Hall Ticket No | | | | | | | | | | | Question Paper Code: AIT006

INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

B.Tech VI Semester End Examinations (Regular) - May, 2019

Regulation: IARE – R16

DATA WAREHOUSING AND DATA MINING

Time: 3 Hours **(Common to CSE | IT)** **Max Marks: 70**

Answer ONE Question from each Unit
All Questions Carry Equal Marks
All parts of the question must be answered in one place only

UNIT – I

1. (a) Explain architecture of Data Warehouse in detail with the help of neat diagram. **[7M]**

   (b) Briefly compare the following concepts. You may use an example to explain your points.
   (i) Snowflake schema, fact constellation, star schema model **[7M]**
   (ii) Data cleaning, data transformation, refresh

2. (a) What is a Data Warehouse? Compare and contrast the operational database systems with Data Warehouses. **[7M]**

   (b) Suppose that a data warehouse consists of the four dimensions, date, spectator, location, and game, and the two measures, count and charge, where charge is the fare that a spectator pays when watching a game on a given date. Spectators may be students, adults, or seniors, with each category having its own charge rate. Draw a star schema diagram for the data warehouse. **[7M]**

UNIT – II

3. (a) Define data cleaning? Express the different techniques for handling missing values? **[7M]**

   (b) Suppose a hospital tested the age and body fat data for 10 randomly selected adults with the following result: **[7M]**

Table 1

Age	23	23	27	27	39	41	47	49	50	52
%fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2	34.6

Calculate the mean, median and standard deviation of age and %fat.

4. (a) Explain different methods used for constructing concept hierarchies for numerical attributes. **[7M]**

   (b) Suppose a group of 12 sales price records has been sorted as follows : 5, 10, 11, 13, 15, 35, 50, 55, 72, 92, 204, 215 Examine the following methods by partition them into three bins **[7M]**
   (i)equal-frequency (equi-depth) partitioning
   (ii)equal-width partitioning

UNIT – III

5. (a) Explain the differences between maximal frequent itemset and closed frequent itemset with an example. **[7M]**

   (b) Explain Association rule mining often generates a large number of rules. Discuss effective methods that can be used to reduce the number of rules generated while still preserving most of the interesting rules. **[7M]**

6. (a) What is Association rule Mining problem? State APRIORI principle? **[7M]**

   (b) Generate multi level association rules for the given data set in Table 2 using Apriori algorithm. Assume confidenceand support=50% **[7M]**

Table 2

TID	T001	T002	T003
List of items	Milk, dal, sugar, bread	Dal, sugar, wheat, jam	Milk, bread, curd, paneer

UNIT – IV

7. (a) Define Classification. Describe the issues regarding Classification. **[7M]**

   (b) Design an efficient method that performs effective naive Bayesian classification over an infinite data stream (i.e., you can scan the data stream only once). If we wanted to discover the evolution of such classification schemes (e.g., comparing the classification scheme at this moment with earlier schemes, such as one from a week ago), Construct modified design would you suggest? **[7M]**

8. (a) How is prediction different from classification? Explain how the classification is done using IF-THEN rules? Justify how Accuracy and Error measures are used for evaluating the classifier models **[7M]**

   (b) It is difficult to assess classification accuracy when individual data objects may belong to more than one class at a time. In such cases, Explain on what criteria you would use to compare different classifiers modeled after the same data. **[7M]**

UNIT – V

9. (a) Explain in detail the major requirements of clustering analysis and various applications of clustering. **[7M]**

   (b) Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36,8): **[7M]**
   Compute the Euclidean distance between the two objects.

10. (a) Write a short note on **[7M]**
    i.Multimedia data mining
    ii. Time series analysis

    (b) State K-means algorithm. Apply K-means algorithm with two iterations to form two clusters by taking the initial cluster centers as subjects 1 and 4 by considering Table 3. **[7M]**

Table 3

Subject	1	2	3	4	5	6	7
A	1.0	1.5	3.0	5.0	3.5	4.5	3.5
B	1.0	2.0	4.0	7.0	5.0	5.0	4.5

– ○ ○ ◯ ○ ○ –