

Time: 3 Hours

DECEMBER 2013

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 data mining technique is used to find correlation among given data?

- | | |
|-----------------------------|--------------------|
| (A) Association rule mining | (B) Classification |
| (C) Clustering | (D) Prediction |

b. ETL does not include

- | | |
|----------------------|-------------------------------|
| (A) Finding data | (B) Deleting data |
| (C) Integrating data | (D) Placing data in warehouse |

c. Data Mining includes

- (A) Analyzing large volumes of data to discover interesting associations or patterns.
- (B) Querying a large data warehouse to uncover undiscovered facts.
- (C) Very complex SQL query operations.
- (D) Slicing and dicing until you uncover interesting details.

d. Which of the following is true?

- (A) Snowflake Schema is normalized star schema.
- (B) Dimension tables are normalized in Snowflake Schema.
- (C) Both (A) and (B) are true
- (D) None of these

e. What is created in association with metadata on inclusion of an external data in the data warehouse?

- | | |
|------------------------|-----------------------|
| (A) Data Mart | (B) Notification data |
| (C) External reference | (D) Structure of data |

Code: AT78 Subject: DATA MINING & WAREHOUSING

- b. How to handle Missing values and Noise in the data cleaning process. **(6)**
- c. Use z – score normalization to normalize the following group of data:
200, 300, 400, 600, 1000 **(4)**
- Q.6** a. Explain classification methods. How it is done by back-propagation? **(8)**
- b. How does tree pruning work? Explain with examples two common approaches to tree pruning – pre pruning and post pruning. **(8)**
- Q.7** a. Discuss any four methods that can be used to improve the efficiency of Apriori – based mining. **(8)**
- b. What is boosting? State why it may improve the accuracy of decision tree induction. **(8)**
- Q.8** a. Briefly describe the following approaches to clustering:
Density – based, grid – based, model – based and partitioning **(8)**
- b. How can you compute the dissimilarity between (i) objects described by ratio – scaled variables (ii) objects of mixed variable types? Give suitable example to explain. **(8)**
- Q.9** Write short notes on any **FOUR** of the following: **(4×4)**
- (i) Graph -based mining
 - (ii) Web mining
 - (iii) Intelligent miner
 - (iv) Audio data mining
 - (v) Collaborative filtering