Set No. 1

IV B.Tech I Semester Regular Examinations, November 2012 DATA WAREHOUSING AND DATA MINING (Computer Science & Engineering)

Time: 3 hours

Code No: M0522/R07

Max Marks: 80

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

- 1. (a) Explain the major issues in data mining.
 - (b) Explain the three-tier datawarehousing architecture. [8+8]
- 2. (a) Briefly discuss about data integration.
 - [8+8](b) Briefly discuss about data transformation.
- 3. The four major types of concept hierarchies are: schema hierarchies, set-grouping hierarchies, operation-derived hierarchies, and rule-based hierarchies.
 - (a) Briefly define each type of hierarchy.
 - (b) For each hierarchy type, provide an example. [16]
- 4. (a) What are the differences between concept description in large data bases and OLAP?
 - (b) Explain about the graph displays of basic statistical class description. [8+8]
- 5. (a) Which algorithm is an influential algorithm for mining frequent item sets for Boolean association rules. Explain.
 - (b) Discuss about association mining using correlation rules. [8+8]
- (a) Write an algorithm for k-nearest neighbor classification given k and n, the 6. number of attributes describing each sample.
 - (b) What is linear regression? Give an example of linear regression using the method of least squares. [8+8]
- (a) What is Cluster Analysis? What are some typical applications of clustering? 7. What are some typical requirements of clustering in data mining?
 - (b) Define data matrix and dissimilarity matrix. Discuss about interval-scaled variables. [2+2+5+3+4]
- 8. (a) Discuss about multidimensional analysis and descriptive mining of complex data objects.
 - (b) Explain text data analysis and information retrieval. [8+8]

1 of 1

www.FirstRanker.com

| Code No: M0522/R07 Set No. 2 |
|--|
| IV B.Tech I Semester Regular Examinations, November 2012 DATA WAREHOUSING AND DATA MINING (Computer Science & Engineering) |
| Time: 3 hours Max Marks: 80 Answer any FIVE Questions All Questions carry equal marks ***** |
| (a) Draw and explain the architecture for on-line analytical mining. (b) Briefly discuss the data warehouse applications. [8+8] |
| 2. (a) Briefly discuss the data smoothing techniques.(b) Explain about concept hierarchy generation for categorical data. [8+8] |
| 3. Write the syntax for the following data mining primitives: (a) Task-relevant data. (b) Concept hierarchies. [16] |
| 4. (a) What is Concept description? Explain. (b) What are the differences between concept description in large data bases and OLAP? [8+8] |
| 5. Explain the Apriori algorithm with example. [16] |
| 6. (a) How scalable is decision tree induction? Explain.(b) Explain backpropagation classification technique |
| 7. What is clustering? Briefly describe the following approaches to clustering methods: partitioning methods, hierarchical methods, density-based methods, grid-based methods, and model-based methods. Give an example in each case. [16] |
| 8. (a) Give an example of generalization-based mining of plan databases by divide-and-conquer. (b) What is sequential pattern mining? Explain |
| (b) What is sequential partern mining: Explain. |

(c) Explain the construction of a multilayered web information base. [8+4+4]

Set No. 3

IV B.Tech I Semester Regular Examinations, November 2012 DATA WAREHOUSING AND DATA MINING (Computer Science & Engineering)

Time: 3 hours

Code No: M0522/R07

Max Marks: 80

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

- (a) Discuss about Concept hierarchy. 1.
 - (b) Briefly explain about classification of database systems. [8+8]
- 2. Write short note on the following data reduction techniques:
 - (a) Data compression.
 - (b) Concept hierarchy generation for categorical data. [16]
- 3. The four major types of concept hierarchies are: schema hierarchies, set-grouping hierarchies, operation-derived hierarchies, and rule-based hierarchies.
 - (a) Briefly define each type of hierarchy.
 - (b) For each hierarchy type, provide an example. [16]
- 4. (a) What are the differences between concept description in large data bases and OLAP?
 - (b) Explain about the graph displays of basic statistical class description. [8+8]
- (a) Explain FP-Growth with example. 5.
 - (b) What are the approaches to mining multilevel association rules? Explain. [8+8]
- (a) Why is tree pruning useful in decision tree induction? What is a draw back 6. of using a separate set of samples to evaluate pruning?
 - (b) How rough set approach and fuzzy set approaches are useful for classification? Explain. [8+8]
- 7. (a) Given the following measurement for the variable age: 16, 25, 28, 46, 29, 44, 38, 37, 54, 27 Standardize the variable by the following:
 - i. Compute the mean absolute deviation of age.
 - ii. Compute the Z-score for the first four measurements.
 - (b) Explain clustering using representatives algorithm with example.
 - (c) Write an algorithm for DBSCAN and give an example of DBSCAN. [4+4+4+4]
- 8. Explain in detail about mining on spatial databases and text databases. [16]

1 of 1

www.FirstRanker.com

www.FirstRank

Code No: M0522/R07

Set No. 4

IV B.Tech I Semester Regular Examinations, November 2012 DATA WAREHOUSING AND DATA MINING (Computer Science & Engineering)

Time: 3 hours

Max Marks: 80

[4+4+4+4]

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

- 1. Briefly compare the following concepts. Use an example to explain your points.
 - (a) Snowflake schema, fact constellation, starnet query model.
 - (b) Data cleaning, data transformation, refresh.
 - (c) Discovery driven cube, multifeature cube, and virtual warehouse. [16]
- 2. Briefly discuss the Discretization and concept hierarchy techniques. [16]
- 3. (a) List and describe any four primitives for specifying a data mining task.
 - (b) Describe why concept hierarchies are useful in data mining. [8+8]
- 4. (a) Discuss the methods for presenting the derived generalization.
 - (b) Explain the limitations for class characterization. [8+8]
- 5. (a) Explain about constraint-based Association mining.
 - (b) Give an example for Association rule mining? Classify Association rules.[8+8]
- 6. (a) Why is tree pruning useful in decision tree induction? What is a draw back of using a separate set of samples to evaluate pruning?
 - (b) How rough set approach and fuzzy set approaches are useful for classification? Explain. [8+8]
- 7. Explain the following:
 - (a) DBSCAN
 - (b) OPTICS
 - (c) DENCLUE
 - (d) BIRCH.
- 8. (a) Explain multidimensional analysis of multimedia data.
 - (b) Define Information retrieval. What are basic measures for text retrieval?
 - (c) What is keyword-based association analysis?
 - [5+4+3+4](d) Briefly discuss about mining the World Wide Web.

www.FirstRanker.com