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

Data Warehousing & Mining

P. Pages : 2

Time : Three Hours

|||||||||||||||||||
0093

NJR/KS/18/4626

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.

1. a) Define Data mining. What are the steps involved in KDD process? **8**

   b) Write a short note on:- **6**
      i) Classification of Data mining.
      ii) Data mining Task Primitive

OR

2. a) Describe the typical architecture of Data mining system? **6**

   b) Why preprocessing is necessary in Data mining. Explain various preprocessing technique in brief. **8**

3. a) Define Data warehousing. State its characteristics features. **3**

   b) Differentiate between OLTP and OLAP. **6**

   c) List different DW schemas. Explain in brief STAR SCHEMA. **4**

OR

4. a) Discuss the different types of OLAP servers. **8**

   b) Write a short note on **any one.** **5**
      i) Cube materialization.   ii) Attribute oriented Induction

5. a) What are the two necessary steps of association rule mining. **3**

   b) Explain closed and maximal frequent item set. **3**

   c) Write the step involved in finding frequent item set using apriori algorithm. **7**

OR

6. a) Discuss various kinds of Association rules. **8**

   b) Explain the concept of constraint based association mining. **5**

7.	a)	Differentiate between classification and prediction.	4
	b)	Give the steps involved in decision tree algorithm. States its advantages and disadvantages.	6
	c)	Write in brief about attribute selection measure.	3

OR

8.	a)	Briefly explain regression methods used in prediction.	4
	b)	Describe Naive Bayesians classification.	6
	c)	Write a short note on:- i) Bagging ii) Boosting	3

9.	a)	What is clustering? How it differs from classification? Also give its application area.	4
	b)	Briefly explain with example, how dissimilarity between object can be computed in the following data types:- i) Interval-scaled variable ii) Binary variable iii) Categorical variable.	9

OR

10.		Write a short note on:-	
	a)	K-means partitioned method	4
	b)	Agglomerative and decisive hierarchical clustering.	2
	c)	Outlier detection	3
	d)	DBCAN clustering	4

11.	a)	Explain in brief the complex data type. i) Data stream ii) Time series data iii) Sequence data	3
	b)	Describe Trend analysis w. r. t. time series data.	6
	c)	Explain the concept of sequence pattern in detail.	5

OR

12.		Write short notes on : **any three.**	
	i)	Methodology of Data streaming.	5
	ii)	Graph mining	5
	iii)	Social Networking	4
	iv)	Task and challenges in link mining.	4

****