Assignment 5

Shashwat Shah
60004220126
C2-2 Div B

1 Construct fp tree with sup = 2

Tid	Items.
1	{b, a}
2	{ b, d, c)
3	{a, d, e}
4	{ d, e, a, c)
5	(c, b, a)
6	(a, c, b, d)
7	(a, f)
8	(b, a, c)
9	(b, d, a)
10	(c, e, b)

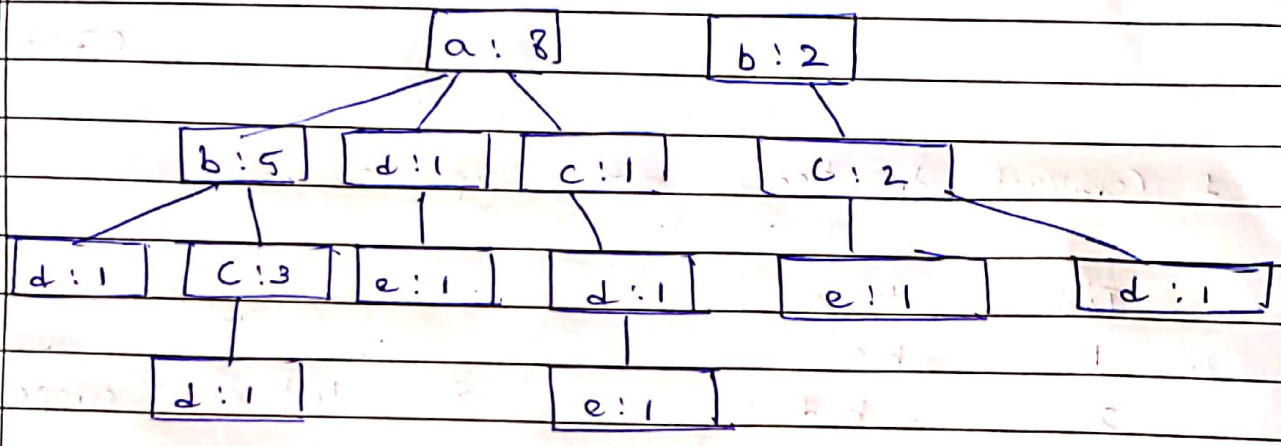
∴ Individual support

a = 8
b = 7
c = 6
d = 9
e = 3
f = 1

∴ Rearranged

Tid	Items.
1	a, b
2	b, c, d
3	a, d, e
4	a, c, d, e
5	a, b, c
6	a, b, c, d
7	a
8	a, b, c
9	a, b, d
10	b, c, e

Tree { }



2) A database has 5 transactions. Let min support = 60 %. and minimum confidence = 80 %.

T	Items
100	MONKEY
200	DONKEY
300	MAKE
400	MUCKY
500	COOKIE

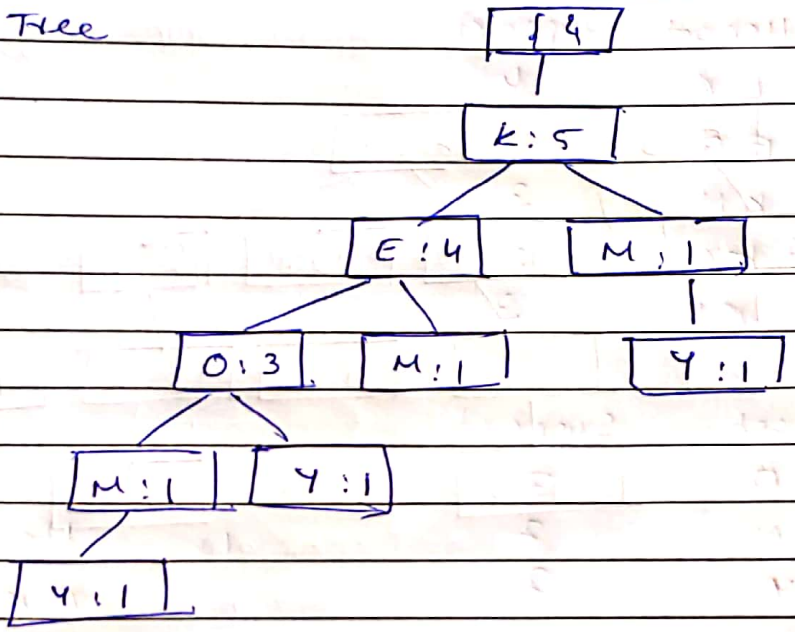
∴ frequency

A = 1	M = 3
C = 2	N = 2
D = 1	O = 3
E = 4	U = 1
I = 1	Y = 3
K = 5	

∴ Rearrange,

T	Items.
100	KEOMY
200	KEOY
300	KEM
400	KMY
500.	KEO,

Tree



(2) Using apriori

C1:

Itemset	Support
A	1
C	2
D	1
E	4
I	1
K	5
M	3
N	2
O	3
V	1
Y	3

→ L1:

Itemset	support
E	4
K	5
M	3
O	3
Y	3

C2:

Itemset	Support
EK	4
EM	2
EO	3
EY	2
KM	3
KO	3
KY	3
MO	1
MY	2
OY	2

L2 :

Itemset	Support
EK	4
EO	3
KM	3
KO	3
KY	3

C3 =

Itemset	Support
EKO	3
EKM	2
EKY	2
EKMO	1
KMY	2
KOY	2

L3	Itemset	Support
	E.K.O	3

No more combinations can be formed.

L = {E.K.O}

Subset = {K} {E} {O} {OK} {OE} {KE}

Associate rules using apriori :-

K → OE = 3/5 = 60%.

OE → K = 3/3 = 100%.

O → KE = 3/3 = 100%.

KE → O = 3/4 = 75%.

E → OK = 3/4 = 75%.

OK → E → 3/3 = 100%.

Min-confidence = 80%.

Strong association rules are

(O,E) → (K)

(O) → (K,E)

(OK) → (E)

Association Rules using FP tree

$L_1 = KY$ $L_2 = KI$ $L_3 = KO$ $L_4 = EO$

$L_5 = KEO$ $L_6 = KE$

i) $L_1 = (K, Y)$ Subset K , Y

Rules	Confidence	%
K → Y	3/5	60%
Y → K	3/3	100%

ii) $L_2 = KM$ Subset K M

Rules	Confid	%
K → M	3/5	60%
M → K	3/3	100%

iii) $L_3 = K, O$ Subset K O

K → O	3/5	60%
O → K	3/3	100%

iv) $L_4 = E$ O Subset E O

Rules	Confidence	%
E → O	3/4	75%
O → E	3/3	100%

v) L_5 : KEO Subset K E O KE KO EO

Rules	Confid	%
K → OE	3/5	60%
KE → O	3/3	100%
O → KE	3/3	75%
OE → K	3/4	75%
E → OK	3/4	100%
OK → E	3/3	

v.) $L_6 = (K, E)$ subset K E

Rules	Conf'd	%
K → E	4/5	75%.
E → K	4/4	100%.

Strong rules-

Y → K M → K O → K O → E

OE → K O → KE OK → E K → E E → K

b) The resulting frequent patterns are similar for both fp tree and the apriori algorithm. However in terms of overal algorithm efficiency for finding frequent patterns amongst dataset. FP tree growth algorithm is better than the apriori algorithm as it does not require candidate generation thus saving time and space. Also fp tree algorithm is able to mine in the conditional pattern base which may substantially reduce the size of the database to be searched. In certain conditions where the dataset to be mined is small it is seen that the efficiency of the apriory algorithm increases.

3) $A_1(2,10)$, $A_2(2,5)$ $A_3(8,4)$, $B_1(5,8)$, $B_2(7,5)$, $B_3(6,4)$
$C_1(1,2)$, $C_2(4,9)$

→ let X_1, X_2, X_3 responses then centroids.

∴ Initial centroids are as follows

$X_1 = A_1(2,10)$ $X_2 = B_1(5,8)$ $X_3 = C_1(1,2)$

∴

First iteration

Cluster 1 = A, (2, 10)

Cluster 2 = (5, 8) (8, 4) (7, 5) (5, 4) (4, 9)

Cluster 3 = (1, 2) (4, 5)

Centroids after first iteration

$X_1 = (2, 10)$

$X_2 (6, 6)$

$X_3 (1.5, 3.5)$

Second iteration

Cluster 1 = (2, 10) (4, 9)

Cluster 2 = (8, 4) (5, 8) (7, 5) (6, 4)

Cluster 3 = (4, 5) (1, 2)

$X_1 = (3, 4.5)$ (6.5, 5.25) (1.5, 3.5)

Third iteration

1 = (2, 10) (5, 8) (4, 9)

2 = (8, 4) (7, 5) (6, 4)

3 = (2, 5) (1, 2)

Centroid $X_1 = (3.67, 9)$

$X_2 = (7, 4.33)$

$X_3 = (1.5, 3.5)$

Fourth iteration

1 = (2, 10) (5, 8) (4, 9)

2 = (8, 4) (7, 5) (6, 4)

3 = (2, 5) (1, 2)

∴ Final Clusters

$C1 = (A_1, B_1, C_2)$

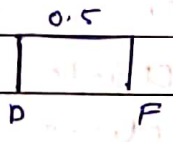
$C2 = (A_3, B_2, B_3)$

$C_3 = (A_2\ C_1)$

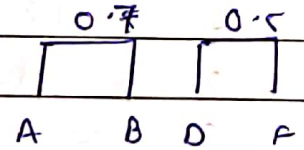
4 a) Single link

At level 0 there are 6 clusters.

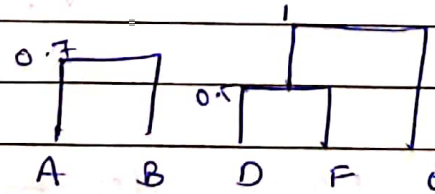
	A	B	C	DF	E
A	0				
B	0.7	0			
C	5.66	4.95	0		
DF	3.2	2.5	2.5	0	
E	4.24	3.54	1.41	2	0



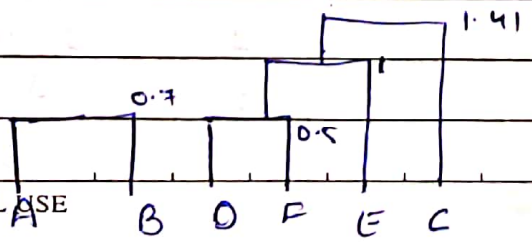
	AB	C	DF	E
AB	0			
C	4.95	0		
DF	2.5	2.5	0	
E	3.54	1.41	1	0



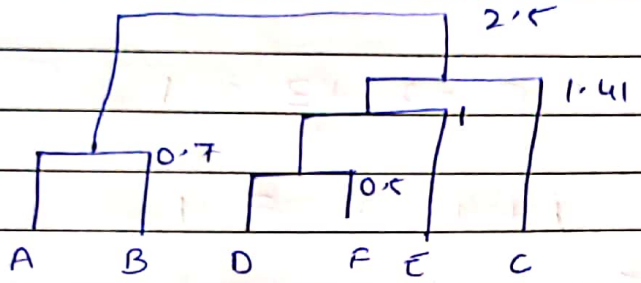
	AB	C	DEF
AB	0		
C	4.95	0	
DEF	2.5	1.41	0



	AB	CDEF
AB	0	
CDEF	2.5	0



∴



5) ~~Average link~~

Item	A	B	C	D
A	0			
B	1	0		
C	4	2	0	
D	5	6	3	0

∴ we use manhatten distance

∴ A → C

= 1 + 0 - 2 - 2 = -3

A → D

= 1 + 0 + 0 - 5 = -4

B → C

= 0 + 1 - 2 - 2 = -3

B → D

= 0 + 1 + 1 - 5 = -3

∴

new mediods = {D, B}

non-mediods = {A, C}

D → A

= -1 + 5 = 4

D → C

= 0 + 0 − 2 + 3 = 1

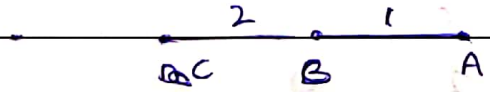
B → A

= 1 + 1 − 1 = 1

B → C

3 + 2 − 2 = 3

∴ Find clusters



Medoids = {D} {B}

non medoids = {A} {C}