

B.Sc. (CBCS) DEGREE SPECIAL SUPPLEMENTARY EXAMINATION, NOVEMBER 2022.

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 × 1 = 10 marks)

Answer ALL questions.

Choose the correct answer :

1. Which of the following features usually applies to data in a data warehouse?
(a) Data are often deleted
(b) Data are rarely deleted
(c) Most applns consists of transactions
(d) Relatively few records are processed by appln.

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

- (a) Clues (b) OLAP
(c) OLTP (d) Algorithm

7. Which of the following is not data mining s/w?

- (a) SPSS (b) Darwin
(c) Mantas (d) CART

8. The association rule notation is _____ used for transition in D

- (a) L (b) t_n
(c) $x \Rightarrow y$ (d) L^2

9. The _____ for an association rule $x \Rightarrow y$ is the ratio of the number of transactions that contain $x \cup y$ to the no. of transaction that contain x.

- (a) Support
(b) Confidence
(c) Support and Confidence
(d) None

2. OLPA stands for

- (a) Online Analytical Processing
(b) Online Analysis Processing
(c) Online Transaction Processing
(d) Online Aggregated Processing

3. Data that can be modeled as dimensional attributes and measure attributes are called _____ data.

- (a) Measured (b) Single dimensional
(c) Dimensional (d) Multi dimensional

4. Give the rule of the form IF X THEN Y rule confidence in defined as a 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 true
(c) X is false when Y is known to be true
(d) X is true when Y is known to be false

5. _____ maps data into subsets with associated simple description.

- (a) Clustering (b) Regression
(c) Classification (d) Summarisation

Page 2 Code No. : 20107 E

10. Expansion of DBSCAN is

- (a) Data based special clustering of appln with noise
(b) Density based spatial clustering of appln with noise.
(c) Density based domain without noise
(d) Domain based spacial noise.

PART B — (5 × 5 = 25 marks)

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

Each answer should not exceed 250 words.

11. (a) Differentiate between Data mining and Data warehousing.

Or

- (b) (i) Define OLAP
(ii) Characteristics of OLAP.

12. (a) What are the types of metadata that is maintained in data warehousing?

Or

- (b) Explain the taxonomy of datamining tasks.

13. (a) Write short notes on system process.

Or

(b) Write short notes on Association Rule.

14. (a) Explain decision tree.

Or

(b) Explain frequent item sets with example.

15. (a) What do you mean by rule based algorithm?

Or

(b) Name the areas of applns of datamining used.

PART C — (5 × 8 = 40 marks)

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

Each answer should not exceed 600 words.

16. (a) Define support and confidence. Explain with example.

Or

(b) Explain various steps in data pre-processing.

Page 5 Code No. : 20107 E

17. (a) Give an example for Apriori with transitions and explain it in detail.

Or

(b) Briefly explain about Naïve Bayesian theorem.

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

Or

(b) Write short notes on :

(i) N/W topology

(ii) Linear regression.

19. (a) Explain Agglomerative algorithm.

Or

(b) What is the purpose of cluster analysis in data warehousing.

20. (a) Discuss about partitioning.

Or

(b) Explain different data types used in cluster analysis.

Page 6 Code No. : 20107 E