Code: AT78   Subject: DATA MINING & WAREHOUSING

AMIETE – IT (NEW SCHEME)

Time: 3 Hours   **DECEMBER 2011**   Max. Marks: 100

NOTE: There are 9 Questions in all.
- **Please write your Roll No. at the space provided on each page immediately after receiving the Question Paper.**
- **Question 1 is compulsory and carries 20 marks. Answer to Q.1 must be written in the space provided for it in the answer book supplied and nowhere else.**
- **The answer sheet for the Q.1 will be collected by the invigilator after 45 minutes of the commencement of the examination.**
- **Out of the remaining EIGHT Questions answer any FIVE Questions. Each question carries 16 marks.**
- **Any required data not explicitly given, may be suitably assumed and stated.**

Q.1   **Choose the correct or the best alternative in the following:**   **(2×10)**

a. Data mining requires

   (A) large quantities of operational data stored over a period of time
   (B) lots of tactical data
   (C) several tape drives to store archival data
   (D) large mainframe computers

b. Characterization and Clustering belong to which of the following Data Mining type?

   (A) Predictive   **(B)** Selective
   (C) Descriptive   **(D)** Additive

c. OLAP stands for_____.

   (A) Online Analytical Program   **(B)** Online Analytical Processing
   (C) Online Application Program   **(D)** Online Analysis Performance

d. How are dimensions in a Multi-Dimensional Database related?

   (A) Hierarchically
   (B) Through foreign keys
   (C) Through a hierarchy and foreign keys
   (D) Through a network

e. Which of the following is a valid method of classification?

   (A) Unsupervised Learning
   (B) Bayesian Approach
   (C) Pattern Discovery and Data Cleaning
   (D) The Decision Tree Approach

f. Which of the following is not a type of Data Mining techniques?

  (A) Neural Network   (B) Decision Tree
  (C) Genetic Algorithm   (D) Classification

g. Clustering is an example of_____.

  (A) Unsupervised Learning   (B) Directed Learning
  (C) Supervised Learning   (D) Focused learning

h. Which of the following is not a component of Data Warehouse?

  (A) Data Warehouse Database
  (B) Sourcing, Acquisition, Cleanup and Transformation Tools
  (C) Data Mart
  (D) Document editor

i. Data Warehouse uses data from_____.

  (A) Daily reports
  (B) Diverse applications and locations
  (C) Diverse applications but from one location
  (D) A single Database

j. Which of the following is related to database structure in the context of a Data Warehouse?

  (A) Data Transformation   (B) Star Schema
  (C) Access Tools   (D) Global Schema

Answer any FIVE Questions out of EIGHT Questions.
Each question carries 16 marks.

Q.2 a. What do you mean by the term Data Mining? List and briefly explain various steps involved in Data Mining Process. **(8)**

  b. Discuss few features on the basis of which a data mining system should be assessed. **(8)**

Q.3 a. What do you mean by data reduction? Explain any two data reduction techniques. **(8)**

  b. What is the role of Data Acquisition, Clean up and Transformation tools in a Data Warehouse? Briefly comment on the role of Meta Data in a Data Warehouse. **(8)**

Q.4   a.   Explain the term Data Warehouse. What are the components of a Data warehouse? Discuss.   **(8)**

   b.   What are the underlying technologies available for creating a Data Warehouse environment? Discuss.   **(8)**

Q.5   a.   Discuss in brief four cube computation methods.   **(10)**

   b.   What is meant by full and partial materialization? Write a short note on Iceberg cube.   **(6)**

Q.6   a.   What do you mean by the term Classification? What is its significance? Discuss various Classification Methods in brief.   **(8)**

   b.   Explain the Decision Trees. What are the applications of a Decision Tree? Discuss.   **(8)**

Q.7   a.   Is there any general strategy for improving classifier and predictor accuracy? If yes, then explain one such technique.   **(8)**

   b.   How does naïve Bayesian classifier work? Explain.   **(8)**

Q.8   a.   What is the need of clustering analysis in Data Mining? Explain various major clustering methods.   **(10)**

   b.   Give the algorithm for k-medoids clustering method.   **(6)**

Q.9   a.   "Data mining as the process of identifying valid, novel, potentially useful, and ultimately comprehensible understandable patterns or models in data to make crucial business decisions". Comment on this statement.   **(8)**

   b.   List and discuss some of the application areas and common applications of Data mining.   **(8)**