

(8 pages)

Reg. No. : .....

**Code No. : 5848**

**Sub. Code : WCSM 21**

M.Sc. (CBCS) DEGREE EXAMINATION, APRIL 2024.

Second Semester

Computer Science — Core

**DATA MINING AND WAREHOUSING**

(For those who joined in July 2023 onwards)

Time : Three hours

Maximum : 75 marks

PART A — (15 × 1 = 15 marks)

Answer ALL questions.

Choose the correct answer:

1. Data mining is used to refer \_\_\_\_\_ stage in knowledge discovery in database.  
(a) selection  
(b) retrieving  
(c) discovery  
(d) coding

2. Which of the following is not a data pre-processing methods?

- (a) Data Visualization
- (b) Data Discretization
- (c) Data Cleaning
- (d) Data Reduction

3. \_\_\_\_\_ is a subject-oriented, integrated, time-variant, nonvolatile collection or data in support of management decisions.

- (a) Data Mining
- (b) Data Warehousing
- (c) Document Mining
- (d) Text Mining

4. \_\_\_\_\_ is the technique which is used for discovering patterns in dataset at the beginning of data mining process.

- (a) Kohonen map.
- (b) Visualization.
- (c) OLAP.
- (d) SQL.

Page 2

**Code No. : 5848**





5. A database containing volatile data used for daily operation of an organization is \_\_\_\_\_

- (a) Historic data
- (b) Metadata
- (c) Knowledge.
- (d) Operational data

6. Query tools and data mining tools are \_\_\_\_\_

- (a) Same
- (b) Different
- (c) Complementary
- (d) Standard.

7. Which of the following function involves data cleaning, data standardizing and summarizing

- (a) Storing data.
- (b) Transforming data
- (c) Data acquisition
- (d) Data Access.

Page 3

Code No. : 5848

8. Classification rules are extracted from \_\_\_\_\_

- (a) Root node
- (b) Decision tree
- (c) Siblings
- (d) Branches.

9. The refers to extracting knowledge from larger amount of data.

- (a) Data abstraction
- (b) Data warehouse
- (c) database
- (d) data mining.

10. In \_\_\_\_\_ approach data ware house is build first and all information needed is selected.

- (a) top-down
- (b) Clientlserver
- (c) bottom-up
- (d) DSS

11. OLAP is used to explore the \_\_\_\_\_ knowledge.

- (a) shallow
- (b) deep
- (c) multidimensional
- (d) hidden.

Page 4

Code No. : 5848

[P.T.O.]





12. The easiest way to gain access to the data and facilitate effective decision making is to setup a

- (a) database.
- (b) data mart.
- (c) data warehouse.
- (d) operational

13. Smaller local data warehouse is called as \_\_\_\_\_

- (a) data mart.
- (b) database.
- (c) data model.
- (d) meta data.

14. Data warehouse is only used for \_\_\_\_\_

- (a) operating the data
- (b) managing the data
- (c) decision making
- (d) queries.

15. The main organizational justification for implementing a data warehouse is to provide \_\_\_\_\_

- (a) cheaper ways of handling transportation.
- (b) decision support.
- (c) storing large volume of data
- (d) access to data.

PART B — ( $5 \times 4 = 20$  marks)

Answer ALL questions, choosing either (a) or (b).  
Each answer should not exceed 250 words.

16. (a) Elucidate the issues in data mining.

Or

(b) Describe the data mining from a database perspective.

17. (a) Elaborate the decision tree algorithm C4.5 improves ID3.

Or

(b) Summarize the propagation in neural network-based algorithms.

18. (a) Point out the concept of outliers in clustering.

Or

(b) Assess the divisive clustering in hierarchical algorithms.





19. (a) Discover the different types of data marts.

Or

- (b) Distinguish between the OLTP and OLAP systems.

20. (a) Express the data warehouse architectural strategies and organizational issues.

Or

- (b) Conclude the performance consideration in data warehouse.

PART C — (5 × 8 = 40 marks)

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

Each answer should not exceed 600 words.

21. (a) Examine the concept of decision trees in data mining.

Or

- (b) Demonstrate the activate functions in neural networks.

22. (a) Outline the regression in statistical-based algorithms.

Or

- (b) Illustrate the K-nearest neighbors of distance-based algorithms

23. (a) Compare and construct parallel and distributed algorithms.

Or

- (b) Enumerate the implementation of PAM algorithm.

24. (a) Discuss the categories of OLAP tools.

Or

- (b) Analysis the Cognos PowerPlay in state of the market.

25. (a) Justify the crucial decisions in designing a data warehouse.

Or

- (b) Explain the census data in national data warehouses.

