

Assignment No → 2

Course Code → CAP 446

Name → KOMAL SINGH

Roll No → RD2112A36

Reg No → 12107402

Section → D2112

Date of Submission → 15th Nov 2021

1 For the following given transaction dataset, generate rules using Apriori algorithm. consider value as Support = 30% and confidence = 60%.

Trans id	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

→ There are various ~~are~~ steps in apriori algorithm.

Step 1 :-

In the first step, we will create a table that contains support count of each itemset in the given dataset. This table is called the Candidate set. We denote through C_1 .

Item Set

Frequency

Supports

Milk

9

$$9/12 = 75\%$$

Egg

3

$$3/12 = 25\%$$

Bread

10

$$10/12 = 83\%$$

Butter

10

$$10/12 = 83\%$$

Ketchup

3

$$3/12 = 25\%$$

Cookies

5

$$5/12 = 41\%$$

Now, we will take out all the itemsets that have the greater support count than the minimum support. Since all the itemsets have greater or equal support count than the minimum support. It will give us the table for the frequent itemset. We denotes it through L1

Frequent Itemset

Frequency

Support

Milk

9

$$9/12 = 75\%$$

Bread

10

$$10/12 = 83\%$$

Butter

10

$$10/12 = 83\%$$

Cookies

5

$$5/12 = 41\%$$

Step 2 :-

In this step, we will 2nd candidate set i.e. L2 with the help of frequent itemset

Teacher's Signature :

i.e. L1. In C2, we will generate create the pair of the itemset L1 in the form of Subsets.

2 nd item sets	Frequency	Supports
{Milk, bread}	7	$7/12 = 58\%$
{Milk, butter}	7	$7/12 = 58\%$
{Milk, cookies}	3	$3/12 = 25\%$
{bread, butter}	9	$9/12 = 75\%$
{bread, cookies}	4	$4/12 = 33\%$
{butter, cookies}	3	$3/12 = 25\%$

After creating the subsets we will again find the support count from the main transaction table or datasets i.e. how many times these pairs have occurred together in the given dataset. We need to compare the C2 support count with the minimum support count & after comparing the itemset with less support count will be eliminated from the table.

Frequent 2-item Set	Frequency	Support
{Milk, bread}	7	$7/12 = 58\%$
{Milk, butter}	7	$7/12 = 58\%$
{Bread, butter}	9	$9/12 = 75\%$
{Bread, cookies}	4	$4/12 = 33\%$

Step 3 :-

In this step we will generate 3rd candidate set i.e. C3 with the help of L2. In C3, we will create the pair of itemset of L2 in the form of subsets.

3 rd item Set	Frequency	Support
{Milk, bread, butter}	6	$6/12 = 50\%$
{Milk, bread, Cookies}	1	$1/12 = 8\%$
{Bread, butter, Cookies}	3	$3/12 = 25\%$
{Milk, butter, Cookies}	2	$2/12 = 16\%$

After creating the subsets we will again find the support count from the main transaction table of dataset. Here we need to compare the C3 support count with the minimum support count and after comparing the itemset with less support count will be eliminated from the table.

Frequent 3-item sets	Frequency	Supports
{Milk, bread, butter}	6	$6/12 = 50$

Step 4 :-

In this step we perform Association Rule. To generate the association rules, first

We will create a new table with the possible rules from the occurred combination i.e. {Milk, Bread, Butter}. For all rules we will calculate confidence.

After calculating the confidence value for all rules we will exclude the rules that have less confidence 60%. Those confidence is less than 60%, they are invalid.

Rule 1:

$$\{Milk\} \rightarrow \{bread, butter\}$$

$$\text{Confidence} = \frac{\text{Support}(Milk, bread, butter)}{\text{Support}(Milk)}$$

$$= \frac{6/12}{9/12} = \frac{6}{9} = 66.66\%$$

66.66% > 60%, so this combination is valid.

Rule 2:

$$\{bread\} \rightarrow \{Milk, butter\}$$

$$\text{Confidence} = \frac{\text{Support}(Milk, bread, butter)}{\text{Support}(bread)}$$

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

$60\% = 60\%$. So this combination is valid.

Rule 3:

$$\{\text{butter}\} \rightarrow \{\text{Milk, bread}\}$$

$$\text{Confidence} = \frac{\text{Support}(\text{Milk, bread, butter})}{\text{Support}(\text{butter})}$$

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

$60\% = 60\%$. So this combination is valid.

Rule 4:

$$\{\text{Milk, bread}\} \rightarrow \{\text{butter}\}$$

$$\text{Confidence} = \frac{\text{Support}(\text{Milk, bread, butter})}{\text{Support}(\text{Milk, bread})}$$

$$= \frac{6/12}{7/12} = \frac{6}{7} = 85.7\%.$$

$85.7\% > 60\%$, So this combination is valid.

Rule 5:

$$\{\text{Milk, butter}\} = \{\text{bread}\}$$

$$\text{Confidence} = \frac{\text{Support}(\text{Milk, bread, butter})}{\text{Support}(\text{Milk, butter})}$$

$$= \frac{6}{12} : \frac{6}{12} = 85.7\%$$

85.7% > 60%. So this combination is valid.

Rule 6:

{bread, butter} \rightarrow {Milk}

$$\text{Confidence} = \frac{\text{Support}(\text{Milk, bread, butter})}{\text{Support}(\text{bread, butter})}$$

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

66.67% > 60%. So this combination is valid.



4) Prepare a FP growth tree and all followed steps for the following transactions. Given Minimum Support = 7

Transaction No.

Products

1	beer, wine, cheese
2	beer, Potato chips
3	Egg, flour, butter, cheese
4	Egg, flour, butter, beer, potato chips
5	wine, cheese
6	Potato chips
7	Eggs, flour, butter, wine, cheese
8	Egg, flour, butter, beer, Potato chips
9	wine, beer
10	beer, Potato chips
11	butter, Egg
12	beer, Potato chips
13	flour, Egg
14	beer, Potato chips
15	Egg, flour, butter, wine, cheese
16	beer, wine, potato chips, cheese
17	wine, cheese
18	beer, Potato chips
19	wine, cheese
20	beer, Potato chips.



Teacher's Signature :

→ There are various steps in FP Growth tree algorithm

Step : 1

The first step is to scan the database to find the occurrences of the itemsets in the database. This step is the same as the first step of apriori. The count of 1-itemsets in the database is called support count or frequency of 1-itemset.

item-Set

Support

beer

11

wine

8

Cheese

8

Potato Chips

10

Egg

7

Flour

6

Butter

6

Now we will take out all the itemset that have the greater support count than the minimum support and discard all the itemset who is less then minimum support. It will give us the table for the frequent itemset.

Frequent itemsetSupport:

10

Beer

11

Lime

8

Cheese

8

Potato Chips

10

Egg

7

Step 2 :

In this step we create a frequent item-set list in which frequent items are stored in the descending order based on the support.

Frequent itemsetSupport

Beer

11

Potato chips

10

Lime

8

Cheese

8

Egg

7

Step 3 :

In this step we rearrange ~~not~~ the original dataset according to his support priority that means whose support is high that are written in first.

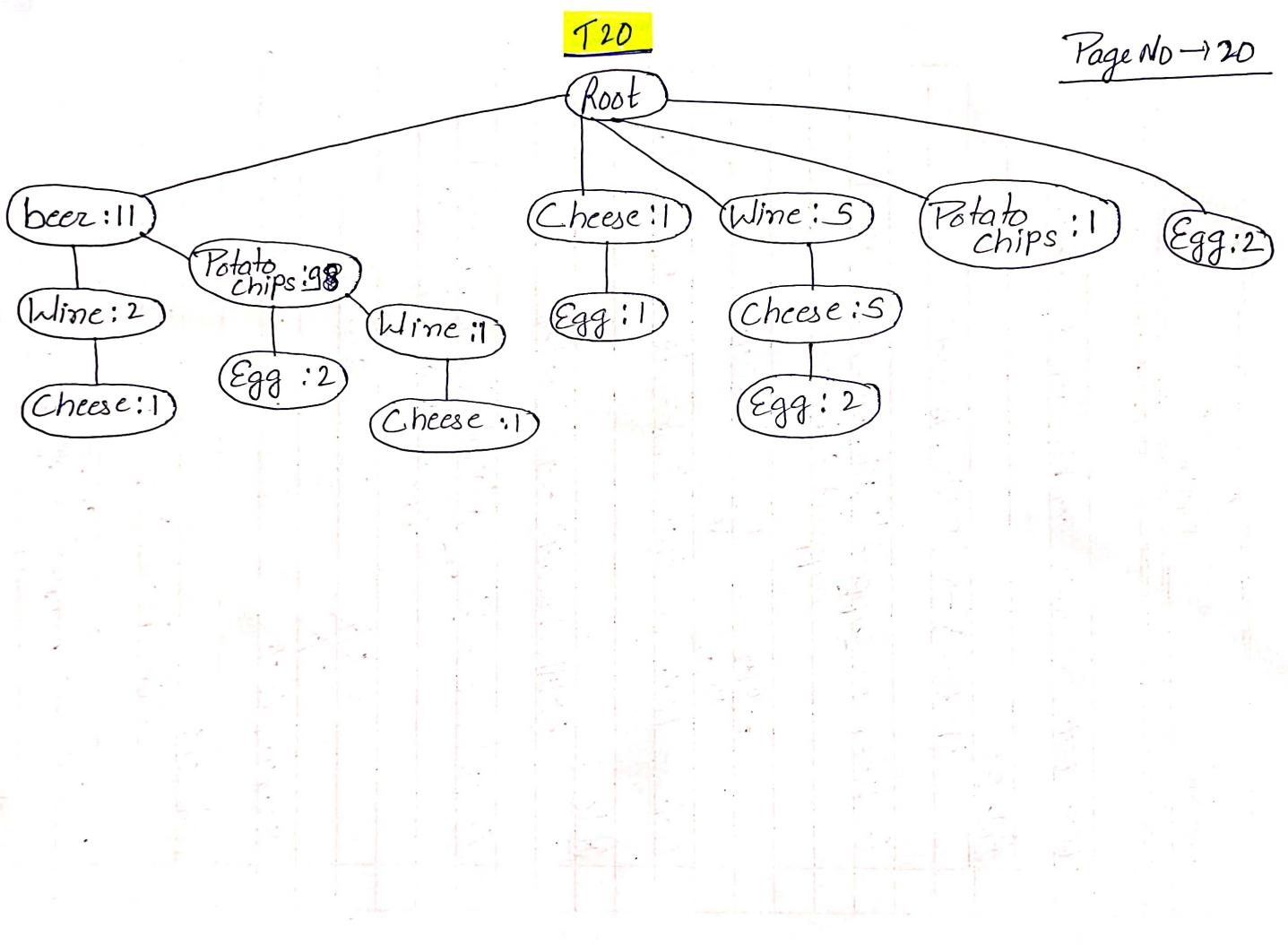


Teacher's Signature :

<u>Trans-id</u>	<u>Product</u>
1	Beer, Wine, Cheese
2	Beer, Potato chips
3	Cheese, Egg
4	Beer, Potato chips, Egg
5	Wine, cheese
6	Potato chips
7	Wine, cheese, Egg
8	Beer, Potato chips, Egg
9	Beer, Wine
10	Beer, Potato chips
11	Egg
12	Beer, Potato chips
13	Egg
14	Beer, Potato chips
15	Wine, cheese, Egg
16	Beer, Potato chips, wine, cheese
17	Wine, cheese
18	Beer, Potato chips
19	Wine, cheese
20	Beer, Potato chips

Step 4:

In this step we construct the FP tree.



Step → 5 :-

In this step we need to trace back the path of each of the item and update with the count. Which item who is directly connected to Root that item now need to go back.

• Cheese

{ Wine, beer : 1 } , { Wine, Potatochips,
beer : 1 } { wine : 5 }

• Egg

{ Potatochips, beer : 2 } { cheese : 1 }
{ cheese, wine : 2 }

• Wine

{ beer : 2 } { Potatochips, beer : 1 }

• Potatochips

{ beer : 9 }

Step → 6 :-

In this step we merge item sets and write down all those dataset who support minimum other all are removed from the list.

• Potatochips

{ beer : 9 }



Teacher's Signature :