

**TEEGALA KRISHNA REDDY ENGINEERING COLLEGE**

(UGC – AUTONOMOUS)

**M TECH I Semester Examinations, July 2021****Data Mining****Time: 3 Hours****Max. Marks: 75****Note:** This question paper contains two parts A and B.

Part A is compulsory which carries 25 marks. Answer all questions in Part A.

Part B consists of 5 Units.

Answer any one full question from each unit. Each question carries 10 marks and may have a, b, c as sub questions.

**---  
PART - A****5 × 5 Marks = 25**

1. a) What are the different types of data mining functionalities? Explain.  
 b) Explain various OLAP operations on a multidimensional database.  
 c) State the relationship between frequent patterns and association rules.  
 d) Discuss the issues for prediction process.  
 e) Derive the equations for Manhattan distance, Minkowski distance and Euclidean distance with an example.

**PART-B**

2. (i) What is KDD? Explain the different stages of KDD process with a diagram. [10]

**OR**

- (ii) What is data preprocessing? Explain how missing data can be handled in detail with an example.

3. (i) Describe typical architecture of a data warehouse with a block diagram? [10]

**OR**

- (ii) Write detailed note on how OLAP technology helps in discovery driven exploration of data cubes.

4. (i) What are the drawbacks of Apriori algorithm? Suggest mechanisms to improve its efficiency in detail [10]

**OR**

- (ii) Explain in brief about Apriori: candidate generation and test approach with an example.

5. (i) Explain classification algorithm using decision tree induction. [10]

**OR**

- (ii) Explain Bayesian classification with suitable examples.

6. (i) Suppose that the data mining task is to cluster the following eight points (with (x,y) representing location) in to three clusters: [10]

A1(2,10),A2(2,5),A3(8,4),B1(5,8),B2(7,5),B3(6,4),C1(1,2),C2(4,9).

The distance function is Euclidean distance. Suppose initially we assign A1, B1, and C1 as the center of each cluster respectively. Use the **k-means** algorithm to show only

- a) The three cluster centers after the first round execution
- b) The final three clusters.

**OR**

- (ii) What are the different clustering methods? Explain in detail.

**--ooOoo--**