Date : 08 / 31 / 2022

Data Mining
Assignment - 1

Sindhuja Yerramalla
U00839259

Q1. Yes, The Apriori Algorithm can be adopted to find these itemsets.

Instead of using total number of rows, we use profit of an itemset here

The reason for using Apriori is, if an itemset has profit ≥ 50, then any proper subset of that itemset has profit ≥ 50

Given that $F(S) = \frac{\sum_{i \in S} P_i}{|S|}$ } used for calculating Frequency and Profit

$P_A = P_B = P_C = P_D = 10$

We can adopt Apriori Algorithm., so

* $L_1 \rightarrow$ All items with profit ≥ 50
* $K \rightarrow 2$
* $L_{K-1} \neq 0$
* $C_K \rightarrow$ will generate L_{K-1} by join and prune step

Date : / /

* Perform a counting step on C_k to obtain L_k

* Return $\cup_i L_i$

	Frequency	Profit
A	0+3+0+1+6 = 10	10 x 10 = 100
B	0+4+0+0+0 = 4	4 x 10 = 40
C	3+0+1+3+0 = 7	7 x 10 = 70
D	2+0+3+5+0 = 10	10 x 10 = 100

So, $L_1 = \{A, C, D\}$, with join and prune we get $C_2 = \{AC, AD, CD\}$ [$\because K=2$]

* Counting step on C_k to obtain L_k.

	Frequency	Profit
AC	0+0+1+0 = 1	1 x 10 = 10
AD	0+0+1+0 = 1	1 x 10 = 10
CD	2+0+1+3+0 = 6	6 x 10 = 60

$\therefore$ $L_2 = \{CD\}$ The output be $\{A, C, D, CD\}$

Date : / /

Q2.

Given $P_A = 5$, $P_B = 10$, $P_C = 6$, $P_D = 4$.

and $f(s) = \sum_{i \in S} P_i$

* No, The Apriori Algorithm cannot be adopted. Because the Apriori propery cannot be satisfied. [If an itemset has profit ≥ 50, then any proper subset has profit ≥ 50 Or If an itemset doesnot have profit ≥ 50, then any proper superset must not have profit ≥ 50.

However, we can use another algorithm

	Frequency	Profit
CD	2+0+1+3+0 = 6	6 × (6+4) = 60
C	3+0+1+3+0 = 7	7 × 6 = 42
D	2+0+3+5+0 = 10	10 × 4 = 40

Algorithm used will be

* $O \leftarrow \phi$
* for each itemset s with frequency ≥ 1 find Profit s, if $S > 50$, $O \leftarrow O \cup \{s\}$
* Return O.

Date : / /

Q3. Given Transaction

T	A	B	C	D
t_1	1	0	0	1
t_2	0	1	0	0
t_3	1	0	1	1
t_4	1	0	1	1

Given, minimum Support threshold = 2.

Step1 → To get the Support count of each item.

Item	Support
A	3
B	1
C	2
D	3

∵ minimum threshold is 2, B will be discarded.

Date : / /

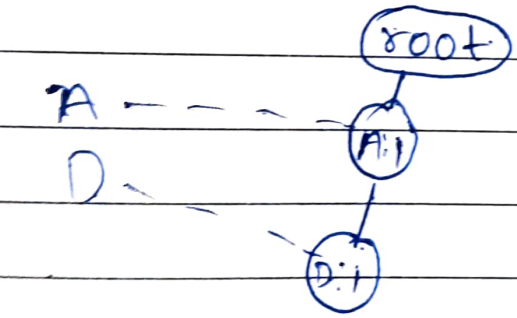
Item	Support
A	3
D	3
C	2

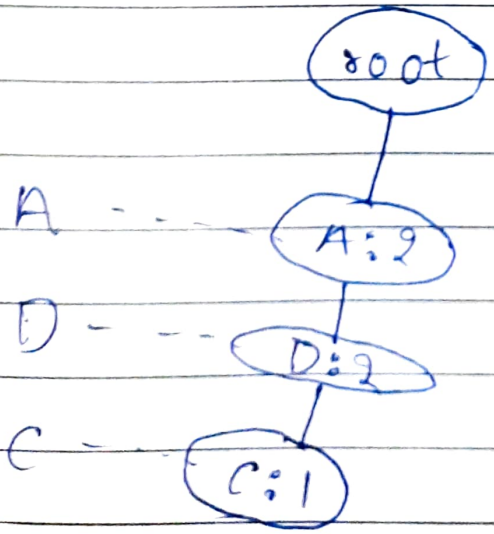
Step 2 - Construct fp tree

t_1 A, D

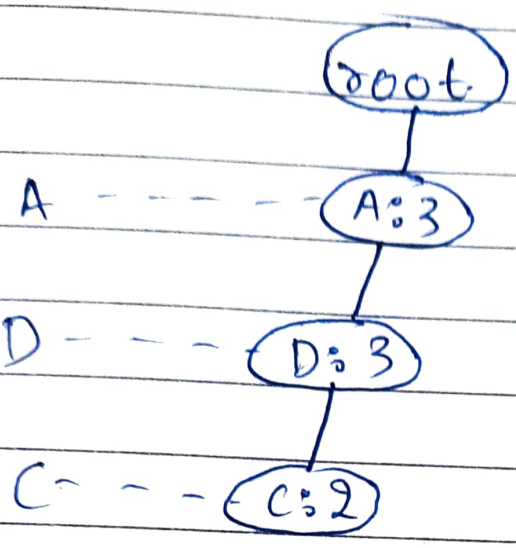
t_3 A, D, C

t_4 A, D, C.





Date : / /



Step3 : Building Conditional FP-tree for each item.

Item.	Conditional FP-tree	Frequent item-sets
C:2	root → (A:2) → (D:2)	{C:2, CD:2, AC:2, AD:2, ACD:2}
D:3	root → (A:3)	{D:3, AD:3}
A:3	root	{A:3}

Step4 : All frequent itemsets are {A, C, D, AC, AD, CD, ACD}

Date : / /

Q4: No, the FP-tree structure obtained remains the same even if the transactions are shuffled. We can clearly show this through contradiction. Suppose the tree structure differs, it follows that at least one transaction will have a different encoding in the prefix tree, implying that the transaction will be reconstructed differently. The transaction, however, remains unique and unchanged This goes against the assumption, Therefore it is proved.