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

AMIETE – IT   (Current Scheme)

Time: 3 Hours   **DECEMBER 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)**

a. An essential process where intelligent methods are applied to extract data patterns is called
   (**A**) Data warehousing   (**B**) Data mining
   (**C**) Text mining   (**D**) Data selection

b. Which of the following is not a data mining functionality?
   (**A**) Characterization and Discrimination (**B**) Classification and regression
   (**C**) Selection and interpretation   (**D**) Clustering and Analysis

c. .... is a summarization of the general characteristics or features of a target class of data.
   (**A**) Data Characterization   (**B**) Data Classification
   (**C**) Data discrimination   (**D**) Data selection

d. Strategic value of data mining is
   (**A**) cost-sensitive   (**B**) work-sensitive
   (**C**) time-sensitive   (**D**) technical-sensitive

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

f. The full form of OLAP is
   (**A**) Online Analytical Processing   (**B**) Online Advanced Processing
   (**C**) Online Advanced Preparation   (**D**) Online Analytical Performance

g. The data is stored, retrieved and updated in
   (**A**) OLAP   (**B**) OLTP
   (**C**) SMTP   (**D**) FTP

h. Which view exposes the information being captured, stored, and managed by operational systems?
   (**A**) top-down view   (**B**) data warehouse view
   (**C**) data source view   (**D**) business query view

i.   Which of the following is not a component of a data warehouse?
   (A) Metadata   **(B)** Current detail data
   (C) Lightly summarized data   **(D)** Component Key

j.   The type of relationship in star schema is ....
   (A) many to many   **(B)** one to one
   (C) one to many   **(D)** many to one

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

Q.2 a. Discuss data mining task primitives in detail.   **(8)**

   b. Describe briefly the Data Mining functionalities, highlighting the kind of patterns that can be mined.   **(8)**

Q.3 a. Describe the process of data cleansing.   **(8)**

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

Q.4 a. Discuss the three-tier data warehouse architecture.   **(8)**

   b. Explain how data mining system can be integrated with database/dataware house system.   **(8)**

Q.5 a. What do you understand by "data cube"? Explain the terms Full cube, Iceberg cube, Closed cube and Shell cube.   **(2+6)**

   b. What is attribute oriented induction? What is its use in data characterization, explain with examples?   **(3+5)**

Q.6 a. Explain Bayesian classification with suitable example.   **(8)**

   b. Give the difference between classification and prediction.   **(4)**

   c. Discuss different types of association rules in data mining.   **(4)**

Q.7 a. What is back propagation? How is it used to classify patterns?   **(4+4)**

   b. Explain linear and non-linear methods of prediction. In which case will you use each, explain with example.   **(4+4)**

Q.8 a. Describe the working of Partitioning Around Medoids (PAM) algorithm.   **(8)**

   b. What is cluster analysis, how is it different from classification? Discuss the types of data that occur in cluster analysis.   **(3+5)**

Q.9 a. Describe various applications of data mining.   **(8)**

   b. What is web mining? Explain the techniques in web mining.   **(8)**