

NAME - ASHIS GHOSH

ROLL - 18101104063

DEPT - Computer Science And Engineering

YEAR - 3rd year

SUBJECT - Data Mining

- 1) Find all frequent item sets using Apriori Algorithm and derive the strong association rules for the given data set. Let the min-support = 20% and min-conf = 60%.

Transaction ID	List of Item ID's
T <sub>1</sub>	A, B, E
T <sub>2</sub>	B, C, D
T <sub>3</sub>	B, D, E
T <sub>4</sub>	C, D, E
T <sub>5</sub>	B, C, D, E
T <sub>6</sub>	B, C, E

Ans:-

min-support = 20%

So, min support count =  $\frac{20}{100} \times 6 = 1.2$

C1

Itemset	Support Count
A	1
B	5
C	4
D	4
E	5

L1

Itemset	Sup. Count
B	5
C	4
D	4
E	5

C2

Item set	Support Count
{BC}	3
{BD}	3
{BE}	4
{CD}	3
{CE}	3
{DE}	3

<u>L2</u>	<u>Itemset</u>	<u>Support Count</u>
	$\{ABC\}$	3
	$\{BD\}$	3
	$\{BE\}$	4
	$\{CD\}$	3
	$\{CE\}$	3
	$\{DE\}$	3

<u>C3</u>	<u>Item Set</u>	<u>Support Count</u>
	$\{B,CD\}$	2
	$\{B,CE\}$	2
	$\{B,D,E\}$	2
	$\{C,D,E\}$	2

<u>L3</u>	<u>Itemset</u>	<u>Support Count</u>
	$\{B,C,D\}$	2
	$\{B,C,E\}$	2
	$\{B,D,E\}$	2
	$\{C,D,E\}$	2

<u>C4</u>	<u>Itemset</u>	<u>Support Count</u>
	$\{B,C,D,E\}$	1

E4 There will be no Itemset in L4.

So, frequent itemsets are —

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

So, taking the frequent itemsets  $\{B,C,D\}$  the rules are —



rule-1  $\{B, C\} \rightarrow \{D\}$ , confidence =  $\frac{\text{sup}\{B, C, D\}}{\text{sup}\{B, C\}}$   
 $= \left(\frac{2}{3} \times 100\right)\%$   
 $= 66.66\%$

rule-2  $\{B, D\} \rightarrow \{C\}$ , confidence =  $\frac{\text{sup}\{B, C, D\}}{\text{sup}\{B, D\}}$   
 $= \left(\frac{2}{3} \times 100\right)\% = 66.66\%$

rule-3  $\{C, D\} \rightarrow \{B\}$ , confidence =  $\frac{\text{sup}\{B, C, D\}}{\text{sup}\{C, D\}}$   
 $= \left(\frac{2}{3} \times 100\right)\% = 66.66\%$

rule-4  $\{B\} \rightarrow \{C, D\}$ , confidence =  $\frac{\text{sup}\{B, C, D\}}{\text{sup}\{B\}}$   
 $= \left(\frac{2}{3} \times 100\right)\% = 66.66\%$

rule-5  $\{C\} \rightarrow \{B, D\}$ , confidence =  $\frac{\text{sup}\{B, C, D\}}{\text{sup}\{C\}}$   
 $= \left(\frac{2}{4} \times 100\right)\% = 50\%$

rule-6  $\{D\} \rightarrow \{B, C\}$ , confidence =  $\frac{\text{sup}\{B, C, D\}}{\text{sup}\{D\}}$   
 $= \left(\frac{2}{4} \times 100\right)\% = 50\%$

As min-conf = 60%, so rule 1, rule 2, rule 3 are strong association rules.

Q2) Find all frequent itemsets using ~~Apriori~~ Apriori Algorithm and derive the strong association rules for the given data set. Let the min-support count = 3 and min-conf = 75%.

Transaction ID	List of Item ID's
T <sub>1</sub>	f, a, e, d, g, i, m, p
T <sub>2</sub>	a, b, c, f, l, m, o
T <sub>3</sub>	b, f, h, j, o
T <sub>4</sub>	b, c, k, s, p
T <sub>5</sub>	a, f, c, e, l, p, m, n

Ans :-

c1

Itemset	Support Count
a	3
b	3
c	4
d	1
e	1
f	4
g	1
h	1
i	1
j	1
k	1
l	<del>1</del> 2
m	3
n	<del>0</del> 1
o	2
p	3
s	1

L1

Itemset	Support Count
a	3
b	3
c	4
<del>o</del> f	4
m	3
p	3

c2

Itemset	Support Count
{a, b}	<del>1</del>
{a, c}	3
{a, f}	3
{a, m}	3
{a, p}	2
{b, c}	2

Itemset	Support count
$\{b, f\}$	2
$\{b, m\}$	1
$\{b, p\}$	1
$\{c, f\}$	3
$\{c, m\}$	3
$\{c, p\}$	3
$\{f, m\}$	3
$\{f, p\}$	2
$\{m, p\}$	2

L2

Itemset	Sup. count
$\{a, c\}$	3
$\{a, f\}$	3
$\{a, m\}$	3
$\{c, f\}$	3
$\{c, m\}$	3
$\{c, p\}$	3
$\{f, m\}$	3

L3

Itemset	Support count
$\{a, c, f\}$	3
$\{a, c, m\}$	3
$\{a, f, m\}$	3
$\{c, f, m\}$	3



Itemset	Support count
$\{a, c, f\}$	3
$\{a, c, m\}$	3
$\{a, f, m\}$	3
$\{c, f, m\}$	3

C4

Itemset	Support count
$\{a, c, f, m\}$	3

L4

Itemset	Sup. count
$\{a, c, f, m\}$	3

So, frequent Itemset  $\rightarrow \{a, c, f, m\}$

rule 1  $\{a, c, f\} \rightarrow \{m\}$ , Confidence =  $(\frac{3}{3} \times 100)\% = 100\%$

rule 2  $\{a, c, m\} \rightarrow \{f\}$  Confidence =  $(\frac{3}{3} \times 100)\% = 100\%$

rule 3  $\{a, f, m\} \rightarrow \{c\}$  Confidence =  $(\frac{3}{3} \times 100)\% = 100\%$

rule 4  $\{c, f, m\} \rightarrow \{a\}$  Confidence =  $(\frac{3}{3} \times 100)\% = 100\%$

rule 5  $\{a, c\} \rightarrow \{f, m\}$  Confidence =  $(\frac{3}{3} \times 100)\% = 100\%$

rule 6  $\{a, f\} \rightarrow \{c, m\}$  Confidence =  $(\frac{3}{3} \times 100)\% = 100\%$

rule 7  $\{a, m\} \rightarrow \{c, f\}$  Confidence =  $(\frac{3}{3} \times 100)\% = 100\%$

rule 8  $\{c, f\} \rightarrow \{a, m\}$  Confidence =  $(\frac{3}{3} \times 100)\% = 100\%$

rule 9  $\{c, m\} \rightarrow \{a, f\}$  Confidence =  $(\frac{3}{3} \times 100)\% = 100\%$

- rule 10  $\{t, m\} \rightarrow \{a, c\}$  confidence =  $(\frac{3}{3} \times 100)\%.$   
rule 11  $\{a\} \rightarrow \{c, t, m\}$  confidence =  $(\frac{3}{3} \times 100)\% = 100\%$   
rule 12  $\{c\} \rightarrow \{a, t, m\}$  confidence =  $(\frac{3}{4} \times 100)\% = 75\%$   
rule 13  $\{t\} \rightarrow \{a, c, m\}$  confidence =  $(\frac{3}{4} \times 100)\% = 75\%$   
rule 14  $\{m\} \rightarrow \{a, c, t\}$  confidence =  $(\frac{3}{3} \times 100)\% = 100\%$

So, all of the rules are strong association rules.

Q3) Find all frequent itemsets using Apriori Algorithm and derive the strong rules for the given data set. Let the min-support = 40% and min-conf = 70%.

Transaction ID	List of Item
T <sub>1</sub>	Diaper, Butter, Detergent
T <sub>2</sub>	Butter, Diaper, Milk
T <sub>3</sub>	Diaper, Baby, Powder
T <sub>4</sub>	Butter, Diaper, Baby Powder, Bread, Umbrella
T <sub>5</sub>	Butter, Milk, Coca-Cola

Ans :- min-support = 40%.

So, min support count =  $(\frac{40}{100} \times 5) = 2$

<u>CI</u>	<u>Itemset</u>	<u>Sup. Count</u>
	Diaper	4
	Butter	4
	Detergent	1
	Milk	2
	Baby Powder	2
	Bread	1
	Umbrella	1
	Coca-Cola	1

<u>L1</u>	<u>Itemset</u>	<u>Sup. Count</u>
	Diaper	4
	Butter	4
	Milk	2
	Baby powder	2

<u>C2</u>	<u>Itemset</u>	<u>Sup. Count</u>
	{Diaper, Butter}	3
	{Diaper, Milk}	1
	{Diaper, Baby powder}	2
	{Butter, Milk}	2
	{Butter, Baby powder}	1
	{Milk, Baby powder}	0

<u>L2</u>	<u>Itemset</u>	<u>Sup count</u>
	{Diaper, Butter}	3
	{Diaper, Baby powder}	2
	{Butter, Milk}	2

C3 C3 doesn't contain any item.

So, all frequent itemsets are —

{Diaper, Butter}

{Diaper, Baby powder}

{Butter, Milk}

So, taking the frequent item  $\rightarrow$  {Diaper, Butter}

The rules are —

rule 1 {Diaper}  $\rightarrow$  {Butter}, Confidence =  $(3/4 \times 100)\%$   
= 75%

rule 2 {Butter}  $\rightarrow$  {Diaper}, Confidence =  $(3/4 \times 100)\%$   
= 75%

So, rule 1, rule 2 are strong association rules



Q4) Find all frequent using FP-Tree and derive the strong association rules for the given data set. Let the min-support = 50% and min-conf = 70%.

Transaction ID	List of Item
T <sub>1</sub>	Shoes, Socks, Tie
T <sub>2</sub>	Shoes, Socks, Tie, Belt
T <sub>3</sub>	Shirts
T <sub>4</sub>	Shoes, Tie
	Shoes, Socks, Belt

Ans.

Q1

Itemset	Sup. Count
Shoes	4
Socks	3
Tie	3
Belt	2
Shirts	1

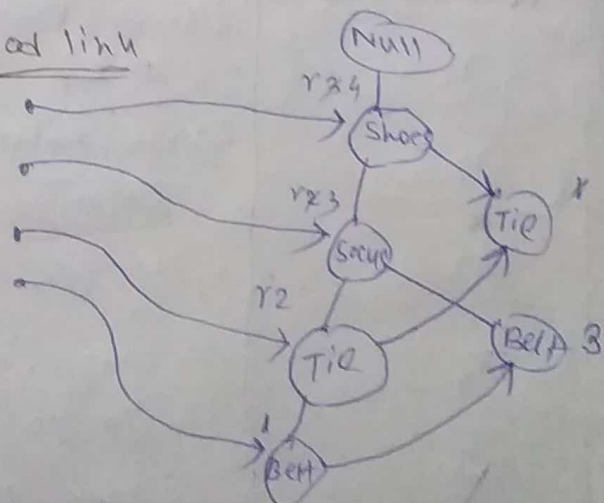
$$\text{min support count} = \frac{50}{100} \times 4 = 2$$

L1

Itemset	Sup. Count
Shoes	4
Socks	3
Tie	3
Belt	2

Item	Count
Shoes	4
Socks	3
Tie	3
Belt	2

Node link



Item	Conditional Pattern Base	Conditional FP tree	Frequent pattern generated
Belt	{Shoe, Socks, Tie:1}	{Shoe, Socks:2}	{Shoe, Belt:2}
Tie	{Shoe, Socks:1} {Shoe, Socks:2}	{Shoe:3}	{Socks, Belt:23}
Socks	{Shoe:13}	{Socks:2}	{Shoe, Socks, Belt:2}
	{Shoe:3}	{Shoe:3}	{Shoe, Tie:3}
			{Socks, Tie:2}
			{Shoe, Socks, Tie:2}
			{Shoe, Socks:3}

taking the itemset  $L = \{Shoes, Socks, Tie\}$

rule1 shoes  $\rightarrow$  Socks & Tie Conf = 50%.

rule2 Socks  $\rightarrow$  shoes & Tie conf = 66.6%.

rule3 Tie  $\rightarrow$  shoes & Socks Conf = 66.66%.

rule4 shoes & Socks  $\rightarrow$  Tie Conf = 66.66%.

rule5 shoes & Tie  $\rightarrow$  Socks Conf = 100%.

rule6 Socks & Tie  $\rightarrow$  shoes Conf = 100%.

Strong association rules are - rule5, rule6.

Q5) Find all frequent itemsets using FP Tree, derive the strong association rules, min-support = 60%, min-conf = 70%.

Transaction ID	Items
T <sub>1</sub>	{Mango, Onion, Nintendo, Key-chain, Eggs, Yo-yo}
T <sub>2</sub>	{Doll, onion, Nintendo, Key-chain, Eggs, Yo-yo}
T <sub>3</sub>	{Mango, Apple, Key-chain, Eggs}
T <sub>4</sub>	{Mango, Umbrella, Corn, Key-chain, Yo-yo}
T <sub>5</sub>	{Corn, Onion, Key-chain, Ice-cream, Eggs}

Ans: min support count =  $\frac{60}{100} \times 5 = 3$

ID  
T<sub>1</sub>  
T<sub>2</sub>  
T<sub>3</sub>  
T<sub>4</sub>  
T<sub>5</sub>

Items  
M, O, N, K, F, Y  
D, O, N, K, E, Y  
M, A, K, E  
M, V, C, K, Y  
C, O, U, K, I, E

Let,  
Mango → M  
Onion → O  
Nintendo → N  
Key-chain → K  
Eggs → E  
Yo-yo → Y  
Doll → D  
Apple → A  
Umbrella → U  
Car = C  
Ice-cream = I

C1

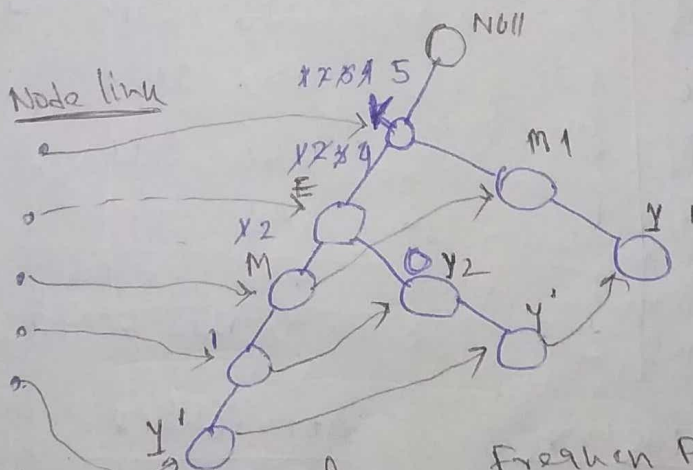
Items	Count
A	1
C	2
D	1
E	4
I	1
K	5
M	3
N	2
O	3
U	1
Y	3

L1

Items	Count
K	5
M	3
O	3
Y	3
E	4

Items	Count
K	5
E	4
M	3
O	3
Y	3

Node link



Item  
Y

Conditional Pattern Base

{K, E, M, O, 1}  
{K, E, O, 1}  
{K, M, 1}  
{K, E, M, 1}  
{K, E, 2}  
{K, E, 2}  
{K, 1}  
{K, 4}

Conditional FP tree

{K: 3}  
{K, E: 3}  
{K: 3}  
{K: 4}

Frequent Pattern Generated

{K, Y: 3}  
{K, O, 1}, {K, E, 1}  
{K, M: 3}  
{K, E, 4}



So, taking itemset  $\{keychain, Eggs, onion\}$   
 or  $\{K, E, O\}$

rule 1  $K \rightarrow E \wedge O$  Conf = 75%

rule 2  $E \rightarrow K \wedge O$  Conf = 75%

rule 3  $O \rightarrow K \wedge E$  Conf = 100%

rule 4  $E \wedge K \rightarrow O$  Conf = 75%

rule 5  $K \wedge O \rightarrow E$  Conf = 100%

rule 6  $E \wedge O \rightarrow K$  Conf = 100%

So, rule 1, 2, 3, 4, 5, 6 are strongly association rules.

Q6)

$K = 3 [P_2, P_4, P_7]$

	X	Y	d <sub>1</sub>	d <sub>2</sub>	d <sub>3</sub>	Cluster
P <sub>1</sub>	2	5	4	7.78	4.12	1
P <sub>2</sub>	2	9	0	7.22	6.40	1
P <sub>3</sub>	8	6	6.7	3.76	2.83	3
P <sub>4</sub>	9	3	9.22	0	3.76	2
P <sub>5</sub>	5	8	3.12	6.40	4.12	1
P <sub>6</sub>	7	5	6.40	2.83	1.41	3
P <sub>7</sub>	6	4	6.40	3.12	0	3
P <sub>8</sub>	1	6	3.12	8.55	5.39	1
P <sub>9</sub>	4	7	2.83	6.40	3.60	1
P <sub>10</sub>	6	8	4.12	5.83	4	3

Cluster

1

2

3

Attributes

P<sub>1</sub>, P<sub>2</sub>, P<sub>5</sub>, P<sub>8</sub>, P<sub>4</sub>

P<sub>9</sub>

P<sub>3</sub>, P<sub>6</sub>, P<sub>7</sub>, P<sub>10</sub>

Mean

$x = 5$

$y = 7 (2, 8, 7)$

$(9, 3)$

$x = 6.8$

$y = 5.8$

$(6.8, 5.8)$

	x	y	d <sub>1</sub>	d <sub>2</sub>	d <sub>3</sub>	cluster
P <sub>1</sub>	2	5	2.16	7.28	4.87	1
P <sub>2</sub>	2	9	2.16	9.22	5.77	1
P <sub>3</sub>	8	6	5.33	3.16	1.22	3
P <sub>4</sub>	9	3	7.98	6	3.56	2
P <sub>5</sub>	5	8	2.41	6.40	2.84	1
P <sub>6</sub>	7	5	4.65	2.83	0.83	3
P <sub>7</sub>	6	4	4.39	3.16	1.97	3
P <sub>8</sub>	1	6	2.01	8.55	5.80	1
P <sub>9</sub>	9	7	1.2	6.46	3.05	1
P <sub>10</sub>	6	8	3.35	5.83	2.84	3

<u>cluster</u>	<u>Attributes</u>	<u>mean</u>
1	P <sub>1</sub> , P <sub>2</sub> , P <sub>5</sub> , P <sub>8</sub> , P <sub>9</sub> P <sub>4</sub>	$\bar{x} = 2.8, \bar{y} = 7$ (2.8, 7) (4, 3)
2	P <sub>3</sub> , P <sub>6</sub> , P <sub>7</sub> , P <sub>10</sub>	$\bar{x} = 6.8, \bar{y} = 5.8$ (6.8, 5.8)
3		