Code No: **20AD4T02** **R20**

II B. TECH II SEMESTER REGULAR EXAMINATIONS, JUNE - 2022
DATA WAREHOUSING AND DATA MINING
(ARTIFICIAL INTELLIGENCE AND DATA SCIENCE)

Time: 3 hours Max. Marks: 70

Note: Answer **ONE** question from each unit **(5 × 14 = 70 Marks)**

UNIT-I

1. a) Describe three-tier data warehouse architecture in detail. [7M]
   b) Define similarity. Explain the commonly used similarity measures. [7M]

(OR)

2. a) Compare and contrast between multi-relational OLAP and multi- dimensional OLAP. [7M]
   b) What is visualization? Discuss, in brief, the different visualization techniques. [7M]

UNIT- II

3. a) Explain the binning methods for data smoothing. [7M]
   b) Write a note on subset selection in attributes for data reduction. [7M]

(OR)

4. a) What is the need of dimensionality reduction? Explain any two techniques for dimensionality reduction. [7M]
   b) Consider the following attribute *age* data (in increasing order): [7M]
   13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.
   Answer the following
   (i) Use z-score normalization to transform the value 35 for *age*, where the standard deviation of *age* is 12.94 years.
   (ii) Use normalization by decimal scaling to transform the value 35 for *age*.

UNIT-III

5. a) What is misclassification rate of a classifier? Describe sensitivity and specificity measures of a classifier. [7M]
   b) State the Bayes' theorem. Discuss how Bayesian classifiers work. [7M]

(OR)

6. a) Write down the criteria to compare and evaluate the classification and prediction methods. [7M]
   b) What is Bayesian Belief network? Explain the two main components in Bayesian belief network. [7M]

UNIT-IV

7. a) The Apriori algorithm makes use of *prior knowledge* of subset support properties. Prove that all nonempty subsets of a frequent itemset must also be frequent. [7M]

   b) Suppose that frequent item sets are saved for a large transactional database, *DB*. Discuss how to efficiently mine the (global) association rules under the same minimum support threshold, if a set of new transactions, denoted as 1*DB*, is *(incrementally) added in*? [7M]

(OR)

8. a) List all frequent item sets and strong association rules with support *'s'* and confidence *'c'* for the following transaction database. [7M]

   I1 = {T1,T4,T5,T7,T8,T9}, I2= {T1,T2,T3,T4,T6,T8,T9}
   I3 = {T3,T5,T6,T7.T8.T9} I4={T2,T4} I5={T1,T8}
   Assume T1 T2 … T9 are Transaction IDs and I1 I2 … I5 are items

   b) A *partitioning* variation of Apriori subdivides the transactions of a database *D* into *n* non-overlapping partitions. Prove that any item set that is frequent in *D* must be frequent in at least one partition of *D*. [7M]

UNIT-V

9. a) Describe the K-means clustering with an example. [7M]

   b) Write and explain the Agglomerative hierarchical clustering algorithm. [7M]

(OR)

10. a) Discuss the working procedure of bisecting K-means algorithm. [7M]

    b) Write the conditions under which density-based clustering is more suitable than partitioning-based clustering and hierarchical clustering. Give application examples to support your argument. [7M]

*** * * * ***