IV B.Tech I Semester Examinations- November 2017

**DATAWAREHOUSING AND DATAMINING**

Time: **3** hours (CSE/IT) Max. Marks: **60**

# SECTION – A

(Short Answer Questions)

**Answer all ten questions 10×1M=10M**

1. The Full form of KDD in Data mining is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

2. The \_\_\_\_\_\_\_\_\_\_\_\_\_ exposes the information being captured, stored, and managed by operational systems.

a) top-down view b) data source view c) data warehouse view d) business query view

3. \_\_\_\_\_\_\_\_\_\_\_\_\_\_is a random error or variance in a measured variable.

4. Which of the following is not a Data reduction Strategy?

a) Data cube aggregation b) Dimensionality reduction

c) Numerosity reduction d) Smoothing

5. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_is a process that abstracts a large set of task-relevant data in a database from a relatively low conceptual level to higher conceptual levels.

6. \_\_\_\_\_\_\_\_\_\_\_\_ is a method of mining frequent item sets without candidate generation.

7. \_\_\_\_\_\_\_\_\_ problem address the problem of Over-fitting the data

a) Back Propagation b) Tree Reduction c) Tree Pruning d) Tree Conjugation

8. \_\_\_\_\_\_is a neural network algorithm for classification that employs a method of gradient descent.

9. Euclidean distance can be defined as \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

10. \_\_\_\_\_\_\_\_\_\_ method hypothesizes a model for each of the clusters and finds the best fit of the data to that model.

a) Density-Based b) Grid Based c) Constrain Based d) Model Based.

**SECTION – B**

**Answer all five questions 5×2M= 10M**

11. List the Data Mining Functionalities.

12. Describe Multidimensional Data Model.

13. Define Data Visualization.

14. What is the Need of Data Transformation?

15. Define Classification and Prediction.

**SECTION – C**

**Answer all four questions 4×5M = 20M**

16. List and Explain the Major Issues in Data Mining.

**(OR)**

17. Describe how Data Mining techniques can be applied on Transactional Data Bases.

18. What are the Ways of handling missing values in Data Cleaning Process? Explain.

**(OR)**

19. Describe the Strategies used in data reduction technique.

20. Explain the Types of Data in Cluster Analysis.

**(OR)**

21. Illustrate K-Means Partitioning Algorithm.

22. With suitable example explain Density-Based Spatial Clustering of Applications with Noise clustering algorithm.

**(OR)**

23. Explain in detail about the Conceptual Clustering.

**SECTION – D**

**Answer all two questions 2×10M= 20M**

1. Design FP-growth method that mines the complete set of frequent item sets with a suitable example.

**(OR)**

1. With a suitable example explain Apriori algorithm for mining frequent item sets for Boolean association rules.
2. Illustrate how Bayesian classifier Works and explain the uses of Bayesian classifier.

**(OR)**

1. Write and explain the basic algorithm for inducing a decision tree from training tuples.