Roll No. :—————————

7E4243

7E4243

B.Tech. (Sem. VII) (Main) Examination, Nov-Dec - 2011
Computer Science
7CS6.2 Data Mining and Ware Housing

Time : 3 Hours

Total Marks : 80
Min. Passing Marks : 24

Attempt any five questions, selecting one question from each unit. All questions carry equal marks. (Schematic diagrams must be shown wherever necessary. Any data you feel missing suitably be assumed and stated clearly. Units of quantities used/calculated must be stated clearly.)

UNIT - I

1. (a) Write down the difference between :

   (i) Operational database data and Data warehouse data. (4)

   (ii) Data mining and Data warehouse. (4)

  (b) (i) Give three example of problems likely to be found when operational data are integrated into the data warehouse. (5)

   (ii) What are schema? Discuss various schemas used in Data warehouse. (3)

OR

1. (a) (i) Define the Data warehouse in terms of subject - oriented, integrated, time - varient and non - volatile collection of data with example. (4)

   (ii) How does data mining works ? Discuss the different phases in the data mining process. (4)

  (b) (i) Define Metadata and Data Mart. Explain how it is useful with suitable example. (4)

   (ii) Discuss the Binning method with suitable example. (4)

UNIT - II

2. (a) (i) What are the characteristics of OLAP? (2)

   (ii) Explain what ROLAP is and why are recommend its use in the Relational Database environment. (4)

(b) (i) Explain the OLAP client - server architecture with suitable diagram. (6)

(ii) Differentiate between ROLAP and MOLAP. (4)

OR

2. (a) (i) Define the Association Rules and Classification Rules. (3)

(ii) Explain basic Data mining tasks with an example. (3)

(b) Explain any one Association Rules algorithm with suitable example. (10)

UNIT - III

3. (a) (i) Describe any one statistical - based algorithm for classification with suitable example. (8)

(ii) Write down the advantages and disadvantages of Neural Networks for classification. (4)

(b) How are decision tree related to the classification rules ? (4)

OR

3. (a) How C 4.5 decision tree algorithm improves algorithm ? Explain. (8)

(b) Explain the major issues faced by most decision tree algorithms. (8)

UNIT - IV

4. (a) (i) Explain the Outlier clustering problem.

(ii) How does clustering differs from classification? Describe about the back = O propagation in neural networks. (8)

(b) Apply the K – means clustering in following items to cluster :

$\{2, 4, 10, 12, 3, 20, 30, 11, 25\}$ suppose that $K = 2$. (8)

OR

4. (a) Explain in brief :

(i) Genetic Algorithm

(ii) Bayesian Classification

(iii) Data Cubes

(iv) Snow Flacks (8)

(b) Explain in brief :

(i) STING and CLIQUE method

(ii) Outlier Analysis

(iii) 3 - Tier Architecture

(iv) Multi - Dimensional Data Model. (8)

UNIT - V

5. (a) (i) What do you mean by tunning normalization and tunning denormalization? (4)

(ii) Explain the data warehouse tunning. (4)

(b) What is Regression ? Explain the Linear Regression Model. (8)

OR

5. (a) What is the techinque through which the performance of database is improved ? Explain it. (8)

(b) What are the advatntages and disdvantages of OLAP. (8)