15529

Reg. No. | | | | | | | | |

V Semester B.C.A. Degree Examination, April - 2022

COMPUTER SCIENCE

DATA MINING

(CBCS Scheme)

Time : 3 Hours Maximum Marks : 100

Instructions to Candidates:

Answer all sections.

SECTION - A

I. Answer any ten questions. Each question carries 2 marks. (10×2=20)

1. What is a Data Warehouse? What is its significance?
2. State the two differences between a database system and a data warehouse.
3. Differentiate between the star schema and snowflake schema.
4. State the three tiers of the 3 tier data warehouse architecture.
5. What is data marting?
6. What ar ROLAP servers?
7. What is data preprocessing? What is its significance?
8. What is data generalization?
9. What are histograms?
10. Differentiate between classification and prediction analysis.
11. What are genetic algorithms? What is the purpose of using these algorithms.
12. What is clustering? State any two clustering algorithms.

SECTION - B

II. Answer any five questions. Each question carries 5 marks. (5×5=25)

13. What are data cubes? Explain with an example.
14. Explain the concept of data aggregation with an example.

P.T.O.

15. What is a Data warehouse? Discuss Data warehouse turing and testing.
16. What is data mining? Discuss its relevance and functionalities.
17. What is Data Reduction? Discuss briefly the techniques for data reduction.
18. Discuss the steps involved in analysis of attribute relevance.
19. Explain the Decision tree classifier with an example.
20. What is Association mining? What is an Association rule? Discuss association rule mining from a transaction database.

SECTION - C

III. Answer any **three** questions. Each question carries **15** marks. (3×15=45)

21. a. Explain the different schemas used for Data warehouse. (8)
    b. Explain the different OLAP functions and tools. (7)
22. a. Explain the MOLAP architecture. Where are its advantages and disadvantages? (8)
    b. What is data normalization? Explain any two techniques of data normalization with appropriate examples. (7)
23. a. Explain any two descriptive statistical measures for mining large databases. (8)
    b. Explain with example any two methods of measuring dispersion of data. (7)
24. a. Discuss the issues of classification and prediction. (8)
    b. What is classification by Back propagation. Briefly explain the back propagation algorithm. (7)
25. a. Explain k-nearest classifier with an example. (8)
    b. Explain different types of data used in cluster analysis. (7)

SECTION - D

IV. Answer any **one** question. Each question carries **10** marks. (1×10=10)

26. Explain the Apriori algorithm with an example.
27. Explain Data cleaning.
