Indian Institute of Information Technology, Allahabad

Course Title: Data Mining and Warehousing

Assignment -1 (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 **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.

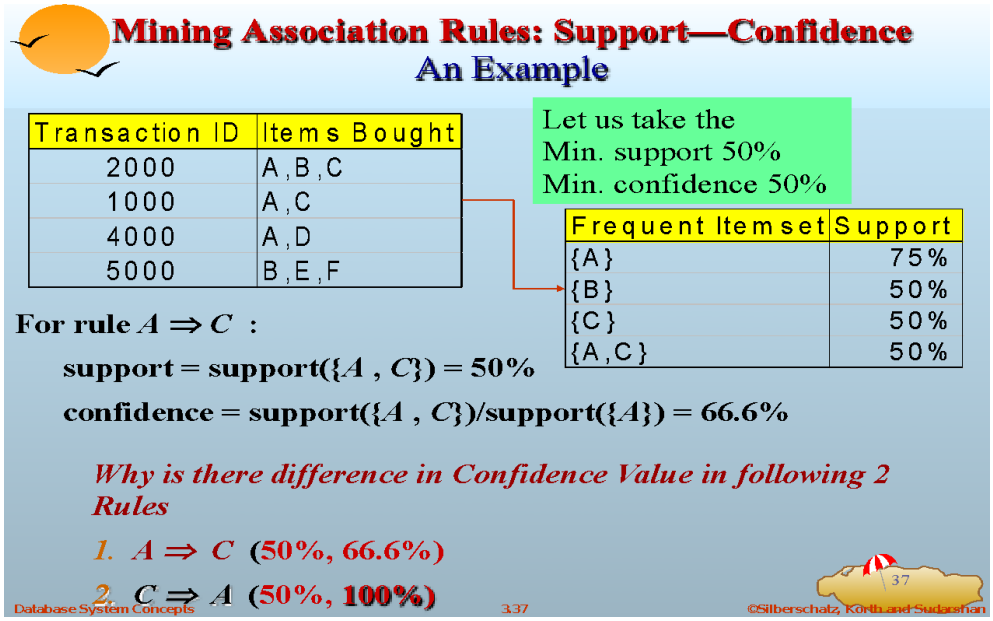
Note: Every student is supposed to submit the solution of the question given in **Group/individual online** and **Alone home** before the **next due tutorial date.**

Classroom Demonstration:

Ques :1 (a) From **DBMS** to **Data Warehouse** to **Data Mining,** identify which one is Retrospective or Prospective Information Delivery and capable of Proactive Information delivery, give examples of Business application in each?

Que.(2) (a) Why do we really need measures like ***Support*** and ***Confidence*** for mining strong associations among the items a customer has purchased in the past. What is the difference between them, what do they exactly represent in the Data.

(b) Will the Association Rule ***Bread*** ⇒ ***Butter*** *have* Support and Confidence Value same as for Association Rule ***Butter*** ⇒ ***Bread.*** Justify your answer with an example. (Elaboration with the Association Rules as shown below)



Group/individual online:

Ques:3 (a) As OLAP can do Analysis as well then why do we need Data Mining, is Data Mining simply extension of OLAP or any significant change, elaborate with example?

(b) When OLAP & Descriptive Analytics both use past data and give retrospective information delivery then what is the significant difference between them?

(c) Which among the following Business Questions a Descriptive Analytics can handle, how it will be done and explain what Business data will be required for it, elaborate with example;

(i) I wish to prepare my future stocks of items so that they are neither too much nor it's too less for the customer demands.
(ii) I wish to devise a discounting package for my Supermarket items.
(iii) With limited Shelf-space in my new Supermarket, I want to place my items for the customers in such a way that it prompts them to buy other items as well.
(iv) Effect of pain killer medicine on my digestion problems

Alone home

Ques :4 In the given dataset calculate support and confidence for the following association rules: {{T-shirt → Trousers}, {Trousers → Belt}, {T-shirt → Belt},{T-shirt, Trousers → Belt}}

Transaction	**Items Bought**
t1	T-shirt, Trousers, Belt
t2	T-shirt, Jacket
t3	Jacket, Gloves
t4	T-shirt, Trousers, Jacket
t5	T-shirt, Trousers, Sneakers, Jacket, Belt
t6	Trousers, Sneakers, Belt
t7	Trousers, Belt, Sneakers