Code No: R05410503 **Set No. 1**

IV B.Tech I Semester Regular Examinations, November 2008
DATA WAREHOUSING AND DATA MINING
(Computer Science & Engineering)

Time: 3 hours **Max Marks: 80**

Answer any FIVE Questions
All Questions carry equal marks

⋆ ⋆ ⋆ ⋆ ⋆

1. (a) Explain data mining as a step in the process of knowledge discovery.
   (b) Differentiate operational database systems and data warehousing. [8+8]
2. (a) Discuss in detail about data transformation.
   (b) Explain about concept hierarchy generation for categorical data. [8+8]
3. Discuss the importance of establishing a standardized data mining query language. What are some of the potential benefits and challenges involved in such a task? List and explain a few of the recent proposals in this area. [16]
4. Write short notes for the following in detail:
   (a) Attribute-oriented induction.
   (b) Efficient implementation of Attribute-oriented induction. [8+8]
5. (a) What is an iceberg query? Give an example.
   (b) Explain about mining distance based association rules.
   (c) How are meta rules useful? Explain with example. [5+6+5]
6. (a) Why naive Bayesian classification called "naive"? Briefly outline the major ideas of naive Bayesian classification.
   (b) Define regression. Briefly explain about linear, non-linear and multiple regressions. [8+8]
7. (a) What is Cluster Analysis? What are some typical applications of clustering? What are some typical requirements of clustering in data mining?
   (b) Discuss about model-based clustering methods. [2+2+5+7]
8. Explain the following:
   (a) Mining time-series and sequence data
   (b) Mining text databases. [8+8]

⋆ ⋆ ⋆ ⋆ ⋆

Code No: R05410503 **Set No. 2**

IV B.Tech I Semester Regular Examinations, November 2008
DATA WAREHOUSING AND DATA MINING
(Computer Science & Engineering)

Time: 3 hours **Max Marks: 80**

Answer any FIVE Questions
All Questions carry equal marks

⋆ ⋆ ⋆ ⋆ ⋆

1. (a) Explain data mining as a step in the process of knowledge discovery.
   (b) Differentiate operational database systems and data warehousing. [8+8]
2. (a) Briefly discuss about data integration.
   (b) Briefly discuss about data transformation. [8+8]
3. Write the syntax for the following data mining primitives:
   (a) The kind of knowledge to be mined.
   (b) Measures of pattern interestingness. [16]
4. (a) Write the algorithm for attribute-oriented induction. Explain the steps involved in it.
   (b) How can concept description mining be performed incrementally and in a distributed manner? [8+8]
5. (a) Explain about iceberg queries with example.
   (b) Can we design a method that mines the complete set of frequent item sets without candidate generation? If yes, explain with example. [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. (a) Give an example of how specific clustering methods may be integrated, for example, where one clustering algorithm is used as a preprocessing step for another.
   (b) Write CURE algorithm and explain. [10+6]
8. (a) Explain about multidimensional analysis and descriptive mining of complex data objects.
   (b) Describe similarity search in time-series analysis.
   (c) What are cases and parameters for sequential pattern mining? [8+4+4]

⋆ ⋆ ⋆ ⋆ ⋆

Code No: R05410503 **Set No. 3**

IV B.Tech I Semester Regular Examinations, November 2008
DATA WAREHOUSING AND DATA MINING
(Computer Science & Engineering)

Time: 3 hours **Max Marks: 80**

Answer any FIVE Questions
All Questions carry equal marks

⋆ ⋆ ⋆ ⋆ ⋆

1. (a) Explain data mining as a step in the process of knowledge discovery.
   (b) Differentiate operational database systems and data warehousing. [8+8]
2. (a) Briefly discuss about data integration.
   (b) Briefly discuss about data transformation. [8+8]
3. Explain the syntax for the following data mining primitives:
   (a) Task-relevant data
   (b) The kind of knowledge to be mined
   (c) Interestingness measures
   (d) Presentation and visualization of discovered patterns. [16]
4. (a) Attribute-oriented induction generates one or a set of generalized descriptions. How can these descriptions be visualized?
   (b) Discuss about the methods of attribute relevance analysis? [8+8]
5. (a) How can we mine multilevel Association rules efficiently using concept hierarchies? Explain.
   (b) Can we design a method that mines the complete set of frequent item sets without candidate generation. If yes, explain with example. [8+8]
6. (a) Explain about basic decision tree induction algorithm.
   (b) Discuss about Bayesian classification. [8+8]
7. (a) Use a diagram to illustrate how, for a constant MinPts value, density-based clusters with respect to a higher density (i.e., a lower value for ε , the neighborhood radius) are completely contained in density- connected sets obtained with respect to a lower density.
   (b) Give an example of how specific clustering methods may be integrated, for example, where one clustering algorithm is used as a preprocessing step for another. [8+8]
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?
   (d) Briefly discuss about mining the World Wide Web. [5+4+3+4]

⋆ ⋆ ⋆ ⋆ ⋆

Code No: R05410503 **Set No. 4**

IV B.Tech I Semester Regular Examinations, November 2008
DATA WAREHOUSING AND DATA MINING
(Computer Science & Engineering)

Time: 3 hours **Max Marks: 80**

Answer any FIVE Questions
All Questions carry equal marks

⋆ ⋆ ⋆ ⋆ ⋆

1. (a) Discuss about Concept hierarchy.
   (b) Briefly explain about - classification of database systems. [8+8]
2. Explain various data reduction techniques. [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 is Concept description? Explain.
   (b) What are the differences between concept description in large data bases and OLAP? [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]
6. (a) How scalable is decision tree induction? Explain.
   (b) Explain about prediction. [8+8]
7. (a) How does the k-means algorithm work? Explain with example.
   (b) Explain about grid-based methods in clustering. [8+8]
8. (a) Describe latent semantic indexing technique with an example.
   (b) Discuss about mining time-series and sequence data. [4+12]

⋆ ⋆ ⋆ ⋆ ⋆