

TCS-833**B. TECH. (CS) (EIGHTH SEMESTER)
MID SEMESTER EXAMINATION, 2019****DATA MINING AND WAREHOUSING****Time : 1 : 30 Hours****Maximum Marks : 50**

Note : (i) This question paper contains two Sections.

(ii) Both Sections are compulsory.

Section—A

1. Fill in the blanks/True-False : (1×5=5 Marks)
 - (a) Sampling is a commonly used approach for selecting a subset of the data objects to be analyzed. (True/False)
 - (b) Regression is an example of supervised learning. (True/False)
 - (c) where irrelevant, weakly relevant or redundant attributes may be deducted and removed.

(2)

TCS-833

- (d) Data warehouse architecture is based on
- (e) Binning is a method to handle the noisy data. (True/False)

2. Attempt any *five* parts : (3×5=15 Marks)

- (a) Difference between DBMS and Data warehouse.
- (b) List any *three* application areas of data mining.
- (c) Distinguish between noise and outliers.
- (d) OLAP vs. OLTP
- (e) What do you mean by descriptive data mining and predictive data mining ?
- (f) Why is data pre-processing required ?

Section—B

3. Attempt any *two* parts of choice from (a), (b) and (c). (5×2=10 Marks)

- (a) Define data warehousing. Explain the various features of data warehouse.
- (b) What is data mining ? Explain the steps involved in data mining process. Give the sketch of the KDD process.

F. No. : a-40

(3)

TCS-833

- (c) Use the two methods below to normalize the following group of data :

200, 300, 400, 600, 1000

- (i) min-max normalization by setting min = 0 and max = 1

- (ii) z-score normalization

4. Attempt any *two* parts of choice from (a), (b) and (c). (5×2=10 Marks)

- (a) Discuss the various operations of OLAP.
- (b) Why processes the data ? What is data cleaning ? How can we go about filling in the missing values for the attribute ?
- (c) Suppose a group of 12 sales price records has been sorted as follows :

5, 10, 11, 13, 15, 35, 50, 55, 72, 92, 204, 215

Partition them into three bins by each of the following methods :

- (i) Equal-frequency partitioning
- (ii) Equal-width partitioning
- (iii) Clustering

F. No. : a-40

P. T. O.

5. Attempt any *two* parts of choice from (a), (b) and (c). (5×2=10 Marks)

- (a) Write short notes on the following :
- (i) Characterization and Discrimination
 - (ii) Discretization
- (b) What are the different components of Data warehouse ? Explain.
- (c) What is fact table ? Explain in brief, the different schemas for multidimensional data models.