Code :R7421504 **1**

IV B.Tech II Semester(R07) Regular Examinations, April 2011
DATA WAREHOUSING & DATA MINING
(Computer Science & Systems Engineering, Electronics & Computer Engineering)
Time: 3 hours **Max Marks: 80**

Answer any FIVE questions
All questions carry equal marks

⋆ ⋆ ⋆ ⋆ ⋆

1. (a) Explain the major issues in Data mining.
   (b) What is the need of data preprocessing? What are various forms of data pre processing? Explain briefly.
2. (a) Describe with the help of a figure the typical process flow with in a Data warehouse.
   (b) What is a Data Warehouse? Give differences between OLIP and OLAP.
3. (a) Explain in detail the Data mining primitives.
   (b) What is data mining query language?
4. (a) What is concept description? Explain .
   (b) Write briefly on 'mining class comparisons'.
5. (a) Give the association rule for the following sentence. "The information that customers who purchase computers also tend to buy antivirus software at the same time".
   (b) Explain the Apriori algorithm.
6. (a) What is classification? What is prediction?
   (b) What is Bayes theorem? Explain about Naive Bayesian classification.
7. (a) Write algorithm for K-means and Explain?
   (b) Discuss about density based methods?
8. (a) Explain mining Associations in multimedia data.
   (b) What is sequential pattern mining? Explain.

⋆ ⋆ ⋆ ⋆ ⋆

Code :R7421504 **2**

IV B.Tech II Semester(R07) Regular Examinations, April 2011
DATA WAREHOUSING & DATA MINING
(Computer Science & Systems Engineering, Electronics & Computer 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) Briefly discuss about data integration and Data Transformation.
2. (a) Differentiate operation data base systems and Data warehousing.
   (b) Explain the three-tire data warehousing architecture.
3. (a) Describe why is it important to have a data mining query language.
   (b) Explain how DMQL is useful in designing GUI?
4. (a) What are the differences between concept description in large databases and OLAP?
   (b) Explain about the graph displays of basic statistical class description.
5. Explain in detail the Apriori algorithm with example.
6. (a) Describe the data classification process with a neat diagram.
   (b) How does the Naive Bayesian classification works? Explain.
7. (a) What are the categories of major clustering methods? Explain?
   (b) Explain about outlier analysis.
8. (a) How does latent semantic indexing reduce the size of the term frequency matrix? Explain.
   (b) What kinds of Association can be mined in multimedia data.

⋆ ⋆ ⋆ ⋆ ⋆

Code :R7421504 **3**

IV B.Tech II Semester(R07) Regular Examinations, April 2011
DATA WAREHOUSING & DATA MINING
(Computer Science & Systems Engineering, Electronics & Computer Engineering)
Time: 3 hours **Max Marks: 80**

Answer any FIVE questions
All questions carry equal marks

⋆ ⋆ ⋆ ⋆ ⋆

1. (a) Discuss in detail about data transformation.
   (b) Explain about concept hierarchy generation for categorical data.
2. (a) What is a Data warehouse? Discuss in detail.
   (b) Explain designing of Fact table.
3. What is DMQL? What is the need of DMQL? Explain the syntax of DMQL? Give one example for DMQL query.
4. (a) What are the differences between concept description in large data bases and OLAP.
   (b) What is concept description? Explain.
5. (a) Discuss about association mining using corelation analysis.
   (b) Which algorithm is an influential algorithm for mining frequent item sets for Boolean association rules. Explain.
6. (a) Write an algorithm for K-nearest neighbor classification given k and n, the number of attributes describing each sample.
   (b) What is linear regression? Give an example of linear regression using the method of least squares.
7. (a) What are the types of data in cluster analysis? Explain.
   (b) Explain about partitioning methods in detail.
8. (a) Write the differences between mining association rules in multimedia databases Versus transactional databases?
   (b) Describe the construction of a multilayered web information base.

⋆ ⋆ ⋆ ⋆ ⋆

Code :R7421504 4

IV B.Tech II Semester(R07) Regular Examinations, April 2011
DATA WAREHOUSING & DATA MINING
(Computer Science & Systems Engineering, Electronics & Computer Engineering)
Time: 3 hours **Max Marks: 80**

Answer any FIVE questions
All questions carry equal marks

⋆ ⋆ ⋆ ⋆ ⋆

1. (a) What is the need of data preprocessing? what are various forms of Data preprocessing Explain briefly.
   (b) Explain data mining as a step in the process of Knowledge discovery.
2. (a) Explain various data reduction techniques.
   (b) Write differences between OLTP and OLAP.
3. (a) What is DMQL? What is the need of DMQL give example.
   (b) Write about task-relavent data and concept hierarchies.
4. (a) What is concept description? Explain.
   (b) Write briefly on 'mining class comparisons'.
5. (a) Discuss about association mining using corelation analysis?
   (b) Write FP-growth algorithm? Explain.
6. (a) What is classification? What is prediction.
   (b) What is Bayes theorem? Explain about Naive Rayesian classification?
7. (a) What are the categories of major clustering methods? Explain.
   (b) Explain K-means algorithm.
8. (a) What is sequential pattern mining? Explain.
   (b) What kinds of association can be mined in multimedia data.

⋆ ⋆ ⋆ ⋆ ⋆