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

Data Warehousing & Mining

P. Pages : 2 **NKT/KS/17/7487**

Time : Three Hours * 0 1 1 3 * 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. Illustrate your answers whenever necessary with the help of neat sketches.

1. a) Explain various data mining functionalities along with examples. **8**

b) What is the need of Data Preprocessing? Explain steps involved in data prepressing. **6**

OR

2. a) Discuss Major issues in Data Mining. **7**

b) Give classification of data mining system and also explain concept hierarchy generation. **7**

3. a) Explain the need of Multidimensional Data Model. **3**

b) List and explain various OLAP operations. **8**

c) Differentiate between OLAP and OLTP. **3**

OR

4. a) Explain three tier architecture of data warehouse with neat sketch. **8**

b) Discuss the architecture of ROLAP and MOLAP in detail with the help of suitable diagram. **6**

5. a) Explain constraint-based Association mining with example. **7**

b) Associate rule mining often generate large number of rules. Discuss effective methods that can be used to reduce the number of rules generated while still preserving most of the interesting rules. **6**

OR

6. a) Define following terms: **6**

i) Frequent Item Sets.
ii) Closed Item sets.
iii) Association rules.

b) Explain in brief Market-Basket analysis using example. **7**

7. a) Why is Bayesian classification called naive? Briefly outline the major ideas of naive Bayesian classification. **7**

b) How the accuracy of a classifier or a predictor is evaluated? Explain. **6**

OR

8. a) What are the different issues regarding classification and prediction. **7**

b) Explain classification by Decision Tree Induction with example. **6**

9. a) Give classification of clustering algo's and explain partition based clustering algorithm namely k-means stating its merits, demerits and application area. **8**

b) Write a short note on SVM (Support Vector M/C). **5**

OR

10. a) Explain various requirements of clustering. **4**

b) Compare agglomerative and divisive hierarchical clustering methods. **6**

c) What is outlier? Why outlier analysis is important? **3**

11. a) Describe the process of mining Time-series data with suitable example. **7**

b) Define following terms: **6**

i) Data streams.
ii) Time series data.
iii) Sequence Data.

OR

12. Write short notes on :

i) Graph Mining. **4**

ii) Network analysis and Multi relational Data Mining. **5**

iii) Mining sequence pattern in Biological Data. **4**

iv) Mining Data streams. **4**

****