

Code No: RT32052

R13**SET - 1****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 compulsory3. 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]

Code No: RT32052

R13**SET - 2****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 compulsory3. 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: [8M]
(a) F-measure (b) Confusion matrix (c) Cross-validation (d) Bootstrap
- 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 [8M]
a) containing at least one Nintendo game,
b) containing items the sum of whose price is less than \$150.
- b) Explain frequent item sets without candidate generation. [8M]
- 7 Explain about K-means algorithm with suitable example. [16M]

Code No: RT32052

R13**SET - 3**

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]

Code No: RT32052

R13**SET - 4**

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.
