

Total No. of Questions :6]

SEAT No. :

P281

Oct./BE/Insem. - 599

[Total No. of Pages : 2

B.E. (Computer Engineering)

DATA MINING AND WAREHOUSING

(2015 Pattern) (Semester - I) (Elective-I) (410244 - D)

Time : 1 Hour]

[Max. Marks : 30

Instructions to the candidates:

- 1) *Answer Q1 or Q2, Q3 or Q4 and Q5 or Q6.*
- 2) *Figure to the right side indicate full marks.*

Q1) a) Explain various data reduction techniques. **[5]**

b) Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70. **[5]**

- i) What is the mean of the data? What is the median?
- ii) What is the mode of the data? Comment on the data's modality (i.e., bimodal, trimodal etc.)
- iii) Give the five-number summary of the data.

OR

Q2) a) Explain the techniques to handle noisy data. **[5]**

b) Make use of the two methods below to normalize the following group of data : 100, 300, 450, 650, 779 **[5]**

- i) Decimal scaling
- ii) Z-score normalization

Q3) a) Explain OLAP operations in Multidimensional data model. **[5]**

b) Compare OLAP and OLTP. **[5]**

OR

P.T.O.

- Q4)** a) Explain three-tier data warehousing architecture. [5]
b) Explain three categories of measures used in multidimensional data. [5]

- Q5)** a) Compute cosine similarity among following documents using term frequency vector. [5]

d1 : “The sun in the sky is bright”

d2 : “We can see the shining sun, the bright sun”

- b) Explain types of attributes. [5]

OR

- Q6)** a) Explain following with example : [5]

i) Minkowski distance

ii) Euclidean distance

- b) Explain the proximity measures for Binary attributes. [5]

