Roll No. ☐☐☐☐☐☐☐☐☐☐☐☐☐                     Total No. of Pages : 02

Total No. of Questions : 09

B.Tech. (CSE) (Sem–7,8)
# DATA MINING AND DATA WAREHOUSING
Subject Code : BTCS702-18
M.Code : 90488
Date of Examination : 21-06-2023

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 FIVE questions carrying FIVE marks each and students have to attempt any FOUR questions.
3. SECTION-C contains THREE questions carrying TEN marks each and students have to attempt any TWO questions.

## SECTION-A

1. Write briefly :

   a) Define Data Warehouse.

   b) Name some open source data mining tools.

   c) Write some data cube operations.

   d) Differentiate between Classification and clustering.

   e) Define Information Gain.

   f) 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.

   g) What is density based methods in clustering?

   h) What is enterprise search?

   i) What are different challenges in clustering the data?

   j) What is web usage mining?

## SECTION-B

2. What is Data Mining and explain the knowledge discovery process.

3. What is classification? Explain decision tree induction by taking a suitable example.

4. What is web content mining? Write its different steps.

5. The distance between some Indian cities are given below, Apply the clustering algorithm to make three clusters. Indicate intermediate steps.

|  | Bathinda | Patiala | Delhi | Amritsar | Mathura |
|---|---|---|---|---|---|
| Bathinda | 0 | 190 | 400 | 250 | 460 |
| Patiala | 190 | 0 | 240 | 225 | 300 |
| Delhi | 400 | 240 | 0 | 450 | 60 |
| Amritsar | 250 | 225 | 450 | 0 | 510 |
| Mathura | 460 | 300 | 60 | 510 | 0 |

6. Explain Naive Bayes Classification.

## SECTION-C

7. Explain the association rule mining problem. Consider an example with the following set of transactions. There are 10 items. https://www.ptustudy.com

| TID | Items bought |
|---|---|
| 001 | B, M, T, Y |
| 002 | B, M |
| 003 | A, T, S, P |
| 004 | A, B, C, D |
| 005 | A, B |
| 006 | T, Y, E, M |
| 007 | A, B, M |
| 008 | B, C, D, T, P |
| 009 | D, T, S |
| 010 | A, B, M |

Assume that we wish to find association rules with at least 30% support and 60% confidence. Find the frequent itemsets and then the association rules.

8. Explain the architecture of search engine in detail. Draw suitable diagrams.

9. What is OLAP? Explain its architecture, characteristics and multi dimensional view.

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