

Roll No. 

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Total No. of Pages : 02

Total No. of Questions : 09

**MCA (Sem.-3)**  
**DATA MINING AND BUSINESS INTELLIGENCE**

Subject Code : PGCA-1972

M.Code : 90803

Date of Examination : 05-01-23

Time : 3 Hrs.

Max. Marks : 70

**INSTRUCTIONS TO CANDIDATES :**

1. SECTION-A is COMPULSORY consisting of TEN questions carrying TWO marks each.
2. SECTION - B & C have FOUR questions each.
3. Attempt any FIVE questions from SECTION B & C carrying TEN marks each.
4. Select atleast TWO questions from SECTION - B & C.

**SECTION-A**

**I. Write short notes on :**

- a) What is the difference between data, information and knowledge?
- b) Draw a snowflake schema by taking a suitable example.
- c) Briefly describe the concept of data cube aggregation.
- d) Mention the broader steps of KDD process.
- e) Identify the relationship between Business Intelligence and Data Warehousing
- f) What is hierarchical clustering? In what form its final output is retrieved?
- g) Draw a clear pipeline of the data analytics life cycle.
- h) What is logistic regression? How can we use it for prediction?
- i) How are rules generated using the Apriori algorithm?
- j) What is a decision tree used for? List down its advantages.

## SECTION-B

2. Describe the essential components of data warehouse and data marts. Also, discuss the concept of metadata and its need for data mining.
3. Explain the star and fact constellation scheme by taking an appropriate database example.
4. Describe the process of data cleaning, data integration and data transformation by elaborating their techniques.
5. Why is dimensionality reduction an essential step in data preprocessing? Explain any one method used to reduce the dimensions with the help of an example.

## SECTION-C

6. A database has four transactions given below:

TID	Bute	Items bought
T100	10/15/04	{K, A, D, B}
T200	10/15/04	{D, A, C, E, B}
T300	10/19/04	{C, A, B, E}
T400	10/22/04	{B, A, D}

Assuming a minimum level of support  $\text{min\_sup}=60\%$  and a minimum level of confidence  $\text{min\_conf}=80\%$ , find all frequent itemsets using the Apriori algorithm. For each iteration show the candidate and acceptable frequent itemsets.

7. What is association rule mining? Elaborate data generalization and summarization-based characterization.
8. Explain in detail how can data mining be helpful in fraud detection? Enlist the whole process.
9. **Write short notes on the following:**
  - a) Bayesian Classification
  - b) Hadoop Architecture.

**NOTE : Disclosure of Identity by writing Mobile No. or Marking of passing request on any paper of Answer Sheet will lead to UMC against the Student.**