Duration: 3hrs **[Max Marks: 80]**

N.B. : (1) Question No 1 is Compulsory.
(2) Attempt any three questions out of the remaining five.
(3) All questions carry equal marks.
(4) Assume suitable data, if required and state it clearly.

1 Attempt any FOUR [20]

A Draw a three tier data warehousing architecture

B Data : 4, 8, 15, 21, 21, 24, 25, 28, 34
Divide data in 3 bins (equal frequency) and perform smoothing by bin means and smoothing by bin boundaries on every bin

C How to calculate correlation coefficient for two numeric attributes and also comment on the significance of this value

D Write a short note on support and confidence

E Explain the concept of information gain which is used in decision tree algorithm?

2 A Describe any two methods of data reduction [10]

B Compare star schema, snowflake schema and fact constellation [10]

3 A Write and explain Bayes classification algorithm [10]

B Write the steps of Ada-boost algorithm [10]

4 A How is data mining used in Business Intelligence? [10]

B Give the overview of partition clustering methods [10]

5 A How can we further improve the efficiency of Apriori-based mining? [10]

B Explain OLAP operations with the examples [10]

6 A Describe the classification performance evaluation measures that are obtained from confusion matrix? [10]

B Use the normalization methods to normalize the following group of data: 200, 300, 400, 600,1000
Use min-max normalization by setting min = 0 and max = 1 and z-score normalization [10]

****