No. of Printed Pages : 3

**BICS-020**

# B.Tech. – VIEP – COMPUTER SCIENCE AND ENGINEERING (BTCSVI)

## Term-End Examination

00563

## June, 2018

## BICS-020 : DATA WAREHOUSING AND MINING

*Time : 3 hours* *Maximum Marks : 70*

***Note :*** *Attempt any **seven** questions. All questions carry equal marks.*

1. Explain the concept of data warehouse. Write the main steps required for constructing a data warehouse. *10*

2. What is data mining ? Describe the steps involved in data mining when viewed as a process of knowledge discovery. *10*

3. Describe various methods for handling missing values in real world data sets. *10*

4. Suppose that the data for analysis includes the attribute age. The age values for the data types 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. *10*

   (a) Use min-max normalization to transform the value 35 for age onto the range [0·0, 1·0].

   (b) Use "smoothing by bin means" to smooth a given data, using a bin depth of 3.

5. Briefly describe the similarities and differences of star and snowflake schemas and analyze their advantages and disadvantages. *10*

6. Compare and distinguish between the following concepts. You may use examples to explain your point(s) : *10*

   (a) Enterprise warehouse, data mart and virtual warehouse

   (b) Tuning and testing data warehouse

7. What is association rule mining ? Explain with an example. What is the difference between uniform support and reduced support in multilevel association ? *10*

8. Briefly explain the following terms to be used in association rule mining : *10*

   (a) Support

   (b) Confidence

   (c) Frequent patterns

   (d) Strong rules

   (e) Correlation

9. Briefly outline the major steps of decision tree classification. *10*

10. Briefly outline how to compute the dissimilarity between the following types of variables : *10*

    (a) Categorical Variables

    (b) Numerical Variables

    (c) Asymmetric Binary Variables

    (d) Ratio-Scaled Variables