Code No: R05321204 | **R05** | **Set No. 2**

III B.Tech II Semester Examinations,December 2010
DATA WAREHOUSING AND DATA MINING
Information Technology

Time: 3 hours   Max Marks: 80

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

1. (a) What are some enhancements to basic decision tree induction? Explain.

   (b) How does a Bayesian belief network learn? Explain.

   (c) What are two strategies to increase classifier accuracy? Explain.   [5+5+6]

2. (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]

3. (a) Briefly explain about data cleaning.

   (b) Why preprocessing of data is needed? Explain.   [8+8]

4. (a) Draw and explain the architecture of typical data mining system.

   (b) Differentiate OLTP and OLAP.   [8+8]

5. A heterogeneous database system consists of multiple database systems that are defined independently, but that need to exchange transform information among themselves and answer global queries. Discuss how to process a descriptive mining query in such a system using a generalization-based approach.   [16]

6. (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.   [8+4+4]

7. (a) Discus about Association rule mining.

   (b) Define multidimensional Association rule. Discuss mining distance-based Association rules.   [8+8]

8. (a) What is Concept description? Explain.

   (b) What are the differences between concept description in large data bases and OLAP?   [8+8]

⋆⋆⋆⋆⋆

1

Code No: R05321204 | R05 | Set No. 4

III B.Tech II Semester Examinations,December 2010
DATA WAREHOUSING AND DATA MINING
Information Technology

Time: 3 hours   Max Marks: 80

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

1. (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.   [8+4+4]

2. (a) What is Concept description? Explain.

   (b) What are the differences between concept description in large data bases and OLAP?   [8+8]

3. (a) Draw and explain the architecture of typical data mining system.

   (b) Differentiate OLTP and OLAP.   [8+8]

4. (a) Briefly explain about data cleaning.

   (b) Why preprocessing of data is needed? Explain.   [8+8]

5. (a) What are some enhancements to basic decision tree induction? Explain.

   (b) How does a Bayesian belief network learn? Explain.

   (c) What are two strategies to increase classifier accuracy? Explain.   [5+5+6]

6. (a) Discus about Association rule mining.

   (b) Define multidimensional Association rule. Discuss mining distance-based Association rules.   [8+8]

7. A heterogeneous database system consists of multiple database systems that are defined independently, but that need to exchange transform information among themselves and answer global queries. Discuss how to process a descriptive mining query in such a system using a generalization-based approach.   [16]

8. (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]

⋆ ⋆ ⋆ ⋆ ⋆

2

Code No: R05321204   **R05**   **Set No. 1**

III B.Tech II Semester Examinations,December 2010
DATA WAREHOUSING AND DATA MINING
Information Technology

Time: 3 hours   Max Marks: 80

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

1. (a) Briefly explain about data cleaning.

   (b) Why preprocessing of data is needed? Explain.   [8+8]

2. (a) Draw and explain the architecture of typical data mining system.

   (b) Differentiate OLTP and OLAP.   [8+8]

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) Discus about Association rule mining.

   (b) Define multidimensional Association rule. Discuss mining distance-based Association rules.   [8+8]

5. (a) What are some enhancements to basic decision tree induction? Explain.

   (b) How does a Bayesian belief network learn? Explain.

   (c) What are two strategies to increase classifier accuracy? Explain.   [5+5+6]

6. (a) What is Concept description? Explain.

   (b) What are the differences between concept description in large data bases and OLAP?   [8+8]

7. A heterogeneous database system consists of multiple database systems that are defined independently, but that need to exchange transform information among themselves and answer global queries. Discuss how to process a descriptive mining query in such a system using a generalization-based approach.   [16]

8. (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.   [8+4+4]

⋆ ⋆ ⋆ ⋆ ⋆

3

Code No: R05321204 | **R05** | **Set No. 3**

III B.Tech II Semester Examinations,December 2010
DATA WAREHOUSING AND DATA MINING
Information Technology

Time: 3 hours

Max Marks: 80

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

1. (a) What are some enhancements to basic decision tree induction? Explain.

   (b) How does a Bayesian belief network learn? Explain.

   (c) What are two strategies to increase classifier accuracy? Explain.   [5+5+6]

2. A heterogeneous database system consists of multiple database systems that are defined independently, but that need to exchange transform information among themselves and answer global queries. Discuss how to process a descriptive mining query in such a system using a generalization-based approach.   [16]

3. (a) What is Concept description? Explain.

   (b) What are the differences between concept description in large data bases and OLAP?   [8+8]

4. (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]

5. (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.   [8+4+4]

6. (a) Briefly explain about data cleaning.

   (b) Why preprocessing of data is needed? Explain.   [8+8]

7. (a) Discus about Association rule mining.

   (b) Define multidimensional Association rule. Discuss mining distance-based Association rules.   [8+8]

8. (a) Draw and explain the architecture of typical data mining system.

   (b) Differentiate OLTP and OLAP.   [8+8]

⋆ ⋆ ⋆ ⋆ ⋆

4