

## III B. Tech II Semester Supplementary Examinations, November/December-2016

## DATA WARE HOUSING AND MINING

(Common to CSE and IT)

Time: 3 hours

Maximum Marks: 70

- Note: 1. Question Paper consists of two parts (**Part-A** and **Part-B**)  
 2. Answering the question in **Part-A** is compulsory  
 3. Answer any **THREE** Questions from **Part-B**

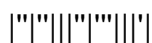
\*\*\*\*\*

**PART -A**

- 1 a) What is outlier mining? Define Data characterization. [3M]  
 b) Define data mining? Mention the steps in the data mining process? [4M]  
 c) What is clustering? What are the requirements of clustering? [4M]  
 d) Define Dimensional Modeling? List out its advantages. [4M]  
 e) Merits of Data Warehouse. What are the characteristics of Data Warehouse? [4M]  
 f) What is support and confidence? What is its purpose in association mining? [3M]

**PART -B**

- 2 a) How is a *data warehouse* different from a database? How are they similar? [5M]  
 b) Explain the OLAP operations in multidimensional model? [7M]  
 c) Discuss the components of Data warehouse? [4M]
- 3 a) How might you determine *outliers* in the data? What other methods are there for *data smoothing*? [5M]  
 b) List out and describe the primitives for specifying a data mining task. [6M]  
 c) i) What are the value ranges of the following *normalization methods*? [5M]  
     (a) min-max normalization  
     (b) z-score normalization  
     (c) normalization by decimal scaling  
 ii) Use the two methods below to *normalize* the following group of data:  
 200; 300; 400; 600; 1000  
     (a) min-max normalization by setting  $min = 0$  and  $max = 1$   
     (b) z-score normalization
- 4 a) Briefly discuss about data mining task premitives. [8M]  
 b) What is data mining? Draw and explain the architecture of a typical data mining system? [8M]
- 5 a) Compare the Advantages and Disadvantages of *Eager Classification* (e.g., decision tree, Bayesian, neural network) versus *Lazy Classification* (e.g., k-nearest neighbor, case-based reasoning). [9M]  
 b) Explain the issues regarding Classification and Prediction? [7M]



**III B. Tech II Semester Regular Examinations, April - 2016**  
**DATA WARE HOUSING AND MINING**  
 (Common to CSE and IT)

Time: 3 hours

Max. Marks: 70

- Note: 1. Question Paper consists of two parts (**Part-A** and **Part-B**)  
 2. Answering the question in **Part-A** is compulsory  
 3. Answer any **THREE** Questions from **Part-B**

\*\*\*\*\*

**PART -A**

- 1 a) What is a Data warehouse? Briefly describe the need for data warehousing. [4M]  
 b) What is Data integration? [4M]  
 c) Describe different types of OLAP servers. [4M]  
 d) Describe random sub sampling. [3M]  
 e) Define a frequent set. [3M]  
 f) Describe different types of clustering. [4M]

**PART -B**

- 2 a) What is data mining? Briefly explain the Knowledge discovery process. [8M]  
 b) Discuss about Data Mining Task Primitives. [8M]  
 3 With examples, discuss in detail about the available techniques for concept hierarchy generation for categorical data. [16M]  
 4 a) Explain the three-tier data warehouse architecture. [8M]  
 b) What is a concept hierarchy? Describe the OLAP operations in the Multidimensional data model. [8M]  
 5 a) Why pruning is useful in decision tree induction? What is a separate set of tuples to evaluate pruning? [8M]  
 b) Why naive Bayesian classification is called naïve? Briefly outline the major ideas of naive Bayesian classification. [8M]  
 6 a) Explain difference between partitions based Apriori and Apriori algorithm. [8M]  
 b) Write an algorithm for finding frequent item-sets using candidate generation. [8M]  
 7 With a suitable example, explain K-Means Clustering algorithm. [16M]

**III B. Tech II Semester Regular Examinations, April - 2016**  
**DATA WARE HOUSING AND MINING**  
 (Common to CSE and IT)

Time: 3 hours

Max. Marks: 70

- Note: 1. Question Paper consists of two parts (**Part-A** and **Part-B**)  
 2. Answering the question in **Part-A** is compulsory  
 3. Answer any **THREE** Questions from **Part-B**  
 \*\*\*\*\*

**PART -A**

- |   |   |      |
|---|---|------|
| 1 | a) Describe about transactional database. | [4M] |
|   | b) What is Data cleaning?                 | [4M] |
|   | c) What is a data ware house?             | [4M] |
|   | d) Describe holdout method.               | [3M] |
|   | e) Define a FP-tree.                      | [4M] |
|   | f) What is Cluster Analysis?              | [3M] |

**PART -B**

- |   |  |       |
|---|--|-------|
| 2 | a) Explain data mining as a step in the process of knowledge discovery.  | [8M]  |
|   | b) What are the major issues in Data Mining? Explain.  | [8M]  |
| 3 | a) What is Data integration? What is entity identification problem and why it is useful?   | [3M]  |
|   | b) What is lossless and lossy dimensionality reduction? Describe any one technique for lossy dimensionality reduction.   | [8M]  |
| 4 | a) Differentiate between operational data base system and data warehouses.   | [8M]  |
|   | b) What is a concept hierarchy? Describe the OLAP operations in the Multidimensional data model.   | [8M]  |
| 5 | a) Explain the classification by decision tree induction with an example.  | [8M]  |
|   | b) Explain the following accuracy measures:<br>(a) F-measure (b) Confusion matrix (c) Cross-validation (d) Bootstrap   | [8M]  |
| 6 | a) The price of each item in a store is non-negative. For each of the following cases, identify the kind of constraint they represent and briefly discuss how to mine such association values efficiently<br>a) containing at least one Nintendo game,<br>b) containing items the sum of whose price is less than \$150. | [8M]  |
|   | b) Explain frequent item sets without candidate generation.  | [8M]  |
| 7 | Explain about K-means algorithm with suitable example.   | [16M] |

-----

**III B. Tech II Semester Regular Examinations, April - 2016**  
**DATA WARE HOUSING AND MINING**  
 (Common to CSE and IT)

Time: 3 hours

Max. Marks: 70

Note: 1. Question Paper consists of two parts (**Part-A** and **Part-B**)  
 2. Answering the question in **Part-A** is compulsory  
 3. Answer any **THREE** Questions from **Part-B**

\*\*\*\*\*

**PART -A**

- |   |  |      |
|---|--|------|
| 1 | a) Describe about object-relational databases.                   | [3M] |
|   | b) What is data reduction? What is dimensionality reduction?     | [4M] |
|   | c) Describe snowflake and fact constellations.                   | [4M] |
|   | d) What is Classification? Describe the need for classification. | [4M] |
|   | e) Define a FP-tree.   | [3M] |
|   | f) Write a note on Hierarchical clustering.                      | [4M] |

**PART -B**

- |   |  |      |
|---|--|------|
| 2 | a) Describe different data mining functionalities.   | [8M] |
|   | b) Draw and explain the architecture of a typical data mining system.                              | [8M] |
| 3 | a) What is noisy data? Explain the binning methods for data smoothening.                           | [8M] |
|   | b) What is data integration? Discuss the issues to be considered for data integration.             | [8M] |
| 4 | a) Differentiate OLTP and OLAP.  | [8M] |
|   | b) Explain the three-tier data warehouse architecture.   | [8M] |
| 5 | a) What is Eager classification and Lazy classification? Write their advantages and disadvantages. | [8M] |
|   | b) Explain the issues regarding classification and prediction.                                     | [8M] |
| 6 | a) Explain difference between partitions based Apriori and Apriori algorithm.                      | [8M] |
|   | b) Write an algorithm for finding frequent item-sets using candidate generation.                   | [8M] |
| 7 | a) What is density based clustering? Describe DBSCAN clustering algorithm.                         | [8M] |
|   | b) What is partitioning method? Describe any one partition based clustering algorithm.             | [8M] |

-----

**III B. Tech II Semester Regular Examinations, April - 2016**  
**DATA WARE HOUSING AND MINING**  
(Common to CSE and IT)

Time: 3 hours

Max. Marks: 70

Note: 1. Question Paper consists of two parts (**Part-A** and **Part-B**)  
2. Answering the question in **Part-A** is compulsory  
3. Answer any **THREE** Questions from **Part-B**  
\*\*\*\*\*

**PART -A**

- 1 a) Describe heterogeneous and legacy databases. [4M]
- b) Describe how correlation coefficient is computed? [3M]
- c) What is a Data warehouse? [4M]
- d) What is Classification? Describe the need for classification. [4M]
- e) Define Apriori property. [3M]
- f) Distinguish between classification and clustering. [4M]

**PART -B**

- 2 a) What are the major issues in Data Mining? Explain. [8M]
- b) Draw and explain the architecture of a typical data mining system. [8M]
- 3 a) What is data cleaning? Describe the approaches to fill missing values. [8M]
- b) Briefly describe various forms of data pre-processing. [8M]
- 4 Briefly discuss about the following data warehouse implementation methods: [16M]  
(a) Indexing OLAP data (b) Metadata Repository.
- 5 a) Describe the criteria used to evaluate classification and prediction methods. [8M]
- b) Explain the following accuracy measures: [8M]  
(i) F-measure (ii) Confusion matrix (iii) Cross-validation (iv) Bootstrap
- 6 a) Briefly explain about FP- growth algorithm. Write its advantages over other mining algorithms. [8M]
- b) Write an algorithm for finding frequent item-sets using candidate generation. [8M]
- 7 a) What is clustering analysis? Give the different types of clustering techniques. [4M]
- b) Consider five points  $\{X_1, X_2, X_3, X_4, X_5\}$  with the following coordinates as a two dimensional sample for clustering : [12M]  
 $X_1 = (0.5, 2.5)$ ;  $X_2 = (0, 0)$ ;  $X_3 = (1.5, 1)$ ;  $X_4 = (5, 1)$ ;  $X_5 = (6, 2)$ ;  
Illustrate the K-means partitioning algorithms using the above data set.

-----

Code No: **R41053**

**R10**

**Set No. 1**

**IV B.Tech I Semester Supplementary Examinations, March/April - 2016**

**DATA WAREHOUSING AND DATA MINING**

**(Computer Science Engineering)**

**Time: 3 hours**

**Max. Marks: 75**

**Answer any FIVE Questions  
All Questions carry equal marks**

**\*\*\*\*\***

- 1 a) What is Data Mining? How it is different from statistical methods? [8]  
b) What are different challenges of data mining? Explore the different types of data attributes. [7]
- 2 a) What are different data pre-processing techniques? Explain. [8]  
b) What are the approaches of feature subset selection? [7]
- 3 a) Differentiate between data base and data warehouse. [8]  
b) What meant indexing OLAP data? [7]
- 4 Write the algorithm for decision tree induction. [15]
- 5 a) Describe the Bayesian classifier. [8]  
b) What is Naïve Bayes classification? [7]
- 6 a) What is association rule mining? How to calculate support and confidence? [8]  
b) What is frequent Item set? Explain the Frequent-growth algorithm. [7]
- 7 a) What are advantage and applications of clustering methods? [8]  
b) Explain K-means and different types of clusters. [7]
- 8 a) Explain the hierarchical clustering algorithm. [8]  
b) What is DBSCAN? Explain its application. [7]

Code No: **R41053**

**R10**

**Set No. 1**

**IV B.Tech I Semester Regular/Supplementary Examinations, Nov/Dec - 2015**

**DATA WAREHOUSING AND DATA MINING**

**(Common to Computer Science & Engineering and Information Technology)**

**Time: 3 hours**

**Max. Marks: 75**

**Answer any FIVE Questions  
All Questions carry equal marks**

\*\*\*\*\*

- 1 a) What are the origins of data mining? Explain with the help of a diagram. [7]  
b) Explain the steps in the data mining process giving example for each step. [8]
- 2 a) Define sampling. What are different types of sampling? [7]  
b) Explain how Principal component Analysis is used for dimensionality reduction. [8]
- 3 a) Explain the star schema of data ware house model with the help diagram. [8]  
b) Differentiate between data ware house and operational database. [7]
- 4 a) Explain how cross validation is useful in classifiers of data mining. [8]  
b) Explain the bayes theorem. [7]
- 5 a) What is classification? Explain decision tree classifier with the help of diagram. [8]  
b) What are the methods for expressing attribute test conditions. [7]
- 6 a) What is support counting? How it is done with hash trees? [8]  
b) Explain pruning?. [7]
- 7 a) What are the advantages of k-means algorithm? [7]  
b) What are the additional issues of k-means algorithm? [8]
- 8 a) Explain DIANA clustering algorithm. [8]  
b) Explain BIRCH clustering algorithm. [7]

Code No: **R41053**

**R10**

**Set No. 2**

**IV B.Tech I Semester Regular/Supplementary Examinations, Nov/Dec - 2015**  
**DATA WAREHOUSING AND DATA MINING**  
**(Common to Computer Science & Engineering and Information Technology)**

**Time: 3 hours**

**Max. Marks: 75**

**Answer any FIVE Questions**  
**All Questions carry equal marks**

\*\*\*\*\*

- 1 a) Explain various data mining tasks giving example for each. [8]  
b) What is the need for preprocessing? List various preprocessing techniques. [7]
- 2 a) What is correlation? Derive an expression for correlation. [8]  
b) Explain how weights are useful for combining similarities. [7]
- 3 a) Explain fact constellation schema with the help of diagram. [8]  
b) Explain partial materialization in data cube implementation. [7]
- 4 a) What are split points and how to find out correct split points? Explain with example. [8]  
b) Explain the criterion for finding the correct tree size. [7]
- 5 a) Explain the nearest neighbor algorithm for classification. [8]  
b) Explain the parameters for the evaluation of classifier. [7]
- 6 a) What are different types of association rules? Give examples. [7]  
b) Explain FP growth algorithm for the generation of frequent item sets. [8]
- 7 a) List different types of clustering approaches. Mentioning example for each. [7]  
b) Explain the k-means algorithm for distance based clustering. [8]
- 8 a) What are the advantages of hierarchical clustering approach? [7]  
b) Explain the basic agglomerative clustering algorithm. [8]



Code No: **R41053**

**R10**

**Set No. 3**

**IV B.Tech I Semester Regular/Supplementary Examinations, Nov/Dec - 2015**

**DATA WAREHOUSING AND DATA MINING**

**(Common to Computer Science & Engineering and Information Technology)**

**Time: 3 hours**

**Max. Marks: 75**

**Answer any FIVE Questions  
All Questions carry equal marks**

**\*\*\*\*\***

- 1 a) What are the challenges in data mining that motivate the mining tasks? [8]  
b) Enumerate the applications of data mining. [7]
- 2 a) Explain similarity between binary vectors. [7]  
b) What is Jaccard coefficient? Derive an expression for Jaccard coefficient. [8]
- 3 a) Draw a data cube and explain OLAP operations. [8]  
b) Explain effective cube computation. [7]
- 4 a) Define the terms Entropy, information gain and gini index. How they are useful for attribute selection? [9]  
b) Explain decision tree induction. [6]
- 5 a) Explain the of naïve Bayesian classifier. [8]  
b) Draw a Bayesian belief network. [7]
- 6 a) Explain any association mining algorithm without generating candidate item sets. [8]  
b) Define support and confidence. What is support threshold? [7]
- 7 a) What are the strengths and weaknesses of k-means clustering? [7]  
b) Justify how K- mediods and PAM clustering algorithms can replace k-means algorithm to overcome its limitations. [8]
- 8 a) What is the basic principle of density based clusteing? How it is advantageous over others? [8]  
b) What are the strengths and weaknesses of DBSCAN? [7]

Code No: **R41053**

**R10**

**Set No. 4**

**IV B.Tech I Semester Regular/Supplementary Examinations, Nov/Dec - 2015**  
**DATA WAREHOUSING AND DATA MINING**  
**(Common to Computer Science & Engineering and Information Technology)**

**Time: 3 hours**

**Max. Marks: 75**

**Answer any FIVE Questions**  
**All Questions carry equal marks**

\*\*\*\*\*

- 1 a) Explain different types of data attributes and their measurements. [8]  
b) Explain different types of datasets giving example for each. [7]
- 2 a) Define similarity and dissimilarity of attributes. [5]  
b) Explain Euclidian distance, Minkowski distance and Mahalanobis distance. [10]
- 3 a) List OLAP operations and explain each with examples. [8]  
b) Explain the 3- tier data ware house architecture with the help of diagram. [7]
- 4 a) Explain the decision tree classifier with the help of a diagram. [8]  
b) Explain overfitting and tree pruning. [7]
- 5 a) Explain Bayes theorem. How it is useful for classification in data mining? [8]  
b) What is Bayes error rate? [7]
- 6 a) What is Apriori property? Explain Apriori algorithm with an example. [8]  
b) What is frequent item set generation? What are candidate itemsets? [7]
- 7 a) Explain a basic k-means algorithm with example. [8]  
b) What are the applications of clustering? [7]
- 8 a) Explain the basic hierarchical clustering strategies, AGNES and DIANA. [7]  
b) Explain DB SCAN clustering algorithm in detail. [8]

Code No:R41053

**R10**

**Set No. 1**

**IV B.Tech I Semester Supplementary Examinations, Feb/Mar - 2015**  
**DATA WAREHOUSING AND DATA MINING**  
(Common to Computer Science & Engineering and Information Technology)

**Time: 3 hours**

**Max. Marks: 75**

**Answer any FIVE Questions**  
**All Questions carry equal marks**

\*\*\*\*\*

- 1 a) What are data mining tasks? Explain. [8]  
b) What are various types of datasets? Explain. [7]
- 2 a) Given a similarity measure with values in the interval  $[0,1]$  describe two ways to transform this similarity value into a dissimilarity value in the interval  $[0,\infty]$ . [6]  
b) What is data preprocessing? Explain in detail. [9]
- 3 a) What is a data warehouse? Distinguish between operational database system and data warehouses. [8]  
b) Describe efficient processing of OLAP queries. [7]
- 4 Explain the following:  
a) Algorithm for decision tree induction [5]  
b) Characteristics of decision tree induction [5]  
c) Evaluating the performance of a classifier [5]
- 5 Write Bayes theorem. Explain classification by using the Bayes theorem. [15]
- 6 a) Describe rule generation in Apriori algorithm with example. [8]  
b) Discuss about computational complexity of frequent itemset generation. [7]
- 7 a) What are different types of clusters? Explain. [7]  
b) Describe strengths and weaknesses of K-means. [8]
- 8 Explain the following w.r.t agglomerative hierarchical clustering:  
a) Specific techniques [12]  
b) Strengths and weaknesses [3]

Code No:R41053

**R10**

**Set No. 2**

**IV B.Tech I Semester Supplementary Examinations, Feb/Mar - 2015**  
**DATA WAREHOUSING AND DATA MINING**  
**(Common to Computer Science & Engineering and Information Technology)**

**Time: 3 hours**

**Max. Marks: 75**

**Answer any FIVE Questions**  
**All Questions carry equal marks**

\*\*\*\*\*

- 1 a) Describe measurement and data collection issues of data quality. [8]  
b) What is Data Mining? Explain its motivating challenges. [7]
- 2 a) Discuss about feature subset selection and feature creation. [8]  
b) What are summary statistics? Explain. [7]
- 3 a) Discuss about efficient computation of data cubes. [7]  
b) Draw a three-tier data warehouse architecture and explain. [8]
- 4 a) What are the methods for expressing attribute test conditions? Explain. [6]  
b) Explain the following:  
(i) Holdout method  
(ii) Handling overfitting in decision tree induction [8]
- 5 Describe Naïve bayes classifier with examples. [15]
- 6 a) Discuss about frequent itemset generation in FP-Growth algorithm. [8]  
b) Explain frequent itemset generation in the Apriori algorithm. [7]
- 7 a) Write the basic K-means algorithm and explain. [10]  
b) What are different types of clustering? Explain. [5]
- 8 Explain the following w.r.t DBSCAN:  
a) Center-based approach [5]  
b) Algorithm [5]  
c) Strengths and weaknesses [5]

Code No:R41053

**R10**

**Set No. 3**

**IV B.Tech I Semester Supplementary Examinations, Feb/Mar - 2015**  
**DATA WAREHOUSING AND DATA MINING**  
(Common to Computer Science & Engineering and Information Technology)

**Time: 3 hours**

**Max. Marks: 75**

**Answer any FIVE Questions**  
**All Questions carry equal marks**

\*\*\*\*\*

- 1 a) What are the core data mining tasks? Explain with examples. [8]
- b) What is an attribute? What are the different types of attributes? [7]
- 2 Discuss in detail about measures of similarity and dissimilarity. [15]
- 3 a) Describe schemes for multidimensional databases. [8]
- b) How to index OLAP data? Explain. [7]
- 4 a) Explain about evaluating the performance of a classifier. [8]
- b) What are the measures for selecting the Best split? Explain. [7]
- 5 a) Give an example of inferencing using BBN. [8]
- b) Describe Bayes error rate with example. [7]
- 6 a) How to represent FP-Tree? Explain. [7]
- b) Discuss about problem definition of association analysis. [8]
- 7 a) What is cluster analysis? Explain additional issues of K-means. [7]
- b) Describe K-means as an optimization problem. [8]
- 8 a) Write the DBSCAN algorithm and explain. [8]
- b) What are the key issues in hierarchical clustering? [7]

Code No:R41053

**R10**

**Set No. 4**

**IV B.Tech I Semester Supplementary Examinations, Feb/Mar - 2015**  
**DATA WAREHOUSING AND DATA MINING**  
**(Common to Computer Science & Engineering and Information Technology)**

**Time: 3 hours**

**Max. Marks: 75**

**Answer any FIVE Questions**  
**All Questions carry equal marks**

\*\*\*\*\*

- 1 a) Define KDD. Explain data mining tasks with example. [8]  
b) Discuss about attributes and measurement. [7]
- 2 a) Define data preprocessing. Explain about Binarization. [7]  
b) Give examples of Proximity measures. [8]
- 3 a) Briefly discuss about data warehouse implementation. [8]  
b) What are OLAP operations in the multidimensional data model? Explain. [7]
- 4 a) How to build a decision tree? Explain. [7]  
b) Discuss about estimation of generalization errors in detail. [8]
- 5 Explain the following w.r.t BBN:  
a) Model representation [8]  
b) Model building [7]
- 6 a) Describe compact representation of frequent itemsets. [8]  
b) Explain about support counting in frequent itemset generation. [7]
- 7 a) Give an overview of cluster analysis. [7]  
b) Explain the following w.r.t K-means:  
(i) Additional issues  
(ii) Strengths and weaknesses [8]
- 8 Discuss in detail about agglomerative hierarchical clustering. [15]

Code No: **R41053**

**R10**

**Set No. 1**

**IV B.Tech I Semester Regular/Supplementary Examinations, Nov/Dec - 2014**  
**DATA WAREHOUSING AND DATA MINING**  
**(Common to Computer Science & Engineering and Information Technology)**

**Time: 3 hours**

**Max. Marks: 75**

**Answer any FIVE Questions**  
**All Questions carry equal marks**

\*\*\*\*\*

- 1 a) What is data mining? List and describe the motivating challenges of data mining. [8]  
b) What is an attribute? Explain different types of attributes with examples. [7]
- 2 a) Define sampling. Mention the different types of approaches for sampling and explain each in detail. [8]  
b) Explain the feature subset of selection with architecture. [7]
- 3 a) Write the differences between operational database systems and data warehouses. [8]  
b) Illustrate and explain the OLAP architecture. [7]
- 4 a) Describe the general approach to solving a classification problem with a suitable example. [8]  
b) List and explain the design issues of decision tree induction. [7]
- 5 a) What are the characteristics of rule based classifier? Explain. [8]  
b) Enumerate on the model representation of Bayesian Belief Networks. [7]
- 6 a) With an example, explain the frequent item set generation in the Apriori algorithm. [8]  
b) Explain the principles of Apriori algorithm. [7]
- 7 a) Write K- Means clustering Algorithm and explain with an example. [10]  
b) Make a comparison of complete and partial clusters. [5]
- 8 a) Derive the time and space complexity of basic agglomerative hierarchical clustering algorithm. [8]  
b) Explain agglomerative hierarchical clustering. [7]

Code No: R41053

**R10**

**Set No. 2**

**IV B.Tech I Semester Regular/Supplementary Examinations, Nov/Dec - 2014**

**DATA WAREHOUSING AND DATA MINING**

**(Common to Computer Science & Engineering and Information Technology)**

**Time: 3 hours**

**Max. Marks: 75**

**Answer any FIVE Questions  
All Questions carry equal marks**

\*\*\*\*\*

- 1 a) Explain the different types of attributes. [8]  
b) Elaborate on the origins of data mining with neat diagram. [7]
- 2 a) What are the similarities and dissimilarities between simple attributes? Explain. [10]  
b) Distinguish between mean and median. Give examples. [5]
- 3 a) Draw and explain the three tier data ware house architecture. [8]  
b) How efficient processing of OLAP queries is done? Give example. [7]
- 4 a) What are the characteristics of decision tree induction? Explain. [8]  
b) Write an algorithm for decision tree induction. [7]
- 5 a) Describe Naïve Bayes classifier with one example. [8]  
b) Give a brief note on Bayes theorem. [7]
- 6 a) Explain the candidate generation and pruning. [8]  
b) Discuss about support counting using Hash Tree. [7]
- 7 a) Explain the K- means clustering algorithm and different types of clusters in detail. [10]  
b) What are the issues in K- Means? [5]
- 8 a) List and describe the specific techniques used in the agglomerative hierarchical clustering. [8]  
b) Write a short note on DBSCAN. [7]



Code No: R41053

**R10**

**Set No. 3**

**IV B.Tech I Semester Regular/Supplementary Examinations, Nov/Dec - 2014**

**DATA WAREHOUSING AND DATA MINING**

**(Common to Computer Science & Engineering and Information Technology)**

**Time: 3 hours**

**Max. Marks: 75**

**Answer any FIVE Questions  
All Questions carry equal marks**

\*\*\*\*\*

- 1 a) Explain data mining tasks with one example for each. [8]  
b) What are the types of data? Explain each with an example. [7]
- 2 a) Explain why computing the proximity between two attributes is often simpler than computing the similarity between two objects. [8]  
b) What is meant by data set? Discuss in detail. [7]
- 3 a) Describe the indexing OLAP data. [8]  
b) Write short notes on efficient computation of data cubes. [7]
- 4 a) Briefly explain the model over fitting. [8]  
b) List and explain the measures for selecting the best split. [7]
- 5 a) Mention the characteristics of Bayesian Belief Networks. Explain. [8]  
b) Give a note on rule growing strategy with a neat diagram. [7]
- 6 a) Write an algorithm for the frequent item set generation using Apriori algorithm. [8]  
b) With an example, explain the FP- growth algorithm. [7]
- 7 a) Mention the strengths and weaknesses of K-means clustering algorithm. [7]  
b) What are outliers? How are they detected? Explain any one technique in detail. [8]
- 8 a) Explain how to the select DBSCAN Parameters. [8]  
b) Describe the center based approach for DBSCAN. [7]

Code No: R41053

**R10**

**Set No. 4**

**IV B.Tech I Semester Regular/Supplementary Examinations, Nov/Dec - 2014**  
**DATA WAREHOUSING AND DATA MINING**  
**(Common to Computer Science & Engineering and Information Technology)**

**Time: 3 hours**

**Max. Marks: 75**

**Answer any FIVE Questions**  
**All Questions carry equal marks**

\*\*\*\*\*

- 1 a) Explain the general characteristics of data sets. [8]  
b) Define data quality? Discuss the measurements and data collection issues. [7]
- 2 a) Write short notes on discretization and binarization. [8]  
b) What are the issues in proximity calculation? Explain. [7]
- 3 a) Define data warehouse. Write the characteristics of data warehouse. [5]  
b) List and explain various OLAP operations with suitable examples [10]
- 4 a) Write notes on evaluating the performance of a classifier. [8]  
b) Explain the normal attributes and ordinal attributes with neat diagrams. [7]
- 5 a) Make a comparison of rule based and class based ordering schemes. [8]  
b) Briefly explain the Bayesian Belief Networks. [7]
- 6 a) Enumerate on the FP – Tree representations with neat diagrams. [8]  
b) With an example, explain the frequent item set generation in FP- Growth algorithm. [7]
- 7 a) Define cluster analysis? What are the different types of clusters? Explain. [8]  
b) What are the differences between hierarchical and partitional clusters? [7]
- 8 a) Write DBSCAN algorithm and derive its time and space complexity. [10]  
b) What are the strengths and weaknesses of DBSCAN? [5]

Code No: **R41053****R10****Set No. 1**

**IV B.Tech I Semester Supplementary Examinations, May/June 2014**  
**DATA WAREHOUSING AND DATA MINING**  
(Common to Computer Science and Engineering and Information Technology)

**Time : 3 hours****Max. Marks: 75**

**Answer any Five Questions**  
**All Questions carry equal marks**

\*\*\*\*\*

1. a) What is Data Mining? Briefly describe the primitives of the data mining.  
b) What are the missing values? Briefly describe any three strategies for dealing with missing data.
2. What is similarity and what is dissimilarity? For a sample numerical and categorical data, briefly describe the computation of similarity and dissimilarity.
3. a) What is OLAP? Briefly describe the indexing of OLAP data.  
b) What is data cube? Describe in detail about data cube materialization.
4. What is decision tree? What is best split? Why one need the best split? Briefly describe the measures for best split.
5. What is a classifier? Why one need a classification technique? What is Bayes theorem? Briefly describe the usage of Bayes theorem for performing classification.
6. What is Apriori principle? For an example data, briefly describe the generation of frequent item set using Apriori algorithm.
7. What is bisecting k-means? In what way it is different from basic k-means algorithm? With a neat diagram, briefly describe the usage of bisecting k-means algorithm and its initialization.
8. What is proximity? For an example data, briefly describe the process of computing *group average* clustering.

Code No: **R41053****R10****Set No. 2**

**IV B.Tech I Semester Supplementary Examinations, May/June 2014**  
**DATA WAREHOUSING AND DATA MINING**  
(Common to Computer Science and Engineering and Information Technology)

**Time : 3 hours****Max. Marks: 75**

**Answer any Five Questions**  
**All Questions carry equal marks**

\*\*\*\*\*

1. What are the origins of data mining? Describe in detail about data mining tasks.
2. What is binary data? With an example, briefly describe the usage of Jaccard and Cosine measures to identify similarity between data objects.
3. a) What is query driven approach? Briefly describe the need for data warehousing.  
b) What is a Data warehouse? Briefly describe the required three strategies to construct the data warehouse.
4. What is model over fitting? Briefly describe the reasons for model over fitting? List the available solutions for solving the over fitting.
5. What are Bayesian classifiers? With an example, describe how to predict a class label using naive Bayesian classification.
6. What is pruning? Why pruning is require? With an example, briefly describe FP-Growth algorithm.
7. What is a cluster? For a sample data and using appropriate dissimilarity measure, briefly describe the creation of clusters using basic k-means clustering algorithm.
8. What is CLIQUE? For an example data, briefly describe the process of computing *complete link* clustering.

Code No: **R41053****R10****Set No. 3****IV B.Tech I Semester Supplementary Examinations, May/June 2014****DATA WAREHOUSING AND DATA MINING****(Common to Computer Science and Engineering and Information Technology)****Time : 3 hours****Max. Marks: 75****Answer any Five Questions  
All Questions carry equal marks**

\*\*\*\*\*

1. a) What is an attribute? What are the hierarchies of attributes (in the form of complexity)? Briefly describe the types of attributes.  
b) What is measure and what is metric? Briefly describe the required criteria for a measure.
2. What is data set? For an example data(set), describe the process of identifying any five summary statistics.
3. a) What are the four keywords of the data ware house? Briefly describe the four key words.  
b) What is OLAP? How organizations are utilizing the data warehouse.
4. a) What is classification? For an example data, briefly describe the decision tree induction.  
b) Briefly describe the methods for expressing attribute test conditions. Why these tests require?
5. What is the performance of Bayesian classifier w.r.t decision tree when applied to large databases? With an example, briefly describe the model building of Bayesian Belief Networks.
6. What is support counting? Why one need support counting? With an example, briefly describe compact representation of frequent item sets.
7. a) What is partitioning method? Describe k-means clustering algorithm  
b) With an example, briefly describe process of viewing k-means as an optimization problem.
8. What is DBSCAN? What is the time and space complexity of DBSCAN algorithm? Briefly describe the process for selection/computing DBSCAN parameters.

Code No: **R41053****R10****Set No. 4**

IV B.Tech I Semester Supplementary Examinations, May/June 2014

**DATA WAREHOUSING AND DATA MINING**

(Common to Computer Science and Engineering and Information Technology)

**Time : 3 hours****Max. Marks: 75**

**Answer any Five Questions**  
**All Questions carry equal marks**

\*\*\*\*\*

1. What is a data set? What is artifacts? Briefly describe the available measures for measuring the quality of the measurement process.
2. What is Correlation? With an example, briefly describe the process of identifying the correlation between data objects.
3. a) What is a Data warehouse? Briefly describe the need for data warehousing.  
b) What is update driven approach? Briefly describe the four key words of the data warehouse.
4. What is a classifier? Why we need a classifier? What are the measures available for evaluating the performance of classifier? For an example data(set), briefly describe the usage of cross – validation.
5. a) What is class conditional independence? With an example, describe in detail Naïve Bayesian classifier.  
b) What is Bayes error rate? For an example data, briefly describe the usage of Bayesian belief networks.
6. What is association analysis? With an example situation, briefly describe the usage of Association analysis in increasing company Return of Investment.
7. What is a cluster? Mention one real time situation where clustering requires? Briefly describe the strength and weaknesses of k-means clustering algorithm.
8. a) What is UPGMA? Briefly describe the key issues in hierarchical clustering.  
b) For an example data, briefly describe the process of defining proximity between clusters.

1 of 1

Code No: R41053

**R10**

**Set No. 1**

**IV B.Tech I Semester Regular Examinations, December 2013**  
**DATA WARE HOUSING AND DATA MINING**  
**(Common to Computer Science & Engineering and Information Technology)**

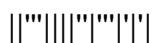
**Time : 3 hours**

**Max. Marks: 75**

**Answer any Five Questions**  
**All Questions carry equal marks**

\*\*\*\*\*

- 1 a) Describe about the motivating challenges for the development of Data Mining [8]  
b) Explain different types of qualitative attributes and quantitative attributes with examples. [7]
- 2 a) What are the examples of proximity measures? Explain [8]  
b) Explain how, simple matching coefficient and Jaccard coefficient are used to find similarities? [7]
- 3 a) Explain the three tier Data Warehouse architecture with neat diagram [8]  
b) What is Multi Dimensional Model? Explain the different schemes in detail. [7]
- 4 a) Mention different characteristics to construct Decision tree. [8]  
b) What is meant by Classification? What are applications of classification Model? [7]
- 5 a) What is meant by Bayesian Classifier? How Bayes theorem is used for classification? [8]  
b) Compare the two methods for estimating probabilities. Which method is better and why? [7]
- 6 a) Write effective candidate generation procedure in detail. [8]  
b) Write the FP-Growth algorithm. [7]
- 7 a) Write the algorithm of bisecting K-means and how it is different from simple K-means? [8]  
b) Explain k-means as an optimization problem in detail. [7]
- 8 a) What is meant by Agglomerative Hierarchical Clustering? How they are different from Density based Clustering. [8]  
b) Describe about strengths and weaknesses of traditional density approach. [7]



Code No: R41053

**R10**

**Set No. 2**

**IV B.Tech I Semester Regular Examinations, December 2013**  
**DATA WARE HOUSING AND DATA MINING**  
(Common to Computer Science & Engineering and Information Technology)

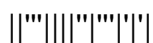
**Time : 3 hours**

**Max. Marks: 75**

**Answer any Five Questions**  
**All Questions carry equal marks**

\*\*\*\*\*

- 1 a) Describe general characteristics of data sets in detail. [8]  
b) Describe how Data Mining technique is different from Traditional techniques. [7]
- 2 a) Differentiate how Pearson's correlation is different from perfect correlation. [8]  
b) Write the algorithm to find out similarities of Heterogeneous Objects. [7]
- 3 a) Difference between operational databases and data warehousing. [8]  
b) What is partial materialization? What are its applications. [7]
- 4 a) What is meant by Model over fitting? How can over fitting done due to presence of noise? [8]  
b) How splitting is done in continuous attributes? [7]
- 5 a) What is meant by Conditional independence? What are advantages of using Naive Classifier? [8]  
b) Compare belief network and Naïve Bayes classifier. [7]
- 6 a) Write the factors that affect the computational complexity of the Apriori algorithm. [8]  
b) Explain how confidence based pruning used in Apriori algorithm. [7]
- 7 a) What is Cluster Analysis? Explain different methods of clusters. [8]  
b) Write the procedure to handle document data for Clustering. [7]
- 8 a) Explain the procedure of selecting the DBSCAN parameters. [8]  
b) Compare the ward's method and Centroid methods. [7]





Code No: R41053

**R10**

**Set No. 3**

**IV B.Tech I Semester Regular Examinations, December 2013**  
**DATA WARE HOUSING AND DATA MINING**  
(Common to Computer Science & Engineering and Information Technology)

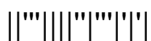
**Time : 3 hours**

**Max. Marks: 75**

**Answer any Five Questions**  
**All Questions carry equal marks**

\*\*\*\*\*

- 1 a) Explain different types of ordered data with examples. [8]  
b) Explain different Data mining tasks. [7]
- 2 a) Discuss about similarity measures for binary data in detail. [8]  
b) Explain different data preprocessing techniques. [7]
- 3 a) Explain the Data Warehouse architecture with neat diagram. [8]  
b) Explain why we need separate Data Warehouse. [7]
- 4 a) What is Hunt's algorithm? How is it helpful to construct Decision Tree? [8]  
b) What are different measures for selecting the best split? [7]
- 5 a) What are the different characteristics of Naïve Bayes classifiers? [8]  
b) Explain how to represent model of Bayesian Belief Network. [7]
- 6 a) Write the procedure of closed frequent item set. [8]  
b) Compare the Apriori algorithm and FP-growth algorithm. [7]
- 7 a) Write K-mean algorithm and also discuss additional issues in k-means. [8]  
b) Differentiate between complete clustering and partial clustering in detail. [7]
- 8 a) Write the algorithm of DBSCAN clustering. [8]  
b) Describe about agglomerative hierarchical clustering algorithm. [7]



Code No: R41053

**R10**

**Set No. 4**

**IV B.Tech I Semester Regular Examinations, December 2013**  
**DATA WARE HOUSING AND DATA MINING**  
**(Common to Computer Science & Engineering and Information Technology)**

**Time : 3 hours**

**Max. Marks: 75**

**Answer any Five Questions**  
**All Questions carry equal marks**

\*\*\*\*\*

- 1 a) What are the different measures to assess the quality of data? [8]  
b) Explain the process of knowledge discovery in database? [7]
- 2 a) Explain about Cosine similarity measure in detail. [8]  
b) Describe about finding similarities and dissimilarities between data objects. [7]
- 3 a) Explain different OLAP operation in multidimensional data. [8]  
b) Describe the OLAP servers. [7]
- 4 a) Explain various methods to evaluate the performance of a classifier. [8]  
b) Write the algorithm of Decision Tree Induction. [7]
- 5 a) Estimate conditional probabilities for continuous attributes with a working example. [8]  
b) What is meant by Bayes error rate? Illustrate the application of Bayesian classifier. [7]
- 6 a) Write Apriori algorithm for generating frequent item set. [8]  
b) Differentiate between maximal and closed frequent item set. [7]
- 7 a) Compare the strengths and weaknesses of different clustering algorithm. [8]  
b) How segmentation and partitioning related to clustering? Explain. [7]
- 8 a) Compare the agglomerative hierarchical clustering and DBSCAN with respect to time and space complexity. [8]  
b) What is meant by cluster proximity? Explain [7]

