

Roll No.

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

Total No. of Pages : 02

Total No. of Questions : 07

B.Sc (IT) (Sem-5)
DATA WAREHOUSE AND MINING
Subject Code : UGCA-1931
M.Code : 90392
Date of Examination: 26-05-2023

Time : 3 Hrs.

Max. Marks : 60

INSTRUCTIONS TO CANDIDATES :

1. SECTION-A is COMPULSORY consisting of TEN questions carrying TWO marks each.
2. SECTION-B contains SIX questions carrying TEN marks each and students have to attempt any FOUR questions.

SECTION-A

1. **Write briefly :**

- a) Define Data Warehouse.
- b) What is data visualization?
- c) What is data summarization?
- d) What is Gini Index?
- e) Differentiate classification and clustering.
- f) What are attribute selection measures?
- g) What is apriori algorithm?
- h) What is classifier's accuracy?
- i) What is data pre-processing?
- j) Let us suppose that there are 200 pages available on Internet for machine learning. The search on this term returns total 210 pages, out of which 190 belongs to Machine Learning, calculate precision and recall for our algorithm.

SECTION-B

2. What is the need for strategic information? Explain the difference between operational and informational data stores.
3. **Write short notes on :**
 - a) Schemas for multidimensional data
 - b) Data mining tasks.
4. Consider the training examples shown in below table for classification problem.

Gender	Car Type	Shirt Size	Class
M	Family	Small	C0
M	Sports	Medium	C0
M	Sports	Medium	C0
M	Sports	Large	C0
M	Sports	Extra Large	C0
M	Sports	Extra Large	C0
F	Sports	Small	C0
F	Sports	Small	C0
F	Sports	Medium	C0
F	Luxury	Large	C0
M	Family	Large	C1
M	Family	Extra Large	C1
M	Family	Medium	C1
M	Luxury	Extra Large	C1
F	Luxury	Small	C1
F	Luxury	Small	C1
F	Luxury	Medium	C1
F	Luxury	Medium	C1
F	Luxury	Medium	C1
F	Luxury	Large	C1

Apply the split algorithm based on information theory to perform classification of the given data set. Indicate all intermediate steps properly.

5. Explain OLAP server architecture. How OLAP data is indexed?
6. Explain k-means clustering method by taking a suitable example.
7. How do we evaluate classifier's accuracy? Explain different techniques for accuracy estimation.

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.