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

AMIETE – IT (Current Scheme)

Time: 3 Hours   **DECEMBER 2018**   **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. Extreme values that occur infrequently are called as
   (A) Rare Values   **(B)** Dimensionality reduction
   (C) Outliers   **(D)** Both **(A)** and **(B)**

b. Mode of the data set {13, 3, 11, 24, 5, 3, 9, 2} is
   (A) 24   **(B)** 3
   (C) 9   **(D)** None of these

c. The output of KDD is
   (A) Useful information   **(B)** Information
   (C) Query   **(D)** Data

d. Data scrubbing is a process to
   (A) upgrade the quality of data after it is moved into a data warehouse
   (B) load the data in the data warehouse and to create the necessary indexes
   (C) upgrade the quality of data before it is moved into a data warehouse
   (D) reject data from the data warehouse and to create the necessary indexes

e. A star schema has what type of relationship between a dimension and fact table?
   (A) many-to-many   **(B)** one-to-many
   (C) one-to-one   **(D)** All of these

f. _____ includes normalization and aggregation as data pre-processing procedures.
   (A) Data reduction   **(B)** Data cleansing
   (C) Data transformation   **(D)** Data integration

g. Which operation performs data aggregation by climbing up a dimensional hierarchy?
   (A) Roll-up   **(B)** Dice
   (C) Slice   **(D)** Drill-down

h. Which of the following is not an OLAP operation?
   (A) Drill-up   **(B)** Drill-down
   (C) Drill-through   **(D)** Drill-across

i. Which of the following is a data smoothing technique?
   (A) Histogram   **(B)** Regression
   (C) Correlation   **(D)** Induction

j. The Apriori Algorithm is used for
   (A) Classification   **(B)** Clustering
   (C) Association   **(D)** Regression

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

Q.2 a. Describe major challenges regarding data mining & society along with efficiency & scalability issue. **(8)**

     b. List and describe the five primitives for specifying a data mining task. **(8)**

Q.3 a. Explain the need of data discretization. How is it performed? In the same context explain concept hierarchy generation. **(2+3+3)**

     b. Explain various data cleaning techniques used in data pre processing phase. **(8)**

Q.4 a. Explain drill-down analysis and event mapping in context of EIS. **(8)**

     b. Explain the procedure for class comparison. **(8)**

Q.5 a. What is Generalization? How it is useful in data transformation? Give an example. **(4)**

     b. Define Iceberg Cube and Shell Cube. **(4)**

     c. Discuss how to support quality drill down although some low level cells may contain empty or too less data for reliable analysis. **(8)**

Q.6 a. How classification is done by back-propagation. Give an example of a general multilayered feed-forward neural network. **(8)**

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

Q.7 a. Why is naive Bayesian classification called "naive"? Briefly outline the major ideas of naive Bayesian classification. **(8)**

   b. Give the underlying principle of neural networks. State its advantages and shortcomings. What are the applications of this technique in today's scenario? **(8)**

Q.8 a. Differentiate between partitioning methods and density based methods of cluster analysis. **(8)**

   b. Give a description of types of data in Cluster Analysis. **(8)**

Q.9 a. Explain the social impact of data mining. **(8)**

   b. Describe few major applications of data mining in the business area. **(8)**