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

Question Paper Code : X 67526

B.E./B.Tech. DEGREE EXAMINATIONS, NOVEMBER/DECEMBER 2020
Seventh Semester
Computer Science and Engineering
CS 1011 – DATA WAREHOUSING AND MINING
(Regulations 2008)

Time : Three Hours | Maximum : 100 Marks

Answer ALL questions

PART – A **(10×2=20 Marks)**

1. State why pattern evaluation is included as a step in the process of knowledge discovery from databases.

2. What does a time-series database store ?

3. Why is data preprocessing an important issue for both data warehousing and data mining ?

4. What do data cleaning routines attempt to fill ?

5. What is association rule mining ?

6. List the two interesting measures of an association rule.

7. How the quality of clustering be assessed ?

8. How do you choose best attribute while constructing a decision tree ?

9. What kind of mining is required in a multimedia database ? State its use.

10. Give an application associated with spatial data mining and time series and sequence analysis of data.

PART – B **(5×16=80 Marks)**

11. a) i) What is a data warehouse and why do we need it ? Discuss with examples. **(8)**

ii) Explain with an example how relational tables can be represented as multidimensional cubes. **(8)**

(OR)

b) Explain with examples the following schemas used to represent multidimensional databases (multidimensional model) :

i) Star schema. **(5)**

ii) Snow flake schema. **(5)**

iii) Fact constellation schema. **(6)**

12. a) Suppose that the data for analysis include the attributed age. The age values for the data tuples are 13, 15, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70. **(16)**

i) Use smoothing by bin means to smooth the above data, using a bin depth of 3. Illustrate your steps.

ii) How will you determine outliers in the data ?

iii) List the various methods for data smoothing.

(OR)

b) i) Explain various methods of data cleaning in detail. **(8)**

ii) Write short notes on Data mining query language. **(8)**

13. a) State the Apriori algorithm and its merits and demerits. Consider the transaction database in Table 1. Determine the strong association rules generated using apriori technique with a minsup = 20% and minconf = 30%. Also prove that support is always ≤ confidence. **(16)**

Table 1

Transaction ID	Items bought
100	X, Y, Z
200	X, Z, W
400	Y, W
700	U, V, W
800	V, Y, Z
900	U, X, Z

(OR)

b) Consider a home finance loan to predict the housing loan prepayment. Design a general hierarchical structure and analyse the factors using rule discovery techniques to accurately predict the number of loan payments in a given quarter/year. Loan is availed for a period of 20 to 25 years, but an average life span of the loan exists for only 7 to 10 years due to prepayment.

Make necessary assumptions : Maintenance record of the customer details and details of the prevailing interest rates, borrower characteristics, account date, fine tune loan prepayment such as interest rates and fees in order to maximize the profits of the company. Elaborately discuss the association rule mining issues. Also discuss on the multilevel association rules and find if you could relate any relation from the above application. **(16)**

14. a) Consider five points $(X_1, X_2, X_3, X_4, X_5\}$ with the following coordinates as a two dimensional sample for clustering :
$X_1 = (0, 2)$; $X_2 = (0, 0)$; $X_3 = (1.5, 0.5)$ $X_4 = (5, 0.5)$; $X_5 = (5, 1.5)$.
Illustrate the K-means partitioning algorithm (clustering algorithm) using the above data set. **(16)**

(OR)

b) With an example explain Bayesian classification. **(16)**

15. a) i) List and explain the three types of dimensions in a spatial data cube with an example. **(8)**

ii) Explain with an example similarity search in multimedia data. **(8)**

(OR)

b) Discuss the following with examples.

i) Time series analysis. **(8)**

ii) Web mining. **(8)**

———————