Association Rule - Apriori Algorithm

Generate rules using Apriori Algorithm. Consider the values as SUPPORT = 50% & CONFIDENCE = 75%

ID	Items
1	Bread, Butter, Jam, Milk
2	Bread, Butter, Milk
3	Bread, Juice, Curd
4	Bread, Milk, Juice
5	Butter, Milk, Juice

Total 5 transactions.

* Itemset Support Count (4) (%)

Item	Count	Support (count/total no. of transaction)	
Bread	4	4/5 = 80%	✓
Butter	3	3/5 = 60%	✓
Jam	1	1/5 = 20%	X
Milk	4	4/5 = 80%	✓
Juice	3	3/5 = 60%	✓
Curd	1	1/5 = 20%	X

Support value = 50% (Given in Question)

2. Itemsets Support Count (L2)

Find the combinations of valid transactions

Item	Count	Support
{Bread, Butter}	2	2/5 = 40%. ✗
{Bread, Milk}	3	3/5 = 60%. ✓
{Bread, Juice}	2	2/5 = 40%. ✗
{Butter, Milk}	3	3/5 = 60%. ✓
{Butter, Juice}	1	1/5 = 20%. ✗
{Milk, Juice}	2	2/5 = 40%. ✗

Valid transactions

{Bread, Milk}, {Butter, Milk}

If we can combine these also.

{Bread, Milk, Butter}

count = 2

support = 2/5 = 40%. ✗

So will consider combination upto 2.

confidence(x → y) = s(x ∪ y) / s(x)

{Bread, Milk}

Bread → Milk

Confidence = support(Bread ∪ Milk) / support(Bread) = 60%/80% = 75%.

Milk → Bread

Confidence = 60%/80% = 75%

Confidence (Butter → Milk) = 60/80 = 100%

Confidence (Milk → Butter) = 60/80 = 75%

In Question the threshold value of Confidence = 75%

So these rules are valid

Bread → Milk

If a person buys bread, 75% probability for that person to buy milk

Milk → Bread

Butter → Milk

Milk → Butter

8 Consider the transaction dataset of a store where each transaction contains the list of items purchased by the customer. Our goal is to find frequent set of items that are purchased by the customers and generate association rules for them.

Transaction id	Item purchased
T_1	I_1, I_2, I_3
T_2	I_2, I_5
T_3	I_4, I_5
T_4	I_1, I_2, I_5
T_5	I_2, I_3, I_5

Minimum support count = 2 and confidence = 50%.

Solution

Step 1: Item set Count

Create a table which has support count of all items in the transaction database

Item	Support Count	
I_1	2	✓
I_2	4	✓
I_3	2	✓
I_4	1	x
I_5	4	✓

So frequent -1 Itemsets = I_1, I_2, I_3, I_5

2. 2- Itemsets.

Check pairs of the frequent 1- Itemsets.

Itemset	Support Count	
I_1, I_2	2	✓
I_1, I_3	1	
I_1, I_5	1	
$I_2 I_3$	2	✓
$I_2 I_5$	3	✓
$I_3 I_5$	1	

valid item set by removing itemset having support count < 2

I_1, I_2 , $I_2 I_3$, $I_2 I_5$

3 3-Itemset

Itemset	Count		
$I_1\ I_2\ I_3$	1	x	All are having
$I_1\ I_2\ I_5$	1	x	count < 2
$I_2\ I_3\ I_5$	1	x	

Since no more frequent items, go back to the previous itemset.

$I_1\ I_2$	$I_2\ I_1$
$I_2\ I_3$	$I_3\ I_2$
$I_2\ I_5$	$I_5\ I_2$

Find confidence of all rules.

Confidence $(I_1 \rightarrow I_2) =$

$$\frac{\text{Support}(I_1 \cup I_2)}{\text{Support}(I_1)}$$

$$= \frac{2}{2} = 1 = 100\% \ \checkmark$$

Confidence $(I_2 \rightarrow I_1)$

$$= \frac{2}{4} = 50\%$$ ✓

Confidence $(I_2 \rightarrow I_3)$

$$= \frac{2}{4} = 50\%$$ ✓

Confidence $(I_3 \rightarrow I_2)$

$$= \frac{2}{2} = 100\%$$ ✓

Confidence $(I_2 \rightarrow I_5)$

$$= \frac{3}{4} = 75\%$$ ✓

Confidence $(I_5 \rightarrow I_2)$

$$= \frac{3}{4} = 75\%$$ ✓

Given that minimum Confidence = 50%.

So all association rules has confidence $\geq 50\%$.
so all rules are strong association rules.

Calculate strong association rules.

$$\text{Lift}(x \to y) = \frac{\text{Confidence}(x \to y)}{\text{support}(y)}$$

$$\text{Lift}(I_1 \to I_2) = \frac{\text{Confidence}(I_1 \to I_2)}{\text{Support}(I_2)}$$

$$= \frac{100}{4} = 25\%$$

Lift

Association Rules	Lift (%)
$I_1 \to I_2$	25%
$I_2 \to I_1$	25%
$I_2 \to I_3$	25%
$I_3 \to I_2$	25%
$I_2 \to I_5$	18.75%
$I_5 \to I_2$	18.75%