



MARRI LAXMAN REDDY INSTITUTE OF TECHNOLOGY AND MANAGEMENT

(AN AUTONOMOUS INSTITUTION)

(Approved by AICTE, New Delhi & Affiliated to JNTUH, Hyderabad)

Accredited by NBA and NAAC with 'A' Grade & Recognized Under Section 2(f) & 12(B) of the UGC act, 1956

II B.Tech II Sem Regular End Examination, July 2022

Data Warehousing and Data Mining

(CSD)

Time: 3 Hours.

Max. Marks: 70

Note: 1. Question paper consists: Part-A and Part-B.

2. In Part - A, answer all questions which carries 20 marks.

3. In Part - B, answer any one question from each unit.

Each question carries 10 marks and may have a, b as sub questions.

PART- A

(10*2 Marks = 20 Marks)

- | | | | | |
|-------|----------------------------------------------------------|----|-----|----|
| 1. a) | What is Data Warehousing? | 2M | C01 | L1 |
| b) | What are apex and base Cuboid? | 2M | C01 | L1 |
| c) | What is Data Mining? | 2M | C02 | L1 |
| d) | What are the measures used for central tendency of Data? | 2M | C03 | L2 |
| e) | What is Information gain? | 2M | C03 | L1 |
| f) | Classification Vs Prediction. | 2M | C04 | L2 |
| g) | Give one example for multi dimension association rule. | 2M | C04 | L3 |
| h) | What is Apriori Property? | 2M | C04 | L1 |
| i) | List various clustering approaches. | 2M | C04 | L1 |
| j) | Is outlier noise or knowledge? Explain. | 2M | C05 | L3 |

PART- B

(10*5 Marks = 50 Marks)

- | | | | | |
|------|-----------------------------------------------|----|-----|----|
| 2 a) | Compare OLTP Vs OLAP. | 5M | C01 | L1 |
| b) | Discuss the steps in building Data warehouse. | 5M | C01 | L1 |

OR

- | | | | | |
|---|--------------------------------------------------------------|-----|-----|----|
| 3 | Discuss the various operations of OLAP with simple examples. | 10M | C01 | L2 |
|---|--------------------------------------------------------------|-----|-----|----|

- | | | | | |
|------|---------------------------------------------------------------------------------------|----|-----|----|
| 4 a) | Why Data preprocessing is needed? What is the role of data cleaning in preprocessing? | 5M | C02 | L2 |
| b) | How Data Discretization is useful in data preprocessing? | 5M | C02 | L2 |

OR

- | | | | | |
|----|--------------------------------------------------------------------------------------------|-----|-----|----|
| 5 | Use the two methods below to normalize the following group of data 200, 300, 400, 600 1000 | 10M | C02 | L3 |
| a) | Min max normalization by setting min=0 and max=1 | | | |
| b) | Z-score normalization. | | | |

- 6 a) Explain the major steps in decision tree construction for classification. 5M C03 L2
 b) Why KNN is called lazy learner? Explain. 5M C03 L3

OR

- 7 Discuss the various criteria used for evaluating classifier models. 10M C03 L2
- 8 a) Discuss "Mining Quantitative Association rules". 5M C04 L2
 b) Strong Association rules are not necessarily Interesting. Explain. 5M C04 L3

OR

A database has five transactions. Let min_sup=60% and min_conf=80%. Find all frequent items using Apriori.

- 9
- | TID | items |
|-----|-------------|
| 1 | M,O,N,K,E,Y |
| 2 | D,O,N,K,E,Y |
| 3 | M,A,K,E |
| 4 | M,U,C,K,Y |
| 5 | C,O,O,K,I,E |
- 10M C04 L3

- 10 a) How to compute the dissimilarity between objects described by the following types of variables 5M C03 L1
 a) Numerical
 b) categorical
- b) Given two objects by the tuples(22,1,42,10) and (20,0,36,8) compute the Euclidean distance and Manhattan distance 5M C03 L2

OR

- 11 Describe each of the following clustering algorithms in terms of the following criteria: 10M C05 L3
 (i) Shapes of clusters that can be determined.
 (ii) Input parameters that must be specified.

---oo0oo---