Roll No. Total No. of Pages : 02

Total No. of Questions : 09

MCA (Sem.–3)

DATA MINING AND BUSINESS INTELLIGENCE

Subject Code : PGCA1972

M.Code : 90803

Date of Examination : 18-12-2023

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

1. **Answer briefly :**

a) What is Data Warehouse? Discuss its features.

b) Define the term Metadata and why it is important?

c) Three persons A, B and C have applied for a job in a private company. The chance of their selection is in the ratio 1 : 2 : 4. The probabilities that A, B and C can introduce changes to improve the profits of the company are 0.8, 0.5 and 0.3, respectively. If the change does not take place, find the probability that it is due to the appointment of C.

d) Define Data Mart and its types.

e) Differentiate between drill-down, roll-up and slice and dice operations.

f) Give two major difference between stars, snow and fact constellations.

g) Describe Data Cube aggregation.

h) Define the term tree pruning.

i) Differentiate between Data Reduction and Dimensionality Reduction.

j) Define spatial mining and web mining.

SECTION-B

2. Explain in detail the KDD process.

3. Why is data pre-processing required? Explain the different steps involved in data pre-processing.

4. Explain OLAP and its types. Discuss various OLAP operations.

5. Explain role of Business Intelligence in Fraud Detection and Market Segmentation, domain. Explain how data mining can be helpful in any of these cases?

SECTION-C

6. Consider a dataset as shown in table below. Let min_support=40% and min confidence=70%

1	{E,F}
2	{A,B}
3	{A,B,C}
4	{C,F,A,B}
5	{C,D,F}

   a) Find all the frequent itemsets using Apriori Algorithm.

   b) List all possible association rules (with support s and confidence c) and identify the all strong association rules.

7. Explain in detail the Data Analytics Life Cycle Phases.

8. Differentiate between classification and prediction. What are the issues associated with classification and prediction? Explain the decision tree method in detail along with advantages and disadvantages.

9. Define Naive Bayesian classification. Using Naive Bayesian classification method, predict class label (Flu) of X = (Chills = Y, Runny_nose = N, Headache = Mild, Fever= Mild, Flue = ?) using following training dataset.

Chills	Runny nose	Headache	Fever	Flu
Y	N	MILD	Y	N
Y	Y	NO	N	Y
Y	N	STRONG	Y	Y
N	Y	MILD	Y	Y
N	N	NO	N	N
N	Y	STRONG	Y	Y
N	Y	STRONG	N	N
Y	Y	MILD	Y	Y

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.