Code No: V3244 **R07** **Set No: 1**

III B.Tech. II Semester Supplementary Examinations, April/May - 2013

DATA WARE HOUSING AND DATA MINING

(Information Technology)

Time: 3 Hours **Max Marks: 80**

Answer any FIVE Questions

All Questions carry equal marks

*****

1. a) How is a data warehouse different from a database? How are they similar? [8M]
   b) Discuss major issues in Data Mining. [8M]

2. a) Differentiate between OLTP and OLAP. [8M]
   b) Summarize various features of Data ware houses. [8M]

3. a) What are interestingness measures. [8M]
   b) Differentiate between loose coupling and semi tight coupling [8M]

4. a) Discuss why analytical characterization is needed and how it can be performed? [8M]
   b) Outline a data cube-based incremental algorithm for mining analytical class comparison. [8M]

5. a) Explain the constraint based association mining. [8M]
   b) Write about distance association rules. [8M]

6. a) Explain the classification by decision tree induction. [8M]
   b) How can you find the accuracy of classifier? [8M]

7. a) Explain different types of data in cluster Analysis. [8M]
   b) What is meant by outlier Analysis? What are various methods for finding outlier analysis? [8M]

8. a) Explain three types of dimensions in a spatial data cube. [8M]
   b) What is meant by spatial classification? What are different approaches for similarity based retrieval in image databases. [8M]

*****

|''|'||'|''|''|''''|

Code No: V3244 **R07** **Set No: 2**

III B.Tech. II Semester Supplementary Examinations, April/May - 2013

DATA WARE HOUSING AND DATA MINING

(Information Technology)

Time: 3 Hours **Max Marks: 80**

Answer any FIVE Questions

All Questions carry equal marks

*****

1. a) In real world data tuples with missing values for some attributes are common occurrence. Describe various methods for handling this problem. [6M]
   b) Discuss issues to be considered during data integration. [4M]
   c) Discuss the principal components Analysis technique for Data Reduction. [6M]

2. a) What is multidimensional data model? How it is different from Databases? [8M]
   b) What are various schemas for multidimensional databases? [8M]

3. a) What are functional component of data mining (GUI)? [10M]
   b) Write the syntax for specifying the kind of knowledge to be mined. [6M]

4. a) Explain different class comparison methods. [8M]
   b) How can you measure the dispersion of data.? [8M]

5. a) What is meant by reduced support? [8M]
   b) Explain mining quantitative association rules. [8M]

6. a) What is Bayesian classifiers? Briefly outline the major ideas of Navie Bayesian classification. [8M]
   b) Write an algorithm for k-nearest neighbor classification given k and n, the number of attributes describing each sample. [8M]

7. a) What are various methods in density-based clustering. Explain. [8M]
   b) Write the k-medoids algorithm. [8M]

8. a) How can you implement spatial association mining? Explain [8M]
   b) What is meant by weblog mining? how it related to web link structures? [8M]

*****

|''|'||'|''|''|''''|

Code No: V3244 **R07** **Set No: 3**

III B.Tech. II Semester Supplementary Examinations, April/May - 2013

DATA WARE HOUSING AND DATA MINING

(Information Technology)

Time: 3 Hours **Max Marks: 80**

Answer any FIVE Questions

All Questions carry equal marks

*****

1. a) Explain the functionalities of Data Mining. [8M]
   b) Explain different Normalization techniques used in data transformation. [8M]

2. a) Explain the data warehouse Architecture with neat diagram [8M]
   b) Discuss various types of OLAP servers. [8M]

3. a) Explain the primitives for specifying a data mining task. [10M]
   b) What are four major types of concept hierarchies? [6M]

4. a) What is concept description .how it is related to class description? Justify. [8M]
   b) Write the algorithm for attribute –oriented induction. [8M]

5. a) What is market basket analysis? How is it used in association rule mining? [8M]
   b) How can you find frequent itemsets using candidate generation? [8M]

6. a) What are various methods for increasing classifier accuracy? Explain. [8M]
   b) Explain other classification methods. [8M]

7. a) Given two objects represented by the tuples (22,1,42,10) and (20,0,36,8) compute the Euclidean distance and manhattan distance between the two objects. [10M]
   b) What are the algorithms used in grid –based method? Explain [6M]

8. a) Differentiate text data mining and sequence data mining. [8M]
   b) Explain the concept of web usage mining [8M]

*****

Code No: V3244 **R07** **Set No: 4**

III B.Tech. II Semester Supplementary Examinations, April/May - 2013

DATA WARE HOUSING AND DATA MINING

(Information Technology)

Time: 3 Hours **Max Marks: 80**

Answer any FIVE Questions

All Questions carry equal marks

*****

1. a) Explain the steps for the process of knowledge discovery. [8M]
   b) Describe the methods for data cleaning. [8M]

2. a) Explain various OLAP operations in the multidimensional Data Model. [8M]
   b) Explain the architecture of three-tier data warehousing. [8M]

3. a) Explain different coupling schemas of data mining to be integrated with a DB/DW system. [12M]
   b) Write the syntax for interestingness measure specification. [4M]

4. a) What is the difference between attribute generalization threshold control and attribute generalization control? [8M]
   b) What is meant by class comparison explain? [8M]

5. a) How can the efficiency of Apriori be improved. [8M]
   b) How FP growth algorithm contributed for improvement of Apriori algorithm. [8M]

6. a) Write Back propagation algorithm [8M]
   b) What is meant by Tree Pruning. Differentiate between pre pruning and post pruning? [8M]

7. a) What is meant by clustering? Explain the partitioning methods with an example. [8M]
   b) How to compute the dissimilarity between objects of Asymmetric binary variables? [8M]

8. a) How can we study time-series data? What are the various components to characterize time-series data? [8M]
   b) Explain about spatial association analysis. [8M]

*****