Q. Consider following dataset, obtain frequent itemsets using Apriori Algorithm. min-sup = 40%.

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

N = 10, min-sup = 40% = 4.

Step 1 – 1-itemset generation.

C_1	count	frequent 1-itemset L_1	Infrequent 1-itemset
{a}	8	{a}	{e}.
{b}	7	{b}	
{c}	6	{c}	
{d}	5	{d}.	
{e}	3		

Step 2 – 2 itemset generation $L_1 \times L_1$

C_2	Pruned itemset		Infrequent 2-itemset	frequent 2-itemset L_2
{a, b}	Nill.	{a, b} - 5	{b, d}	{a, b}
{a, c}		{a, c} - 5	{c, d}	{a, c}
{a, d}		{a, d} - 4		{a, d}
{b, c}		{b, c} - 5		{b, c}.
{b, d}		{b, d} - 3		
{c, d}		{c, d} - 3		

Step 3 – 3 itemset generation $L_2 \times L_2$

C_3	Pruned itemset	Infrequent 3-itemset	frequent 3-itemset L_3
{a, b, c}	{a, b, d}.	{a, b, c} – 3	Nill.
{a, b, d}	{a, c, d}	{a, b, c}	
{a, c, d}			

All frequent itemset

{a}	{a, b}
{b}	{a, c}
{c}	{a, d}
{d}	{b, c}

Q. Consider following set of 3-frequent itemsets. Assume there are only five items in data set.

(a) List all candidate 4-itemsets obtained by candidate generation procedure in Apriori.

(b) List all candidate 4-itemsets that survive the candidate pruning step of Apriori Algorithm.

{1,2,3}, {1,2,4}, {1,2,5}, {1,3,4}, {1,3,5},
{2,3,4}, {2,3,5}, {3,4,5}

(a) $L_3 \times L_3$. 4-Itemset generation. - 4-Itemset candidates

{1,2,3,4} { ~~1,2,4~~ 2,3,4,5 }
{1,2,3,5}
{1,2,4,5}
{1,3,4,5}

(b) 4 itemsets that survive candidate pruning step of Apriori

{1,2,3,4} - All 3-item subset are frequent, survived
{1,2,3,5} - ——— " ———, survived
{1,2,4,5} - subset {2,4,5} infrequent - Pruned. {1,2,4,5}
{1,3,4,5} - All 3-item subset are frequent, survived
{2,3,4,5} - ——— " ———, survived.

Only {1,2,4,5} from step (a) is pruned in step (b)

Q. Use Apriori to find all frequent itemsets for the data given. Also give non-frequent-itemsets (N), itemsets not generated (NG) and itemsets pruned at each step (P), min-sup = 25%.

Txn	Items
1	K, S, T
2	J, T
3	J, S, T
4	J, S
5	K, S
6	J, K
7	J, K, S, T
8	J, S, T

→ N = 8, items → J, K, S, T, min-sup = 25% = 2

Step 1 – 1 itemset generation.

Candidate itemset C_1

{J} – 6
{K} – 4
{S} – 6
{T} – 5

non-frequent itemset (N): Nill. All are frequent

NG: Nill.

Itemset pruned (P): Nill. All 1-itemset are frequent

Step 2 – $L_1 \times L_1$

Candidate itemset (C_2)

{J, K} – 2
{J, S} – 4
{J, T} – 4
{K, S} – 3
{K, T} – 2
{S, T} – 4

Non-frequent itemset (N): Nill. All 2-itemset are frequent

NG: Nill.

Itemset pruned (P): Nill. No subset of each 2-itemset is infrequent

Step 3 – $L_2 \times L_2$.

C_3

{J, K, S} – 1
{J, K, T} – 1
~~{J, K, S, T}~~ –
{J, S, T} – 3
{K, S, T} – 2

N: {J, K, S}, {J, K, T}.

NG: Nill.

P: Nill. No subset of each 3-itemset is infrequent.

Step 4 → $L_3 \times L_3$.

C

{J, K, S, T} –.

N: Nill.

NG: Nill.

P: {J, K, S, T}.

Q. A database consists of 4 transactions. min-sup = 60%. find all frequent itemsets using Apriori Algorithm?

Txn	date	items.
1	10/15/99	K, A, D, B
2	10/15/99	D, A, C, E, B
3	10/19/99	C, A, B, E
4	10/22/99	B, A, D.

N = 4

$\text{min-sup} = \frac{60}{100} \times 4 = 2.4$

→ Rearrange dataset with ~~increasing~~ alphabetical order of items.

Txn	Items
1	A, B, D, K
2	A, B, C, D, E
3	A, B, C, E
4	A, B, D.

Step 1 → 1 itemset - generation

C_1		~~Pruned~~ itemset Infrequent	frequent 1-itemset L_1.
{A}	4	{C}	{A}
{B}	4	{E}	{B}
{C}	2	{K}.	{D}.
{D}	3		
{E}	2		
{K}	1		

Step 2 - 2 itemset generation $L_1 \times L_1$

C_2		Pruned itemset	Infrequent 2-itemset	frequent 2-itemset L_2
{A, B}	4	Nill.	Nill.	{A, B}
{A, D}	3			{A, D}
{B, D}	3			{B, D}.

Step 3 - 3 itemset generation $L_2 \times L_2$

C_3		frequent 3-itemset	Infrequent 3-itemset
{A, B, D}	3	{A, B, D}	Nill.

Pruned itemset. Nill.

Q. Consider transaction data in given dataset.

TID	List of item-IDs
1	I_1, I_2, I_5
2	I_2, I_4
3	I_2, I_3
4	I_1, I_2, I_4
5	I_1, I_3
6	I_2, I_3
7	I_1, I_3
8	I_1, I_2, I_3, I_5
9	I_1, I_2, I_3

1. find frequent itemsets using
   (a) Apriori algorithm
   (b) FP-Growth algorithm.
2. Generate association rules from frequent itemsets.

min_support = 2

min-confidence = 70%.

→ 1(a) Frequent itemsets generation using Apriori Algorithm -

Step 1 — Get frequency count support

unique items - I_1, I_2, I_3, I_4, I_5.

C1 — Candidate set C_1.

Item	Support
I_1	6
I_2	7
I_3	6
I_4	2
I_5	2

Compare candidate support count with min support count →

L1

Itemset	Sup. cnt
$\{I_1\}$	6
$\{I_2\}$	7
$\{I_3\}$	6
$\{I_4\}$	2
$\{I_5\}$	2

Nothing is removed as all are frequent

Step 2 — Generate C_2 candidates from L_1, $L_1 \times L_1$

C2

Itemset	Sup-cnt.
$\{I_1, I_2\}$	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

Compare candidate sup count with min sup cnt. →

L2

Itemset.	Sup. cnt
$\{I_1, I_2\}$	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

Step 3 — Generate C_3 candidates from L_2.

Itemset	Are subset frequent?	Pruned	Sup-cnt
$\{I_1, I_2, I_3\}$	Yes	No	2
$\{I_1, I_2, I_5\}$	Yes	No	2
$\{I_1, I_2, I_4\}$	$\{I_1, I_4\}$ not frequent	Yes.	
$\{I_1, I_3, I_5\}$	$\{I_3, I_5\}$ not frequent	Yes	
$\{I_2, I_3, I_4\}$	$\{I_3, I_4\}$ not frequent	Yes.	
$\{I_2, I_3, I_5\}$	$\{I_2, I_3\}$ not frequent	Yes	
$\{I_2, I_4, I_5\}$	$\{I_4, I_5\}$ not frequent	Yes.	

L3

Itemset	Sup-cnt
$\{I_1, I_2, I_3\}$	2
$\{I_1, I_2, I_5\}$	2

1. (a) Candidate generation using FP-Growth algorithm -

Step 1 - Generate frequent 1-itemsets.

{I_1} - 6
{I_2} - 7
{I_3} - 6
{I_4} - 2
{I_5} - 2

All are frequent.

Arrange frequent itemsets in descending order of support count

L = [{I_2:7}, {I_1:6}, {I_3:6}, {I_4:2}, {I_5:2}]

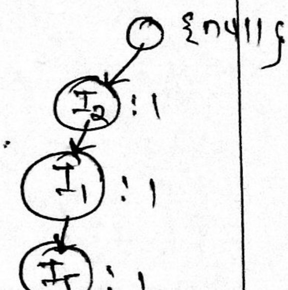
Step 2 - Construct FP-tree.

Item_Id	sup cnt	~~Node-Link~~
I_2	7	
I_1	6	
I_3	6	
I_4	2	
I_5	2	

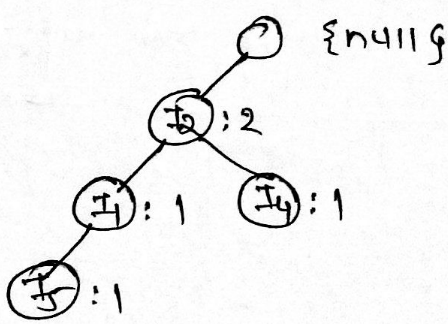
Arrange txn in L order.

1	I_2, I_1, I_5
2	I_2, I_4
3	I_2, I_3
4	I_2, I_1, I_4
5	I_1, I_3
6	I_2, I_3
7	I_1, I_3
8	I_2, I_1, I_3, I_5
9	I_2, I_1, I_3.
10	

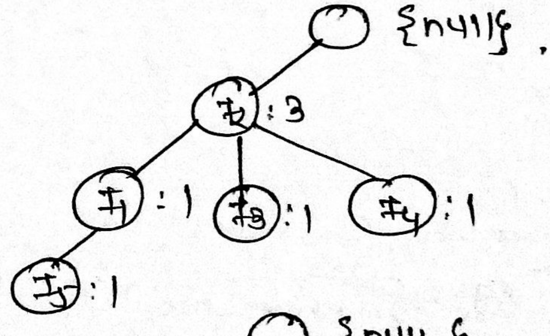
(a) Start with first transaction T_1 = {I_2, I_1, I_5}.



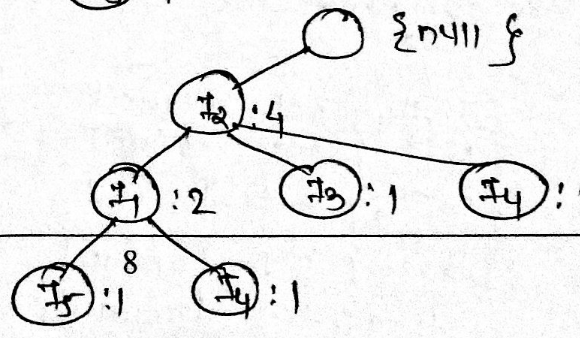
(b) Take T_2 = {I_2, I_4}.



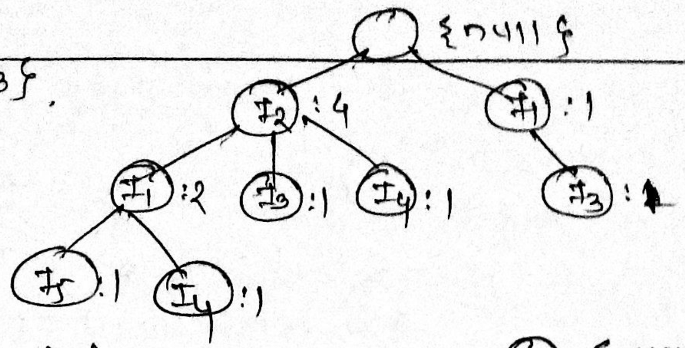
(c) Take T_3 = {I_2, I_3}.



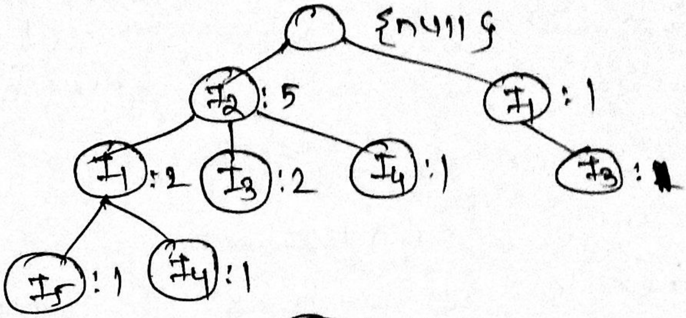
(d) Take T_4 = {I_2, I_1, I_4}.



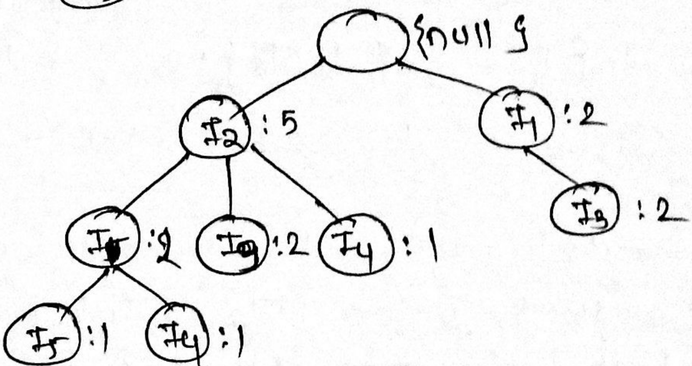
(e) Take $T_5 = \{I_1, I_3\}$.



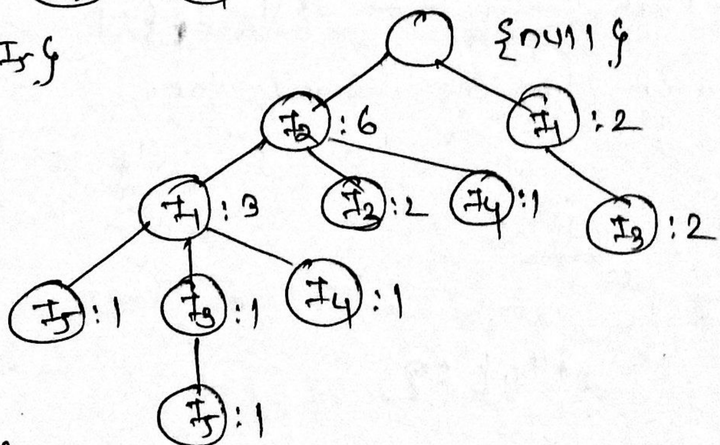
(f) Take $T_6 = \{I_2, I_3\}$



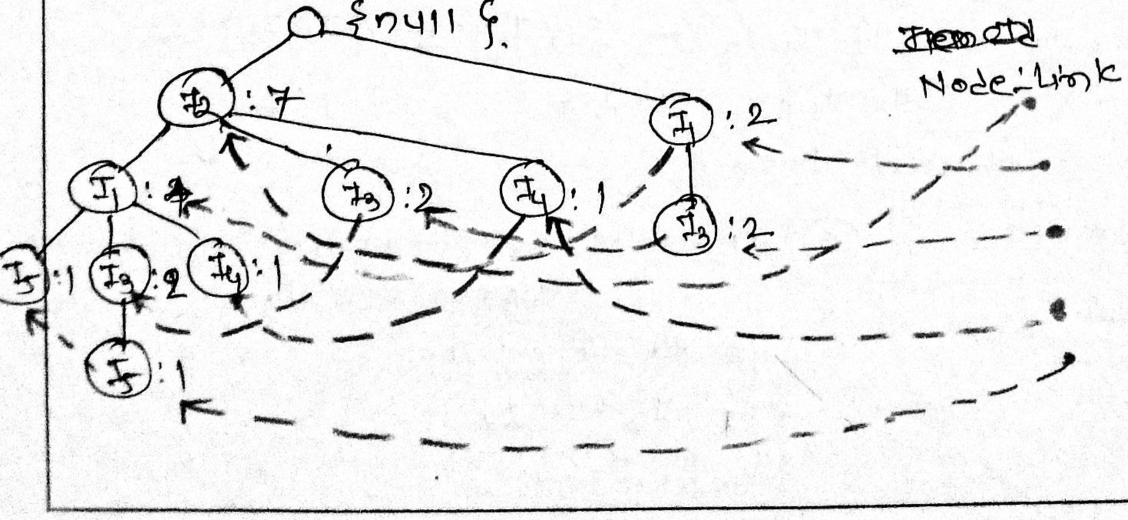
(g) Take $T_7 = \{I_1, I_3\}$.



(h) Take $T_8 = \{I_2, I_1, I_3, I_5\}$



(i) Take $T_9 = \{I_2, I_1, I_3\}$.



Item-Id	Sup. cnt
I_2	7
I_1	6
I_3	6
I_4	2
I_5	2

Step3 – Construct conditional Pattern ~~node~~ base for each ~~frequent~~ frequent 1-itemset

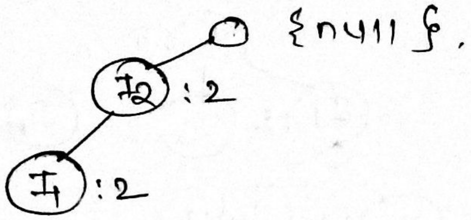
(a) Start with last item in L, i.e. I_5.

Paths involving I_5 $\rightarrow$ $\{I_2, I_1, I_5 : 1\}$ & $\{I_2, I_1, I_3, I_5 : 1\}$.

Prefix Paths for I_5 $\rightarrow$ $\{I_2, I_1 : \rangle$ & $\langle I_2, I_1, I_3 : 1\}$

= conditional pattern base.

(b) ~~Step~~ – Construct I_5 – conditional FP tree.



I_3 is ignored as its support count is 1 less than min-sup

(c) Generate combinations of frequent pattern for I_5.

$\{I_2, I_5\} : 2$, $\{I_1, I_5\} : 2$, $\{I_2, I_1, I_5\} : 2$.

Repeat 3a, 3b, 3c for each 1-frequent itemset.

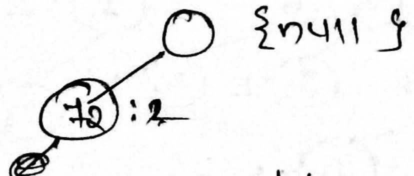
for I_4.

(a) ~~Generate~~ Generate conditional pattern base

Paths involving I_4 $\rightarrow$ $\{I_2, I_1, I_4\} : 1$, $\{I_2, I_4\} : 1$

Prefix Paths I_4 $\rightarrow$ $\{I_2, I_1\} : 1$, $\{I_2\} : 1$

(b) Construct I_4 – conditional tree.



I_1 is ignored as its sup-count 1 is less than min-sup.

(c) Generate combinations of frequent pattern for I_4.

$\{I_2, I_4\} : 2$

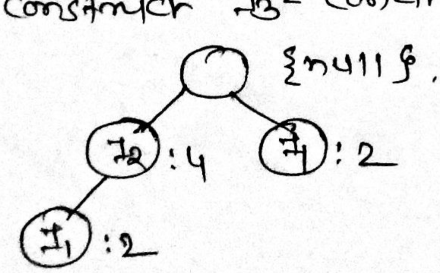
for I_3.

(a) Generate conditional pattern base.

Paths involving I_3 $\rightarrow$ $\{I_2, I_1, I_3\} : 2$, $\{I_2, I_3\} : 2$, $\{I_1, I_3\} : 2$

Prefix paths I_3 $\rightarrow$ $\{I_2, I_1\} : 2$, $\{I_2\} : 2$, $\{I_1\} : 2$

(b) Construct I_3 – conditional tree.



(c) Generate combinations of frequent pattern for I_3.

$\{I_2, I_3\} : 4$

$\{I_1, I_3\} : 2$

$\{I_2, I_1, I_3\} : 2$

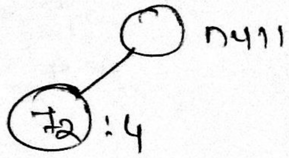
For [illegible]

(a) Generate conditional pattern base.

Paths involving [illegible] → $\langle I_2, I_1 \rangle : 4$, $\langle I_1 \rangle : 2$

Prefix paths [illegible] → $\langle I_2 \rangle : 4$

(b) Construct [illegible]-conditional tree.



(c) Generate combinations of frequent pattern for I_3.

$\{I_2, I_1\} : 4$

Mining FP-tree by creating conditional sub pattern bases

Item	Conditional Pattern base.	frequent pattern Generated
I_5	$\{I_2, I_1\}:1$, $\{I_2, I_1, I_3\}:1$	$\{I_2, I_5\}:2$, $\{I_1, I_5\}:2$, $\{I_2, I_1, I_5\}:2$
I_4	$\{I_2, I_1\}:1$, $\{I_2\}:1$	$\{I_2, I_4\}:2$
I_3	$\{I_2, I_1\}:2$, $\{I_2\}:2$, $\{I_1\}:2$	$\{I_2, I_3\}:4$, $\{I_1, I_3\}:4$, $\{I_2, I_1, I_3\}:$
I_2	$\{I_2\}:4$	$\{I_2, I_1 : 4\}$.

Q.2 (c) Generate association rules from frequent itemsets

frequent itemsets.

$\{I_2, I_5\}$, $\{I_1, I_5\}$, $\{I_2, I_1, I_5\}$, $\{I_2, I_4\}$, $\{I_2, I_1\}$.

$\{I_2, I_3\}$, $\{I_1, I_3\}$, $\{I_2, I_1, I_3\}$.

FP_1 : $\{I_2, I_5\}$. $\quad I_2 \rightarrow I_5$; $I_5 \rightarrow I_2$

FP2 : $\{I_1, I_5\}$ $\quad I_1 \rightarrow I_5$; $I_5 \rightarrow I_1$

FP3 : $\{I_2, I_1, I_5\}$. $\quad I_2 \rightarrow I_1 I_5$; $I_1 \rightarrow I_2 I_5$; $I_5 \rightarrow I_2 I_1$; $I_1 I_5 \rightarrow I_2$; $I_5 I_2 \rightarrow I_1$; $I_2 I_1 \rightarrow I_5$

FP4 : $\{I_2, I_4\}$: $I_2 \rightarrow I_4$; $I_4 \rightarrow I_2$

FP5 : $\{I_2, I_3\}$: $I_2 \rightarrow I_3$; $I_3 \rightarrow I_2$

FP6 : $\{I_1, I_3\}$ $\quad I_1 \rightarrow I_3$; $I_3 \rightarrow I_1$

FP7 : $\{I_2, I_1\}$ → $I_2 \rightarrow I_1$; $I_1 \rightarrow I_2$

FP8 : $\{I_2 I_1 I_3\}$ $\quad I_2 \rightarrow I_1 I_3$; $I_1 \rightarrow I_2 I_3$; $I_3 \rightarrow I_2 I_1$; $I_1 I_3 \rightarrow I_2$; $I_2 I_3 \rightarrow I_1$; $I_2 I_1 \rightarrow I_3$.

Check confidence of each generated rule, If it is less than min confidence discard that rule

Q. Consider following transaction dataset, obtain frequent itemsets using FP-Growth algorithm, min-sup = 40%.

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

min.sup = 40%.
$= \frac{40}{100} \times 10$
$= 4$.

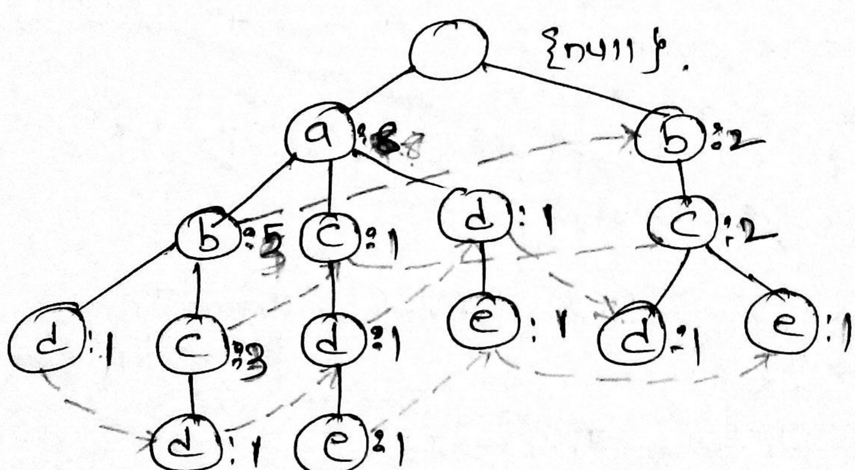
Step 1 – Generate frequent 1-itemset

Item	Sup. cnt
a	8
b	6
c	6
d	5
e	3

frequent 1-itemset	
a	8
b	6
c	6
d	5

L = [{a}:8, {b}:6, {c}:6, {d}:5].

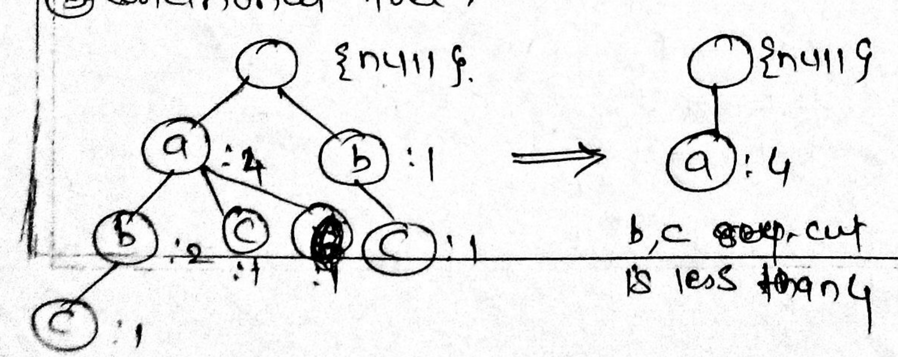
Step 2 – Construct fp tree.



Step 3 – for each 1-itemset, generate (a) conditional pattern base
(b) conditional tree,
(c) frequent patterns.

for. d (a) Conditional pattern base. <a,c,d>:1 , <a,d>:1
Paths ending with d ⇒. <a,b,d>:1 , <a,b,c,d>:1 , <b,c,d>:1
Prefix paths ⟶ <a,b>:1 , <a,b,c>:1 , <b,c>:1
<a,c>:1 , <a,d>:1 .

(b) Conditional tree.



b,c sup. cnt
is less than 4

(c) Generate frequent pattern

{a,d}:4.

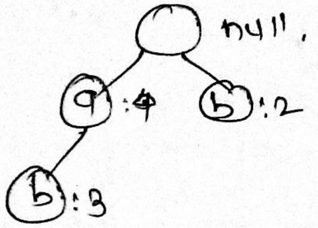
for c

(a) Conditional Pattern base,

Paths involving c → $\langle a,b,c\rangle:3$, $\langle a,c\rangle:1$, $\langle b,c\rangle:2$

Prefix paths with c → $\langle a,b\rangle:3$, $\langle a\rangle:1$, $\langle b\rangle:2$

(b) Conditional FP-tree, (c) Generate frequent patterns involving a & b



~~{a,b,c}~~

{a,c}:4

{b,c}:5

{a,b,c}:3

for b.

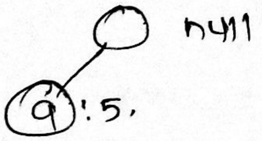
(a) Conditional Pattern base,

Paths involving b → $\langle a,b\rangle:5$, $\langle b\rangle:2$

Prefix paths with b → $\langle a\rangle:5$.

(b) Conditional FP tree (c) Generate frequent pattern involving a

{a,b}:5.



for a

(a) Conditional Pattern base,

Paths involving a → $\langle a\rangle:8$.

Prefix paths with a → ∅.

As no prefix paths involving a are there, FP tree can not be constructed.

Item	Conditional FP Base	Frequent Pattern
d	$\langle a,b\rangle:1$, $\langle a,b,c\rangle:1$, $\langle b,c\rangle:1$, $\langle a,c\rangle:1$, $\langle a,d\rangle:1$	{a,d}:4
c	$\langle a,b\rangle:3$, $\langle a\rangle:1$, $\langle b\rangle:2$	{a,c}:4 {b,c}:5 {a,b,c}:3
b	{a}:5.	{a,b}:5.
a	∅	∅.

min_conf = 70,

(B) Determine strong association rules from frequent itemsets.

FP: {a,d}

a→d, $s(a\to d) = \frac{3}{10} = 30\%$. $c(a\to d) = \frac{3}{8} = 37.5\%$.

d→a, $s(d\to a) = \frac{3}{10} = 30\%$. $c(d\to a) = \frac{3}{4} = 75\%$.

~~In order to say rule is strong,~~ Both support & confidence needs to be more than min_sup & min_conf. Hence, both rules are discarded.

FP: {a,c} $s(a \to c) = \frac{4}{10} = 40\%$. $c(a \to c) = \frac{4}{8} = 50\%$.

$s(c \to a) = \frac{4}{10} = 40\%$. $c(c \to a) = \frac{4}{6} = 66\%$.

Both are discarded, as min-conf criteria not satisfied.

FP: {b,c} $s(b \to c) = \frac{5}{10} = 50\%$. $c(b \to c) = \frac{5}{6} = 83.33\%$.

$s(c \to b) = \frac{5}{10} = 50\%$. $c(c \to b) = \frac{5}{6} = 83.33\%$.

Both rules are strong, as min-sup & min-conf criteria fulfilled.

for: {a,b}. $s(a \to b) = \frac{5}{10} = 50\%$. $c(a \to b) = \frac{5}{8} = 62.5\%$.

$s(b \to a) = \frac{5}{10} = 50\%$. $c(b \to a) = \frac{5}{6} = 83.33\%$.

Only IInd rule satisfies criteria, hence strong rule.

for {a,b,c}.

$s(ab \to c) = \frac{3}{10} = 30\%$. $c(ab \to c) = \frac{3}{5} = 60\%$.

$s(c \to bc) = \frac{3}{10} = 30\%$. $c(c \to ab) = \frac{3}{6} = 50\%$.

$s(ac \to b) = \frac{3}{10} = 30\%$. $c(ac \to b) = \frac{3}{4} = 75\%$.

$s(b \to ac) = \frac{3}{10} = 30\%$. $c(b \to ac) = \frac{3}{6} = 50\%$.

$s(bc \to a) = \frac{3}{10} = 30\%$. $c(bc \to a) = \frac{3}{5} = 60\%$.

$s(a \to bc) = \frac{3}{10} = 30\%$. $c(a \to bc) = \frac{3}{8} = 37.5\%$.

None of the above rules meets both min-sup & min-conf criteria, hence are discarded.

Strong A.R $b \to c$

$c \to b$.

$b \to a$.

Q. Using binary database from below table, with minsup = 4 & mincomf = 0.5

(a) List down frequent item set.

(b) find confidence of A.R. BC → D, CD → EF, AC → F, CE → F, ABC → DEF

(c) find support for {ABCDEF}, {EF}, {BDE}, {AEF}.

(d) find maximal frequent item set or set.

Txn.	A	B	C	D	E	F
1	1	1	1	1	1	0
2	0	1	1	0	0	1
3	0	1	0	1	1	0
4	0	1	1	0	1	1
5	0	1	1	1	1	0
6	1	1	0	1	1	1
7	1	0	1	1	1	0

Ans. →

(a) frequent item sets.

	Support Count
A →	3
B →	6
C →	5
D →	5
E →	6
F →	3

$\xRightarrow{minsup = 4}$ frequent 1-itemset: B, C, D, E

⇒ 2-itemset

BC →	4
BD →	4
BE →	5
CD →	3
CE →	4
DE →	5

⇒ frequent 2-itemset: BC, BD, BE, CE, DE

3-itemset

BCD →	2
BCE →	3
BDE →	4
CDE →	3

⇒ frequent 3-itemset: BDE.

frequent item sets ⇒ {B}, {C}, {D}, {E}, {BC}, {BD}, {BE}, {CE}, {DE} & {BDE}.

(b)

$$c(BC \to D) = \frac{s(BCD)}{s(BC)} = \frac{2}{4} = 0.5$$

$$c(CD \to EF) = \frac{s(CDEF)}{s(CD)} = \frac{0}{2} = 0$$

$$c(AC \to F) = \frac{s(ACF)}{s(AC)} = \frac{0}{2} = 0$$

$$c(CE \to F) = \frac{s(CEF)}{s(CE)} = \frac{1}{4} = 0.25$$

$$c(ABC \to DEF) = \frac{s(ABCDEF)}{s(ABC)} = \frac{0}{1} = 0$$

(c) $sup(ABCDE) = \frac{1}{7} = 0.142$

$sup(EF) = \frac{2}{7} = 0.285$

$sup(BDE) = \frac{4}{7} = 0.571.$

$sup(AEF) = \frac{1}{7} = 0.142$

(D) Maximal frequent item set –

frequent itemsets →	immediate ~~frequent~~ superset			is maximal?
B (6)	BC(4)	BD(4)	BE(5)	No
C (5)	BC(4)	CD(3)	CE(4)	No
D (5)	BD(4)	CD(3)	DE(5)	No
E (6)	BE(5)	CE(4)	DE(5)	No
BC(4)	BCD(2)	BCE(3)		Yes,
BD(4)	BDE(4)			No
BE(5)	BDE(4)			No
CE(4)	CDE(3)	BCE(3)		Yes,
DE(5)	BDE(4)	CDE(3)		No.
BDE(4)	ABDE(2)	BCDE(1)	BDEF(1)	Yes.

Maximal frequent itemsets are CE, BDE, BC

Maximal frequent itemset is frequent itemset for which none of its immediate superset are frequent

Q. Suppose there are 20 items, numbered 1 to 20. & also 20 baskets, numbered 1 to 20. Item i is in basket b if & only if i divides b with no reminder.

(a) If support threshold is 5, what item are frequent. ?

(b) If support threshold is 5, which pairs of items are frequent

(c) Which basket is largest. ?

→

Basket (b)	Item (i)	Basket (b)	Item (i)
1	1	11	1, 11
2	1, 2,	12	1, 2, 3, 4, 6, 12
3	1, 3,	13	1, 13
4	1, 2, 4,	14	1, 2, 7, 14
5	1, 5	15	1, 3, 5, 15
6	1, 2, 3, 6	16	1, 2, 4, 8, 16
7	1, 7	17	1, 17
8	1, 2, 4, 8,	18	1, 2, 3, 6, 9, 18
9	1, 3, 9	19	1, 19
10	1, 2, 5, 10	20	1, 2, 4, 5, 10, 20

(a) min_sup = 5

frequent 1-itemsets - ~~{12}~~, ~~{16}~~, ~~{18}~~, ~~{20}~~.
{1}, {2}, {3}, {4}, {

(b) min_sup = 5.

Using Apriori principle, only superset of frequent itemset will be frequent so creating 2-itemset from 1-itemset obtained in (a).

{1, 2} ~~{12, 16}~~ — support count 10 — frequent 2-itemset
{1, 3} ~~{12, 18}~~ — 6 — 〃
{1, 4} ~~{12, 20}~~ — 5 — 〃.
{2, 3} ~~{16, 18}~~ — 3 — not 〃
{2, 4} ~~{16, 20}~~ — 5 — frequent 2-itemset
{3, 4} ~~{18, 20}~~ — 1 — not 〃.

frequent 2-itemset = {1, 2}, {1, 3}, {1, 4}, {2, 4}.

(c) Largest basket → 12, 18, 20.

(d) What is confidence of AR {2, 4} → {3}.

$$c(\{2,4\} \rightarrow \{3\}) = \frac{\sigma(2,4,3)}{\sigma(2)} = \frac{1}{10} = 0.1 = 10\%.$$

Consider following set of 3-frequent itemsets. Assume there are only ~~five~~ six items in transaction sets.

(a) List all candidate 4-itemsets obtained by candidate generation procedure in Apriori.

(b) List all candidate 4-itemsets that survive the candidate pruning step of Apriori algorithm.

→ (A) {1,2,3}, {1,2,6}, {1,3,4}, {2,3,4}, {2,4,5}
{3,4,6}, {4,5,6}.

→ (a) Candidate 4-itemset generation $C_4 = L_3 \times L_3$

{1,2,3,6}
{1,2,3,4}
{1,2,4,5}
{1,3,4,6}
{2,3,4,5}
{2,3,4,6}
{2,4,5,6}
{3,4,5,6}

(b) 4-itemsets that survive candidate pruning step of Apriori

{1,2,3,6} →	subset {1,3,6} not frequent →	Pruned.
{1,2,3,4} →	{1,2,4} not frequent →	Pruned.
{1,2,4,5} →	{1,2,5} not frequent →	Pruned.
{1,3,4,6} →	{1,3,6} not frequent →	Pruned.
{2,3,4,5} →	{2,3,5} not frequent →	Pruned
{2,3,4,6} →	{2,3,6} not frequent →	Pruned
{2,4,5,6} →	{2,4,6} not frequent →	Pruned
{3,4,5,6} →	{3,4,5} not frequent →	Pruned

All candidate 4-itemsets are pruned.

Q. following contigency table summarizes supermarket txn data where Bread refers to transactions containing bread, ~ Bread refers to transaction that do not contain bread. Same syntax applies for coke as well.

ⓐ Suppose AR Bread $\Rightarrow$ Coke is mined, min-sup = 25%, min-conf = 50%, Is AR strong ?

ⓑ Based on given data, is purchase of Bread independent of purchase of Coke? If not, what kind of corelation relationship exists between two ?

	Bread	~ Bread	Row Sum
coke	2000	500	2500
~ coke	1000	1500	2500
	3000	2000	5000

	b	~b	
a	f_{11}	f_{10}	f_{1+}
~a	f_{01}	f_{00}	f_{0+}
	f_{+1}	f_{+0}	N

ⓐ AR Bread $\Rightarrow$ coke.

$$supp(\text{Bread} \Rightarrow \text{Coke}) = \frac{2000}{5000} = 40\% > \text{min-sup}.$$

$$conf(\text{Bread} \Rightarrow \text{coke}) = \frac{2000}{3000} = 67\% > \text{min-conf}$$

As both support & confidence of rule are greater than respective thresholds, AR is strong.

ⓑ $\cancel{I(A,B)} = \frac{\cancel{N f_{11}}}{\cancel{f_{1+} f_{+1}}}$

$f_{1+} = 2500$ $\quad$ $f_{0+} = 2500$

$f_{+1} = 3000$ $\quad$ $f_{+0} = 2000$

$N = 5000$

$\cancel{I(\text{Bread, Coke})} = \frac{\cancel{5000 \times 2000}}{\cancel{2500 \times 3000}} = \cancel{\frac{5 \times 2 \times 100}{75}} = \cancel{1.333}$

$$IS(A,B) = \frac{N \cdot f_{11}}{f_{1+} \, f_{+1}}.$$

$$I(\text{Bread, coke}) = \frac{5000 \times 2000}{2500 \times 3000} = 1.333 > 1.$$

$I(A,B) > 1$, if A & B are positively correlated.

Purchase of Bread is not independent of purchase of coke. There is exists a positive correlation between them.

Q. Suppose we have market basket data for 100 transaction & 20 items. Support for item a is 25%, support for item b is 90%. Support for itemset {a, b} = 20%. min-sup = 10%. min-conf = 60%.

(a) Compute confidence of AR {a} → {b}. Is the rule interesting according to confidence measure?

(b) Compute interest factor for association pattern {a, b}. Describe nature of relationship between item a & b in terms of interest measure.

(c) What conclusion can you draw from results of part (a) & part (b)?

→ (a) $\cancel{s(a\to b) = \frac{s(ab)}{s(a)\, s(b)} = \frac{0.24}{0.25\times 0.9} = 0.88 > \text{min-sup}}$

$$c(a\to b) = \frac{s(ab)}{s(a)} = \frac{0.2}{0.25} = 0.8 > \text{min-conf}$$

∴ Confidence exceeds minimum confidence value hence its interesting.

(b) $$I(A,B) = \frac{s(A,B)}{s(A)\, s(B)}$$

$$I(a,b) = \frac{s(a,b)}{s(a)\, s(b)} = \frac{0.2}{0.25\times 0.9} = 0.889$$

$I(A,B) < 1$, If A & B are negatively correlated.

Hence, items a, & b are negatively correlated.

(c) High confidence rules may not be interesting.