DMBI Assignment 1

Chapter 5 - Frequent Pattern Mining
Chapter 6 - Business Intelligence

Name: Atharva Lotankar
Class: D15C; RNum: 27
Batch - B; DMBI

Q1. What are the multiple level and multi dimensional rules. Explain with suitable examples for each.

→ Association rule mining is a key technique in data mining used to uncover interesting relations between variables in large datasets.

(I) Multiple Level Association Rules

→ Multiple level association rules involving mining association patterns across multiple levels of abstraction or hierarchies in the data.

Many real-life datasets have attributes arranged in a hierarchy.

Example → A product can be classified at different granularity levels:

- Level 1 → Electronics (High Lvl)
- Level 2 → Computers (More specific)
- Level 3 → Laptops (Even more specific)

A multiple level association rule might find patterns at any of these levels — for instance:

> Customers who buy Electronics, also buy Home_Appliances, or at a more detailed level, customers who buy Laptops also buy Laptop Bags.

Rule @L1: "If customer buys Electronic, then they buy accessories"

Rule @ L3: "If customer buys a Gaming Laptop, then they buy RGB Mouse"

(II) Multidimensional Association Rules

Multidimensional assoc^n rules extend traditional single-dimensional rules by involving multiple attributes or dimensions of the data, such as customer demographics, product details, or transaction time.

These rules capture correlations across multiple dimensions, not just item co-occurences.

Characteristics:- Dimensions can be nominal or quantitative.
- These attributes are often discretized into intervals for mining.
- Rules consider combinations like "AgeGrp" + "Occupation" + "Purchases".

Example → Suppose a retail dataset has attributes: Age, Occupation & Product purchased.

A multidimensional rule could be:
"Customers aged 20-30 AND occupation = student → likely to buy HP Inkject Printer"

This rule involves multiple dimensions (Age, Occupation, Product) to reveal more contextual associations than a single-dimensional rule.

Q2. Explain Market Basket Analysis with example.

→ Market Basket analysis is a modeling technique which is also called 'affinity analysis', it helps identifying which items are likely to be purchased together.

It is used in deciding the location of items inside a store, for e.g. if a customer buys a packet of bread he is more likely to buy a packet of butter too, keeping bread-butter next to each other in store would result in customers getting tempted to buy one item with the other.

All market-based problem assumes we have some large number of items, e.g. of "bread", "milk", etc. Customer buy subset of items as per his need & marketer gets the info that which things customer has taken together. So the marketers use this info to put items on diff. position.

* Examples:

→ Credit card transactions done by customer may be analysed.

→ Phone calling patterns may be analysed.

→ Fraudulent Medical Insurance claims can be identified.

→ Special combo offers may be offered to the customers on the products sold together.

Q3. Explain Business Intelligence Issues.
&

Issues in BI.

(i) Organizations & People

- Management within an organization do not agree that decision taken based on data or evidence work for them, they prefer to run the operation from instinct.
- In order to assess business progress there is no overall business strategy laid out with objectives & measures for those objectives.
- The data needed for BI system can't be obtained from IT personnel as they are overloaded & they have no resources available.
- For performance improvement of business either making use of BI or not, there is no incentives provided to staff within the organisation.
- There is no obvious time to establish a BI system. The business is in a state of high change or flux.

(2) Data and Technology

- The Data of the organization is not clean. The time & effort needed to correct or handle this type of data leads to an unsuccessful delivery of BI project.

  For e.g. there could be diff defn^s for same item → customer may be coded differently on sales system to that held in accounts system.

- The technology chosen for BI turns out to be so particular that is ultimately resulted in time consuming process which leads to delay in project completion.

- The BI Technology discourages the use of system due to following:

  - The info. presentation quality is poor or limited.
  - Response time for data presentation is too slow or not acceptable.
  - It's too difficult or limited to ask new questions of BI technology for either of them, end users or BI expert.

Q4. Consider transaction db given in table below. Apply Apriori Algorithm with minimum support of 50% & confidence of 50%. Find all frequent item set & all the association rules.

Tid	Items
100	1, 3, 4
200	2, 3, 5
300	1, 2, 3, 5
400	2, 5
500	1, 2, 3
600	3, 5
700	1, 2, 3, 5
800	1, 5
900	1, 3

→ Total Frequent Itemsets : {1}, {2}, {3}, {4}, {5}

→ Association rules:
From individual dataset (itemset):

Item	SupportCount	Support%
{1}	6	6/9 x 100 = 66.67%
{2}	5	5/9 x 100 = 55.56%
{3}	7	7/9 x 100 = 77.78%
{4}	1	1/9 x 100 = 11.11%
{5}	6	6/9 x 100 = 66.67%

Frequent 1-itemsets = {1}, {2}, {3}, {5}

Similarly for 2-itemsets

Items	Support Count	Support %
{1, 2}	3	3/9 x 100 = 33.33 %
{1, 3}	5	5/9 x 100 = 55.55 %
{1, 5}	3	3/9 x 100 = 33.33 %
{2, 3}	4	4/9 x 100 = 44.44 %
{2, 5}	4	4/9 x 100 = 44.44 %
{3, 5}	4	4/9 x 100 = 44.44 %

Frequent 2-itemsets : {1, 3}

∴ since only one two itemset exists, we cannot generate any valid 3-itemsets using Apriori property.

So;

From frequent itemset = {1, 3}

Rule 1: {1} → {3}

$$\text{Confidence} = \frac{\text{Sup}(\{1,3\})}{\text{Sup}(\{1\})} = \frac{55.55}{66.67} \times 100$$

$$= 83.32\%$$

Rule 2: {3} → {1}

$$\text{Confidence} = \frac{\text{Sup}(\{1,3\})}{\text{Sup}(\{3\})} = \frac{55.55 \times 100}{77.78}$$

$$= 71.42\%$$

∴ Both rules satisfy the minimum conf of 50%