(6 Pages) **Reg. No. :** .....

Code No. : 20608 E **Sub. Code : SMCS 63**

B.Sc. (CBCS) DEGREE EXAMINATION, APRIL 2021.

Sixth Semester

Computer Science — Core

DATA WAREHOUSING AND DATA MINING

(For those who joined in July 2017 onwards)

Time : Three hours Maximum : 75 marks

PART A — ($10 \times 1 = 10$ marks)

Answer ALL questions.

Choose the correct answer :

1. The following technology is not well suited for determining

   (a) Expert system technology

   (b) Data visualization

   (c) Technology limited to specific data types such as numeric data types

   (d) Parallel architecture

2. OLAP stands for

   (a) Online Analytical Processing

   (b) Online Analysis Processing

   (c) Online Transaction Processing

   (d) Online Aggregation Processing

3. Task of interfering a model from labeled training date is called

   (a) Un supervised learning

   (b) Supervised learning

   (c) Reinforcement learning

   (d) None of the above

4. Give the role of the form IF X THEN Y rule confidence is defined as the conditional probability that select one.

   (a) Y is false when X is known to be false

   (b) Y is true when X is known to be false

   (c) X is true when Y is known to be true

   (d) X is false when Y is known to be false

5. A ———— is a single neuron with multiple inputs and output.

   (a) Back propagation (b) Perceptron

   (c) Propagation (d) None

6. Deep knowledge can be found only by using ————————

(a) Clues (b) SQL

(c) OLAP (d) Algorithm

7. ———————— is a creative activity that has to be performed repeatedly in order to get best results.

(a) Selection (b) Reporting

(c) Cleaning (d) Coding

8. Variance is a function of ————————

(a) Complexity (b) Score function

(c) Score matrix (d) Sample size

9. Meta data is a ————————

(a) Data about mining

(b) Data about query

(c) Data about data

(d) Data about no data

10. Which of the following is required by K-means clustering?

(a) defined distance matrix

(b) no clusters

(c) initial guess as to cluster centroids

(d) all the above

PART B — (5 × 5 = 25 marks)

Answer ALL questions, choosing either (a) or (b).

Each answer should not exceed 250 words.

11. (a) What is data mining? Explain.

Or

(b) Describe the features of Data warehouse and how it might be defined?

12. (a) (i) Define OLAP.

(ii) Characteristics of OLAP.

Or

(b) What do you think of data mining from a database perspective?

13. (a) Explain the architecture of data mining system.

Or

(b) Write short notes on Association Rule.

14. (a) Discuss on Apriori algorithm.

Or

(b) What are the issues regarding classification and prediction?

15. (a) Explain K-means algorithm.

Or

(b) What are the types of data in cluster analysis?

PART C — (5 × 8 = 40 marks)

Answer ALL questions, choosing either (a) or (b).

Each answer should not exceed 600 words.

16. (a) Explain data mining matrices.

Or

(b) Explain basic data mining tasks.

17. (a) Briefly explain about Naïve Baysian theorem.

Or

(b) Explain various steps in data pre processing.

18. (a) Explain the process architecture with neat diagram.

Or

(b) Give an example for Apriori with transations.

19. (a) Explain about the association rule generation process.

Or

(b) Discuss decision tree based algorithm.

20. (a) Explain the method of clustering the large database.

Or

(b) Explain the account of clustering of categorical attributes.

———————