R13

Code No: 117CD

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
B. Tech IV Year I Semester Examinations, November/December - 2016
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)   Define Data ware housing.   [2]
   b)   Differentiate OLAP, ROLAP and HOLAP.   [3]
   c)   Discuss about subset selection   [2]
   d)   Mention any three measures of Similarity.   [3]
   e)   Define Association rule mining two step processes.   [2]
   f)   Write short note on support and confidence measures.   [3]
   g)   Mention types of classifier techniques.   [2]
   h)   Define Pre pruning and post pruning.   [3]
   i)   Discuss on Agglomerative and Divisive clustering techniques. [2]
   j)   Mention the various types of clustering methods.   [3]

PART-B
                                                                 (50 Marks)

2.   Explain data mining as a step process of knowledge discovery. Mention the Functionalities of Data mining.   [10]

OR

3.   Differentiate Operational database systems and data warehousing. Explain the star schema and fact constellation schemas.   [10]

4.   Explain the various Data pre-processing techniques. How data reduction helps in data pre-processing.   [10]

OR

5.   How can the data cube be efficiently constructed for discovery-driven Exploration? Explain various operations of a Data Cube.   [10]

6.   How can we mine multilevel Association rules efficiently using concept hierarchies? Explain. Illustrate with an A-priori algorithm for the given dataset below.   [10]

TID	List of items
001	milk, dal, sugar, bread
002	Dal, sugar, wheat,jam
003	Milk, bread, curd, paneer
004	Wheat, paneer, dal, sugar
005	Milk, paneer, bread
006	Wheat, dal, paneer, bread

OR

7. Can we design a method that mines the complete set of frequent item sets without candidate generation? If yes, explain with example table mentioned above.   [10]

8. Describe the data classification process with a neat diagram. How does the Naive Bayesian classification works? Explain.   [10]

OR

9. What is prediction? Explain the various prediction techniques. Explain about Decision tree Induction classification technique.   [10]

10. What are outliers? Discuss the methods adopted for outlier detection.   [10]

OR

11. State K-means algorithm. Apply k-means algorithm with two iterations to form two clusters by taking the initial cluster centers as subjects 1 and 4.   [10]

Subject	A	B
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

--ooOoo--