Date : / /

Data Mining
Assignment-1

Q1 Given

T	A	B	C	D	E
t_1	0	1	1	0	1
t_2	1	0	1	1	0
t_3	1	1	1	0	1
t_4	0	0	0	1	1
t_5	0	0	1	0	0

given, minimum support ≥ 2.

Itemsets	Frequency	Support
A	0+1+1+0+0	2
B	1+0+1+0+0	2
C	1+1+1+0+1	4
D	0+1+0+1+0	2
E	1+0+1+1+0	3

∴ Frequent itemsets after 1st iteration

$L_1 = \{A, B, C, D, E\}$

Date : / /

estimation C_2.

Itemset pairs	Frequency/Support
(A, B)	1
(A, C)	2
(A, D)	1
(A, E)	1
(B, C)	2.
(B, D)	0
(B, E)	2.
(C, D)	1
(C, E)	2.
(D, E)	1

$L_2 = \{(A, C), (B, C), (B, E), (C, E)\}$

Estimate C_3 Itemset Pairs	Freq/support
(A, B, C) ~~(A, C)~~	1
(A, B, E) ~~(B, E)~~	1
(A, C, E) ~~(B, E)~~	1
(B, C, E) ~~(C, E)~~	2

Date : / /

$L_3 = \{(B, C, E)\}$

The frequent itemsets after third iteration is

$L_3 = \{(B, C, E), (B), (C), (E), (B, C), (C, E), (B, E)\}$

$\therefore L_3 = \{(B\ C\ E), (B), (C), (E), (B\ C), (B\ E), (E\ B)\}$

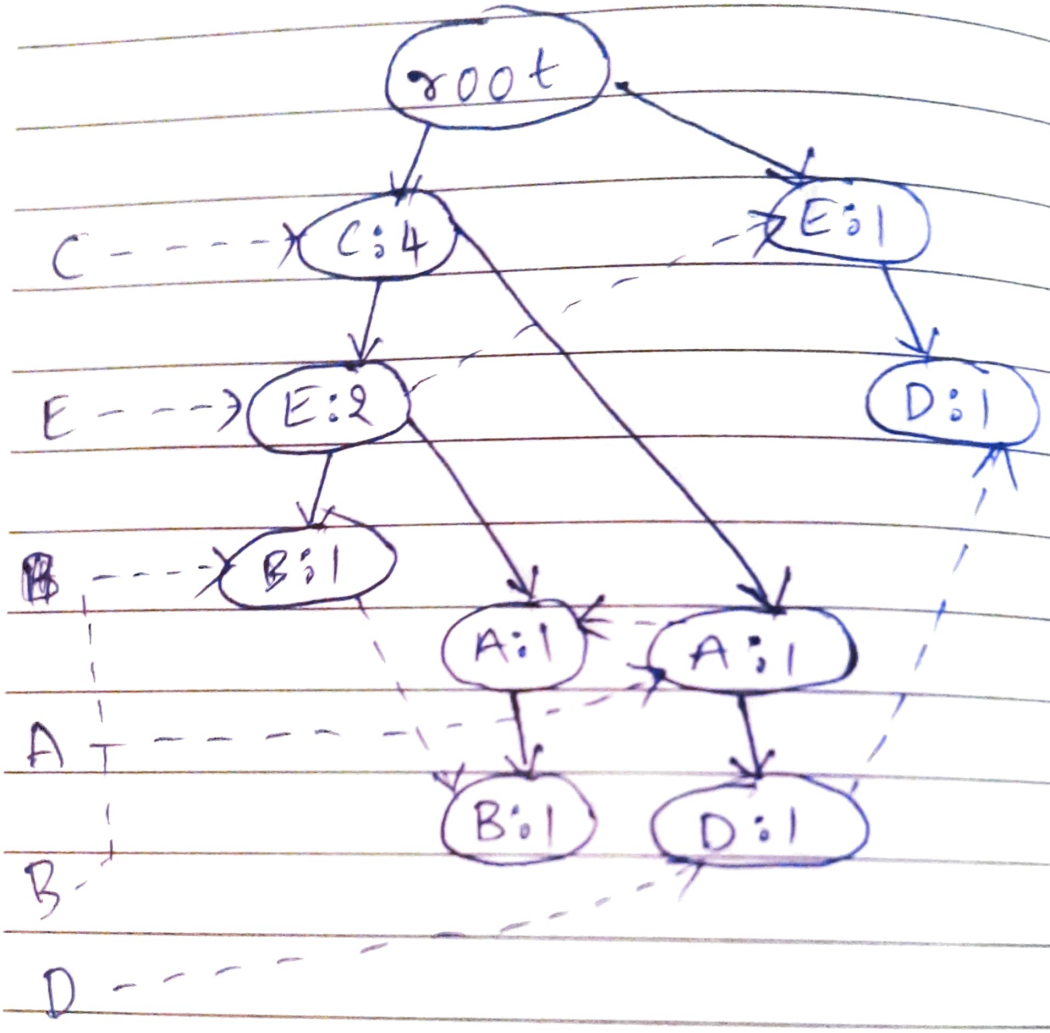
Q2 FP-tree.

T	Items
t_1	{B, C, E}
t_2	{A, C, D}
t_3	{A, B, C, E}
t_4	{D, E}
t_5	{C}

Item	frequency
A	2
B	2
C	4
D	2
E	3

Transaction	Items bought	(ordered) frequent Items
t_1	{B, C, E}	~~{B, C, E}~~ {C, E, B}
t_2	{A, C, D}	~~{A, C, D}~~ {C, A, D}
t_3	{A, B, C, E}	~~{A, B, C, E}~~ {C, E, A, B}
t_4	{D, E}	~~{D, E}~~ {E, D}
t_5	{C}	~~{C}~~ {C}

Date : / /



Date : / /

Q3

1) $\{A\} \rightarrow \{C\}$

$$\text{Confidence} = \frac{f(A,C)}{f(A)} = \frac{2}{2} = 100\%$$

2) $\{C\} \rightarrow \{A\}$

$$\text{confidence} = \frac{f(A,C)}{f(C)} = \frac{2}{4} = 50\%$$

3) $\{B\} \rightarrow \{CE\}$

$$\text{confidence} = \frac{f(B,C,E)}{f(B)} = \frac{2}{2} = 100\%$$

4) $\{CE\} \rightarrow \{B\}$

$$\text{Confidence} = \frac{f(BCE)}{f(CE)} = \frac{2}{2} = 100\%$$

Date : / /

Q4

Yes, we can still apply the Apriori Algorithm to find the frequent itemsets even when support value of a transaction is sum of all item's weights (positive or negative weight).

But we can't say that derived frequent itemsets are reliable because, This weighted method is the dominance of heavy weighted items, irrespective of their occurrence frequency over the other low-weight most frequent items. Even the most frequently occured items are pruned if their support value is below the agreed minimum support, due to their lower weight.

If a particular subset of a superset is having support value less than the minimum support, due to this weighted approach, it needs to be pruned. So does the superset.