

Assignment - 2

Course Code :- CAP446

Name :- Sirjanpreet Kaur

Section :- D2112

Roll no :- B56

Reg. No:- 12107974

Assignment -2

1.

Ques: For the following given transaction dataset, generate rules using Apriori algorithm. Consider values of SUPPORT = 30% and CONFIDENCE = 60%.

TID	PRODUCTS			
1.	Milk	Egg	Bread	Butter
2.	Milk	Butter	Egg	Ketchup
3.	Bread	Butter	Ketchup	
4.	Milk	Bread	Butter	
5.	Bread	Butter	Cookies	
6.	Milk	Bread	Butter	Cookies
7.	Milk	Cookies		
8.	Milk	Bread	Butter	
9.	Bread	Butter	Egg	Cookies
10.	Milk	Butter	Bread	
11.	Milk	Bread	Butter	
12.	Milk	Bread	Cookies	Ketchup

Ans: SUPPORT = 30% CONFIDENCE = 60%

TID	PRODUCTS			
1	Milk	Egg	Bread	Butter
2	Milk	Butter	Egg	Ketchup
3	Bread	Butter	Ketchup	
4	Milk	Bread	Butter	
5	Bread	Butter	Cookies	
6	Milk	Bread	Butter	Cookies
7	Milk	Cookies		
8	Milk	Bread	Butter	

2.

TID		PRODUCTS		
9	Bread	Butter	Egg	Cookies
10	Milk	Butter	Bread	
11	Milk	Bread	Butter	
12	Milk	Bread	Cookies	Ketchup

Now, we are going to count each item and write their occurrences in each transaction:-

Frequent Items :-

Items	Frequency
Milk	9
Egg	3
Bread	10
Butter	10
Ketchup	3
Cookies	5

We have to calculate support of each item, the formula of support is:-

$$\text{Support} = \frac{\text{Frequency of item}}{\text{Total Transactions}}$$

Items	Frequency	Support
Milk	9	$(9/12) \times 100 = 75\%$
Egg	3	$(3/12) \times 100 = 25\%$
Bread	10	$(10/12) \times 100 = 83.33\%$
Butter	10	$(10/12) \times 100 = 83.33\%$
Ketchup	3	$(3/12) \times 100 = 25\%$
Cookies	5	$(5/12) \times 100 = 41.66\%$

3.

We now eliminate or discard items Egg and ketchup as those are not supporting MINSUPP 30%. So we will not consider these items.

Now, we are going to create Frequent 2-item sets:

Item-set	Frequency	Support
{Milk, Bread}	7	(7/12) X 100 = 58.3%
{Milk, Butter}	7	(7/12) X 100 = 58.3%
{Milk, Cookies}	3	(3/12) X 100 = 25%
{Bread, Butter}	9	(9/12) X 100 = 75%
{Bread, Cookies}	4	(4/12) X 100 = 33.33%
{Butter, Cookies}	3	(3/12) X 100 = 25%

We are going to eliminate item-sets {Milk, Cookies} and {Butter, Cookies} as these are not supporting MINSUPP i.e. 30%. These itemsets have SUPPORT less than MINSUPP. So, we are discarding those itemsets.

Now, we are going to create Frequent 3-item sets:

Item-Set	Frequency	Support
{Milk, Bread, Butter}	6	(6/12) X 100 = 50%
{Milk, Bread, Cookies}	2	(2/12) X 100 = 16.66%
{Milk, Butter, Cookies}	1	(1/12) X 100 = 8.33%
{Bread, Butter, Cookies}	3	(3/12) X 100 = 25%

We are going to eliminate which do not support MINSUPP i.e. 30%. We only consider item-set {Milk, Bread, Butter}.

Rule Generation:-

Now, we have frequent 3 itemset i.e.
 $\{\text{Milk}, \text{Bread}, \text{Butter}\}$. We need to calculate
 confidence. The given CONFIDENCE = 60%.

$$\text{Confidence}(A \rightarrow B) = \frac{\text{support}(A \cup B)}{\text{support}(A)}$$

1. Confidence ($\{\text{Milk}\} \rightarrow \{\text{Bread}, \text{Butter}\}$) -

$$= \frac{\text{support}(\{\text{Milk}\} \cup \{\text{Bread}, \text{Butter}\})}{\text{support}(\{\text{Milk}\})}$$

$$= \frac{6/12}{9/12} = 66.67\%$$

It supports confidence (Min confidence). So we can consider this frequent itemset.

2. Confidence ($\{\text{Bread}\} \rightarrow \{\text{Milk}, \text{Butter}\}$) = $\frac{\text{support}(\{\text{Bread}\} \cup \{\text{Milk}, \text{Butter}\})}{\text{support}(\{\text{Bread}\})}$

$$= \frac{6/12}{10/12} = 60\%$$

It supports min. confidence. So we can consider this itemset.

3. Confidence ($\{\text{Butter}\} \rightarrow \{\text{Milk}, \text{Bread}\}$) -

$$\frac{\text{support}(\{\text{Butter}\} \cup \{\text{Milk}, \text{Bread}\})}{\text{support}(\{\text{Butter}\})}$$

5.

$$\begin{aligned}
 &= \frac{6}{12} \\
 &= \frac{10}{12} \\
 &= 60\%
 \end{aligned}$$

It supports min confidence. So we can consider this frequent itemset.

4. Confidence ($\{\text{Milk, Bread}\} \rightarrow \{\text{Butter}\}$) =

$$\frac{\text{support}(\{\text{Milk, Bread}\} \cup \{\text{Butter}\})}{\text{support}(\{\text{Milk, Bread}\})}$$

$$\begin{aligned}
 &= \frac{6}{12} \\
 &= \frac{7}{12} \\
 &= 85.7\%
 \end{aligned}$$

It supports min confidence. So we can consider this frequent itemset

5. Confidence ($\{\text{Milk, Butter}\} \rightarrow \{\text{Bread}\}$) =

$$\frac{\text{support}(\{\text{Milk, Butter}\} \cup \{\text{Bread}\})}{\text{support}(\{\text{Milk, Butter}\})}$$

$$\begin{aligned}
 &= \frac{6}{12} \\
 &= \frac{7}{12}
 \end{aligned}$$

$$= 85.7\%$$

It supports min confidence. So, we can consider this frequent itemset.

6.

6. Confidence (of Bread, Butter \rightarrow Milk) =

support (of Bread, Butter) / support (of Milk)

support (of Bread, Butter)

$$= \frac{6/12}{9/12}$$

$$= 66.67\%$$

It supports min confidence, so we can consider this frequent item-set.

Ques 2: Prepare a FP growth tree and all followed steps for the following transactions. Given Minimum support = 7

<u>Transaction No.</u>	<u>Products</u>
1	beer, wine, cheese
2	beer, potato chips
3	eggs, flour, butter, cheese
4	eggs, flour, butter, beer, potato chips
5	wine, cheese
6	potato chips
7	eggs, flour, butter, wine, cheese
8	eggs, flour, butter, beer, potato chips
9	wine, beer
10	beer, potato chips

7.

<u>Transaction</u>	<u>Products</u>
<u>No.</u>	
11	butter, eggs
12	beer, potato chips
13	flour, eggs
14	beer, potato chips
15	eggs, flour, butter, wine, cheese
16	beer, wine, potato chips, cheese
17	wine, cheese
18	beer, potato chips
19	wine, cheese
20	beer, potato chips

Ans: Minimum support = 7.

<u>Transaction</u>	<u>Products</u>
<u>No.</u>	
1	beer, wine, cheese
2	beer, potato chips
3	eggs, flour, butter, cheese
4	eggs, flour, butter, beer, potato chips
5	wine, cheese
6	potato chips
7	eggs, flour, butter, wine, cheese
8	eggs, flour, butter, beer, potato chips
9	wine, beer
10	beer, potato chips
11	butter, eggs
12	beer, potato chips
13	flour, eggs

<u>Transaction No.</u>	<u>Products</u>
14	beer, potato chips
15	eggs, flour, butter, wine, cheese
16	beer, wine, potato chips, cheese
17	wine, cheese
18	beer, potato chips
19	wine, cheese
20	beer, potato chips

To Count the number of items which are present i.e. frequency.

Items	Frequency
beer	11
wine	8
cheese	8
potato chips	10
eggs	7
flour	6
butter	6

We have 6 items. Minimum support is 7. We have two items that do not support min. supp. Their value is less than 7. So, we need to discard flour and butter as their values are less than minimum support i.e. 7.

To Now, we are going to give priority to items from highest to lowest. We provide highest priority to that item which has high occurrence i.e. highest frequency.

<u>Items</u>	<u>Frequency</u>	<u>Priority</u>
beer	11	1
wine	8	3
cheese	8	4
potato chips	10	2
eggs	7	5

In this, the frequency of wine and cheese is same, so we give priority according to First Come First Serve

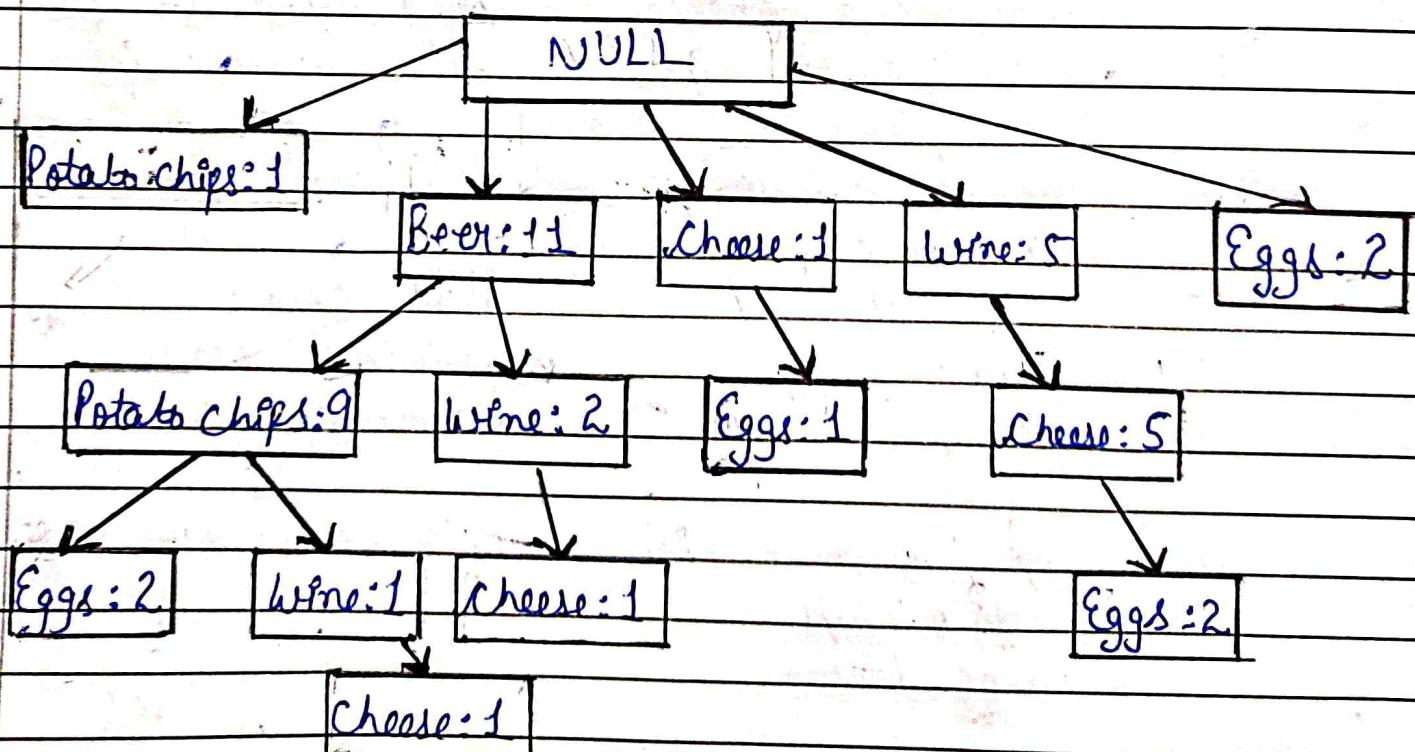
3. According to priority, we need to rearrange the items according to priority i.e. highest priority item place first, then lower than highest priority and so on

<u>Transcription</u>	<u>Items</u>	<u>Ordered Items</u>
No		
1	beer, wine, cheese	beer, wine, cheese
2	beer, potato chips	beer, potato chips
3	eggs, flour, butter, cheese	cheese, eggs
4	eggs, flour, butter, beer, potato chips	beer, potato chips, eggs
5	wine, cheese	wine, cheese
6	potato chips	potato chips
7	eggs, flour, butter, wine, cheese	wine, cheese, eggs
8	eggs, flour, butter, beer, potato chips	beer, potato chips, eggs
9	wine, beer	beer, wine

10.

<u>Transactions</u>	<u>Stems</u>	<u>Ordered Items</u>
10	beer, potato chips	beer, potato chips
11	butter, eggs	eggs
12.	beer, potato chips	beer, potato chips
13	flour, eggs	eggs
14	beer, potato chips	beer, potato chips
15	eggs, flour, butter, wine, cheese	wine, cheese, eggs
16	beer, wine, potato chips, cheese	beer, potato chips, wine, cheese
17	wine, cheese	wine, cheese
18	beer, potato chips	beer, potato chips
19	wine, cheese	wine, cheese
20	beer, potato chips	beer, potato chips

4. Now, we are going to create FP growth tree. The root node is always taken as NULL



we need to traceback the path of Hemets.

6. Eggs:-

↳ cheese, wine: 2 γ , cheese: 1 γ , potato chips, beer: 2 γ

7. cheese:-

↳ wine: 5 γ , wine, beer: 1 γ , wine, potato chips, beer: 1 γ

8. wine:-

↳ beer: 2 γ , potato chips, beer: 1 γ

9. potato chips:-

↳ beer: 9 γ

In this, we can't consider beer because the item which is directly connected to root or null cannot be considered.

Now, we are going to equate item-set. We consider only those items which has occurrence equal to or more than min support i.e. 7

potato chips:-

↳ Beer: 9 γ

There is only one association i.e. beer. beer with association of potato chips as the occurrence of Beer is more than min support.