Name : Pavani Rangineni

CWID: A20516359

DATA MINING HW5

1.1. 15.Answer the following questions using the data sets shown in Figure 5.34.Note that each data set contains 1000 items and 10,000 transactions.Dark cells indicate the presence of items and white cells indicate the absence of items. We will apply the Apriori algorithm to extract frequent itemsets with minsup = 10% (i.e., itemsets must be contained in at least 1000 transactions).

(a) Which data set(s) will produce the most number of frequent itemsets?

Data set e will produce the most number of frequent itemsets.

(b) Which data set(s) will produce the fewest number of frequent itemsets?

Data set d will produce the fewest number of frequent itemsets.

 (c) Which data set(s) will produce the longest frequent itemset?

Data set e will produce the longest frequent itemsets.

(d) Which data set(s) will produce frequent itemsets with highest maximum support?

Data set b will produce the frequent itemsets with highest maximum support.

(e) Which data set(s) will produce frequent itemsets containing items with wide-varying support levels (i.e., items with mixed support ranging from less than 20% to more than 70%)?

Data set e will produce the frequent itemsets containing items with wide-varying support levels.

1.2 1. Given the database in Table 8.2.

(a) Using minsup = 3/8, show how the Apriori algorithm enumerates all frequent patterns from this dataset.

	A	B	C	D	E	F	G
T1	1	1	1	1	0	0	0
T2	1	0	1	1	0	1	0
T3	1	0	1	1	1	0	1
T4	1	1	0	1	0	1	0
T5	0	1	1	0	0	0	1
T6	0	0	0	1	0	1	1
T7	1	1	0	0	0	0	1
T8	0	0	1	1	0	1	1

C1

Itemset	Sup. Count
A	5
B	4
C	5
D	6
E	1
F	4
G	5

L1

Itemset	Sup. Count
A	5
B	4
C	5
D	6
F	4
G	5

C2

Itemset	Sup.count
AB	3
AC	3
AD	4
AF	2
AG	2
BC	1
BD	2
BF	1
BG	2
CD	4
CF	2
CG	3
DF	4
DG	3
FG	2

L2

Itemset	Sup. Count
AB	3
AC	3
AD	4
CD	4
CG	3
DF	4
DG	3

C3

Itemset	Sup. Count
ABC	1
ABD	2
ACD	3
CDG	2
DFG	2

L3

Itemset	Sup. Count
ACD	3

Frequent Itemset Final Tree:
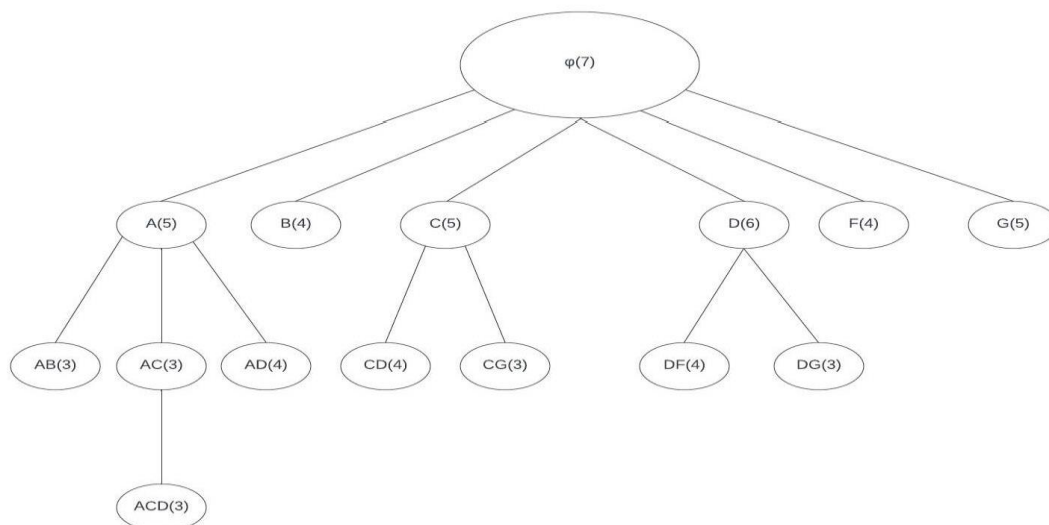
Frequent Itemsets with minsup = 3

Sup	itemsets
6	D
5	A, C, G
4	B, F, AD, CD, DF
3	AB, AC, CG, DG, ACD

1.2 4. Given the database in Table 8.4. Show all rules that one can generate from the set ABE.

	A	B	C	D	E
T1	1	0	1	1	0
T2	0	1	1	0	1
T3	1	1	1	0	1
T4	0	1	0	1	1
T5	1	1	1	0	1
T6	1	1	1	1	0

Itemset	Support count
ABE	2
AB	3
AE	2
BE	4
A	4
B	5
E	4

So here, by taking a frequent itemset ABE, the rules can be

Confidence(X->Y) = _____

$\{AB\} \rightarrow \frac{Sup(ABE)}{Sup(AB)} = \frac{2}{3}$ E == 0.66 = 66%

$\{AE\} \rightarrow \frac{Sup(ABE)}{Sup(AE)} = \frac{2}{2}$ B == 1 = 100%

$\{BE\} \rightarrow \frac{Sup(ABE)}{Sup(BE)} = \frac{1}{2}$ A == 0.5 = 50%

$\{A\} \rightarrow \frac{Sup(ABE)}{Sup(A)} = \frac{1}{2}$ BE == 0.5 = 50%

$\{B\} \rightarrow \frac{Sup(ABE)}{Sup(B)} = \frac{2}{5}$ AE == 0.4 = 40%

$\{E\} \rightarrow \frac{Sup(ABE)}{Sup\ E} = \frac{2}{3}$ AB == 0.5 = 50%