Roll No. ☐☐☐☐☐☐☐☐☐☐☐   Total No. of Pages : 01

Total No. of Questions : 08

M.Tech. (CSE) (2019) (Sem.–2)
DATA PREPARATION AND ANALYSIS
Subject Code : MTCS-206-18
M.Code : 76057

Time : 3 Hrs.   Max. Marks : 60

INSTRUCTIONS TO CANDIDATES :
1.Attempt any FIVE questions out of EIGHT questions.
2.Each question carries TWELVE marks.

1. A. Suppose that the data for analysis includes the attribute age. The age values for the data tuples are 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.

   Explain with example, how might you apply data smoothing?

   B. Explain association rule mining and its parameters such as support and confidence with the help of example.

2. How scalability can be measured in various dimensions during data gathering process? Explain the issues related to scalability with the explanation and example.

3. How to perform data consistency and what are the set of measures to check data inconsistency? Take an example of application consistency in databases for data consistency.

4. A. What is missing data? When and why is data missed? How to handle missing data in real time?

   B. What are the various data transformational languages?

5. Data segmentation is the process of taking your data and segmenting it so that you can use it more efficiently within marketing and operations. How can data segmentation help with personalisation?

6. What is clustering? Explain various applications of cluster analysis. What are the requirements of Clustering in Data analysis?

7. Imagine you were building a product to help people buy shoes online. How to generate and validate hypothesis for shoe-buying module?

8. Explain data visualisation. What are the methods to visualize the data to track time series data and geo-located data?

NOTE : Disclosure of Identity by writing Mobile No. or Making of passing request on any page of Answer Sheet will lead to UMC against the Student.