

Code No: V3244

R07

Set No: 1

III B.Tech. II Semester Supplementary Examinations, December - 2012

**DATA WERHOUSING AND DATA MINING**

(Information Technology)

**Time: 3 Hours****Max Marks: 80**

Answer any FIVE Questions  
All Questions carry equal marks

\*\*\*\*\*

- 1
  - a) Are all of the patterns interesting? Explain.
  - b) Describe four challenges to data mining regarding data mining methodology and user interaction issues.
  - c) How can we find a 'good' subset of the original attributes? (6+6+4)
  
- 2
  - a) What are metadata? Explain metadata repository.
  - b) Describe efficient computation of data cubes?
  - c) What is a concept hierarchy? Give examples. (5+6+5)
  
- 3 Explain the following:
  - (a) Syntax for specifying the kind of knowledge to be mined.
  - (b) Coupling schemes
  - (c) Interestingness measures (6+6+4)
  
- 4
  - a) Why perform attribute relevance analysis? Explain.
  - b) Give an example of analytical characterization.
  - c) Explain graph displays of basic statistical class descriptions. (4+6+6)
  
- 5
  - a) Discuss about constraint-based association mining.
  - b) Explain mining frequent item sets without candidate generation. (8+8)
  
- 6
  - a) What is Boosting? State why it may improve the accuracy of decision tree induction.
  - b) Explain rough set approach and Fuzzy set approaches.
  - c) What are issues regarding classification and prediction? (4+6+6)
  
- 7
  - a) Describe deviation-based outlier detection.
  - b) Explain about a hierarchical clustering algorithm using dynamic modeling.
  - c) Define nominal, ordinal, and ratio-scaled variables. (5+5+6)
  
- 8 Explain the following:
  - (a) Web based mining
  - (b) Keyword-based and similarity-based retrieval
  - (c) Similarity search in multimedia data
  - (d) Dimensions and measures of a spatial data cube (4+4+4+4)

\*\*\*\*\*

1 of 1

Code No: V3244

R07

Set No: 2

III B.Tech. II Semester Supplementary Examinations, December - 2012

**DATA WERHOUSING AND DATA MINING**

(Information Technology)

**Time: 3 Hours****Max Marks: 80**

Answer any FIVE Questions  
All Questions carry equal marks

\*\*\*\*\*

- 1 a) What are various forms of data preprocessing? Explain.  
b) Describe segmentation by natural partitioning with example.  
c) What is data mining? Describe knowledge discovery steps with a neat diagram. (4+6+6)
- 2 a) Give examples for defining star, snowflake, and fact constellation schemas.  
b) What are the steps for the design and construction of data warehouses? Explain.  
c) Describe discovery-driven exploration of data cubes. (6+5+5)
- 3 a) What are primitives for specifying a data mining task? Explain in detail.  
b) Give an example of a DMQL query. (12+4)
- 4 a) How is attribute-oriented induction actually implemented? Explain.  
b) Explain mining descriptive statistical measures in large databases. (8+8)
- 5 a) What is an iceberg query? Give example.  
b) Describe mining distance-based association rules.  
c) Explain metarule-guided mining of association rules. (5+5+6)
- 6 a) How does tree pruning work?  
b) How does backpropagation work? Explain with example.  
c) Describe classifier accuracy. (3+7+6)
- 7 a) Write k-means algorithm. How does the k-means algorithm work?  
b) What is an outlier? Explain statistical-based outlier detection.  
c) Write and explain about the CURE algorithm. (6+5+5)
- 8 Explain the following:  
(a) Spatial classification and spatial trend analysis  
(b) Mining associations in multimedia data  
(c) Sequential pattern mining  
(d) Latent semantic indexing (4+4+4+4)

\*\*\*\*\*

Code No: V3244

**R07****Set No: 3**

III B.Tech. II Semester Supplementary Examinations, December - 2012

**DATA WERHOUSING AND DATA MINING**

(Information Technology)

**Time: 3 Hours****Max Marks: 80**

Answer any FIVE Questions  
All Questions carry equal marks

\*\*\*\*\*

- 1
  - a) What motivated data mining? Why is it important?
  - b) What is noise? Describe data smoothing techniques.
  - c) Can we reduce the data volume by choosing alternative, 'smaller' forms of the data representation? (5+5+6)
- 2
  - a) Describe a 3-tier data warehouse architecture.
  - b) What are OLAP operations in multidimensional data model? Explain.
  - c) Explain about multi feature cubes. (6+5+5)
- 3
  - a) Define schema, set-grouping, operation-derived and rule-based hierarchies.
  - b) Describe various forms of presenting and visualizing the discovered patterns.
  - c) What are the desired architectures for data mining systems? Explain. (4+6+6)
- 4
  - a) What is concept description?
  - b) Give analysis of attribute relevance.
  - c) Explain attribute-oriented induction. (3+6+7)
- 5
  - a) Explain in detail about mining single-dimensional Boolean association rules from transactional databases.
  - b) Describe mining quantitative association rules. (12+4)
- 6
  - a) Explain classification based on concepts from association rule mining.
  - b) How Naïve Bayesian classification works? Explain.
  - c) What is attribute selection measure? Explain. (6+6+4)
- 7
  - a) Define clustering. What are density-based methods? Explain in detail.
  - b) Describe interval-scaled and binary variables. (10+6)
- 8
 

Explain the following:

  - (a) Document classification analysis
  - (b) Periodicity analysis
  - (c) Components of time-series data
  - (d) Example of plan mining process (4+4+4+4)

\*\*\*\*\*

Code No: V3244

**R07****Set No: 4**

III B.Tech. II Semester Supplementary Examinations, December - 2012

**DATA WERHOUSING AND DATA MINING**

(Information Technology)

**Time: 3 Hours****Max Marks: 80**

Answer any FIVE Questions  
All Questions carry equal marks

\*\*\*\*\*

- 1 a) Describe various methods for the generation of concept hierarchies for categorical data.  
b) Explain about data compression.  
c) Discuss about advanced database systems and advanced database applications. (5+5+6)
- 2 a) List the differences between operational database systems and data warehouses.  
b) What are various types of OLAP services? Explain.  
c) Describe data warehouse usage. (6+5+5)
- 3 a) What are the functional components of data mining GUI? Explain.  
b) Describe syntax for concept hierarchy specification.  
c) Explain about task-relevant data with example. (5+6+5)
- 4 a) What are methods of attribute relevance analysis? Explain.  
b) Discuss about the presentation of the derived generalization.  
c) Describe measuring the dispersion of data. (5+5+6)
- 5 a) Explain about mining multidimensional association rules from relational databases and data warehouses.  
b) Discuss about constraint-based association mining. (8+8)
- 6 a) What is classification? What is prediction?  
b) Describe training Bayesian belief networks.  
c) Write Back propagation algorithm. (4+6+6)
- 7 a) What are grid-based methods in clustering? Explain.  
b) What is an outlier? Describe various approaches of outlier detection. (8+8)
- 8 Explain the following:  
(a) Set-valued and list-valued attributes  
(b) Spatial association analysis  
(c) Cases and parameters for sequential pattern mining  
(d) Inverted index (4+4+4+4)

\*\*\*\*\*

1 of 1