R15

Code No: 127CD

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

B. Tech IV Year I Semester Examinations, November/December - 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) Differentiate Data warehouse and DBMS. [2]
b) Write two differences between OLAP and OLTP . [3]
c) What is meant by KDD? [2]
d) Discuss about data cleaning. [3]
e) Differentiate multiple and multilevel association rule mining. [2]
f) Discuss about the concept hierarchy generation. [3]
g) Explain about Entropy and Information gain. [2]
h) Define two step classification process. [3]
i) Mention the Key issues in hierarchical clustering. [2]
j) Differentiate OPTICS and DBSCAN. [3]

PART-B

(50 Marks)

2. What is a Data Warehouse? Explain three types of schemas that are used for modeling data warehouse with examples. State the applications of Data mining. What is its need in Business? [10]

OR

3. Define Data Cube computation. Explain the various methods for Data Cube Computation. Discuss Construction of Multi-dimensional model. And its operations. [10]

4. What is the need for Data preprocessing? Discuss briefly various forms of Data - preprocessing. [10]

OR

5.a) How to handle missing values in data sets?
b) Discuss attribute subset selection for dimensionality reduction. [5+5]

6. Construct the FP-Tree from the given Transactional Database. Explain the procedure in detail with minimum support = 3. [10]

TID	Items
100	F, A, C, D, G, I, M, P
200	A, B, C, F, L, M, O
300	B, F, H, J, O, W
400	B, C, K, S, P
500	A, F, C, E, L, P, M, N

OR

7. Explain market basket analysis and its relevance to association rule. Explain the Apriori algorithm using the following transactional data assuming that the support count is 22%. Illustrate with an example. [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

8. How Neural Networks can be used for Data classification? Which algorithm is suitable? Explain them with example? Discuss the role of Information Gain in classification.[10]

OR

9. What are the various methods of evaluating accuracy of a classifier or predictor? Explain Bagging and Boosting techniques? [10]

10. Explain the partitioning methods? Solve the following problem using Partition methods (K-means, K-mediods) for {2, 4, 10, 12, 8, 20, 30, 11, 25} where k = 2. [10]

OR

11. How Density based clustering algorithms are different from partitioning based cluster algorithms. Compare both. Explain DBSCAN algorithm with suitable example. [10]

--ooOoo--