

Code No. M0522

R07

Set No.1

IV B.Tech I Semester Supplementary Examinations, February, 2012

DATA WAREHOUSING AND DATA MINING

(Computer Science and Engineering)

Time: 3 hours

Max. Marks: 80

Answer any FIVE Questions
All Questions carry equal marks

1. a) What is data mining? Briefly explain the Knowledge discovery process.
b) Explain the three-tier data warehouse architecture.
2. a) With an example, describe any two schema (*star/snowflake/fact constellation*) definitions using DMQL statements.
b) What is data integration? Discuss the issues to be considered for data integration.
3. a) Briefly describe data generalization, summarization and analytical characterization
b) What is association and correlation? With an example describe classification and prediction.
4. What is constraint-based mining? Describe in detail about the possible constraints in high-level declarative DMQL and user interface.
5. a) What is backpropagation? Describe backpropagation algorithm.
b) Discuss about multidimensional association rule mining from relational databases.
6. a) Describe how categorization of major clustering methods is being done.
b) What is Hierarchical clustering? Describe any one Hierarchical clustering algorithm.
7. a) What is text mining? Describe about basic measures for text retrieval.
b) Briefly describe document cluster analysis.
8. What is conceptual clustering? Describe in detail about COBWEB.

Code No. M0522**R07****Set No.2****IV B.Tech I Semester Supplementary Examinations, February, 2012****DATA WAREHOUSING AND DATA MINING****(Computer Science and Engineering)****Time: 3 hours****Max. Marks: 80****Answer any FIVE Questions
All Questions carry equal marks***********

1. a) What is a concept hierarchy? Describe the OLAP operations in the Multidimensional data model.
b) What is association and correlation? With an example describe classification and prediction.
2. a) What is data cleaning? Describe the approaches to fill missing values.
b) What is noisy data? Explain the binning methods for data smoothing.
3. a) Draw and explain the architecture of a typical data mining system.
b) Briefly discuss about functional components of GUI based data mining system.
4. a) Discuss the concept involved in designing GUI based on DMQL.
b) What is a data ware house? Differentiate between operational data base system and data warehouses.
5. a) What is Decision tree? With an example, briefly describe the algorithm for generating decision tree.
b) What is a concept hierarchy? Describe the OLAP operations in the Multidimensional data model.
6. a) What is cluster analysis? Describe the dissimilarity measures for interval-scaled variables and binary variables.
b) What are Bayesian classifiers? With an example, describe how to predict a class label using naive Bayesian classification.

Code No. M0522

R07

Set No.2

7. a) What is misclassification rate of a classifier? Describe sensitivity and specificity measures of a classifier.
b) Describe Constraint based association mining.
8. What is spatial data mining? What is spatial data cube, and what are the three dimensions in a spatial data cube?

FirstRanker

Code No. M0522**R07****Set No.3****IV B.Tech I Semester Supplementary Examinations, February, 2012
DATA WAREHOUSING AND DATA MINING****(Computer Science and Engineering)****Time: 3 hours****Max. Marks: 80****Answer any FIVE Questions
All Questions carry equal marks***********

1. a) What is transactional database? Describe any five advanced database systems.
b) Draw and explain the architecture of a typical data mining system.
2. a) What is data normalization? Explain any two normalization methods.
b) What is data dispersion? Describe the common measures for data dispersion.
3. With examples, describe in detail about the available techniques for concept Hierarchy Generation for categorical data.
4. a) What is Analytical characterization? What is the need to perform attribute relevance analysis?
b) Describe the procedure for mining class comparisons.
5. a) What is attribute selection measure? Briefly describe the attribute selection measures for decision tree induction.
b) Briefly outline the major steps of decision tree classification.
6. a) Describe the dissimilarity measures for interval-scaled variables and binary variables.
b) Describe how categorization of major clustering methods is being done.
7. a) What is multimedia data mining? What kind of associations can be mined in multimedia data?
b) What is Grid based clustering? Describe any one Grid based clustering algorithm.
8. What is WWW? What is meant by authoritative web page? Describe web usage mining.

Code No. M0522**R07****Set No.4****IV B.Tech I Semester Supplementary Examinations, February, 2012****DATA WAREHOUSING AND DATA MINING****(Computer Science and Engineering)****Time: 3 hours****Max. Marks: 80**

Answer any FIVE Questions
All Questions carry equal marks

1. a) Describe the three challenges to data mining regarding data mining methodology and user interaction issues.
b) With an example, describe snowflake and fact constellations.
2. a) Briefly describe various forms of data pre-processing.
b) What is a measure? How measures are computed? Describe the organization of measures.
3. a) What is a concept hierarchy? Briefly describe the OLAP operations in the Multidimensional data model.
b) Briefly describe the primitives for specifying a data mining task.
4. a) What are the quantitative association rules? What is ARCS and discuss the involved steps.
b) What is transactional database? With an example, explain multilevel association rule mining.
5. a) Describe the criteria used to evaluate classification and prediction methods.
b) With an example, explain the classification by decision tree induction.
6. a) What is Density based clustering? Describe DBSCAN clustering algorithm.
b) What is partitioning method? Describe any one partition based clustering algorithm.
7. a) What is informational data store? Briefly describe the characteristics of informational data.
b) What is Association rule mining? Briefly describe the criteria for classifying association rules
8. What is multimedia data mining? How similarity search can be performed on multimedia data? Describe the contents of a multimedia data cube.