

Code: AT78

Subject: DATA MINING &amp; WAREHOUSING

AMIETE – IT

Time: 3 Hours

DECEMBER-2014

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. 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
- b. Bayes Theorem is:
- (A)  $P(H|X)=P(X|H)P(H)/P(X)$       (B)  $P(H|X)=P(X|H)P(X)/P(H)$   
 (C)  $P(X|H)=P(X|H)P(H)/P(X)$       (D)  $P(X|H)=P(X|H)P(X)/P(H)$
- c. 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
- d. Time variant nature of the data in the data warehouse includes:
- (A) Allows for the analysis of the past  
 (B) Relate information to the present  
 (C) Enables forecast the future  
 (D) All of these
- e. A goal of data mining includes which of the following?
- (A) To explain some observed event or condition  
 (B) To confirm that data exists  
 (C) To analyze data for expected relationships  
 (D) To create a new data warehouse
- f. A snowflake schema is which of the following types of tables?
- (A) Fact      (B) Dimension  
 (C) Helper      (D) All of these

- g. 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
- h. \_\_\_\_\_ includes normalization and aggregation as data pre-processing procedures.
- (A) Data transformation
  - (B) Data cleansing
  - (C) Data reduction
  - (D) Data integration
- 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 which of the following?
- (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.**

---

- |            |   |            |
|------------|---|------------|
| <b>Q.2</b> | a. Explain the architecture of data mining system?                  | <b>(8)</b> |
|            | b. Why are decision tree classifiers so popular? Explain.           | <b>(8)</b> |
| <b>Q.3</b> | a. Explain the snowflake schema with the help of examples.          | <b>(8)</b> |
|            | b. What are various steps of data preprocessing?                    | <b>(8)</b> |
| <b>Q.4</b> | a. Explain how the data warehouse acts as a basis for EIS.          | <b>(8)</b> |
|            | b. Explain drill-down analysis and event mapping in context of EIS. | <b>(8)</b> |

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

- Q.5** a. Discuss the problem related to use and storage of unstructured data in data warehouse? What are the methods to capture and store external information? (8)
- b. Explain multiway array aggregation scheme for full cube computation. (8)
- Q.6** a. Discuss the criteria used to compare and evaluation of the classification and prediction method. (4)
- b. How classification is done by back-propagation. Give an example of a general multilayered feed-forward neural network. (8)
- c. Association rule mining often generates a large number of rules. Discuss effective methods that can be used to reduce the number of rules generated while still preserving most of the interesting rules. (4)
- Q.7** a. Describe Market Basket Analysis with suitable example. (10)
- b. For class characterization, what are the major differences between a data cube-based implementation and a relational implementation such as attribute-oriented induction? Discuss which method is most efficient and under what conditions this is so. (6)
- Q.8** a. Differentiate between partitioning methods and density based methods of cluster analysis. (8)
- b. State why, for the integration of multiple heterogeneous information source, many companies in industry prefer the update-driven approach (which constructs and uses data warehouse), rather than the query-driven approach (which applies wrappers and integrators)? Describe situation where the query-driven approach is preferable over the update-driven approach. (8)
- Q.9** a. What is web mining? Explain the types of web mining. (8)
- b. What are the major challenges faced in bringing data mining research to market? Illustrate one data mining research issue that, in your view, may have a strong impact on the market and on society. Discuss how to approach such a research issue. (8)