Indian Institute of Information Technology, Allahabad

Course Title: Data Mining

Assignment -2 (January-2022)

Instructor: Prof. O.P.Vyas, Dr Manish Kumar and Dr. Muneendra Ojha

---

This exercise sheet will be covered in the google online meetings. There are three different types of exercises:

Classroom Demonstration: The assigned TA in your classroom will discuss the one or two questions with their solution so that the next few questions from the Group/**individual online** can be attempted smoothly.

Group/individual online: This exercise will be solved during the online meeting. We will give you some time where you can discuss and solve the exercise in a small group/individually. Afterwards, we will discuss possible solutions.

Alone home :This exercise should be solved by you alone after the online meeting. The exercise and possible solutions will be discussed in the next online meeting/tutorial session. The correct approach / solution will be awarded with bonus points.

Note: Every student is supposed to submit the solution of the question given in **Group/individual online** and **Alone home** before the **next due tutorial date.**

Classroom Demonstration:

Ques.(1) Calculate Support of {apple, beer, rice} and confidence of {apple -> beer}.

Transaction 1	
Transaction 2	
Transaction 3	
Transaction 4	
Transaction 5	
Transaction 6	
Transaction 7	
Transaction 8	

Que.(2) Given below association rules are mined from a transactional database of an electronic company. Discuss, what the following association rules represent?

(a) buys(X, “computer”) ⇒ buys(X, “software”)
[support = 1%,confidence = 50%]

(b) age(X, “20..29”) ∧ income(X, “40K..49K”) ⇒ buys(X, “laptop”)
[support = 2%, confidence = 60%].

where X is a variable representing a customer

Group/individual online:

Ques:3 (a) Calculation of support in Apriori Algorithm is very expensive. Explain?

(b) Why Association Rule Mining is important in data mining? Explain the process of Association Rule Mining . Where do the mining results are used?

(c) Consider the market basket transactions given in the following table.
Let min_support= 40% and min_confidence= 40%.
Find all the frequent itemsets using Apriori Algorithm.

Transaction ID	Items Bought
T1	A, B, C
T2	A,B,C,D,E
T3	A,C,D
T4	A,C,D,E,
T5	A,B,C,D

Alone home

Ques (4) You are given the transaction data shown in the Table below from the electronic shop. There are 10 distinct transactions and each transaction involves between2 to 4 electronic items. There are a total of 5 electronic items that are involved in the transactions. For all of the parts below the minimum support is 30% and the minimum confidence is 70%. Show your work for full credit.

a. Apply the *Apriori Algorithm* to the dataset of transactions and identify all frequent k-item sets.

b. Find all strong association rules of the form: X ∧ Y-->Z and note their confidence values.

Transaction ID	Items Bought
1	{Laptop, Printer, Tablet, Headset}
2	{Printer, Monitor, Tablet}
3	{Laptop, Printer, Tablet, Headset}
4	{Laptop, Monitor, Tablet, Headset}
5	{Printer, Monitor, Tablet, Headset}
6	{Printer, Tablet, Headset}
7	{Monitor, Tablet}
8	{Laptop, Printer, Monitor}
9	{Laptop, Tablet, Headset}
10	{Printer, Tablet}

Ques (5) Apply the Apriori Algorithm to the dataset of transactions and identify all frequent k-item-sets with minimum support is 30% and the minimum confidence is 70%.

S.No	Item ID	Item Set
1	Tid-1	F,A,C,D,G,M,P
2	Tid-2	A,B,C,F,L,M,O
3	Tid-3	B,F,N,O
4	Tid-4	B,K,C,P
5	Tid-5	A,F,C,L,P,M,N