

2 0 2 1

(July)

COMPUTER SCIENCE

(Honours)

(Data Mining)

(CS-602 CT)

Marks : 75

Time : 3 hours

The figures in the margin indicate full marks for the questions

Answer **one** question from each Unit

UNIT—I

1. Explain two major issues in data mining. 2½+2½=5
2. Explain two data mining tasks/ functionalities. 2½+2½=5

UNIT—II

3. Suppose the data for analysis include the attribute income (in ₹). The values for income are (in increasing order) 10,000, 12,000, 13,000, 15,000, 15,000, 15,500, 16,000, 17,000 and 20,000. Answer the following :
 - (a) Use min-max normalization to transform the value of income ₹ 15,000 onto the range [0, 100]. 3

- (b) Find standard deviation and use z-score normalization to transform the value 15,000 of income. 3+3=6
- (c) Use normalization by decimal scaling to transform the value 15,000. 3
- (d) Comment on which method you would prefer to use for the given data, giving reasons as to why. 3

4. (a) What is a data warehouse? Elaborate on the characteristics of a data warehouse. 4+5=9
- (b) With the help of examples, differentiate between star schema and snowflake schema. 3+3=6

UNIT—III

5. (a) Use a priori algorithm to find the frequent item sets from the following data set. Show the results of every step. Assume minimum support as 40% : 10

Trans ID	Data items
1	D, E
2	A, B, C, D, E
3	B, C, D, E
4	B, C
5	A, B, C

(3)

- (b) Write short notes on metarule guided association rule mining and constraint-based association rule mining. 5+5=10
6. (a) Explain the FP-tree algorithm to find frequent item sets. 15
- (b) Write a short note on constraint-based association rule mining. 5

UNIT—IV

7. (a) Explain how prediction can be done using linear regression. 6
- (b) Explain how DBSCAN algorithm works. 12
- (c) Explain the BIRCH algorithm for cluster analysis. 12
8. (a) Give two applications of cluster analysis. Use k-means clustering algorithm to find three clusters from the following data set :
- 100, 30, 40, 60, 60, 10, 30, 20, 90, 50, 30, 10, 65, 10
- Also show the result of each step. 4+16=20
- (b) Explain two methods of outlier detection. 5+5=10

(4)

UNIT—V

9. Explain how data mining can be used for ATM card fraud detection. 5
10. Give five factors to choose a data mining system. 5
