

(6 pages)

Reg. No. : \_\_\_\_\_

Code No. : 10107 E Sub. Code : SMCS 63

B.Sc. (CBCS) DEGREE EXAMINATION, APRIL 2023.

Sixth Semester

Computer Science – Core

DATA WAREHOUSING AND DATA MINING

(For those who joined in July 2017-2019)

Time : Three hours Maximum : 75 marks

PART A — (10 × 1 = 10 marks)

Answer ALL questions.

Choose the correct answer :

1. \_\_\_\_\_ databases are owned by particular departments or business groups.
- (a) Informational  
(b) Operational  
(c) Both informational and operational  
(d) Flat

6. \_\_\_\_\_ may be defined as the data objects that do not comply with the general behavior or model of the data available.
- (a) Outlier Analysis (b) Evolution Analysis  
(c) Prediction (d) Classification
7. From where are classification rules extracted?
- (a) Branches (b) Decision Tree  
(c) Siblings (d) Root node
8. How to define Classification accuracy?
- (a) A subdivision of a set of examples into a number of classes  
(b) The task of assigning a classification to a set of examples  
(c) Measure of the accuracy of the classification of a concept that is given by a certain theory  
(d) None of these
9. Which is needed by K-means clustering?
- (a) defined distance metric  
(b) number of clusters  
(c) initial guess as to cluster centroids  
(d) all of these

Page 3 Code No. : 10107 E

2. Business Intelligence and data warehousing is used for \_\_\_\_\_
- (a) Forecasting  
(b) Data Mining  
(c) Analysis of large volumes of product sales data  
(d) All of the above
3. Online Analytical Processing (OLAP) is a technology that is used to create software.
- (a) Decision support (b) Forecasting  
(c) Informational (d) None
4. Data warehouses and OLAP tools are based on \_\_\_\_\_ data model.
- (a) Single dimension (b) Two dimension  
(c) Multidimensional (d) None
5. Which of the following is the right approach to Data Mining?
- (a) Infrastructure, exploration, analysis, exploitation, interpretation  
(b) Infrastructure, exploration, analysis, interpretation, exploitation  
(c) Infrastructure, analysis, exploration, interpretation, exploitation  
(d) None of these

Page 2 Code No. : 10107 E

10. A \_\_\_\_\_ is the complete set of genes of an organism.
- (a) Genome (b) proteome  
(c) nucleotides (d) Genomics

PART B — (5 × 5 = 25 marks)

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

11. (a) Write note on (i) meta data (ii) Data marts
- Or
- (b) What are the design considerations of building a data warehouse?
12. (a) What are the five decision support tools?
- Or
- (b) Mention the OLAP guidelines.
13. (a) Explain Architecture of data mining.
- Or
- (b) Describe about data transformation.
14. (a) What is Mining Multilevel Association Rules?
- Or
- (b) Write about Baye's theorem.

Page 4 Code No. : 10107 E

[P.T.O.]

15. (a) Mention A Categorization of Major Clustering Methods.

Or

(b) Write note on Data Mining for Financial Data 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 the Overall Architecture of data warehouse.

Or

(b) Describe Data ware house administration and management.

17. (a) Illustrate categorization of OLAP Tools.

Or

(b) Discuss about multi dimensional versus multi relational OLAP.

18. (a) What is Data mining? Explain.

Or

(b) Summarize about data integration and transformation.

19. (a) Write in detail about Naïve Bayesian classification.

Or

(b) Explain linear regression.

20. (a) What are the Types of Data in Cluster Analysis?

Or

(b) Expand and explain about DBSCAN.