Exercise-2

A.

Compute the support for itemsets {e}, {b, d}, and {b, d, e} by treating each transaction ID as a market basket.

$\{e\}: s = \frac{8}{10} = 0.8$

$\{b, d\}: s = \frac{2}{10} = 0.2$

$\{b, d, e\}: s = \frac{2}{10} = 0.2$

B.

Use the results in part (a) to compute the confidence for the association rules {b,d} -> {e} and {e} -> {b,d} Is confidence a symmetric measure?

$\{b, d\} \rightarrow \{e\}: c = \frac{0.2}{0.2} = 1$

$\{e\} \rightarrow \{b, d\}: C = \frac{0.2}{0.8} = 0.25$

C.

Repeat part (a) by treating each customer ID as a market basket. Each item should be treated as a binary variable (1 if an item appears in at least one transaction bought by the customer, and 0 otherwise).

$\{e\}: S = \frac{4}{5} = 0.8$

$\{b, d\}: S = \frac{5}{5} = 1$

$\{b, d, e\}: S = \frac{4}{5} = 0.8$

D.

Use the results in part (c) to compute the confidence for the association rules {b,d} -> {e} and {e} -> {b,d}.

$\{b, d\} \rightarrow \{e\}: C = \frac{0.8}{1} = 0.8$

$\{e\} \rightarrow \{b, d\}: C = \frac{0.8}{0.8} = 1$

E.

Suppose s1 and c1 are the support and confidence values of an association rule r when treating each transaction ID as a market basket. Also, let s2 and c2 be the support and confidence values of r when treating each customer ID as a market basket. Discuss whether there are any relationships between s1 and s2 or c1 and c2.

no clear difference in treating transaction IDs or customer IDs as market baskets.

Exercise 6

A.

What is the maximum number of association rules that can be extracted from this data (including rules that have zero support)?

The total number of possible rules, R, extracted from a data set that

$R = 3^d + 2^{d+1} + 1$

There are 6 items in the table.

So, R = 602

B.

What is the maximum size of frequent itemsets that can be extracted (assuming minsup > 0)?

With minsup > 0, we only need to look for the largest itemset in the data set. Itemsets with ID 6 and 9 have the maximum size of 4 in the data set.

C.

Write an expression for the maximum number of size-3 itemsets that can be derived from this data set.

Disregarding the support threshold, there are $\frac{6!}{3!}$ possible 3 itemset (with duplicate). The number of distinct 3-itemsets is therefore:

$$= \frac{6!}{3!\,3!} = 20$$

D.

Find an itemset (of size 2 or larger) that has the largest support.

the itemset with the largest support is {bread; butter} that is 5. (Ignoring the 1-itemsets Null node.)

Exercise 8

A.

List all candidate 4-itemsets obtained by a candidate generation procedure using the Fk+1 * F1 merging strategy.

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

B.

List all candidate 4-itemsets obtained by the candidate generation procedure in Apriori.

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

C.

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

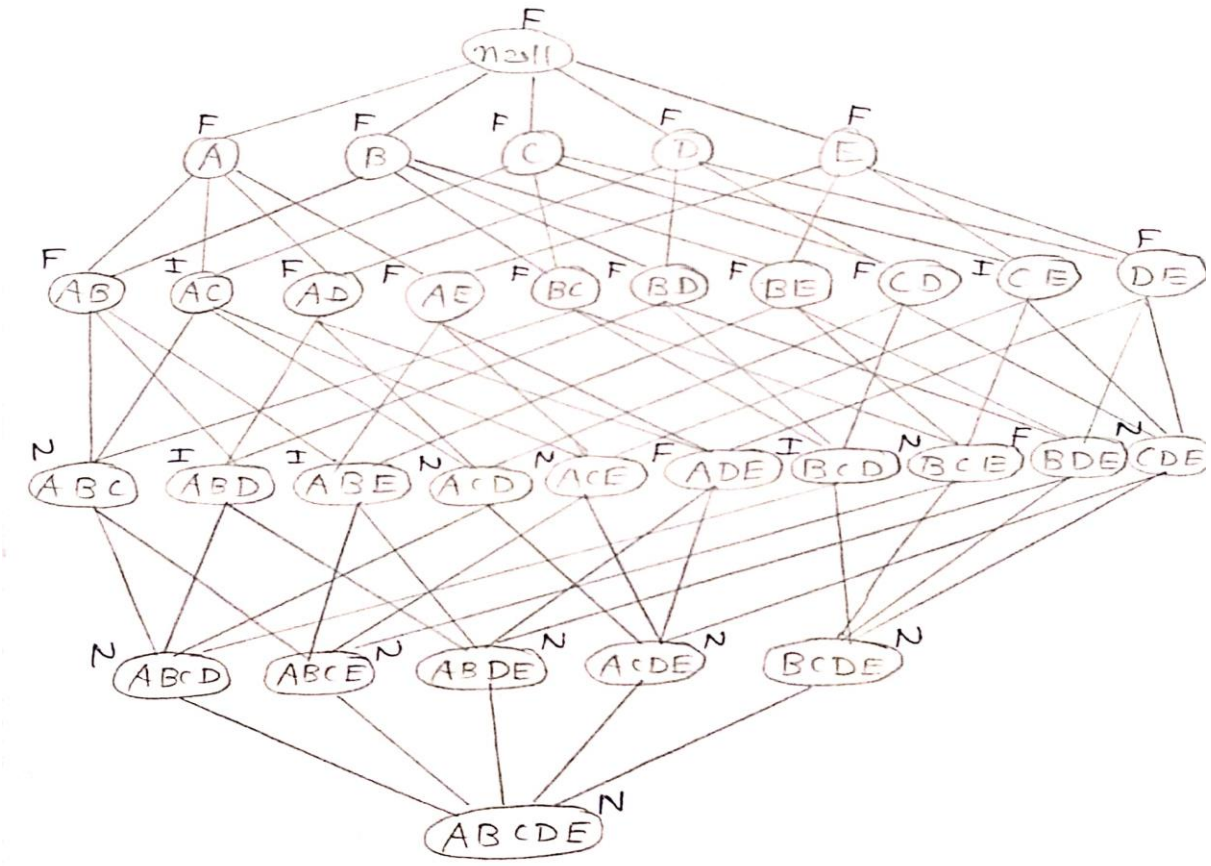
All the subset {1,2,3}, {1,2,4}, {1,3,4}, {2,3,4} of {1,2,3,4} are frequent.

All the subset {1,2,3}, {1,2,5}, {1,3,5}, {2,3,5} of {1,2,4,5} are frequent.

Other 4 items are pruned.

Exercise 9

A.



B.

Frequent Itemset = $\frac{16}{32} = 50\ \%$

C.

Total N = 11

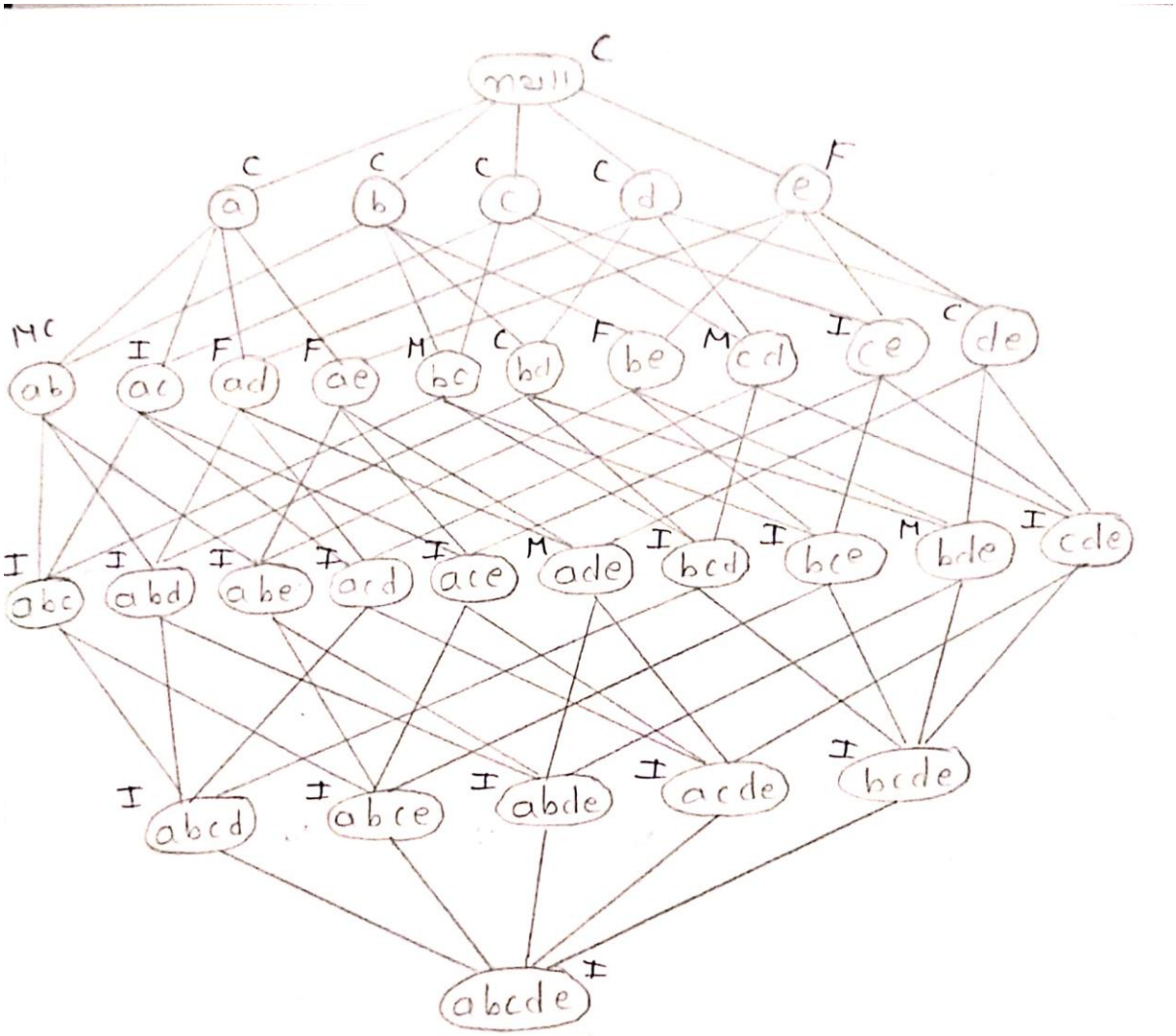
So, pruning ratio is $\frac{N}{Total\ itemset} = \frac{11}{32} = 34.4\ \%$

D.

Total I = 5

So, False alarm rate is $\frac{I}{Total\ Itemset} = \frac{5}{32} = 15.6\ \%$

Exercise 12



Exercise 13

A.

$\{b\} \rightarrow \{c\}$

	C	!c
b	3	4
!b	2	1

$\{a\} \rightarrow \{d\}$

	a	!d
a	4	1
!a	5	0

$\{b\} \rightarrow \{d\}$

	d	!d
b	6	1
!b	3	0

$\{e\} \rightarrow \{c\}$

	c	!c
e	2	4
!e	3	1

$\{c\} \rightarrow \{a\}$

	a	!a
c	2	3
!c	3	2

B.

1.

For example, association rule between two itemsets is.

$\{x\} \rightarrow \{y\}$

Total Itemset is T

$$Support = \frac{Support\ count(x\ U\ y)}{|\ T\ |}$$

Where Support count = In how many transaction x and y available

Rules	Support	Rank
$\{b\} \rightarrow \{c\}$	0.3	3
$\{a\} \rightarrow \{d\}$	0.4	2
$\{b\} \rightarrow \{d\}$	0.6	1
$\{e\} \rightarrow \{c\}$	0.2	4
$\{c\} \rightarrow \{a\}$	0.2	4

2.

For example, association rule between two itemsets is.

$\{x\} \rightarrow \{y\}$

$$Confidence\ (c) = \frac{support(x\ U\ y)}{support\ (x)}$$

Rules	Confidence	Rank
$\{b\} \rightarrow \{c\}$	0.4	3
$\{a\} \rightarrow \{d\}$	0.8	2
$\{b\} \rightarrow \{d\}$	0.85	1
$\{e\} \rightarrow \{c\}$	0.3	5
$\{c\} \rightarrow \{a\}$	0.4	4

3.

For example, association rule between two itemsets is.

$\{x\} \rightarrow \{y\}$

$$Interest\ (x \rightarrow y) = \frac{p(x,y)}{p(x)p(y)}$$

Rules	Interest	Rank
$\{b\} \rightarrow \{c\}$	0.214	3
$\{a\} \rightarrow \{d\}$	0.72	2
$\{b\} \rightarrow \{d\}$	0.771	1
$\{e\} \rightarrow \{c\}$	0.167	5
$\{c\} \rightarrow \{a\}$	0.2	4

4.

For example, association rule between two itemsets is.

$\{x\} \rightarrow \{y\}$

$$IS\ (x \rightarrow y) = \frac{p(x,y)}{sqrt(p(x)p(y))}$$

Rules	IS	Rank
$\{b\} \rightarrow \{c\}$	0.507	3
$\{a\} \rightarrow \{d\}$	0.596	2
$\{b\} \rightarrow \{d\}$	0.756	1
$\{e\} \rightarrow \{c\}$	0.365	5
$\{c\} \rightarrow \{a\}$	0.4	4

5.

$$Klosgen\ (x \rightarrow y) = sqrt(x,y) * \big(p(y|x) - p(y)\big)\ where\ \ p(y|x) = \frac{p(x,y)}{p(x)}$$

Rules	Klosgen	Rank
$\{b\} \rightarrow \{c\}$	-0.039	2
$\{a\} \rightarrow \{d\}$	-0.063	4
$\{b\} \rightarrow \{d\}$	-0.033	1
$\{e\} \rightarrow \{c\}$	-0.075	5
$\{c\} \rightarrow \{a\}$	-0.045	3

6.

$$Odds\ ratio\ (x \rightarrow y) = \frac{p(x,y)p(!\,x\,,!\,y)}{p(x,!\,y)p(!\,x,y)}$$

Rules	Odd Ratio	Rank
$\{b\} \rightarrow \{c\}$	0.375	2
$\{a\} \rightarrow \{d\}$	0	4
$\{b\} \rightarrow \{d\}$	0	4
$\{e\} \rightarrow \{c\}$	0.167	3
$\{c\} \rightarrow \{a\}$	0.444	1

Exercise 20

	b	! b
a	9	1
! a	1	89

	b	! b
a	89	1
! a	1	9

A.

- Support(A) = 10/100 = 0.1
- Support (B) = 10/100 = 0.1
- Support(A,B) = 9/100 = 0.09
- Correlation coefficient for the association pattern {A, B}

  $A \rightarrow B = \frac{9*89-1}{\sqrt{89*9}} = 28.266$

  $B \rightarrow A$ = 28.266
- $Confidence(A \rightarrow B)$ = 9/10 = 0.9
- $Confidence\ (B \rightarrow B)$ = 9/10 = 0.9

B.

- Support(A) = 90/100 = 0.9
- Support (B) = 90/100 = 0.9
- Support(A,B) = 89/100 = 0.89
- Correlation coefficient for the association pattern {A, B}

  $A \rightarrow B = \frac{9*89-1}{\sqrt{89*9}} = 28.266$

  $B \rightarrow A$ = 28.266
- $Confidence(A \rightarrow B)$ = 89/90 = 0.98
- $Confidence\ (B \rightarrow B)$ = 89/10 = 0.98

C.

Interest, support and confidence are not consider as invariant whereas the ϕ-coefficient is invariant. This is due to the fact that ϕ-coefficient takes absence and presence of an item in a transaction into consideration.

Reference:

https://www.solver.com/xlminer/help/association-rules#:~:text=Confidence%20is%20the%20ratio%20of,all%20items%20in%20the%20antecedent.&text=Lift%20is%20nothing%20but%20the%20ratio%20of%20Confidence%20to%20Expected%20Confidence.