No. of Printed Pages : **2**

Sl. No.

BE6-R4 : DATA WAREHOUSING AND DATA MINING

NOTE :

1. **Answer question 1 and any FOUR from questions 2 to 7.**
2. **Parts of the same question should be answered together and in the same sequence.**

Time : 3 Hours **Total Marks : 100**

1. (a) Differentiate between Operational Database System and Data Warehouse.

(b) List any eight applications of Data Mining.

(c) Write the difference between supervised learning and unsupervised learning.

(d) Explain following terms :

Recall, Accuracy, Precision, F1-measure

(e) Which are the ways to handle missing values in data cleaning process ?

(f) What is Cuboid ? List the OLAP operations.

(g) Explain the following terms with respect to Density-based clustering algorithm :

(i) Core-distance

(ii) Reachability-distance **(7x4)**

2. (a) Discuss the KDD (Knowledge Discovery in Databases) process with appropriate diagram.

(b) What is data normalization ? Explain different methods of data normalization.

(c) Find mean, median, mode and range of following data :

8, 4, 5, 1, 5, 9, 4, 6, 7, 4, 2 **(8+6+4)**

3. (a) What is Apriori property ? Explain in detail with example.

(b) What is Multilevel Association Rules ? Explain with example. **(10+8)**

4. (a) Write steps to perform k-means clustering algorithm and list the weaknesses of k-means algorithm.

(b) What is outlier ? Which are methods to identify outlier ? Which are the applications of outlier detection ?

(c) Which are the types of OLAP servers ? Differentiate them. **(6+6+6)**

5. (a) Explain discretization and Concept Hierarchy Generation in brief.

(b) What is the difference between star and snow flake schemas ?

(c) Which are two types of hierarchical clustering methods ? Explain any one of them. **(5+7+6)**

6. (a) Discuss the major issues in data mining.

(b) What are time series databases and why are these important now-a-days ?

(c) Explain the major steps of decision tree classification technique. **(6+6+6)**

7. (a) What is Back propagation in Neural Networks ? How does it work ?

(b) Explain Naïve base classification algorithm with a suitable example.

(c) What is the purpose to use Data Cube Aggregation and Dimensionality Reduction strategies ? **(7+7+4)**

- o O o -