FP tree algorithm:

to identify frequent patterns in the area of Data Mining

AIM :

- To Draw a FP tree
- To identify frequent patterns

Problem : Find all frequent itemsets and frequent patterns in the following database using FP-growth algorithm. **Take minimum support as 30%.**

TID	Items
1	E, A, D, B
2	D, A, C, E, B
3	C, A, B. E
4	B, A, D
5	D
6	D,B
7	A,D,E
8	B,C

Table 1 - Snapshot of the Database

Step 1 - Calculate Minimum support

First should calculate the minimum support count. Question says minimum support should be 30%. It calculate as follows:

Minimum support count(30/100 * 8) = **2.4**

As a result, 2.4 appears but to empower the easy calculation it can be rounded to to the ceiling value. Now,

Minimum support count is **ceiling**(30/100 * 8) = **3**

Step 2 - Find frequency of occurrence

Now time to find the frequency of occurrence of each item in the Database table.

Item	Frequency
A	5
B	6
C	3
D	6
E	4

Table2 -Frequency of Occurrence

Step 3 - Prioritize the items

Item B got the highest priority (**1**)

At the same time you have opportunity to drop the items which not fulfill the minimum support requirement.

E.g. If Database contain **F** which has frequency 1, then you can drop it.

Note : We have no items with MinSup

*Some people display the frequent items using list instead of table. The frequent item list for the above table will be **B:6, D:6, A: 5, E:4, C: 3**.

Step 4 -Order the items according to priority

As you see in the Table 3 new column added to the Table 1.

Last column indicate the ordered list

TID	Items	Ordered Items
1	E, A, D, B	B,D,A,E
2	D, A, C, E, B	B,D,A,E,C
3	C, A, B. E	B,A,E,C
4	B, A, D	B,D,A
5	D	D
6	D,B	B,D
7	A,D,E	D,A,E
8	B,C	B,C

Table 3 - New version of the Table 1

Step 5 –Draw the FP-Tree

As a result of previous steps we got a ordered items table (Table 3). Now it's time to draw the FP tree.

Row 1:

FP trees have 'null' node as the root node.

TID	Items	Ordered Items
1	E, A, D, B	B,D,A,E

Start with root node first and attach the items of the row 1 one by one respectively. (See the Figure 1) And write their occurrences in front of it.

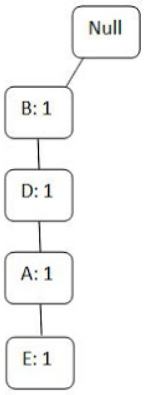


Figure 1- FP tree for Row 1

Row 2:

Then update the above tree (Figure 1) by entering the items of row 2.

TID	Items	Ordered Items
1	E, A, D, B	B,D,A,E
2	D, A, C, E, B	B,D,A,E,C

Update the frequency count of repeated visit of nodes

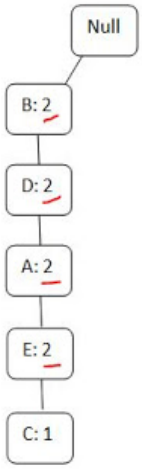


Figure 2- FP tree for Row 1,2

Row 3:

TID	Items	Ordered Items
1	E, A, D, B	B,D,A,E
2	D, A, C, E, B	B,D,A,E,C
3	C, A, B. E	B,A,E,C

- In row 3 you have to visit B,A,E and C respectively. Can't connect B to existing A overtaking D in existing path
- draw another A and connect it to B and then connect new E to that A and new C to new E. See Figure 3.

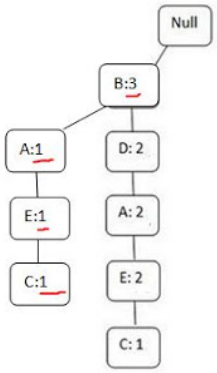


Figure 3 - After adding third row

Row 4:

TID	Items	Ordered Items
1	E, A, D, B	B,D,A,E
2	D, A, C, E, B	B,D,A,E,C
3	C, A, B. E	B,A,E,C
4	B, A, D	B,D,A
5	D	D

Then row 4 contain B,D,A. Now we can just rename the frequency of occurrences in the existing branch. As B:4,D,A:3.

Row 5:

In fifth raw have only item D. Now we have opportunity draw new branch from 'null' node. See Figure 4.

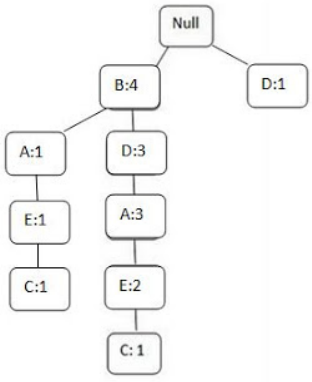


Figure 4- Connect D to null node

Row 7:

TID	Items	Ordered Items
1	E, A, D, B	B,D,A,E
2	D, A, C, E, B	B,D,A,E,C
3	C, A, B. E	B,A,E,C
4	B, A, D	B,D,A
5	D	D
6	D,B	B,D
7	A,D,E	D,A,E
8	B,C	B,C

Attach two new nodes A and E to the D node which hanging on the null node.
Then mark D,A,E as D:2,A:1 and E:1.

Row 8 :
Attach new node C to B. Change the traverse times.(B:6,C:1)

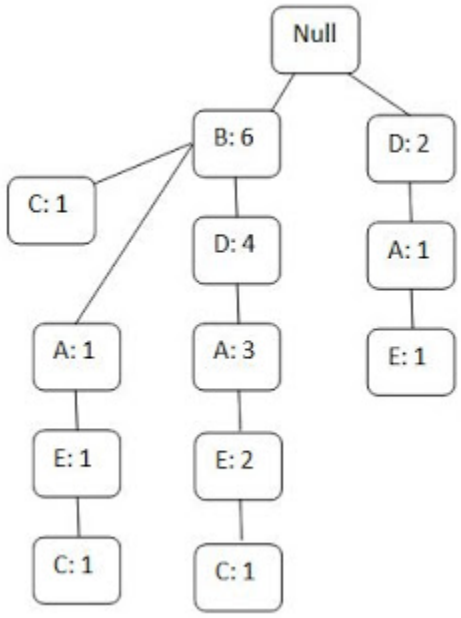


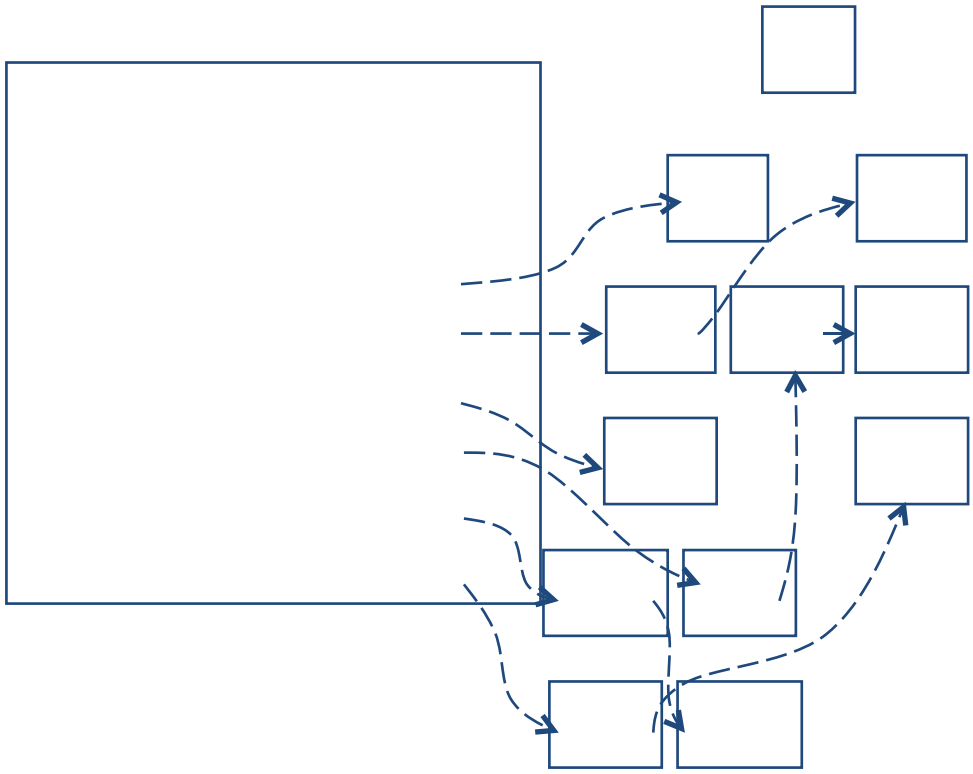
Figure 5 - Final FP tree

Step 6 - Validation

How we know is this correct?

Count the frequency of occurrence of each item of the FP tree and compare it with Table 2. If both counts equal, then it is positive point to indicate your tree is correct.

Linked based DS



How to identify frequent patterns from FP tree?

TID	Items	Ordered Items
1	E, A, D, B	B,D,A,E
2	D, A, C, E, B	B,D,A,E,C
3	C, A, B. E	B,A,E,C
4	B, A, D	B,D,A
5	D	D
6	D,B	B,D
7	A,D,E	D,A,E
8	B,C	B,C

Item	Frequency
A	5
B	6
C	3
D	6
E	4

Table2 -Frequency of Occurrence

B:6, D:6, A: 5, E:4, C: 3.

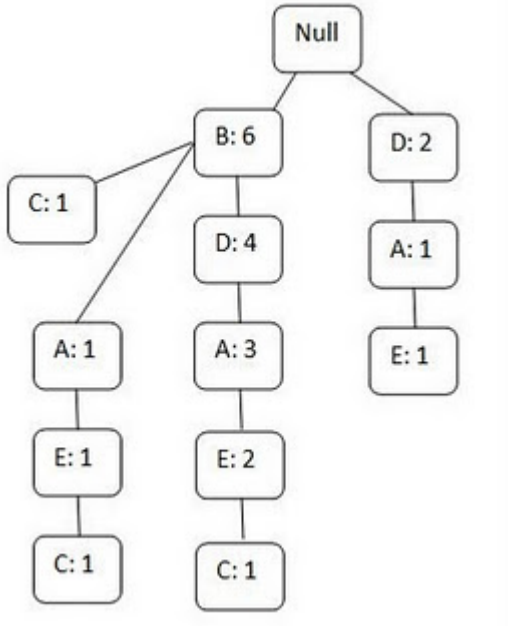


Figure 1

The Final FP tree vreated is as in Fig1.

Minimum support count is 3.

Orders list of Items of the DB in order of their frequency :

B:6
D:6
A:5
E:4
C:3

To observe the FPs we have to go through bottom to top, from C to B.

FP using C

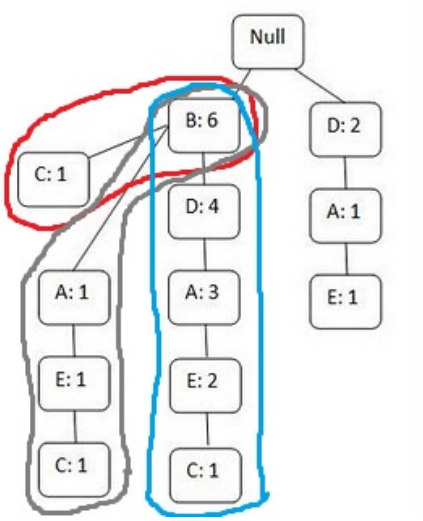


Figure 2

Step 1: Find the conditional pattern base for C:3.

Here c.count = 3 which is mentioned as C:3

In Figur 2 there are 3 Cs with count = 1 for each.

Traverse bottom to top and get the branches which have Cs with the occurrence of C.

3 branches are derived

- **BDAE: 1 [in** Blue]
  B: 1 - [in Red]
  BAE: 1 – [in Brown]

These branches regard as Conditional Pattern Base for C. You have to note three facts.

1. Even we traverse bottom to top we write the branches in top to bottom manner.
2. C is not there.
3. 1 came after each branch to represent the frequency of occurrence of C in each branch.

Note: To ensure you correctly got all the occurrences of C in FP tree, add occurrences of C in each branch and compare with all the occurrences of C in FP tree; for this case **1+1+1 = 3** it is correct.

Step 2: Find F-list from the Conditional Pattern Base of C.

The list of items are B:3,D:1,A:2,E:2
B:3 only eligible for F-list due (Minimum Support Count=3)

F-list = **B:3**

Step 3: Draw the Conditional FP tree for C (figure 3)

When drawing the conditional FP tree you have to make sure to get Conditional Pattern Base as Ordered Items as described in or Data Table Table 3.

TID	Items	Ordered Items
1	E, A, D, B	B,D,A,E
2	D, A, C, E, B	B,D,A,E,C
3	C, A, B. E	B,A,E,C
4	B, A, D	B,D,A
5	D	D
6	D,B	B,D
7	A,D,E	D,A,E
8	B,C	B,C

Now we can find **C:3, BC:3** as identified FPs from tree.

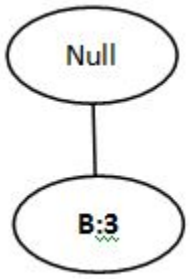


Figure 3

If you answer a question paper the answer will be **Conditional pattern base for C:3** **BDAE: 1; B: 1; BAE: 1** **F-list: B:3**
Conditional FP-tree for C (Draw Figure 3.) **Frequent patterns** **C:3, BC:3**

FP using E
B:6
D:6
A:5
E:4
C:3

Step 1: Find Conditional Pattern Base for E:4.

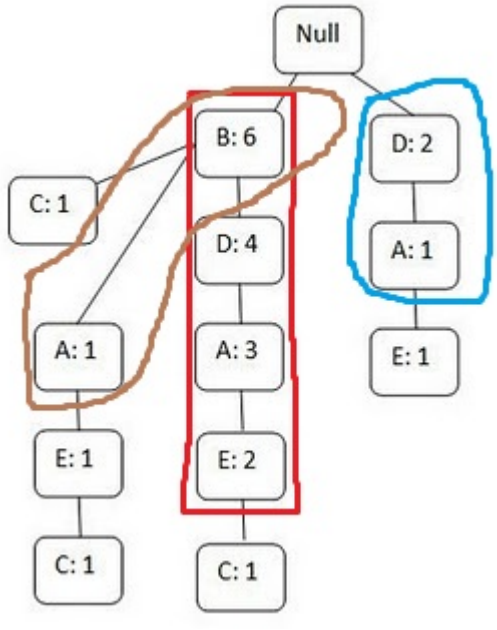


Figure 4

BDA: 2 - Red line
BA:1 - Brown line
DA:1 - Blue line

Step 2: Prepare F-List

F-list =
A:4
B:3
D:3

Final Conditional Pattern Base = ABD : 2, AB: 1, AD: 1
Note:

- 3 comes to B by 2 Bs coming from BDA: 2 and 1 B coming from BA:1.
- And 4 As and 3 Ds also according to that manner.
- Why A appears before the B in F-list even B appears before previously? That have a reason. We got a list from this Conditional Pattern Base as B:3, D:3, A:4. But in F-list

they should appear according to their frequency of occurrences in Conditional Pattern Base. So A coming to front and others going to back. Then it is easy to use for creating conditional FP tree for E.

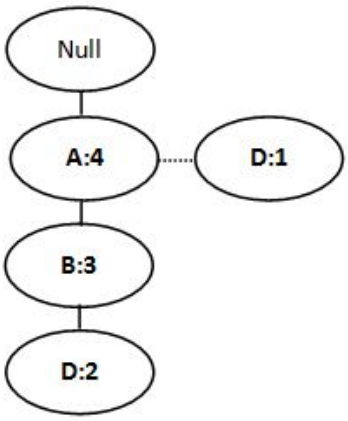


Figure 5

Repeat the same task for this tree (Conditional tree for E so E is silent)
Items in this FP tree are
A:4
B:3
D:3
Let's go bottom to top.

DE

We are going to identify FPs for DE. Because of we use conditional FP tree for E, E comes to back.
Conditional pattern base for DE:3
A:1, AB:2
List of Items are L A:3, B:2
F-list: A:3 [B.count < min Supp]

Conditional FP-tree for DE : Figure 6
Frequent patterns
DE:3, ADE:3 [This final list of paeern should follow the original sequence of table 3]

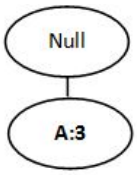


Figure 6

Same way

BE

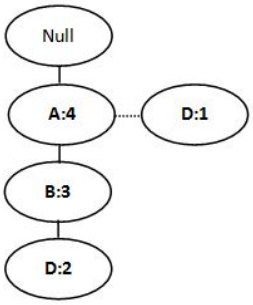


Figure 5

Conditional pattern base for BE:3
A:3
F-list: A:3

Conditional FP-tree for DE : Figure 6

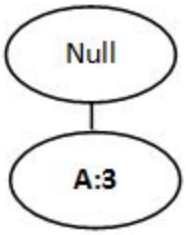


Figure 6

Frequent patterns
BE:3, ABE:3

AE (from fig 5)

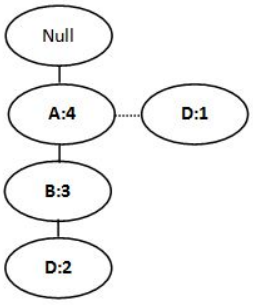


Figure 5

Conditional pattern base for AE:4
NULL

Conditional FP-tree for AE
NULL
Frequent patterns
AE:4

We find all FPs of conditional FP tree for E. Next step is remind the previous or main tree. And before leave E we can write E:4 also as a FP. Let's move to A.

A

B:6
D:6
A:5
E:4
C:3

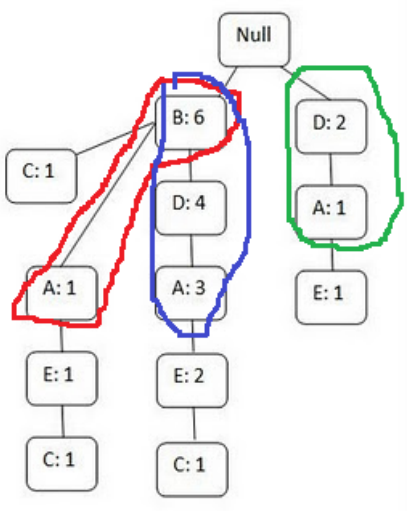


Figure 1

Conditional pattern base for A:5
BD: 3; B:1; D:1
F-list: B:4, D:4
Conditional FP-tree for A : Figure 7

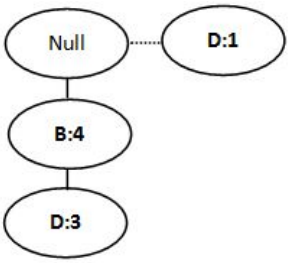


Figure 7

Then we will go on the conditional FP tree for A as like as previous trip.

AD

Conditional pattern base for AD:4

B:3
F-list :B:3
Conditional FP-tree for AD : Figure 3

Frequent patterns
AD:4, ABD:3

AB

Conditional pattern base for AB:4
NULL

Frequent patterns
AB:4

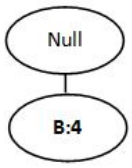


Figure 8

And also can get A:5 as a FP.

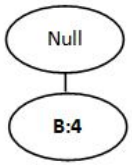
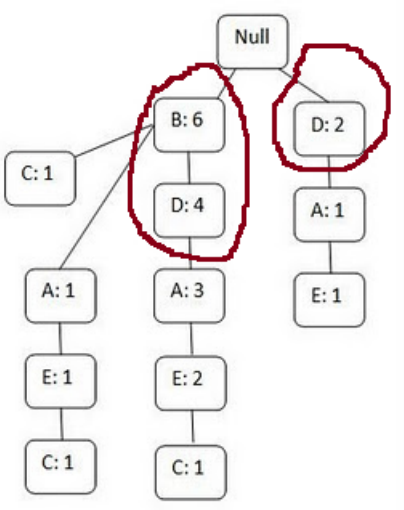


Figure 8

D

B:6
D:6
A:5
E:4
C:3



Conditional pattern base for D:6
B: 4
F-list : B:4

Conditional FP-tree for D : Figure 8

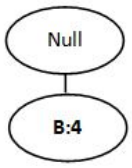


Figure 8

Frequent patterns
D:6, BD:4

Then last item, B.

B

B:6
D:6
A:5
E:4
C:3

Conditional pattern base for B:6
NULL

Conditional FP-tree for B
NULL
Frequent patterns
B:6

The identified frequent itemsets from the Database with minimum support as 30% are **C:3, BC:3, DE:3, ADE:3 , BE:3, ABE:3, AE:4, E:4, AD:4, ABD:3, AB:4, A:5, D:6, BD:4, B:6.**