Name: Kavya Kodishala
UNT ID: 11614043.

Data Mining.

Homework No: 5
CSCE 5380 Section: 002.

Q2)

a)

(i) FALSE, Because Support of {A,B} = support of {A,B,C} meaning that transaction containing {A,B} also has {A,C} but since there might be another record that only contains {A,C} Support of {A,B} and {A,C} are not equal as {A,C} will be greater than {A,B}

(ii) TRUE we know that the transaction that Contains {A,B} also Contains C as the Support of {A,B} = Support of {A,B,C}. This proves that the confidence of the Rule {A,B} → {C} is 100%.

(iii) FALSE, we know that Support of {A,B} = Support of {A,B,C} so a Seperate transaction for {A,B,D} is not possible which meakes the support of {A,B} equal to support of {A,B,C,D} but as the Support of {A,B} is not equal to {B,C} we can Say that there is another Record with a transaction {B,C,D} In Such a

a case the Support of {B,C,D} will be greater than Support {A,B,D}

b) i) TRUE, Based on given all the transactions that contain {A,B} also contain {B,C} This means that {A,B,C} must occur together in a transaction from this we can say that wherever {A,B} occurs in a records also occurs in it. Therefore the confidence of the Rule {A,B} → {C} is 100%.

(ii) FALSE, As it is given that all the transactions that contains {A,B} also contain {B,C} but we cannot assume that the vice-versa that is we cannot say that all the transactions that contains {B,C} also contain {A,B} so there can be another record which contains the item {B,C,D} only making the Support of {B,C,D} greater than {A,B}.

(iii) TRUE; An item is said to be closed item set if none of its supersets have the same support as the item set. Here if we consider {A,B,C,D} as Superset of {A,B,D} we know that all transactions that contain {A,B,C,D}

also known contains {B,C} making the occurance of {A,B,D} seperately from {A,B,C,D} not possible. Therefore, the support of {A,B,D} is equal to support of {A,B,C,D} so {A,B,D} is not a closed item set'.

Q.4)

a) Find all frequent itemset using Apriori algorithm.

Items	Support	Support percentage
A	1/5	20
C	2/5	40
D	1/5	20
E	4/5	80
I	1/5	20

K	5/5	100
M	3/5	60
N	2/5	40
O	4/5	60
U	1/5	20
Y	3/5	60

As it is given that minimum Support is equal to 60 so, we can Seperate all the items whose minimum Support = 60.

Items	Min Support = 60
E	4
K	5
M	3
O	4
Y	3

Now for 2 frequent itemsets.

2-Itemsets	Support	Percentage
E, K	4/5	80
E, M	2/5	40
E, O	3/5	60
E, Y	2/5	40
K, M	3/5	60
K, O	3/5	60
K, Y	3/5	60
M, O	1/5	20
M, Y	2/5	40
O, Y	2/5	40

Again taking all the itemsets with minimum Support = 60

Items	Min support
E,K	4
E,O	3
K,M	3
K,O	3
K,Y	3

For 3 frequent itemsets

3 Itemsets	Support	Percentage
E,K,M	2|5	40
E,K,O	3|5	60
E,K,Y	2|5	40
K,M,O	1|5	40
K,M,Y	2|5	40
K,O,Y	2|5	40
E,M,O	1|5	20
E,M,Y	1|5	20

Itemsets with minimum Support = 60

Items	Min Support = 60
E, K, O	3

b) Frequent Itemset : {O, K, E}

OK → E	K → OE
O → KE	KE → O
OE → K	E → OK

Let us Say k is the number of items in the frequent itemset ; Therefore the association rule becomes $2^k - 2^1 = 2^3 - 2 = 6$

The 6 possible association Rules are

Rules	Support	confidence
OK → E	3/3	100
O → KE	3/3	100
OE → K	3/4	75
K → OE	3/3	100
KE → O	3/5	60
E → OK	3/4	75

Since the minimum confidence = 80% we can find the 3 strong association rule:

$$OK \longrightarrow E$$
$$OE \longrightarrow K$$
$$O \longrightarrow KE$$

we need a E transaction, buys(x, item) ∩ buys(x, item2)

$\Rightarrow$ buys(x, item3) [S,C]

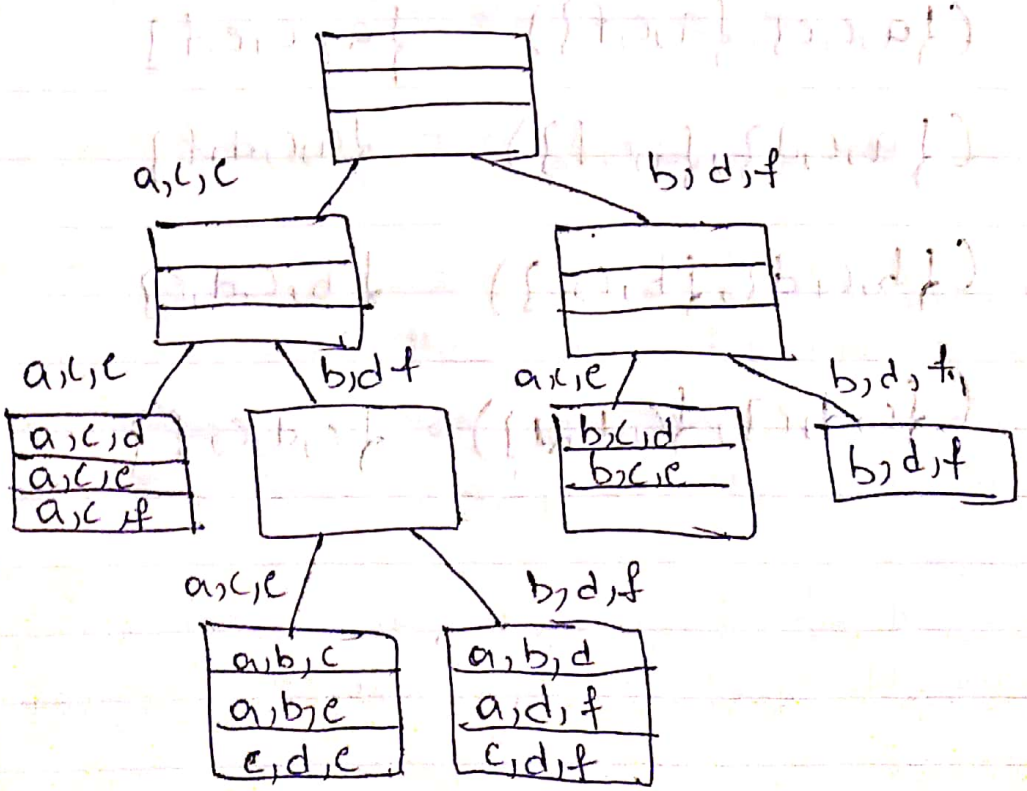
$\{O,K\} \longrightarrow \{E\}$

$\{O,E\} \longrightarrow \{K\}$

Are the final strong association Rules

Q5)

a)



b) All the candidate 4 itemsets that can be generated from the frequent 3-itemsets using the candidate generation procedure for Apriori Algorithm.

when we merge

$(\{a,b,c\},\{a,b,d\}) = \{a,b,c,d\}$

$(\{a,b,c\},\{a,b,e\}) = \{a,b,c,e\}$

$(\{a,b,d\},\{a,b,e\}) = \{a,b,d,e\}$

$(\{a,c,d\},\{a,c,e\}) = \{a,c,d,e\}$

$(\{a,c,e\},\{a,c,f\}) = \{a,c,e,f\}$

$(\{a,c,d\},\{a,c,f\}) = \{a,c,d,f\}$

$(\{b,c,d\},\{b,c,e\}) = \{b,c,d,e\}$

$(\{c,d,e\},\{c,d,f\}) = \{c,d,e,f\}$

{a, b, c, d}	{a, c, c, f}
{a, b, c, c}	{a, c, d, f}
{a, b, d, c}	{b, c, d, e}
{a, c, d, c}	{c, d, e, f}

c)

For pruning we need to check if the 3-item sets that can be obtained from 4-item sets are frequent for {a, b, c, d} we can get {a, b, c}, {a, b, d}, {a, c, d} {b, c, d} which are are frequent item set according to the question

So we can say that {a, b, c, d} is frequent.

→ For {a, b, c, c} → {a, b, c} | {a, b, c}, {b, c, c} {a, c, c} are there are frequent item sets according to the question. From this we can say that {a, b, c, c} is frequent as well.

→ For {a, b, d, c} → {a, b, d}, {a, b, c}, {a, d, c} {b, d, c} where {a, d, c} and {b, d, c} are

are infrequent. From this we can say that {a, b, d, e} is not frequent and can be ~~pruned~~. pruned.

⇒ For {a, c, d, e} → {a, c, d}, {a, c, e}, {a, d, e} {c, d, e} where {a, d, e} is infrequent from this we can say that {a, c, d, e} is not frequent and can be pruned.

⇒ For {a, c, d, f} → {a, c, d}, {a, c, f}, {a, d, f}, {c, d, f} all these frequent item set according to the question. From this we can say that {a, c, d, f} is frequent.

⇒ For {a, c, e, f} → {a, c, e}, {a, c, f}, {a, e, f}, {c, e, f} where {c, e, f} and {b, e, f} are infrequent. and From this we can say that {a, c, e, f} is not frequent and can be pruned.

⇒ For {b, c, d, e} → {b, c, d}, {b, c, e}, {b, d, e} {c, d, e} where {b, d, e} is infrequent. From this we can say that {b, c, d, e} is not frequent and can be pruned

⇒ For {a, d, e, f} ⟶ {a, d, e}, {a, d, f}, {a, e, f}, {d, e, f} since {a, e, f} and {d, e, f} are infrequent. From this we can say that {a, d, e, f} is not frequent and can be pruned.

So, after the above pruning step we can get the following 4-itemsets {a, b, c, d}, {a, b, c, e} and {a, c, d, f}

All the candidate 4-itemsets after the pruning.

{a, b, c, d}
{a, b, c, e}
{a, c, d, f}

Q7)

a) Support of {A} is equal to support of {A, B, C} based on the given.

we can say that A only occurs when {A, B, C} is present in the transaction

the as the support of A's equal

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

b) Given All the transaction that contain {A} is a subset of transactions that contain {B} transaction that contain {A, B}, {A, B, C}, {A, B, D}, {A, B, E}, {A, B, C, D}, {A, B, C, E}, {A, B, D, E}, {A, B, C, D, E}.