ที่มีค่า support $<$ minimum support

Transaction ID	Items
1	Apple, Cereal, Diapers
2	Beer, Cereal, Eggs
3	Apple, Beer, Cereal, Eggs
4	Beer, Eggs

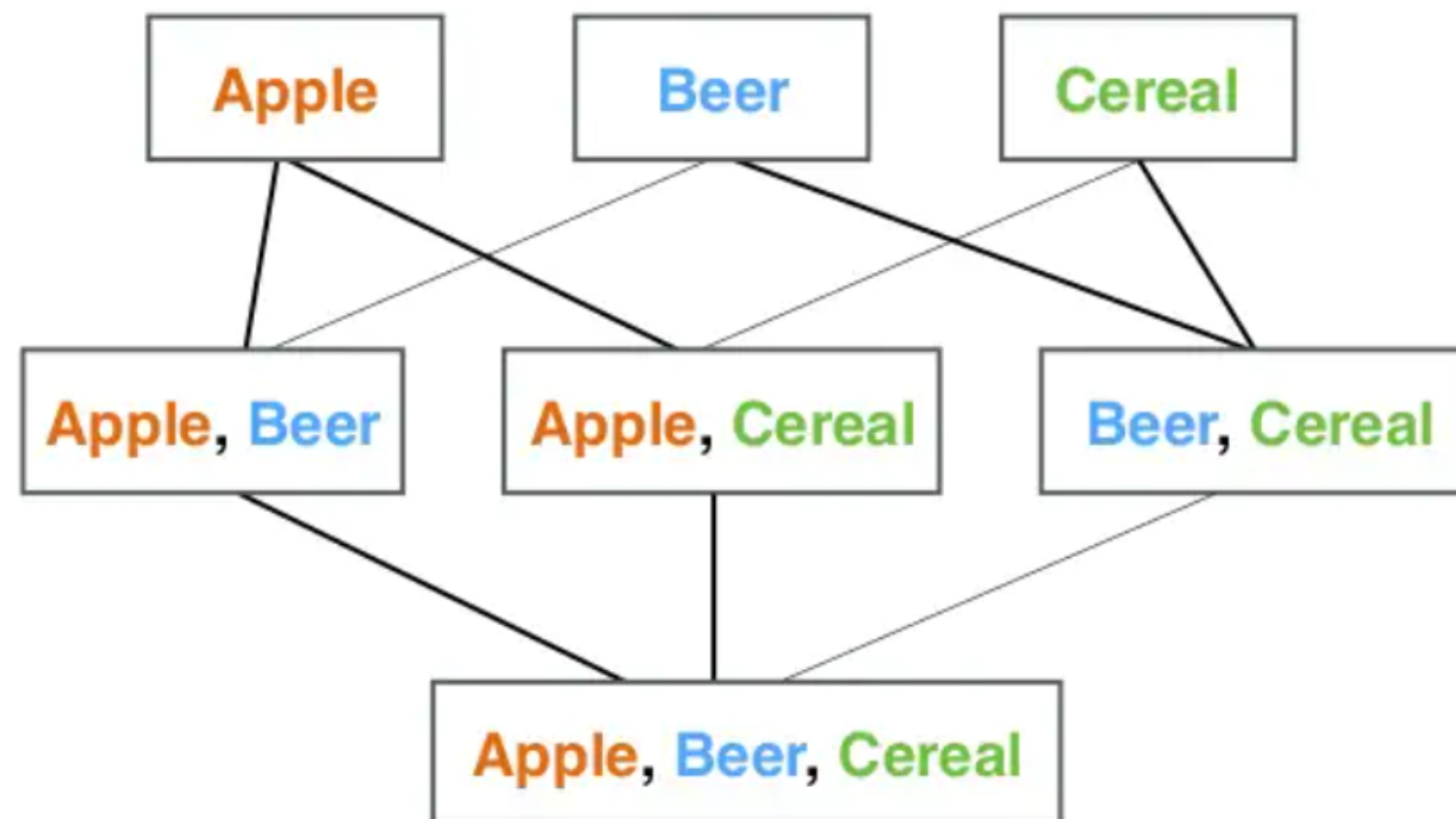


Items	Transaction ID				Support
	1	2	3	4	
Apple	1	0	1	0	2/4 = 50%
Beer	0	1	1	1	3/4 = 75%
Cereal	1	1	1	0	3/4 = 75%
Diapers	1	0	0	0	1/4 = 25%
Eggs	0	1	1	1	3/4 = 75%



Market Basket Analysis

- สินค้าใน supermarket มีจำนวน 3 ชนิด
- **Apple**, **Beer**, **Cereal**
- มีโอกาสที่เป็นไปได้ทั้งหมด (หรือ subset ทั้งหมด) คือ $2^3 - 1 = 7$ รูปแบบ หรือคำนวณได้จาก $2^N - 1$ โดยที่ N คือ จำนวนของสินค้า (item)





Market Basket Analysis

- การหากฎความสัมพันธ์ (association rules) มี 2 ขั้นตอนใหญ่ๆ
 - **ขั้นตอนที่ 1** หา frequent itemset ซึ่งใช้เวลานานกว่าขั้นตอนที่ 2
 - มี 2 เทคนิคที่นิยมใช้ คือ
 - เทคนิค Apriori
 - เทคนิค FP-Growth
 - **ขั้นตอนที่ 2** สร้างกฎความสัมพันธ์จาก frequent itemset ที่หาได้จากขั้นตอนที่ 1



Market Basket Analysis

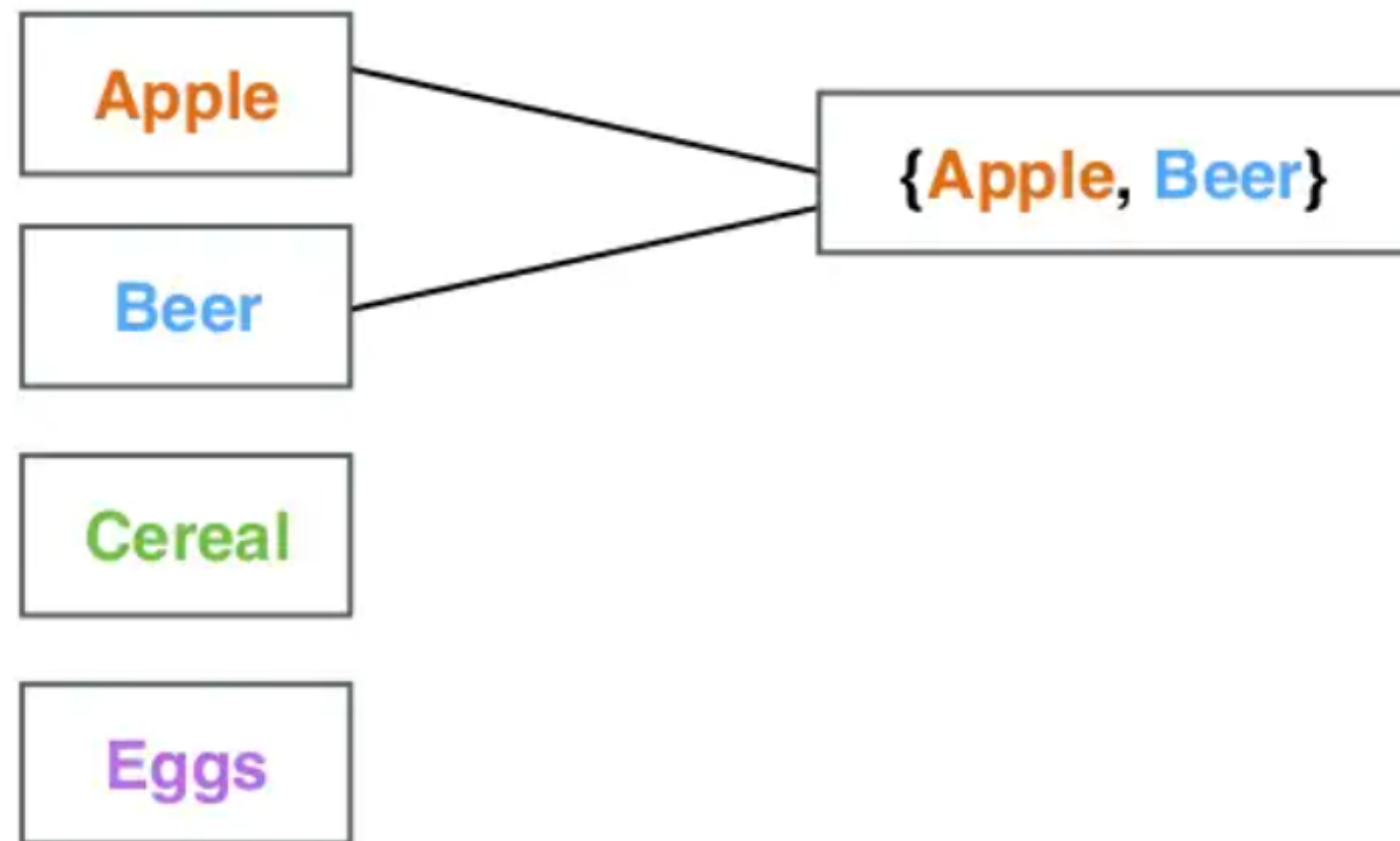
- เทคนิคการหากฎความสัมพันธ์ (Association Rules)
 - **Apriori** (Agrawal and Srikant, 1994)
 - สร้างรูปแบบของสินค้า (itemset) ที่มีจำนวนเพิ่มขึ้นทีละ 1
 - คำนวณค่า support จากในฐานข้อมูล
 - ข้อเสียคือต้องดึงข้อมูลจากฐานข้อมูลหลายรอบทำให้ทำงานช้า
 - FP-Growth (Han, Pei and Yin, 2000)
 - อ่านข้อมูลในฐานข้อมูลและสร้าง FP-tree
 - คำนวณค่า support จาก FP-tree
 - ทำงานได้เร็วกว่าวิธี Apriori

- R. Agrawal and R. Srikant, Fast algorithms for mining association rules in large databases, Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pages 487-499, Santiago, Chile, September 1994
- J. Han, H. Pei, and Y. Yin. Mining Frequent Patterns without Candidate Generation. In: Proc. Conf. on the Management of Data (SIGMOD'00, Dallas, TX). ACM Press, New York, NY, USA 2000



Apriori

- รูปแบบ (pattern) ที่มีสินค้า 2 ชนิด
 - รูปแบบที่มี **Apple** อยู่



Items	Transaction ID				Support
	1	2	3	4	
Apple	1	0	1	0	2/4 = 50%
Beer	0	1	1	1	3/4 = 75%

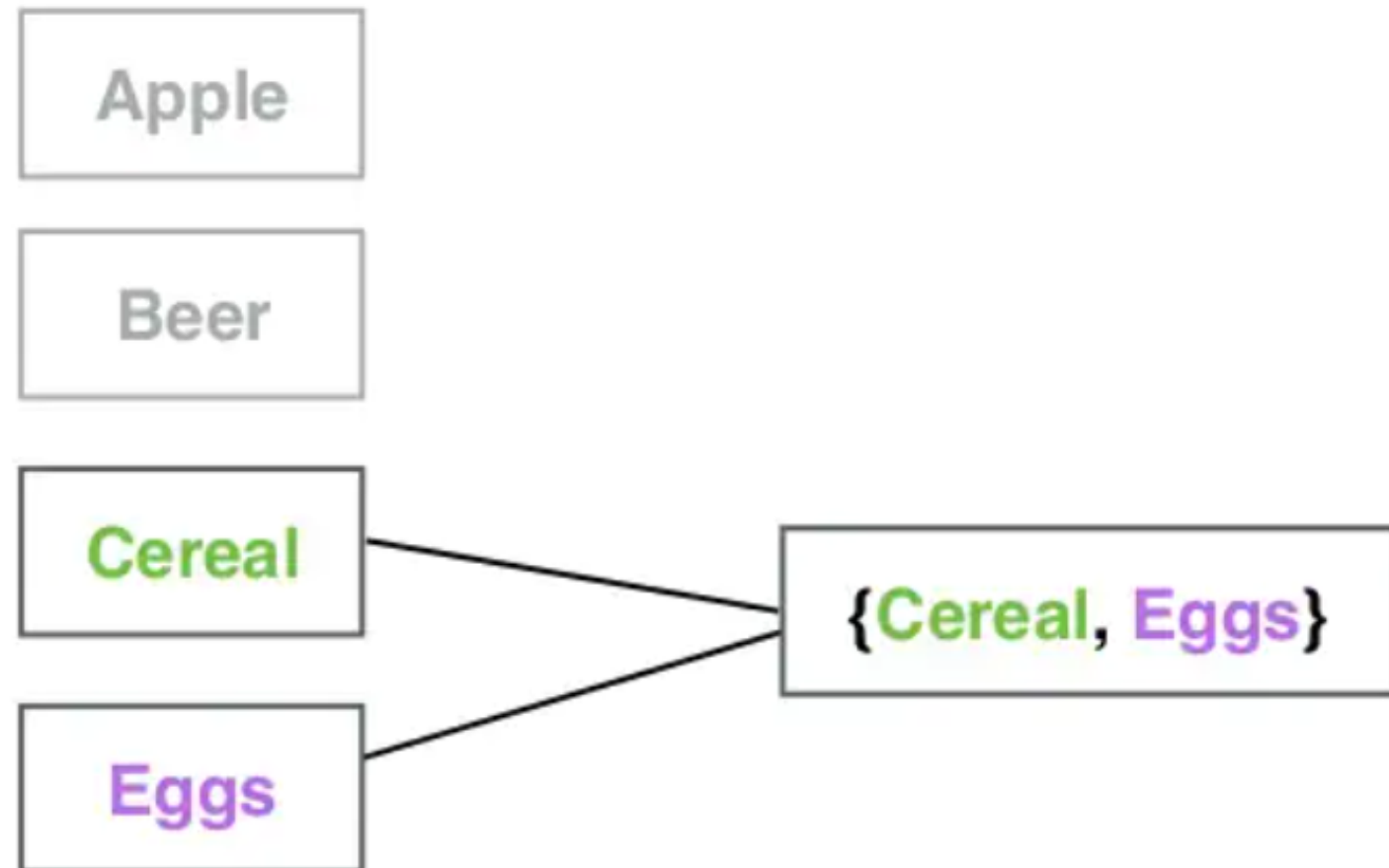


Itemset	Transaction ID				Support
	1	2	3	4	
{Apple, Beer}	0	0	1	0	1/4 = 25%



Apriori

- รูปแบบ (pattern) ที่มีสินค้า 2 ชนิด
 - รูปแบบที่มี **Cereal** อยู่



Items	Transaction ID				Support
	1	2	3	4	
Cereal	1	1	1	0	3/4 = 75%
Eggs	0	1	1	1	3/4 = 75%

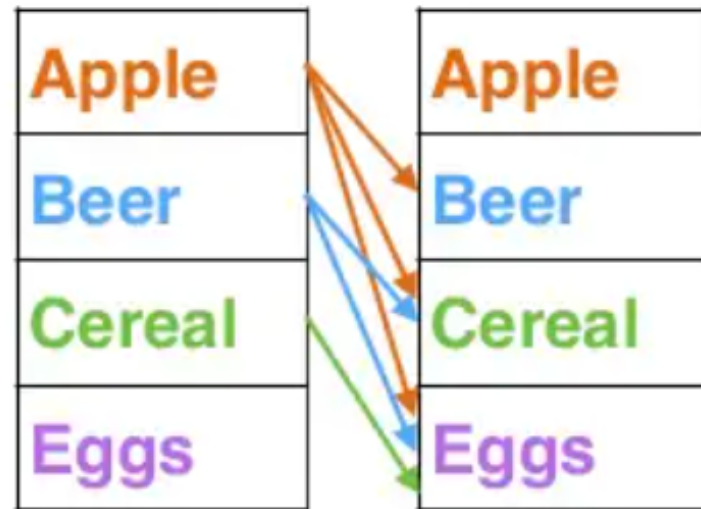


Itemset	Transaction ID				Support
	1	2	3	4	
{Cereal, Eggs}	0	1	1	0	2/4 = 50%



Apriori

- นับจำนวนครั้งการซื้อสินค้า 2 ชนิด



- {Apple, Beer}, {Apple, Cereal}, {Apple, Eggs}
- {Beer, Cereal}, {Beer, Eggs}
- {Cereal, Eggs}

Itemset	Transaction ID				Support
	1	2	3	4	
{Apple, Beer}	0	0	1	0	1/4 = 25%
{Apple, Cereal}	1	0	1	0	2/4 = 50%
{Apple, Eggs}	0	0	1	0	1/4 = 25%
{Beer, Cereal}	0	1	1	0	2/4 = 50%
{Beer, Eggs}	0	1	1	1	3/4 = 75%
{Cereal, Eggs}	0	1	1	0	2/4 = 50%



Itemset	Transaction ID				Support
	1	2	3	4	
{Apple, Beer}	0	0	1	0	1/4 = 25%
{Apple, Cereal}	1	0	1	0	2/4 = 50%
{Apple, Eggs}	0	0	1	0	1/4 = 25%
{Beer, Cereal}	0	1	1	0	2/4 = 50%
{Beer, Eggs}	0	1	1	1	3/4 = 75%
{Cereal, Eggs}	0	1	1	0	2/4 = 50%



Apriori

- สินค้าที่มีการซื้อมากกว่าหรือเท่ากับ 50% เรียกว่า frequent itemset

Frequent itemset	Support	Size
{Apple}	$2/4 = 50\%$	1
{Beer}	$3/4 = 75\%$	1
{Cereal}	$3/4 = 75\%$	1
{Eggs}	$3/4 = 75\%$	1
{Apple, Cereal}	$2/4 = 50\%$	2
{Beer, Cereal}	$2/4 = 50\%$	2
{Beer, Eggs}	$3/4 = 75\%$	2
{Cereal, Eggs}	$2/4 = 50\%$	2
{Beer, Cereal, Eggs}	$2/4 = 50\%$	3

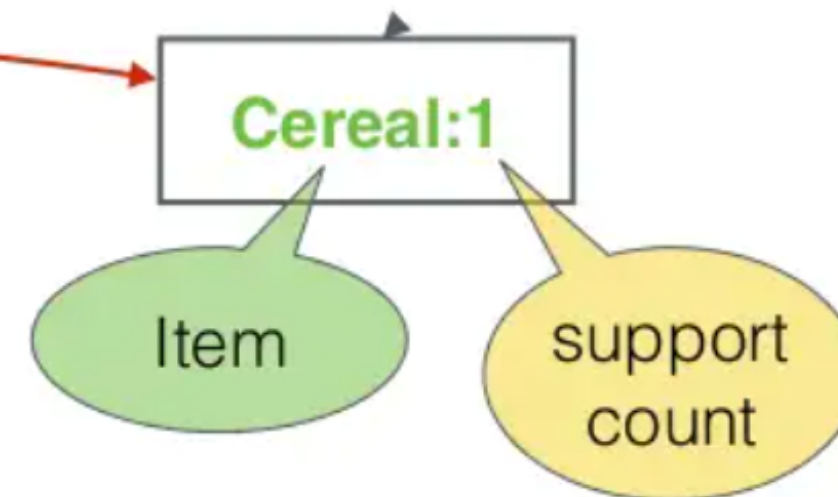


FP-Growth

- **ขั้นตอนที่ 1** สร้าง compact data structure ที่เรียกว่า FP-tree
 - การสร้าง FP-Tree อ่านข้อมูลจากฐานข้อมูลเพียงแค่ 2 ครั้งเท่านั้น
 - หาค่า support ของรูปแบบการซื้อสินค้า (item) ที่มีความยาว 1
 - เรียงลำดับตามค่า support ที่คำนวณได้
 - สร้าง FP-Tree จากการอ่านข้อมูลในฐานข้อมูลเพียงหนึ่งรอบ

NULL

Transaction ID	Items
1	Cereal, Apple, Diapers
2	Beer, Cereal, Eggs
3	Beer, Cereal, Eggs, Apple
4	Beer, Eggs





FP-Growth

- **ขั้นตอนที่ 1** สร้าง compact data structure ที่เรียกว่า FP-tree
 - การสร้าง FP-Tree อ่านข้อมูลจากฐานข้อมูลเพียงแค่ 2 ครั้งเท่านั้น
 - หาค่า support ของรูปแบบการซื้อสินค้า (item) ที่มีความยาว 1
 - เรียงลำดับตามค่า support ที่คำนวณได้
 - สร้าง FP-Tree จากการอ่านข้อมูลในฐานข้อมูลเพียงหนึ่งรอบ

Transaction ID	Items
1	Cereal, Apple, Diapers
2	Beer, Cereal, Eggs
3	Beer, Cereal, Eggs, Apple
4	Beer, Eggs



NULL

Cereal:1

Apple:1

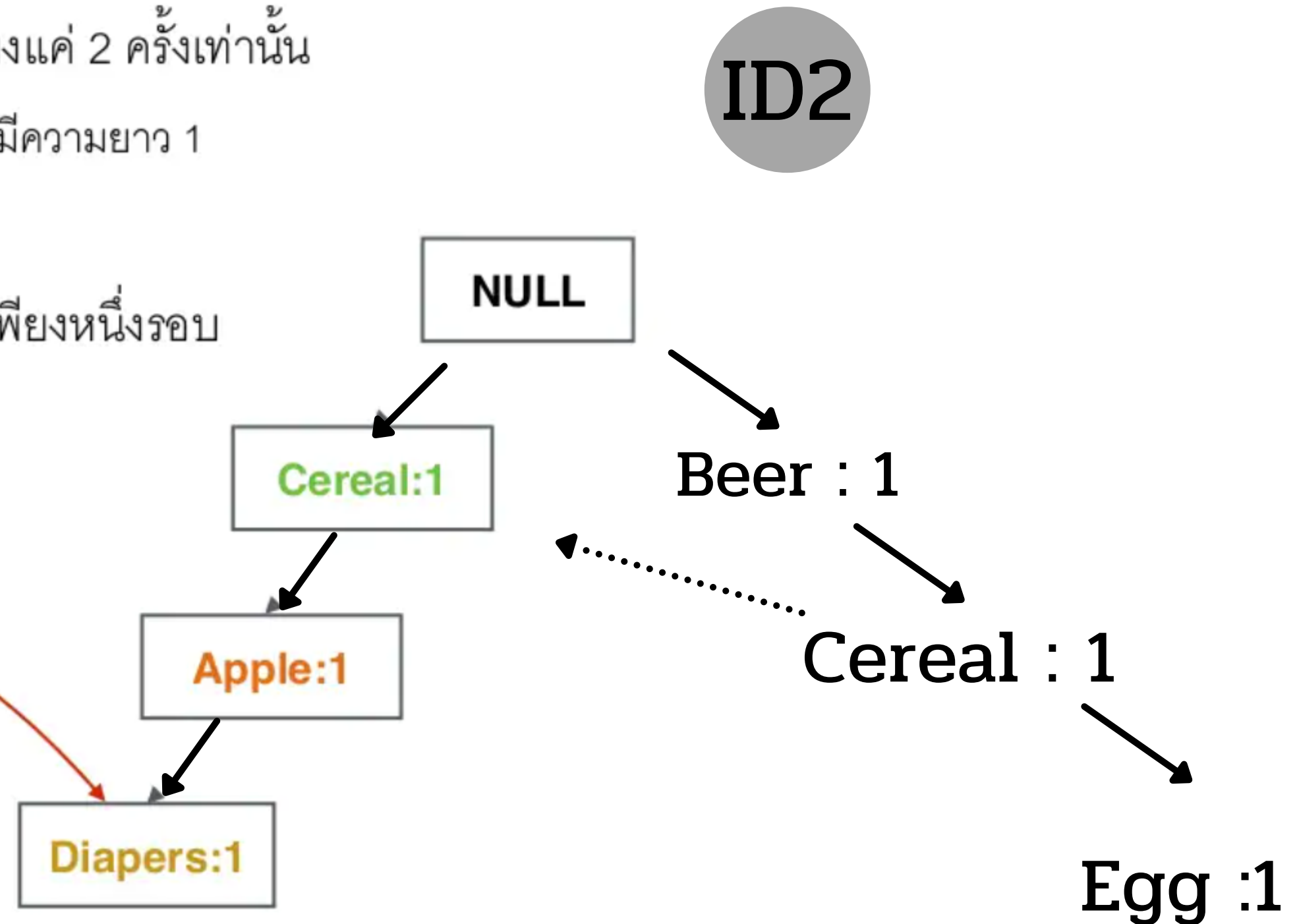
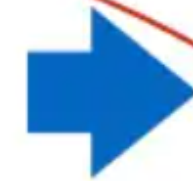




FP-Growth

- **ขั้นตอนที่ 1** สร้าง compact data structure ที่เรียกว่า FP-tree
 - การสร้าง FP-Tree อ่านข้อมูลจากฐานข้อมูลเพียงแค่ 2 ครั้งเท่านั้น
 - หาค่า support ของรูปแบบการซื้อสินค้า (item) ที่มีความยาว 1
 - เรียงลำดับตามค่า support ที่คำนวณได้
 - สร้าง FP-Tree จากการอ่านข้อมูลในฐานข้อมูลเพียงหนึ่งรอบ

Transaction ID	Items
1	Cereal, Apple, Diapers
2	Beer, Cereal, Eggs
3	Beer, Cereal, Eggs, Apple
4	Beer, Eggs

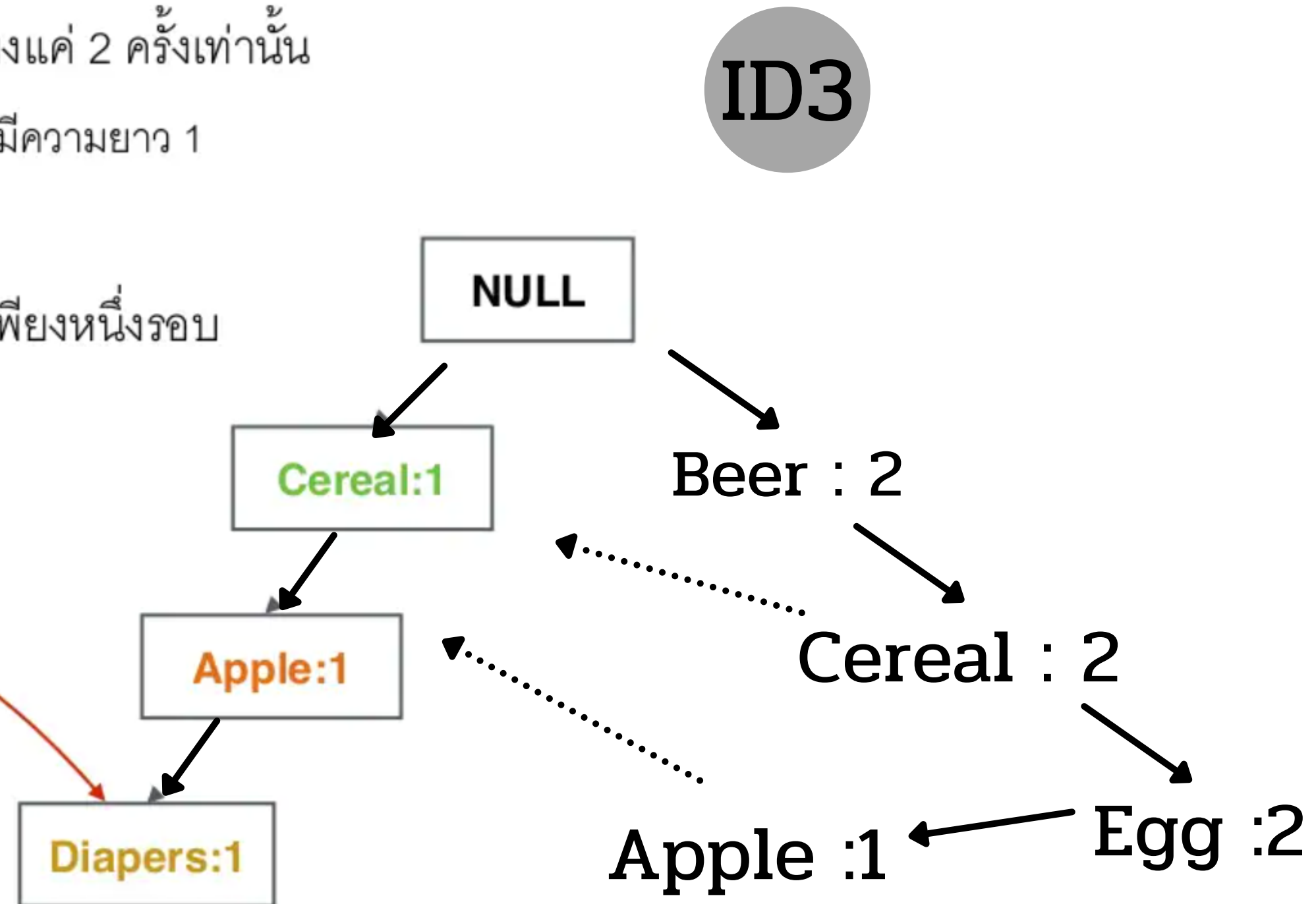




FP-Growth

- **ขั้นตอนที่ 1** สร้าง compact data structure ที่เรียกว่า FP-tree
 - การสร้าง FP-Tree อ่านข้อมูลจากฐานข้อมูลเพียงแค่ 2 ครั้งเท่านั้น
 - หาค่า support ของรูปแบบการซื้อสินค้า (item) ที่มีความยาว 1
 - เรียงลำดับตามค่า support ที่คำนวณได้
 - สร้าง FP-Tree จากการอ่านข้อมูลในฐานข้อมูลเพียงหนึ่งรอบ

Transaction ID	Items
1	Cereal, Apple, Diapers
2	Beer, Cereal, Eggs
3	Beer, Cereal, Eggs, Apple
4	Beer, Eggs

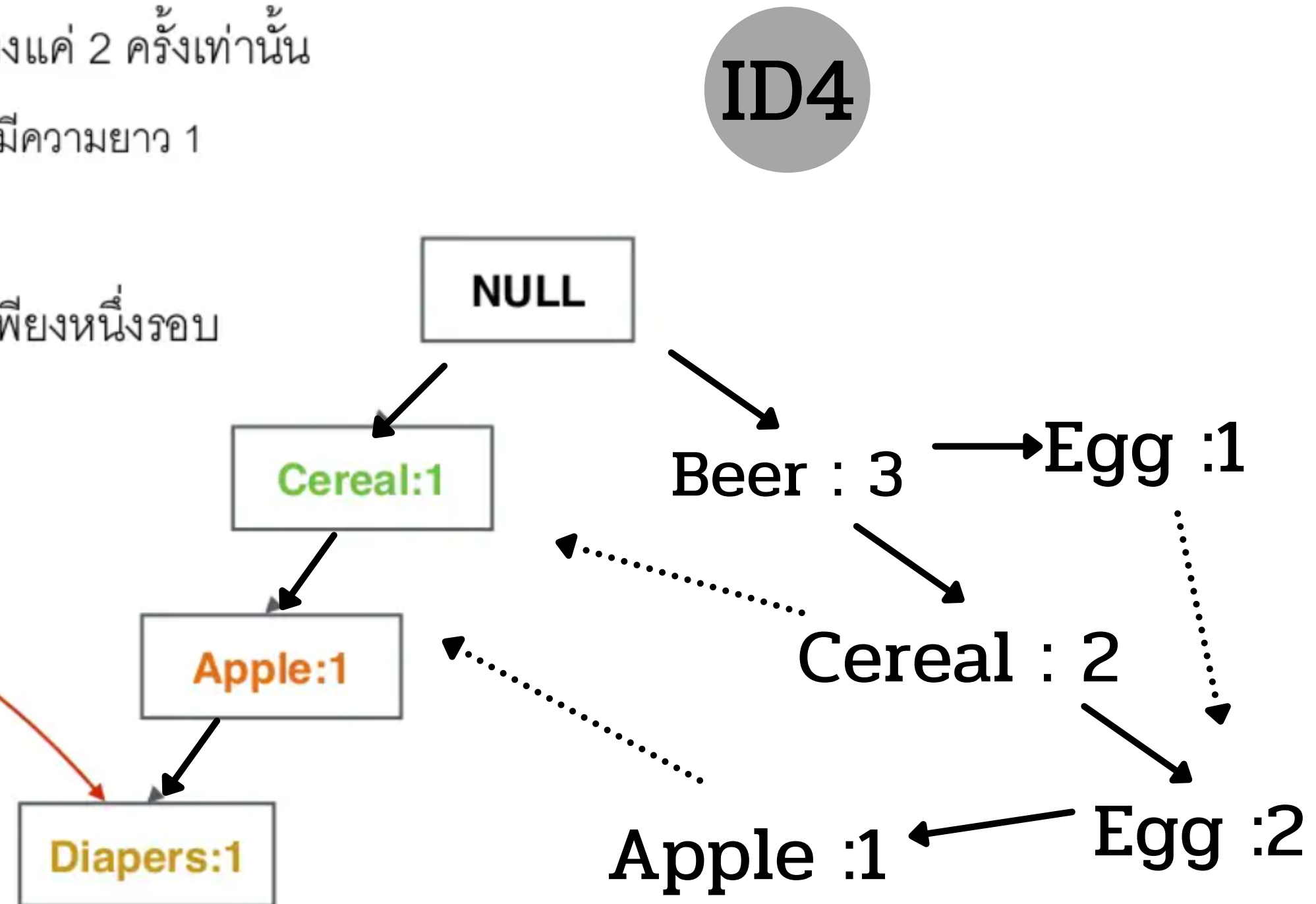
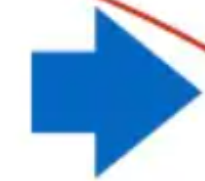




FP-Growth

- **ขั้นตอนที่ 1** สร้าง compact data structure ที่เรียกว่า FP-tree
 - การสร้าง FP-Tree อ่านข้อมูลจากฐานข้อมูลเพียงแค่ 2 ครั้งเท่านั้น
 - หาค่า support ของรูปแบบการซื้อสินค้า (item) ที่มีความยาว 1
 - เรียงลำดับตามค่า support ที่คำนวณได้
 - สร้าง FP-Tree จากการอ่านข้อมูลในฐานข้อมูลเพียงหนึ่งรอบ

Transaction ID	Items
1	Cereal, Apple, Diapers
2	Beer, Cereal, Eggs
3	Beer, Cereal, Eggs, Apple
4	Beer, Eggs





Apriori & FP-Growth

Frequent itemset	Support		Frequent itemset	Support
{Apple}	2/4 = 50%		{Apple}	2/4 = 50%
{Beer}	3/4 = 75%		{Apple, Cereal}	2/4 = 50%
{Cereal}	3/4 = 75%		{Eggs}	3/4 = 75%
{Eggs}	3/4 = 75%		{Cereal, Eggs}	2/4 = 50%
{Apple, Cereal}	2/4 = 50%		{Beer, Eggs}	3/4 = 75%
{Beer, Cereal}	2/4 = 50%		{Beer, Cereal, Eggs}	2/4 = 50%
{Beer, Eggs}	3/4 = 75%		{Cereal}	3/4 = 75%
{Cereal, Eggs}	2/4 = 50%		{Beer, Cereal}	2/4 = 50%
{Beer, Cereal, Eggs}	2/4 = 50%		{Beer}	3/4 = 75%

ผลลัพธ์จาก Apriori

ผลลัพธ์จาก FP-Growth



Association rules

- กฎความสัมพันธ์ (association rules)
 - สร้างจากสินค้าที่ลูกค้าซื้อบ่อยๆ
 - รูปแบบของกฎความสัมพันธ์ คือ **LHS** **RHS**
 - LHS** คือ Left Hand Side สินค้าที่ซื้อพร้อมกันบ่อยๆ ด้านซ้ายของกฎ
 - RHS** คือ Right Hand Side สินค้าที่ซื้อพร้อมกันบ่อยๆ ด้านขวาของกฎ

Frequent itemset	Support	Size
{Apple, Cereal}	2/4 = 50%	2
{Beer, Cereal}	2/4 = 50%	2
{Beer, Eggs}	3/4 = 75%	2
{Cereal, Eggs}	2/4 = 50%	2
{Beer, Cereal, Eggs}	2/4 = 50%	3



Apple	Cereal
Beer	Cereal
Beer	Eggs
Cereal	Apple
Cereal	Eggs
Cereal, Eggs	Beer
Eggs	Beer



Association rules

- ตัววัดประสิทธิภาพของกฎความสัมพันธ์
- Confidence
 - แสดงความเชื่อมั่นของกฎความสัมพันธ์ที่เมื่อรูปแบบ **LHS** เกิดขึ้นแล้วรูปแบบ **RHS** จะเกิดขึ้นด้วยเป็นจำนวนกี่เปอร์เซ็นต์

$$\text{confidence}(\mathbf{LHS} \quad \mathbf{RHS}) = \frac{\text{support}(\mathbf{LHS}, \mathbf{RHS})}{\text{support}(\mathbf{LHS})}$$

$$\begin{aligned} \text{confidence}(\mathbf{Apple} \quad \mathbf{Cereal}) &= \frac{\text{support}(\mathbf{Apple}, \mathbf{Cereal})}{\text{support}(\mathbf{Apple})} \\ &= \frac{2/4}{2/4} \\ &= 100\% \end{aligned}$$



Data preparation in RM 7

- โหลดไฟล์ pos.csv จาก thumb drive เข้าไปใน Repositories
- แปลงข้อมูลจาก POS database เป็นตาราง

TID	Transaction time	Product
1	01-13-2014 20:04	Apple
1	01-13-2014 20:04	Cereal
1	01-13-2014 20:04	Diapers
2	01-14-2014 11:30	Beer
2	01-14-2014 11:30	Cereal
2	01-14-2014 11:30	Eggs
3	01-15-2014 14:30	Apple
3	01-15-2014 14:30	Beer
3	01-15-2014 14:30	Cereal
3	01-15-2014 14:30	Eggs
4	01-16-2014 14:15	Beer
4	01-16-2014 14:15	Eggs

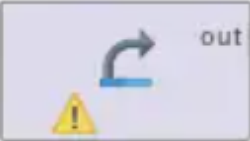
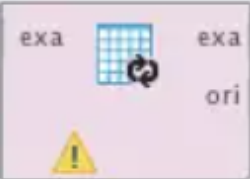
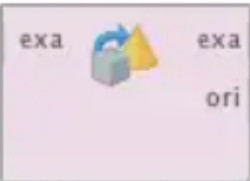


TID	Apple	Beer	Cereal	Diapers	Eggs
1	TRUE	FALSE	TRUE	TRUE	FALSE
2	FALSE	TRUE	TRUE	FALSE	TRUE
3	TRUE	TRUE	TRUE	FALSE	TRUE
4	FALSE	TRUE	FALSE	FALSE	TRUE



Data preparation in RM 7

- ไอเปอเรเตอร์ที่เกี่ยวข้อง

ไอเปอเรเตอร์		คำอธิบาย
	Retrieve	ใช้สำหรับดึงข้อมูลออกจาก Repositories
	Pivot	ใช้สำหรับสร้างตารางในรูปแบบ Pivot
	Numerical to Binominal	ใช้สำหรับแปลงข้อมูลตัวเลขให้เป็นประเภท Binominal



Data preparation

- เปลี่ยนพารามิเตอร์ของโอเปอเรเตอร์ Pivot ดังนี้
 - group attribute เป็นค่า id
 - index attribute เป็นค่า name

The screenshot displays the Datacube interface for configuring a Pivot operator. It is divided into three main sections:

- Operators:** A search bar contains 'pivot'. A tree view shows 'Blending (2)' > 'Table (2)' > 'Rotation (2)' > 'Pivot' (highlighted with a blue circle '2').
- Process:** A workflow diagram showing 'Retrieve pos' (1) connected to 'Pivot' (3). The Pivot operator has 'exa' and 'ori' ports. A blue circle '4' points to the output 'res'.
- Parameters:** A configuration panel for the Pivot operator with the following settings:
 - group attribute: id
 - index attribute: name
 - consider weights:
 - weight aggrega...: sum



Data preparation

<new process*> – RapidMiner Studio Educational 9.9.002 @ TANAWAT

File Edit Process View Connections Settings Extensions Help

Repository

+ Import Data

New - Repository - TANAWAT (Local)

- Connections
- data
 - 4. Replace Value - Done (8/25)

Operators

num

- Attributes (12)
- Types (11)
 - Numerical to Binominal
 - Numerical to Polynominal

Import Data - Format your columns.

Format your columns.

Replace errors with missing values ⓘ

	TID <i>integer id</i>	Apple <i>polynominal</i>	Beer <i>polynominal</i>	Cereal <i>polynominal</i>	Diapers <i>polynominal</i>	Eggs <i>polynominal</i>
1	1	true	false	true	true	false
2	2	false	true	true	false	true
3	3	true	true	true	false	true
4	4	false	true	false	false	true

PREVIEW



Data preparation in RM 7

- เลือกโอเปอเรเตอร์ Numerical to Binominal มาใส่ใน Process
 - เนื่องจากการหาความสัมพันธ์ไม่สามารถทำงานกับตัวเลข (numeric) ได้ครับ
 - เพื่อแปลงข้อมูลตัวเลข ให้เป็นค่า TRUE/FALSE

The screenshot displays the Rapid Miner 7 interface with three main panels:

- Operators:** A search bar contains 'numerical to binominal'. Under the 'Types (1)' category, the 'Numerical to Binominal' operator is highlighted with a blue circle containing the number 5.
- Process:** A workflow diagram showing three operators: 'Retrieve pos', 'Pivot', and 'Numerical to Binominal'. The 'Numerical to Binominal' operator is highlighted with a blue circle containing the number 6.
- Parameters:** The configuration panel for the 'Numerical to Binominal' operator. It shows 'attribute filter ty...' set to 'all', 'invert selection' and 'include special attributes' as unchecked options, and 'min' and 'max' values both set to '0.0'.

A watermark 'facebook.com/datacube.th' is visible at the bottom of the Process panel.



Data preparation in RM 7

- เลือกโอเปอเรเตอร์ Numerical to Binominal มาใส่ใน Process
 - เนื่องจากการหาความสัมพันธ์ไม่สามารถทำงานกับตัวเลข (numeric) ได้ครับ
 - เพื่อแปลงข้อมูลตัวเลข ให้เป็นค่า TRUE/FALSE

The screenshot displays the Rapid Miner 7 interface with three main panels:

- Operators:** A search bar contains 'numerical to binominal'. Under the 'Types (1)' category, the 'Numerical to Binominal' operator is highlighted with a blue circle labeled '5'.
- Process:** A workflow diagram showing three operators: 'Retrieve pos', 'Pivot', and 'Numerical to Binominal'. The 'Numerical to Binominal' operator is highlighted with a blue circle labeled '6'.
- Parameters:** The configuration panel for the 'Numerical to Binominal' operator. It shows 'attribute filter ty...' set to 'all', 'invert selection' and 'include special attributes' as unchecked options, and 'min' and 'max' values both set to '0.0'.

A watermark 'facebook.com/datacube.th' is visible at the bottom of the Process panel.



Data preparation

- ผลการแปลงจาก POS database

ExampleSet (4 examples, 0 special attributes, 5 regular attributes)

Row No.	value_Apple	value_Beer	value_Cereal	value_Diap...	value_Eggs
1	true	false	true	true	false
2	?	true	true	?	true
3	true	true	true	?	true
4	?	true	?	?	true

Filter (4 / 4 examples): all



Data preparation

- ไอเปอเรเตอร์ที่เกี่ยวข้อง

ไอเปอเรเตอร์		คำอธิบาย
	FP-Growth	ใช้สำหรับหา frequent itemset ด้วยเทคนิค FP-Growth
	Create Association Rules	ใช้สำหรับสร้างกฎความสัมพันธ์จาก frequent itemset ที่ค้นพบ



Data preparation

- ใช้ข้อมูลที่แปลงได้เพื่อหากฎความสัมพันธ์
- ใช้โอเปอเรเตอร์ FP-Growth เพื่อหา frequent itemset
 - เปลี่ยนพารามิเตอร์ min support เป็น 0.5 หรือ 50%

The screenshot displays the Orange3 interface for configuring the FP-Growth operator. On the left, the 'Operators' panel shows the 'FP-Growth' operator selected under 'Associations (1)', indicated by a blue circle with the number 7. The main 'Process' window shows a workflow with three operators: 'Retrieve pos', 'Pivot', and 'Numerical to Binom...', followed by the 'FP-Growth' operator. A blue circle with the number 8 is positioned below the 'FP-Growth' operator. On the right, the 'Parameters' panel for 'FP-Growth' is shown, with the 'min support' parameter set to '0.5', highlighted by a blue circle with the number 9. A watermark 'facebook.com/datacube.th' is visible at the bottom of the interface.



Data preparation in RM 7

- ใช้ไอคอนเรเตอร์ Create Association Rules เพื่อสร้างกฎความสัมพันธ์
- เปลี่ยนพารามิเตอร์ min confidence เป็น 0.6 หรือ 60%

The screenshot displays the RM 7 interface for creating association rules. On the left, the 'Operators' panel shows 'Create Association Rules' selected under 'Modeling (1) > Associations (1)', indicated by a blue circle '10'. The central 'Process' window shows a workflow: 'Retrieve pos' (out) -> 'Pivot' (exa, ori) -> 'Numerical to Binom...' (exa, ori) -> 'FP-Growth' (exa, fre) -> 'Create Association Rules' (ite, rul, ite). A blue circle '11' points to the 'Create Association Rules' operator. On the right, the 'Parameters' panel for 'Create Association Rules' shows 'criterion' set to 'confidence' and 'min confidence' set to '0.6', with a blue circle '12' pointing to the 'min confidence' field.



Data preparation

- แสดง frequent itemset ที่หาได้

The screenshot shows the RapidMiner Studio Professional 7.0.000 interface. The 'FrequentItemSets (FP-Growth)' tab is active, displaying a table of frequent itemsets. The table has columns for 'Size', 'Support', and 'Item 1'. The 'Item 1' column is highlighted with a yellow box and an arrow pointing to the text 'คลิกที่ header ของแต่ละคอลัมน์เพื่อ sort'. The 'Update View' button is highlighted with an orange box and the text 'ส่วน filter'. The table content is highlighted with a green box and the text 'ส่วน frequent itemset'.

Size	Support	Item 1
1	0.750	value_Eggs
1	0.750	value_Cereal
1	0.750	value_Beer
1	0.500	value_Apple
2	0.500	value_Eggs value_Cereal
2	0.750	value_Eggs value_Beer
2	0.500	value_Cereal value_Beer
2	0.500	value_Cereal value_Apple
3	0.500	value_Eggs value_Cereal value_Beer



Data preparation

- แสดงกฎความสัมพันธ์ที่หาได้

Result History x FrequentItemSets (FP-Growth) x AssociationRules (Create Association Rules) x

Views: Design Results Questions?

Show rules matching
all of these conclusions: value_Eggs, value_Cereal, value_Beer, value_Apple

ส่วน filter

No.	Premises	Conclusion	Support	Confidence
1	value_Eggs	value_Cereal	0.500	0.667
2	value_Cereal	value_Eggs	0.500	0.667
3	value_Cereal	value_Beer	0.500	0.667
4	value_Beer	value_Cereal	0.500	0.667
5	value_Cereal	value_Apple	0.500	0.667
6	value_Eggs	value_Cereal, value_Beer	0.500	0.667
7	value_Cereal	value_Eggs, value_Beer	0.500	0.667
8	value_Beer	value_Eggs, value_Cereal	0.500	0.667
9	value_Eggs, value_Beer	value_Cereal	0.500	0.667
10	value_Eggs	value_Beer	0.750	1
11	value_Beer	value_Eggs	0.750	1
12	value_Apple	value_Cereal	0.500	1
13	value_Eggs, value_Cereal	value_Beer	0.500	1
14	value_Cereal, value_Beer	value_Eggs	0.500	1

ส่วนแสดงกฎความสัมพันธ์



Model

The screenshot displays the RapidMiner Studio Educational 9.9.002 @ TANAWAT interface. The main workspace shows a workflow model in the Design view, consisting of the following operators: Retrieve, Pivot, Numerical to Binomial, FP-Growth, and Create Association Rules. The process is currently at 75% completion. The left sidebar contains the Repository and Operators panels. The right sidebar shows the Parameters panel for the selected operator, with the following settings:

Parameter	Value
logverbosity	init
logfile	
resultfile	
random seed	2001
send mail	never
encoding	SYSTEM



Step 1

TID	Apple	Beer	Cereal	Diapers	Eggs
1	TRUE	FALSE	TRUE	TRUE	FALSE
2	FALSE	TRUE	TRUE	FALSE	TRUE
3	TRUE	TRUE	TRUE	FALSE	TRUE
4	FALSE	TRUE	FALSE	FALSE	TRUE



	A	B	C	D	E	F
1	TID	Apple	Beer	Cereal	Diapers	Eggs
2	1	1	0	1	1	0
3	2	0	1	1	0	1
4	3	1	1	1	0	1
5	4	0	1	0	0	1
6						



Step 2 - Import Data

Import Data - Select the data location.

Select the data location.

Rapid Miner

Bookmarks	File Name	Size	Type	Last Modified
★ -- Last Directory	1. Book1.csv	1 KB	Microsoft Excel Co...	Aug 24, 2021
	1. CSV 1.csv	1 KB	Microsoft Excel Co...	Aug 24, 2021
	1. CSV 1.txt	1 KB	Text Document	Aug 24, 2021
	1. Sunny Play.txt	1 KB	Text Document	Aug 11, 2021
	2. Book Excel.xlsx	3 KB	Microsoft Excel Wor...	Aug 24, 2021
	2. Write CSV.csv	1 KB	Microsoft Excel Co...	Aug 24, 2021
	3. Replace Value.xlsx	8 KB	Microsoft Excel Wor...	Aug 25, 2021
	4. Replace Value - Done.xlsx	8 KB	Microsoft Excel Wor...	Aug 25, 2021
	5. Outlier.xlsx	8 KB	Microsoft Excel Wor...	Aug 25, 2021
	6. Outlier - update.xlsx	8 KB	Microsoft Excel Wor...	Aug 25, 2021
	7. Asso Rule.xlsx	8 KB	Microsoft Excel Wor...	Aug 25, 2021
	8. Asso Rule - Copy.xlsx	8 KB	Microsoft Excel Wor...	Aug 25, 2021
	Book Excel.xlsx	3 KB	Microsoft Excel Wor...	Aug 24, 2021
	customer churn data.xlsx	24 KB	Microsoft Excel Wor...	Oct 17, 2019

8. Asso Rule - Copy.xlsx

All Files

The selected file will be imported as: Excel [Change](#)

← Previous → Next ✖ Cancel



Step 3 - Set TID = id

//New - Repository - TANAWAT/process/Ass* - RapidMiner Studio Educational 9.9.002 @ TANAWAT

File Edit Process View Connections Settings Extensions Help

Import Data - Format your columns.

Format your columns.

Replace errors with missing values ⓘ

	TID <i>integer</i> <i>id</i>	Apple <i>integer</i>	Beer <i>integer</i>	Cereal <i>integer</i>	Diapers <i>integer</i>	Eggs <i>integer</i>
1	1	1	0	1	1	0
2	2	0	1	1	0	1
3	3	1	1	1	0	1
4	4	0	1	0	0	1

Repository

+ Import Data

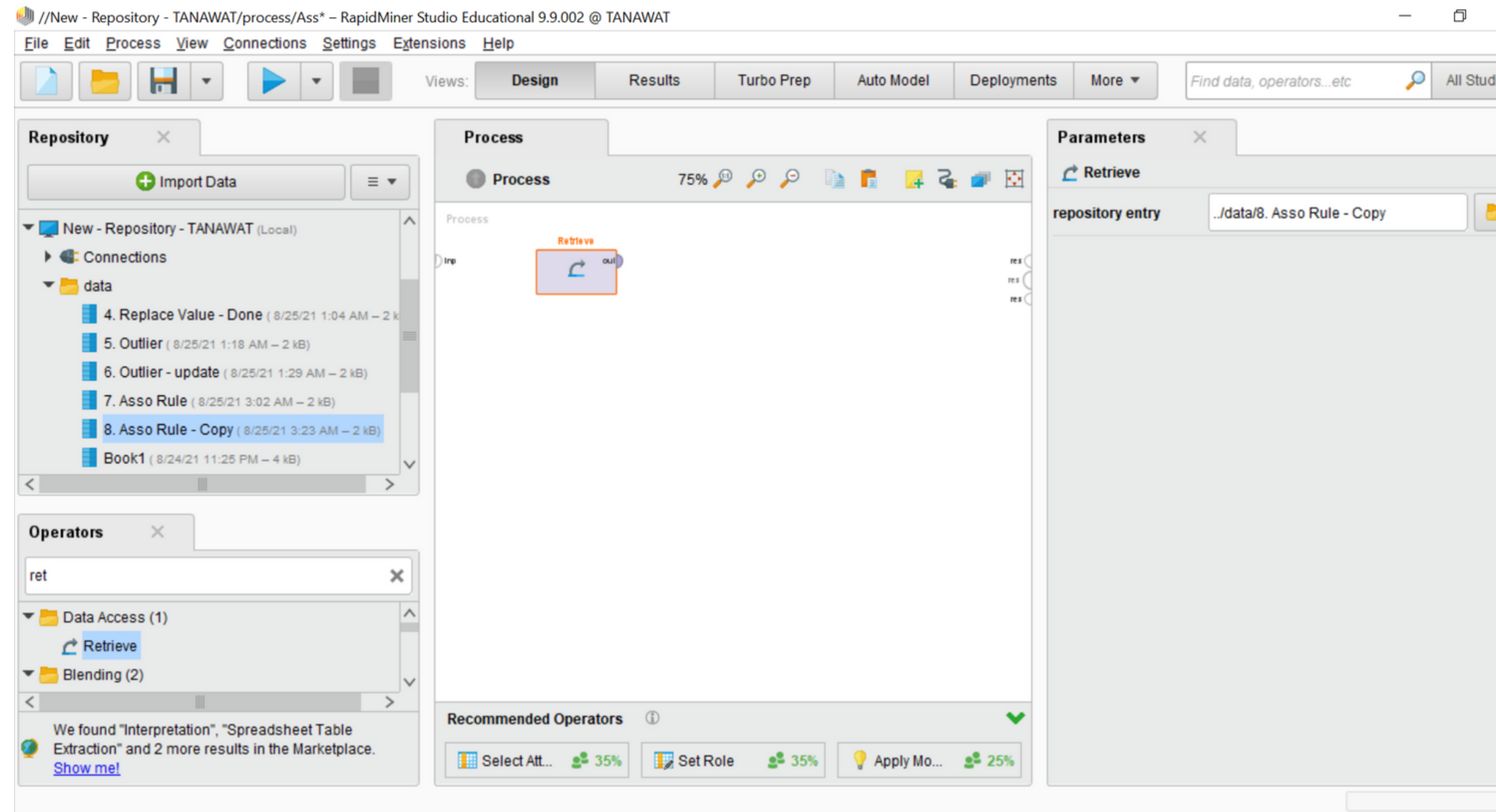
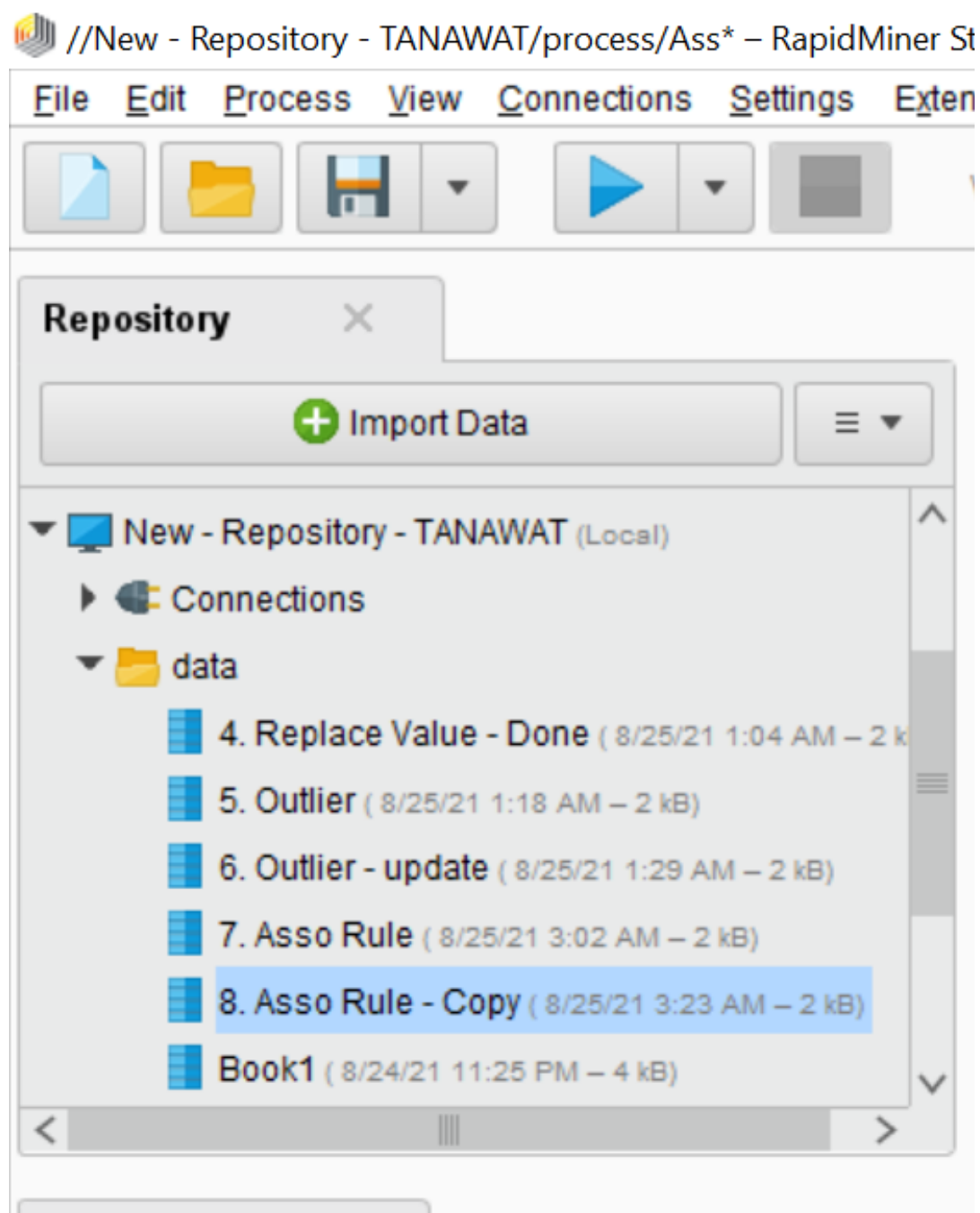
- 6. Outlier - update (8/25/21 1:2
- 7. Asso Rule (8/25/21 3:02 AM
- 8. Asso Rule - Copy (8/25/21 3
- Book1 (8/24/21 11:25 PM - 4 kB

Operators

Create Ass



Step 4 - ดึงข้อมูลมาใช้ด้วย Retrieve





Step 5 - ใช้ Pivot ตั้งค่าตารางข้อมูลที่จำเป็น

The screenshot displays the RapidMiner Studio interface. The main workspace shows a process flow with a 'Retrieve' operator connected to a 'Pivot' operator. The 'Pivot' operator is highlighted with an orange box, and its configuration panel is open on the right. The 'Parameters' panel for the 'Pivot' operator is visible, showing options for 'group by attributes', 'column grouping attribute', and 'aggregation attributes'. The 'Select Attributes...' button is highlighted with an orange box, indicating the step of selecting the data for the pivot operation.

Repository: New - Repository - TANAWAT (Local)

- Connections
- data
 - 4. Replace Value - Done (8/25/21 1:04 AM - 2 kB)
 - 5. Outlier (8/25/21 1:18 AM - 2 kB)
 - 6. Outlier - update (8/25/21 1:29 AM - 2 kB)
 - 7. Asso Rule (8/25/21 3:02 AM - 2 kB)
 - 8. Asso Rule - Copy (8/25/21 3:23 AM - 2 kB)
 - Book1 (8/24/21 11:25 PM - 4 kB)

Process: Retrieve -> Pivot

Parameters: Pivot

- group by attributes
- column grouping attribute: Enter value...
- aggregation attributes: Edit List (0)...
- use default aggregation



Step 5 - ใช้ Pivot ตั้งค่าตารางข้อมูลที่จำเป็น

The screenshot displays the RapidMiner Studio interface with the 'Pivot' operator selected in the Operators panel. A 'Select Attributes: group by attributes' dialog box is open, showing the configuration for the pivot operation. The dialog has two main sections: 'Attributes' and 'Selected Attributes'. The 'Attributes' section contains a search bar and a list with one item, '# TID'. The 'Selected Attributes' section also has a search bar and a list with five items: '# Apple', '# Beer', '# Cereal', '# Diapers', and '# Eggs'. The dialog includes 'Apply' and 'Cancel' buttons at the bottom right. The background shows the 'Repository' and 'Operators' panels, with the 'Pivot' operator highlighted in the 'Rotation' category.



Step 6 - นำ Num to Binominal แปลงค่า 0 และ 1

The screenshot displays the RapidMiner Studio interface. The main workspace shows a process flow with three operators: Retrieve, Pivot, and Numerical to Binominal. The 'Numerical to Binominal' operator is selected, and its parameters are visible in the right-hand pane. The parameters are:

- attribute filter type: all
- invert selection:
- include special attributes:
- min: 0.0
- max: 0.0

The 'Repository' pane on the left shows a list of data files, with '8. Asso Rule - Copy' selected. The 'Process' pane at the top shows the current process is 'Process' with a progress indicator at 75%.



Step 7 - นำ FP-Growth หาคความถี่การใช้งาน ที่ 50% ขึ้น

The screenshot shows the RapidMiner Studio interface with the following components:

- Repository:** Lists data files such as '4. Replace Value - Done', '5. Outlier', '6. Outlier - update', '7. Asso Rule', '8. Asso Rule - Copy', 'Book1', and 'CSV 1'.
- Process:** A workflow diagram showing the sequence: Retrieve (input: inp) → Pivot (output: out) → Numerical to Binomial (input: inp, output: out) → FP-Growth (input: inp, output: res).
- Parameters:** Configuration for the FP-Growth operator:
 - input format: items in dummy coded ...
 - positive value: (empty)
 - min requirement: support
 - min support: 0.5
 - min items per itemset: 1
 - max items per itemset: 0
 - max number of itemsets: 1000000
 - find min number of itemsets:
 - must contain list: Edit Enumeration (0)...
 - must contain regexp: (empty)
- Operators:** A search bar with 'FP' and a list of operators including 'Modeling (1)', 'Associations (1)', and 'FP-Growth'.

กำหนด 0.5

เอาคลิกออก

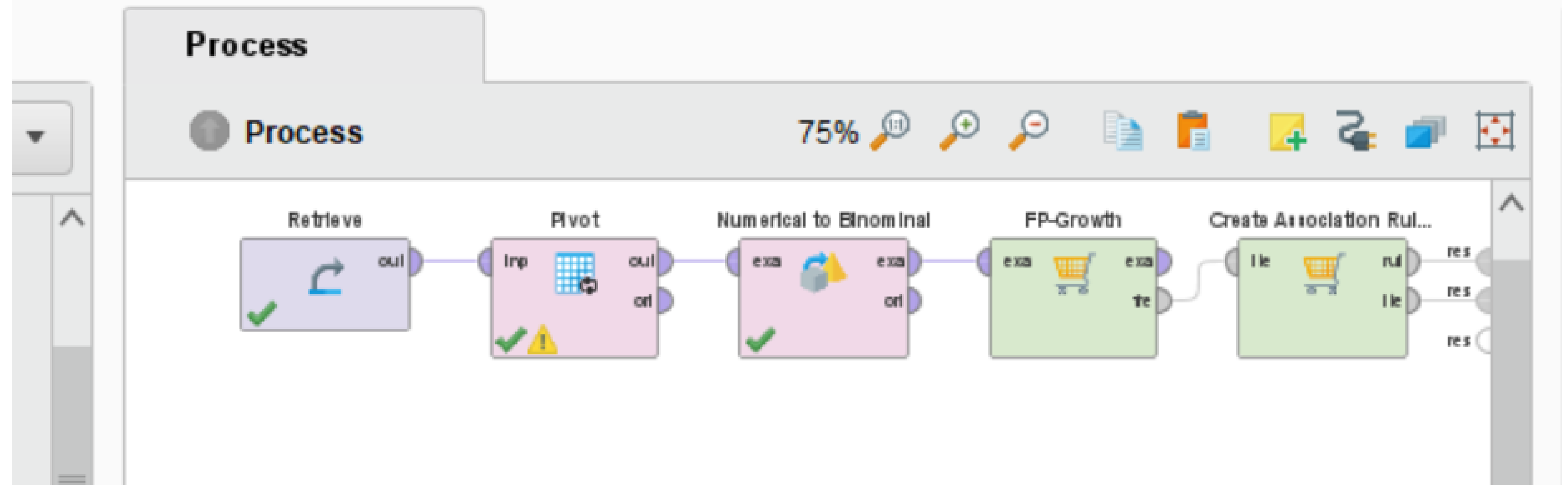


Step 8 - นำ Create Association Rules หารูปแบบความสัมพันธ์

Educational 9.9.002 @ TANAWAT

s [Help](#)

Design Results Turbo Prep Auto Model Deployments More ▾





Step 9 - อ่านค่าความสัมพันธ์ FP

//New - Repository - TANAWAT/process/Ass* - RapidMiner Studio Educational 9.9.002 @ TANAWAT

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Deployments More

ExampleSet (//New - Repository - TANAWAT/data/8. Asso Rule - Copy)

Result History **FrequentItemSets (FP-Growth)** AssociationRules (Create Association Rules)

Data

No. of Sets: 9
 Total Max. Size: 3

Min. Size:
 Max. Size:
 Contains Item:

Size	Support	Item 1	Item 2	Item 3
1	0.750	Beer		
1	0.750	Cereal		
1	0.750	Eggs		
1	0.500	Apple		
2	0.500	Beer	Cereal	
2	0.750	Beer	Eggs	
2	0.500	Cereal	Eggs	
2	0.500	Cereal	Apple	
3	0.500	Beer	Cereal	Eggs

Annotations



Step 10 - อ่านค่าความสัมพันธ์ Asso

The screenshot shows the RapidMiner Studio interface. The main window displays the results of an Association Rules process. The 'Results' view is active, showing a table of association rules. The table has three columns: 'Premises', 'Conclusion', and 'Sup'. The rules are as follows:

Premises	Conclusion	Sup
Beer	Eggs	0.75
Eggs	Beer	0.75
Apple	Cereal	0.50
Beer, Cereal	Eggs	0.50
Cereal, Eggs	Beer	0.50

On the left side, there is a sidebar with icons for 'Data', 'Graph', 'Description', and 'Annotations'. Below the 'Data' icon, there is a section titled 'Show rules matching' with a dropdown menu set to 'all of these conclusions:'. Below this, the items 'Beer', 'Cereal', and 'Eggs' are listed. At the bottom left, there is a 'Min. Criterion:' label with an input field.

On the right side, there is a 'Repository' panel showing a tree view of the data source. The 'data' folder is expanded, showing several files, including '8. Asso Rule - Copy' which is highlighted.



Step 10 - อ่านค่าความสัมพันธ์ Asso

The screenshot displays the RapidMiner Studio interface. The main window shows the results of an Association Rules process. The central area contains a graph with nodes for 'Apple', 'Eggs', 'Beer', 'Cereal', and five rules. The rules are labeled with their confidence and support values: Rule 1 (0.750 / 1.000), Rule 2 (0.750 / 1.000), Rule 3 (0.500 / 1.000), Rule 4 (0.500 / 1.000), and Rule 5 (0.500 / 1.000). The left sidebar shows the 'AssociationRules (Create Association Rules)' process, and the right sidebar shows the 'Repository' with a list of data sources and processes.

Repository

- Import Data
- Training Resources (connected)
- Samples
- Community Samples (connected)
- Local Repository (Local)
- New - Repository - TANAWAT (Local)
 - Connections
 - data
 - 4. Replace Value - Done (8/25/21)
 - 5. Outlier (8/25/21 1:18 AM - 2 kB)
 - 6. Outlier - update (8/25/21 1:29 AM - 2 kB)
 - 7. Asso Rule (8/25/21 3:02 AM - 2 kB)
 - 8. Asso Rule - Copy (8/25/21 3:21 AM - 2 kB)
 - Book1 (8/24/21 11:25 PM - 4 kB)
 - CSV 1 (8/24/21 11:02 PM - 4 kB)
 - Sunny Play (8/24/21 11:03 PM - 2 kB)
 - process
- Temporary Repository (Local)
- DB (Legacy)