



Vidyavardhini's College of Engineering and Technology

Department of Artificial Intelligence & Data Science

AY: 2025-26

Class:	TE	Semester:	V th
Course Code:	CSC504	Course Name:	Data warehousing & mining

Name of Student:	Pranita Kumbhar
Roll No. :	70
Assignment No.:	05
Title of Assignment:	Frequent Patterns
Date of Submission:	
Date of Correction:	

Evaluation

Performance Indicator	Max. Marks	Marks Obtained
Completeness	5	5
Demonstrated Knowledge	3	3
Legibility	2	2
Total	10	10

Performance Indicator	Exceed Expectations (EE)	Meet Expectations (ME)	Below Expectations (BE)
Completeness	5	3-4	1-2
Demonstrated Knowledge	3	2	1
Legibility	2	1	0

Checked by

Name of Faculty : Ms. Neha Raut

Signature :

Date :

Q.1] Consider the transaction data given below. Use apriori algorithm with min-sup count = 2 & min-confidence = 70% to find all frequent items sets & strong association rules.

→

TID	List of items
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

→ Given : min-sup = 2.

1] Count of Each Itemset (C_1) by scanning the database.

Itemset	Count
I1	6
I2	7
I3	6
I4	2
I5	2

2] Prune Step : There is no item which have count less than 2, thus no item will delete.

3] Join Step : Form C_2 from C_1 using $C_1 \bowtie C_1$ & find out their occurrences.

Itemset	Count
$\{I_1, I_2\}$	4
$\{I_1, I_3\}$	4
$\{I_1, I_4\}$	1
$\{I_1, I_5\}$	2
$\{I_2, I_3\}$	4
$\{I_2, I_4\}$	2
$\{I_2, I_5\}$	2
$\{I_3, I_4\}$	0
$\{I_3, I_5\}$	1
$\{I_4, I_5\}$	0

4] Prune Step (L_2) : C_2 shows that itemset $\{I_1, I_4\}$, $\{I_3, I_4\}$, $\{I_3, I_5\}$, $\{I_4, I_5\}$ does not meet min-sup, thus it is deleted.

Itemset	Count
$\{I_1, I_2\}$	4
$\{I_1, I_3\}$	4
$\{I_1, I_5\}$	2
$\{I_2, I_3\}$	4
$\{I_2, I_4\}$	2
$\{I_2, I_5\}$	2

5] Join step: Form C_3 from L_2 using $L_2 \bowtie L_2$ & find out their occurrences.

Itemset	Count
$\{I_1, I_2, I_3\}$	2
$\{I_1, I_2, I_4\}$	1
$\{I_1, I_2, I_5\}$	2
$\{I_1, I_3, I_4\}$	0
$\{I_1, I_3, I_5\}$	1
$\{I_1, I_4, I_5\}$	0
$\{I_2, I_3, I_4\}$	0
$\{I_2, I_3, I_5\}$	1
$\{I_2, I_4, I_5\}$	0
$\{I_3, I_4, I_5\}$	0

6] Prun Step (L_3): C_3 shows that itemset $\{I_1, I_2, I_4\}$, $\{I_1, I_3, I_4\}$, $\{I_1, I_3, I_5\}$, $\{I_1, I_4, I_5\}$, $\{I_2, I_3, I_4\}$, $\{I_2, I_3, I_5\}$, $\{I_2, I_4, I_5\}$, $\{I_3, I_4, I_5\}$ does not meet min-sup, thus it is deleted.

Itemset	Count
$\{I_1, I_2, I_3\}$	2
$\{I_1, I_2, I_5\}$	2

Thus $\{I_1, I_2, I_3\}$ & $\{I_1, I_2, I_5\}$ is frequent.

7] Generate Association Rules: From the frequent itemset discovered above, the association could be:

- $\{I_1, I_2\} \rightarrow \{I_3\}$

$$\text{Confidence} = \text{support } \{I_1, I_2, I_3\} / \text{support } \{I_1, I_2\}$$

$$= (2/4) \times 100 = 50\%$$

- $\{I_1, I_3\} \rightarrow \{I_2\}$

$$\text{Confidence} = \text{support } \{I_1, I_2, I_3\} / \text{support } \{I_1, I_3\}$$

$$= (2/4) \times 100 = 50\%$$

- $\{I_2, I_3\} \rightarrow \{I_1\}$

$$\text{Confidence} = \text{support } \{I_1, I_2, I_3\} / \text{support } \{I_2, I_3\}$$

$$= (2/4) \times 100 = 50\%$$

- $\{I_1, I_2\} \rightarrow \{I_5\}$

$$\text{Confidence} = \text{support } \{I_1, I_2, I_5\} / \text{support } \{I_1, I_2\}$$

$$= (2/4) \times 100 = 50\%$$

- $\{I_1, I_5\} \rightarrow \{I_2\}$

$$\text{Confidence} = \text{support } \{I_1, I_2, I_5\} / \text{support } \{I_1, I_5\}$$

$$= (2/2) \times 100 = 100\%$$

- $\{I_2, I_5\} \rightarrow \{I_1\}$

$$\text{Confidence} = \text{support } \{I_1, I_2, I_5\} / \text{support } \{I_2, I_5\}$$

$$= (2/2) \times 100 = 100\%$$

- $\{I_1\} \rightarrow \{I_2, I_3\}$

$$\text{Confidence} = \text{support } \{I_1, I_2, I_3\} / \text{support } \{I_1\}$$

$$= (2/6) \times 100 = 33.33\%$$

- $\{I_2\} \rightarrow \{I_1, I_3\}$

$$\text{Confidence} = \text{support } \{I_1, I_2, I_3\} / \text{support } \{I_2\}$$

$$= (2/7) \times 100 = 28.6\%$$

- $\{I_3\} \rightarrow \{I_1, I_2\}$

$$\text{Confidence} = \text{support } \{I_1, I_2, I_3\} / \text{support } \{I_3\}$$

$$= (2/6) \times 100 = 33.33\%$$

- $\{I_1\} \rightarrow \{I_2, I_3\}$

$$\text{Confidence} = \text{support } \{I_1, I_2, I_3\} / \text{support } \{I_1\}$$

$$= (2/6) \times 100 = 33.33\%$$

- $\{I_2\} \rightarrow \{I_1, I_3\}$

$$\text{Confidence} = \text{support } \{I_1, I_2, I_3\} / \text{support } \{I_2\}$$

$$= (2/7) \times 100 = 28.6\%$$

- $\{I_3\} \rightarrow \{I_1, I_2\}$

$$\text{Confidence} = \text{support } \{I_1, I_2, I_3\} / \text{support } \{I_3\}$$

$$= (2/2) \times 100 = 100\%$$

- This shows that association rules $\{I_1, I_3\} \rightarrow \{I_2\}$, $\{I_2, I_3\} \rightarrow \{I_1\}$, $\{I_3\} \rightarrow \{I_1, I_2\}$ are strong as they satisfy minimum confidence threshold of 70%.

Q.2] Generate frequent pattern tree for the following transaction with 30% minimum support.

Transaction ID	Items
T1	E, A, D, B
T2	D, A, C, E, B
T3	C, A, B, E
T4	B, A, D
T5	D
T6	D, B
T7	A, D, E
T8	B, C

→ Given : $\text{min-sup} = 30\%$

$\therefore \text{sup-count to be satisfied} = 8 \times 0.3 = 2.4$

1] Scan the database for count of each itemset

Itemset	sup-count
{A}	5
{B}	6
{C}	3
{D}	6
{E}	4

2] Sort the set frequent itemsets in order of descending support count and denote that lists as L.

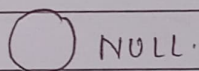
Itemset	sup-count
{B}	6
{D}	6
{A}	5
{E}	4
{C}	3

3] Scan the database for second time & sort items in each transaction according to descending support count.

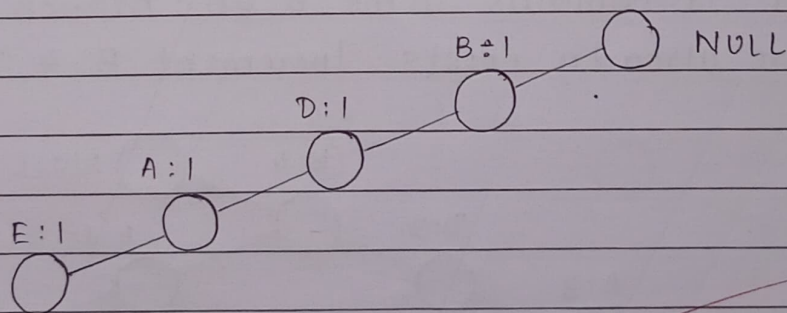
Transaction ID	Items
T1	{B, D, A, E}
T2	{B, D, A, E, C}
T3	{B, A, E, C}
T4	{B, D, A}
T5	{D}
T6	{B, D}
T7	{D, A, E}
T8	{B, C}

4] Construct the FP-tree.

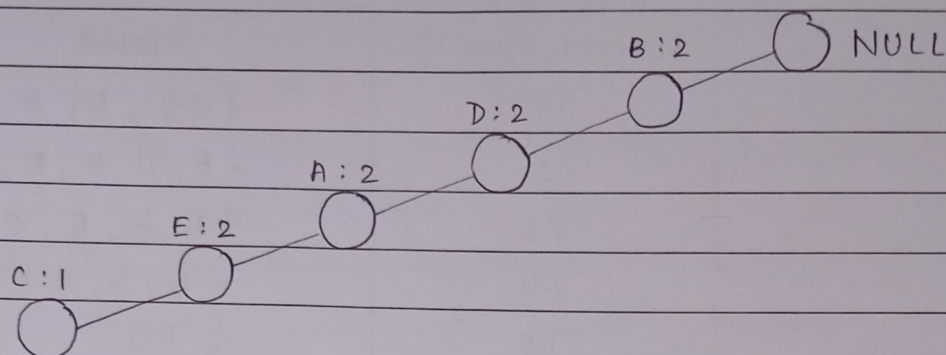
a] Create a root node with LABEL "NULL".



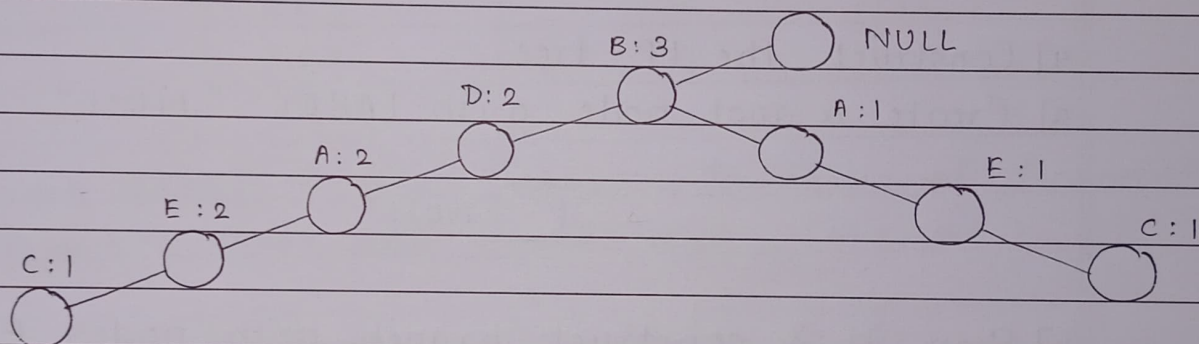
b] Scan T1 & construct branch with nodes B:1, D:1, A:1, E:1 linked to each other from root node.



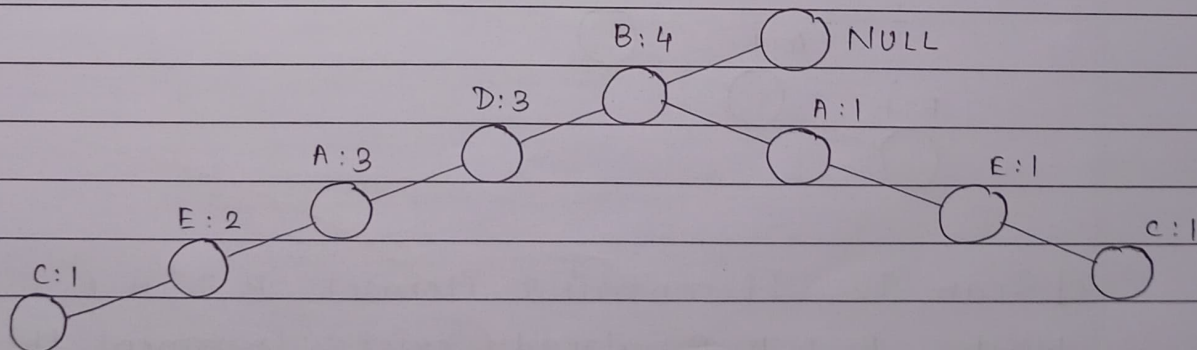
c] Scan T2. It contains itemset B, D, A, E, C in order. Nodes B, D, A, E already exists. Increment their count as B:2, D:2, A:2, E:2 & make a branch for C:1.



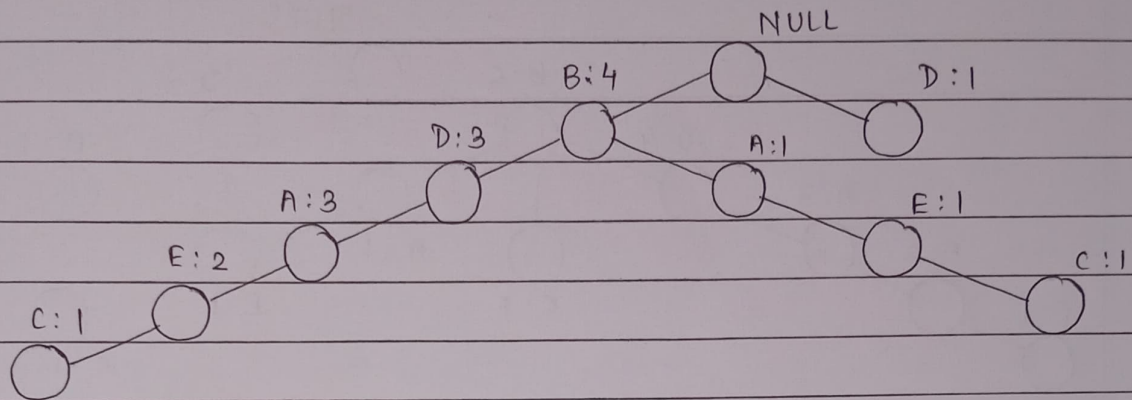
d] Scan T3. It contains items B, A, E, C. Branch with nodes which are not in order. Increment B:3.



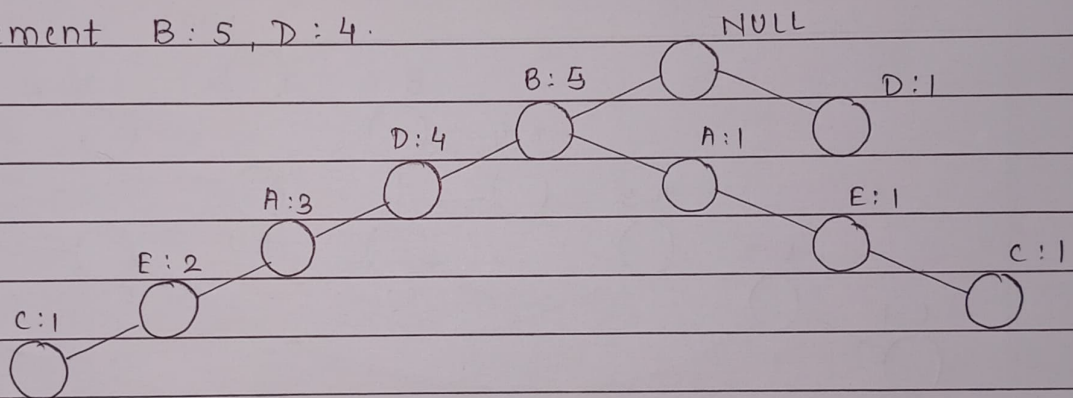
e] Scan T4. It contains items B, D, A. Branch with nodes which are already exists. Increment B:4, D:3, A:3.



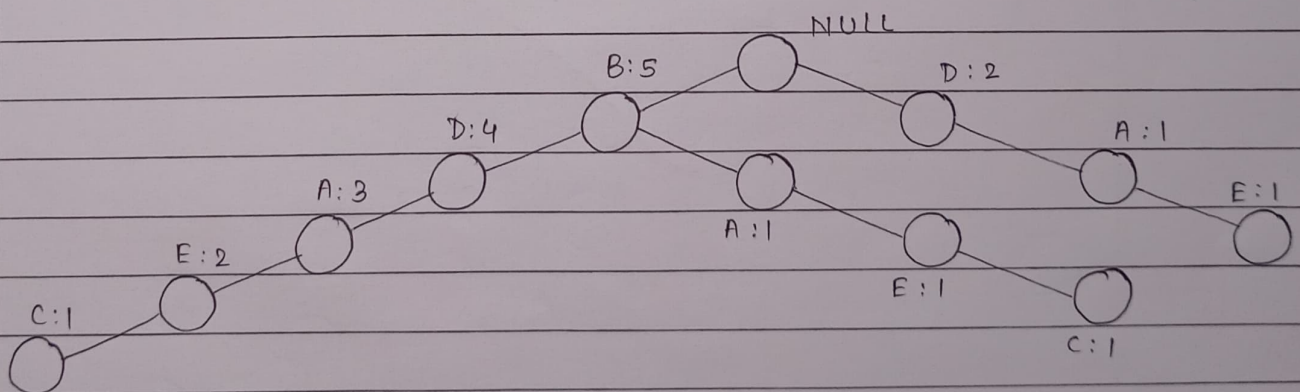
f] Scan T5. It contains item D.



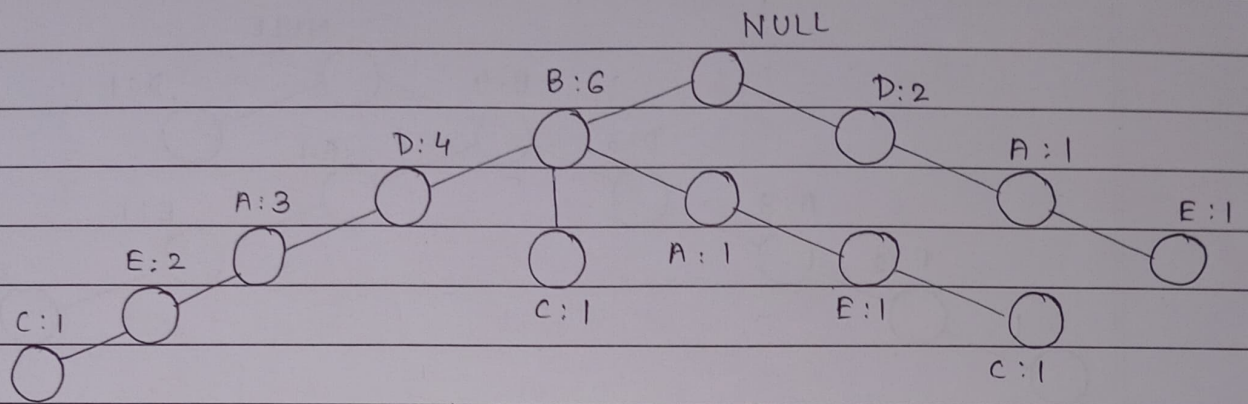
g] Scan T6. It contains items B, D which already exists.
Increment B: 5, D: 4.



h] Scan T7. It contains items D, A, E, Increment D: 2.



i] Scan T8. It contains items B, C. Node B is already exists. Increment B: 6.



j] Connect all similar nodes.

