

AMIETE – IT (NEW SCHEME) - Code: AT78

Subject: DATA MINING & WAREHOUSING

Time: 3 Hours

Max. Marks: 100

DECEMBER 2010

NOTE: There are 9 Questions in all.

- **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 half an hour 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. Which of the following is not a Data Mining Functionality:

- (A) Association analysis (B) Cluster Analysis
(C) Classification & Prediction (D) Randomization

b. Which of the key words does not distinguish warehouses from other repositories:

- (A) Subject-oriented (B) Time-variant
(C) Volatile (D) Integrated

c. Which of the following is not an OLAP server

- (A) ROLAP (B) DOLAP
(C) MOLAP (D) HOLAP

d. Which of the following is a data smoothing technique

- (A) Histogram (B) Regression
(C) Correlation (D) Induction

e. Which of the following is not a basis for classification for an association rule:

- (A) Types of values (B) Dimensions of data involved
(C) Volume of data involved (D) Levels of abstractions involved

f. A decision tree is a flow chart like structure where:

- (A) each internal node denotes a test on an attribute
(B) each internal node is of type “yes” or “no”
(C) each branch represents an outcome of a test
(D) the leaf nodes represents the classes or class distribution

- g. Which of the following is not a category of clustering methods
- (A) Partitioning methods
 - (B) Relational method
 - (C) Hierarchical methods
 - (D) Grid-based methods
- h. Which of the following is not associated with data cubes:
- (A) fact table
 - (B) dimension table
 - (C) location
 - (D) dimension
- i. In which of the following areas the data mining techniques cannot be applied:
- (A) Medical-diagnosis
 - (B) Prediction in share market
 - (C) Retail Industry
 - (D) All of these
- j. Which of the following is not a commercial Data Mining System:
- (A) Mineset
 - (B) Clementine
 - (C) D2kMiner
 - (D) Enterprise Miner

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

- Q.2** a. Draw the architecture of a typical data mining system and briefly explain all of its components. (9)
- b. Discuss various issues related to mining methodology and user-interaction, in the field of data mining. (7)
- Q.3** a. Define OLAP? What are the four different types of OLAP Servers from implementational point of view? Explain briefly. (9)
- b. Explain the Three-tier datawarehouse architecture. What are the three data warehouse models from architecture point of view? (7)
- Q.4** a. Discuss various problems in data, which the data-cleaning methods can deal with. What are the different methods to deal with “missing values”? (9)
- b. What do you mean by data reduction? Explain any two data reductions techniques. (7)
- Q.5** a. How does data mining relate to information processing and OLAP? Discuss in detail. (9)
- b. What do you mean by association rule mining? Give an example of market basket analysis from the real-world. (7)

-
- Q.6** a. Compare classification and prediction methods. (6)
- b. How does tree pruning work? Explain with examples. (5)
- c. How classification is done by back-propagation. Give an example of a general multilayered feed-forward neural network. (6)
- Q.7** a. Discuss the various types of typical requirements for clustering in data-mining. (9)
- b. What are model-based clustering methods? Discuss the two major approaches viz. statistical approach and neural network based approach used here. (7)
- Q.8** a. Since many data mining products are available in the market, mention any five multiple dimensional features which shall be kept in mind before choosing such a system. (10)
- b. Briefly discuss how data mining may incorporate intelligent query answering. (6)
- Q.9** Write short notes on any **FOUR**:
- (i) Snowflake schema
 - (ii) Equal-width and Equal-frequency binning.
 - (iii) Attribute oriented induction
 - (iv) DBSCAN
 - (v) Audio data mining
 - (vi) Evolution analysis (16)