R13

Code No: 117CD

**JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD**

**B. Tech IV Year I Semester Examinations, April/May - 2018**

**DATA WAREHOUSING AND DATA MINING**

**(Computer Science and Engineering)**

**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

(25 Marks)

| | | |
|---|---|---|
| 1.a) | List out the operations of OLAP. | [2] |
| b) | What is fact table? Write its uses. | [3] |
| c) | Define discretization. | [2] |
| d) | What is predictive mining? Explain it briefly. | [3] |
| e) | Write the purpose of Apriori algorithm. | [2] |
| f) | Define support and confidence measure. | [3] |
| g) | What is boosting? | [2] |
| h) | Define decision tree. | [3] |
| i) | Write the strengths of hierarchical clustering. | [2] |
| j) | Compare agglomerative and divisive methods. | [3] |

## PART-B

(50 Marks)

| | | |
|---|---|---|
| 2.a) | With a neat sketch, Explain three tier architecture of data ware housing. | |
| b) | Explain various data warehouse models. | [5+5] |
| | **OR** | |
| 3. | Write a note on<br>a) Relational OLAP<br>b) Multi dimensional OLAP. | [5+5] |
| 4.a) | Discuss in detail about the steps of knowledge discovery? | |
| b) | Write a note on subset selection in attributes for data reduction. | [5+5] |
| | **OR** | |
| 5.a) | Explain various data mining tasks. | |
| b) | Discuss briefly about data cleaning techniques. | [5+5] |
| 6.a) | Write FP- growth algorithm. | |
| b) | Explain how association rules are generated from frequent item sets. | [5+5] |
| | **OR** | |
| 7.a) | Explain the procedure to mining closed frequent data item sets. | |
| b) | Explain, how can you improve the performance of Apriori algorithm. | [5+5] |

8.a) What is Bayesian belief network? Explain in detail.
b) Write a note attribute selection measures. [5+5]

**OR**

9.a) Write k-nearest neighbor classification algorithm and its characteristics.
b) Write decision tree induction algorithm. [5+5]

10.a) What is outlier detection? Explain distance based outlier detection.
b) Write partitioning around mediods algorithm. [5+5]

**OR**

11.a) Write K-means clustering algorithm.
b) Write the key issue in hierarchical clustering algorithm. [5+5]

--ooOoo--