

**Code No: 5405AR****JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD****M. Tech II Semester Examinations, January - 2020****DATA WAREHOUSING AND DATA MINING****(Computer Science)****Time: 3hrs****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) Differentiate between regression and classification. [5]
- b) What is rough set? What are the advantages and applications of rough sets? [5]
- c) Mention types of clustering techniques for clustering graph and network data. [5]
- d) Write short notes on web structure mining. [5]
- e) What is a spatial database? What are its features? [5]

**PART - B****5 × 10 Marks = 50**

2. How can efficiency of Apriori-based algorithm be improved? Describe briefly any three variations of the Apriori algorithm. [10]
- OR**
3. Explain Pattern Mining in Multi-level and Multi-Dimensional Space. [10]
4. Explain in detail the Genetic Algorithm and its limitations when compared to alternative optimization algorithms. [10]
- OR**
5. What is Classification? With an example, explain how Support Vector machines can be used for classification, evaluating the accuracy of a classifier or a predictor? Justify this statement with suitable illustrations. [10]
6. Explain exception – maximization algorithm. [10]
- OR**
- 7.a) Explain any one Grid-based clustering algorithm.
- b) Explain DBSCAN, density based clustering algorithm to discover cluster with arbitrary shape. [5+5]
8. Why current search engines are not sufficient for web resource discovery? Differentiate web content mining, web usage mining. [10]
- OR**
9. Explain about Text Clustering with an illustrative example. [10]
- 10.a) Explain the methods of mining spatial database.
- b) Write in detail about the Temporal Data Mining tasks. [5+5]
- OR**
11. Explain Mining time-series and sequence data. [10]

---0000---