**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 Section2(f) & 12(B)of the UGC act,1956

### IV B.Tech I Sem Regular End Examination, Nov/Dec 2022
### Data Mining
(CSE/IT)

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) | Distinguish between Noise and Outliers. | 2M | CO1 | BL2 |
| | b) | What do you understand by Subjective Measures of Pattern Interestingness? | 2M | CO1 | BL1 |
| | c) | Give the classification of Rule Constraints. | 2M | CO2 | BL1 |
| | d) | How are Frequent Itemsets, Closed Itemsets, and Max Itemsets related? | 2M | CO2 | BL1 |
| | e) | Differentiate Classification and Prediction. | 2M | CO3 | BL2 |
| | f) | How does the Naïve Bayes classifier handle the Zero frequency Problems? | 2M | CO3 | BL1 |
| | g) | List out the different Types of Outliers. | 2M | CO4 | BL1 |
| | h) | Compare Agglomerative and Divisive Hierarchical Clustering. | 2M | CO4 | BL2 |
| | i) | What is Web Usage Mining? | 2M | CO5 | BL1 |
| | j) | What do you meant by Trend Analysis? | 2M | CO5 | BL1 |

### PART- B
(10*5 Marks = 50 Marks)

| | | | | | |
|---|---|---|---|---|---|
| 2 | a) | What do you think of Data Mining from a Database perspective? | 5M | CO1 | BL1 |
| | b) | What is the significance of Data Transformation in preprocessing? Illustrate the Normalization techniques for Data Transformation. | 5M | CO1 | BL1 |

**OR**

| | | | | | |
|---|---|---|---|---|---|
| 3 | a) | List and Explain the Task primitives for specifying a Data Mining Task. | 5M | CO1 | BL4 |
| | b) | Explain in brief about Issues to be considered during Data Integration. | 5M | CO1 | BL4 |

4  a)  Apply Apriori algorithm to extract the frequent itemsets on the following data (Use min_sup=50%)          5M   CO2   BL3

| TID | ITEMS |
|-----|-------|
| 1   | I1, I3, I4 |
| 2   | I2, I3, I5 |
| 3   | I1, I2, I3, I5 |
| 4   | I2, I5 |

   b)  Define Graph Pattern Mining. Give it's applications.          5M   CO2   BL1

**OR**

5  a)  What are the advantages of FP-Growth algorithm?          5M   CO2   BL1

   b)  Categorize the Techniques used for Mining Multidimensional Association rules based on treatment of Quantitative Attributes.          5M   CO2   BL3

6  a)  Illustrate the major steps of Decision Tree Classification.          5M   CO3   BL4

   b)  Write the differences between Eager Learning and Lazy learning approaches.          5M   CO3   BL1

**OR**

7  a)  Distinguish between Naïve Bayesian classification and Bayesian Belief networks.          5M   CO3   BL2

   b)  How would you solve a Classification problem using Rule-based Classification? Explain.          5M   CO3   BL4

8  a)  What is the Objective of Clustering? Categorize Clustering approaches.          5M   CO4   BL1

   b)  Do you Agree with "WaveCluster and CLIQUE are two clustering algorithms that are both gridbased and density-based."? Justify your Answer.          5M   CO4   BL3

**OR**

9  a)  Write the basic K-means algorithm and Illustrate.          5M   CO4   BL1

   b)  Explain the Parametric Statistical based approaches for Outlier Detection.          5M   CO4   BL4

10 a)  Explain in brief about Classification in Data Streams.          5M   CO5   BL

   b)  Discuss the Various steps involved in Text Mining.          5M   CO5   BL

**OR**

11 a)  What is Spatial Database? Explain the methods of Mining Spatial Databases?          5M   CO5   BL4

   b)  Demonstrate the Generalized Sequential Patterns methods for SPM.          5M   CO5   BL5

---oo0oo---

**CO - Course Outcome**                    **BL - Blooms Taxonomy Levels**