B.E. (Computer Science & Engineering) Seventh Semester (C.B.S.)

Data Warehousing & Mining

P. Pages : 2 **NRT/KS/19/3571**

Time : Three Hours * 0 2 5 6 * Max. Marks : 80

Notes :
1. All questions carry marks as indicated.
2. Solve Question 1 OR Questions No. 2.
3. Solve Question 3 OR Questions No. 4.
4. Solve Question 5 OR Questions No. 6.
5. Solve Question 7 OR Questions No. 8.
6. Solve Question 9 OR Questions No. 10.
7. Solve Question 11 OR Questions No. 12.
8. Due credit will be given to neatness and adequate dimensions.
9. Assume suitable data whenever necessary.
10. Illustrate your answers whenever necessary with the help of neat sketches.
11. Use of non programmable calculator is permitted.

1. a) Draw and explain KDD architecture with neat sketch. **6**

b) What is data mining? What are technical issues to be considered when designing and implementing data ware house environment. **7**

OR

2. a) How data ware house is different from database? **5**

b) What do you mean by preprocessing in data ware house? Explain two data reduction strategies in brief. **8**

3. a) Draw and explain 3 tier architecture of data ware house. **8**

b) Explain the various operations to be performed on multidiamentional cube. **6**

OR

4. a) Discuss the design of ROLAP & MOLAP. **8**

b) Suppose your task as a software engineer at ABC university is to design a data mining system to examine the university course database, which contains following information. Name, Address, Status (UG OR PG) of each students, the course taken, & CGPA. Describe the architecture you would choose. **6**

5. a) What do you mean by frequent item set, and closed item set elaborate with example. **6**

b) Explain Apriori algorithm & state how to improve the efficiency of same. **7**

OR

6. a) What is Market Basket Analysis? **5**

b) Generate FP – Tree for following transaction database. Assume min – sup = 40%. **8**

TID	Itemset
T_1	K, D, B, T
T_2	D, G, T, W
T_3	B, K, D, T
T_4	K, W, G, D, T
T_5	W, S, D, T

7. a) Briefly outline the major ideas of naive Bayesian Classification. **7**

b) State and explain the various issues regarding classification and prediction. **6**

OR

8. a) Why association classification gives higher classification accuracy than the classical decision tree method. **7**

b) What do you mean by lazy learner? Why they called as lazy learners? **6**

9. a) Write k – means algorithm. Apply it to classify the following data in 2 – clusters. **8**
Data : 3, 3.5, 5, 3.3, 4.5, 5.2, 6.4
Use Euclidean distance method.

b) Why is it that BIRCH encounters difficulties in finding clusters of arbitrary shape but OPTICS does not? **6**

OR

10. a) Explain agglomerative (AGNES) and divisive (DIANA) algorithm in brief. **10**

b) Differentiate between supervised and unsupervised learning. **4**

11. a) Explain the techniques for mining time series data. **7**

b) Elaborate data mining applications **any two.** **6**

1) Science & Engineering

2) Social Network Analysis

3) Biological Data Analysis

4) Telecommunication Industry.

OR

12. a) Discuss the basic measure for text retrieval. Explain various retrieval methods. **8**

b) Graph and network mining have become increasingly important and heavily researched. Justify the same. **5**

****