Exercise I (Association Rules)

Consider the mining of association rules on the transactions:

transaction id	items
1	A, B, E
2	A, B, D, E
3	B, C, D, E
4	B, D, E
5	A, B, D
6	B, E
7	A, E

A. What is the support of the itemset {B,D,E}?

B. What is the support and confidence of the association rule BD → E?

C. Consider the application of the Apriori algorithm to find all the frequent itemsets whose counts are at least 3. Recall that the algorithm scans the transaction list a number of times, where the i^{th} scan generates the set F_i of all size-i frequent itemsets from a candidate set C_i. Show C_i and F_i for each possible i.

D. Find all the association rules with support at least 3 and confidence at least 3/4. For your convenience, all the itemsets with support at least 3 are {{A}, {B}, {D}, {E}, {A,B},{A,E},{B,D},{B,E},{D,E}, {B,D,E}}.

Exercise II (Association Rules)

The following is an example of customer purchase transaction data set.

CID	TID	Date	Items Purchased
1	1	01/01/2001	10,20
1	2	01/02/2001	10,30,50,70
1	3	01/03/2001	10,20,30,40
2	4	01/03/2001	20,30
2	5	01/04/2001	20,40,70
3	6	01/04/2001	10,30,60,70
3	7	01/05/2001	10,50,70
4	8	01/05/2001	10,20,30
4	9	01/06/2001	20,40,60
5	10	01/11/2001	10,20,30,60

Note: CID = Customer ID and TID = Transactions ID

Q.1 Calculate the *support*, *confidence* and *lift* of the following association rule. Indicate if the items in the association rule are independent of each other or have negative or positive impacts on each other.

{10} -> {50,70}

Q.2 The following is the list of large two item sets. Show the steps to apply the Apriori property to generate and prune the candidates for large three itemsets. Describe how the Apriori property is used is in the steps. Give the final list of candidates large three item sets.

{10,20} {10,30} {20,30} {20,40}