

Code: AT78 Subject: DATA MINING &amp; WAREHOUSING

**AMIETE – IT (Current Scheme)**

Time: 3 Hours

**JUNE 2015**

Max. Marks: 100

**PLEASE WRITE YOUR ROLL NO. AT THE SPACE PROVIDED ON EACH PAGE IMMEDIATELY AFTER RECEIVING THE QUESTION PAPER.**

**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 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)**

- Which of the following represent information granularity?  
(A) The extent of accuracy within the information can be updated by users  
(B) The extent of subjectivity within the information  
(C) The extent of detail within the information  
(D) The extent of strategy within the information
- A star schema has what type of relationship between a dimension and fact table?  
(A) Many-to-many  
(B) One-to-one  
(C) One-to-many  
(D) All of these
- \_\_\_\_\_ stores multidimensional aggregate information.  
(A) Data cube  
(B) Data Mart  
(C) Both (A) & (B)  
(D) None of these
- The active data warehouse architecture includes which of the following?  
(A) At least one data mart.  
(B) Data that can be extracted from numerous internal and external sources.  
(C) Near real time updates  
(D) All of these
- All of the following terms describe OLAP, except:  
(A) The gathering of input information  
(B) Processing input information  
(C) Updating existing information to reflect to the gathered and processed information  
(D) None of these

**Code: AT78      Subject: DATA MINING & WAREHOUSING**

- f. \_\_\_\_\_ includes normalization and aggregation as data pre-processing procedures.
- (A) Data transformation                      (B) Data cleansing  
(C) Data reduction                            (D) Data integration
- g. Data scrubbing is
- (A) A process to reject data from the data warehouse and to create the necessary indexes  
(B) A process to load the data in the data warehouse and to create the necessary indexes  
(C) A process to upgrade the quality of data before it is moved into a data warehouse  
(D) A process to upgrade the quality of data after it is moved into a data warehouse
- h. The generic two-level data warehouse architecture includes which of the following?
- (A) At least one data mart  
(B) Data that can be extracted from numerous internal and external sources  
(C) Near real-time updates  
(D) All of these
- i. Data transformation includes which of the following
- (A) A process to change data from a detailed level to a summary level  
(B) A process to change data from a summary level to a detailed level  
(C) Joining data from one source into various sources of data  
(D) Separating data from one source into various sources of data
- j. The load and index is
- (A) A process to reject data from the data warehouse and to create the necessary indexes  
(B) A process to load the data in the data warehouse and to create the necessary indexes  
(C) A process to upgrade the quality of data after it is moved into a data warehouse  
(D) A process to upgrade the quality of data before it is moved into a data warehouse

---

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

---

- Q.2** a. Where is data mining is used? (4)
- b. Describe the steps in data mining. (4)
- c. Explain the different criteria based on which the data mining systems can be categorized. (8)

**Code: AT78      Subject: DATA MINING & WAREHOUSING**

- Q.3** a. Explain the following concepts:  
(i) Methods for Data transformation  
(ii) Various Strategies for Data Reduction (2×4)
- b. Explain the following data discretization technique  
(i) Entropy-Based Discretization  
(ii) Histogram Analysis (2×4)
- Q.4** a. Explain how the data warehouse acts as a basis for EIS. (8)
- b. Explain drill-down analysis and event mapping in context of EIS. (8)
- Q.5** a. Discuss various OLAP operations in the multidimensional data modal. (7)
- b. What is generalization and its approaches? (5)
- c. Define Iceberg Cube and Shell Cube. (4)
- Q.6** a. What is Data Classification Process? How it is differ than Predication? (8)
- b. Why is tree pruning useful in decision tree induction? (2)
- c. Prove that all nonempty subsets of a frequent itemset must also be frequent. (6)
- Q.7** a. Why is naive Bayesian classification called “naive”? Briefly outline the major ideas of naive Bayesian classification. (8)
- b. What are the technology changes in bringing the system-of-record data into the data warehouse? (8)
- Q.8** a. What do you mean by cluster analysis? What are the typical requirements of clustering in data mining? (7)
- b. Explain the PAM (partition around medoids) algorithm and its phases. (6)
- c. Define outlier mining. (3)
- Q.9** a. Explain how data mining used in retail industry and in biomedical field. (8)
- b. What are mining? Explain the techniques in web mining. (8)