Printed Pages: 3 | 481 | ECS-075

(Following Paper ID and Roll No. to be filled in your Answer Book)

Paper ID : 110755 | Roll No. | | | | | | | | | |

B.Tech.

(SEM. VII) THEORY EXAMINATION, 2015-16

DATA MINING & DATA WAREHOUSING

[Time:3 hours] **[Total Marks:100]**

Note : Attempt questions from all Sections as per directions.

Section-A

Attempt **all** parts of this section. Answer in brief.(2x10=20)

1. (a) Write some of the tasks of the data mining?

   (b) Briefly explain the concept of Frequent Itemsets and Closed Itemsets.

   (c) Define CLARA and CLARANS.

   (d) What are the steps to build the data warehouse?

   (e) Discuss various OLAP operations.

   (f) Describe the market basket analysis.

   (g) Write short notes on linear regression.

https://quescol.com | Free PYQs, Solutions, Syllabus, Notes etc

(h) Why data warehouse is maintained separately from database?

(i) What is star schema?

(j) Name and describe the main features of genetic algorithm.

Section-B

Attempt **any five** questions from this section : (10×5=50)

2. Define KDD. Identity and describe the phases in KDD process.

3. Explain attribute subset selection method for data reduction with examples.

4. Describe the difference between the following approaches for the integration of data mining system with database or data warehouse: no coupling, loose coupling and semi tight coupling.

5. Draw a box-and-whisker plot for the following data set:

   126, 132, 138, 140, 141, 141, 142, 143, 144, 144, 144, 145, 146, 147, 148, 148, 149, 149, 150, 150, 150, 154, 155, 158, 158.

   Also find the outliers.

https://quescol.com | Free PYQs, Solutions, Syllabus, Notes etc

6. Explain Principal Component Analysis (PCA) in detail.

7. What are Outliers? How Outlier analysis can be done?

8. Distinguish between Dimensionality reduction and Numerosity reduction.

9. Discuss why analytical characterization and attribute relevance analysis are needed.

Section-C

Attempt **any two** questions in this section : (15x2=30)

10. Describe the various types of OLAP servers. How OLAP is different from OLTP?

11. Explain with diagram the starts, snowflake and fact constellation schemas for multidimensional databases. Also write their advantages and disadvantages.

12. Explain Chi-square test method. Show using chi-square test that gender and preferred reading are independent or not from given table. (Given are the observed counts).

	Male	Female	Total
Fiction	250	200	450
Non-Fiction	50	1000	1050
Total	300	1200	1500

—x—

https://quescol.com | Free PYQs, Solutions, Syllabus, Notes etc