

AMIETE – IT (NEW SCHEME)

Time: 3 Hours

JUNE 2012

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)

- a. Data mining is the process of:
- (A) retrieving subjective data
 - (B) extracting hidden pattern in historical data
 - (C) classifying the data
 - (D) None of the above
- b. Facts tables are which of the following:
- (A) Completely Denormalized
 - (B) Partially Denormalized
 - (C) Completely Normalized
 - (D) Partially Normalized
- c. 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
- d. A snowflake schema is which of the following types of tables?
- (A) Fact
 - (B) Dimension
 - (C) Helper
 - (D) All of the above
- e. A goal of data mining includes which of the following?
- (A) To explain some observed event or condition
 - (B) To confirm
 - (C) To analyze data for expected relationships
 - (D) To create a new data warehouse
- f. 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 the above

Code: AT78

Subject: DATA MINING & WAREHOUSING

- g. Which of the following is not correct for the Data in a Data Warehouse?
- | | |
|------------------|------------------|
| (A) Integrated | (B) Time variant |
| (C) Non-volatile | (D) Volatile |
- h. In ID3 Algorithm ID stands for'
- | | |
|----------------------------|----------------------------|
| (A) Iterative Dichotomiser | (B) Iterative Dichotometer |
| (C) Implied Definition | (D) Interactive Data |
- i. Which of the following is not an example of Data Mining Tools?
- | | |
|-----------------------|---------------|
| (A) Intelligent Miner | (B) Data Mind |
| (C) Discovery Server | (D) MinerData |
- j. Which of the following best describes the transient data:
- (A) Data in which changes to existing records cause the previous version of the records to be eliminated
 - (B) Data in which changes to existing records do not cause the previous version of the records to be eliminated
 - (C) Data that are never altered or deleted once they have been added
 - (D) Data that are never deleted once they have been added

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

- Q.2** a. What kind of process a data mining application can perform? List and explain the functions of a data mining application in brief. **(8)**
- b. Give the classification of data mining systems. **(4)**
- c. Mention any four data mining task primitives. **(4)**
- Q.3** a. Explain various data cleaning techniques used in data pre processing phase. **(6)**
- b. Explain various phases of data reduction and data transformation in data pre processing phase. **(10)**
- Q.4** a. Discuss the concept of cube and cuboids. What are the operations that can be performed on a data cube? List and explain each in brief. **(4+4)**
- b. Explain data generalization in attribute oriented induction. Explain various mining class comparisons in attribute oriented induction. **(4+4)**
- Q.5** a. What is the need of a Data Warehouse application? Draw the architecture of data warehouse. Mention back-end tools used in data warehousing. **(8)**

- b. How OLAP Application help in business decision making? Compare and contrast OLAP and OLTP. (8)
- Q.6** a. Explain mining rules for the following:
- (i) Multi-level association
 - (ii) Multi-dimensional association (4+4)
- b. What is the significance of classification? Discuss various types of classification techniques in decision tree induction. (8)
- Q.7** a. Explain the terms classification and Prediction. Give a brief note on conceptual clustering with example. (4)
- b. Give an example to illustrate:
- (i) Bayesian Belief networks
 - (ii) Rule extraction from decision tree. (4+4)
- c. Explain the ensembling methods like bagging and boosting increase accuracy. (4)
- Q.8** a. How Data Mining can be helpful in the information available through World Wide Web? Discuss. (8)
- b. Explain privacy and data security in social impacts of data mining. (4)
- c. Mention any **TWO** features for each of the following:
- (i) Statistical data mining
 - (ii) Audio data mining. (4)
- Q.9** a. What is Clustering? What are different clustering methods? Discuss Partition method of clustering. (8)
- b. Explain cluster analysis for each of the following:
- (i) Density based methods
 - (ii) Hierarchical methods (4+4)