MCSE First Year EXAMINATION, 2024

Second Semester

Data Warehousing and Data Mining

Time : 3 hours   Full Marks : 100

Answer question no.1 and any four from the rest

All questions carry equal marks

1. Answer true or false stating reasons   2 x10 = 20

   (a) Soft computing techniques are mainly used in a data mining system for data preprocessing.
   (b) A heterogeneous database consists of a group of legacy databases that combines different kind of data systems.
   (c) Predicting sales amounts of a new product based on advertising expenditure is a data mining problem.
   (d) A data warehouse is a department subset of a data mart.
   (e) 'Summarization' provides the general characteristics of a target class of data.
   (f) The performance of an OLAP system is measured by its transaction throughput.
   (g) Text data mining methods using keyword do not always provide satisfactory results.
   (h) A sequence database stores sequences of values or events obtained over repeated measurements of time.
   (i) Pattern evaluation is a process that identifies only the interesting patterns.
   (j) Data warehouses are designed for real time business operations.

2. (a) Describe with a neat diagram the architecture and operation of a typical data mining system.
                                                                                    12

   (b) Describe briefly with examples the difference between Predictive and Descriptive data mining.
                                                                                    8

[Turn over

3. (a) Discuss with reasons whether each of the following activities is a data mining task or not.

2 X 5

    (i) Dividing the customers of a company according to their profitability.

    (ii) Predicting the future stock price of a company using historical records.

    (iii) Monitoring the heart rate of a patient for abnormalities.

    (iv) Monitoring seismic waves for earthquake activities.

    (v) Extracting the frequencies of a sound wave.

  (b) Describe how data mining can be employed in two different domains of applications.

2 X 5

4. Suppose a Banking institution wants to predict if their client will subscribe to a new long- term deposit plan. They conducted a marketing campaign over telephone to sale this long- term deposit policy to their clients. The goal is to identify those clients (who may be more interested) for further campaign. Assume that a set of training date is available. For each telephonic conversation the following attributes were recorded.

Bank client data:
1. Age
2. Type of job
3. Marital status
4. Education
5. Whether defaulter
6. Whether availed Housing loan
7. Whether availed personal loan
8. Date of Contact
9. Duration of Call

Answer following questions:

  (a) Formulate the problem as a data mining problem

  (b) Describe the necessary steps involved in solving the problem

  (c) Discuss the merits and demerits of the method you have suggested

5 + 12 + 3

5. (a) What is text data mining? Suppose you need to classify a given set of news articles on various faculties of games and sports where the name of each such faculty is the respective class title. Assume that the number of classes (from which the given set of news articles is derived) is not known *a priori*. Describe the following necessary steps to solve the problem. Use any suitable artificial neural network architecture as a soft computing tool for classification.

(i) Feature selection

(ii) Feature reduction

(iii) Analytical formulation of the problem

(iv) Algorithm for classification

3 + 3 + 4 + 10

6. (a) What is a data warehouse? Unlike a conventional database, why is it suitable for data mining operation? 3 + 7

(b) Describe with an example how data are organized in a typical data warehouse.

10

7. Write short notes on any two of the following. 2 X 10

(a) Pattern evaluation

(b) Data mining for market analysis

(c) Data Cubes