

Roll No.

Total Pages : 03

CMCSQ/D-23

24043

DATA MINING AND ANALYTICS USING R
MS-20-31

Time : Three Hours]

[Maximum Marks : 75

Note : Attempt *Five* questions in all, selecting *one* question from each Unit. Q. No. 1 is compulsory.

(Compulsory Question)

1. (a) What are the key features that distinguish a data warehouse from other data repository systems ?
- (b) What is meant by mining associations ?
- (c) Differentiate between qualitative and quantitative data.
- (d) Enlist the various data types and data structures in R.
- (e) What are the various mining algorithm interfaces available in R ?
- 5×3=15**

Unit I

2. (a) Discuss the multi-layered architecture of a data warehouse in detail.
- 7.5**

- (b) What are the major issues in data mining ? Explain in detail. 7.5
- 3. (a) What are the various ways for data cleaning ? Explain using suitable examples. 7.5
- (b) Discuss the various ways for data discretization in brief using suitable examples. 7.5

Unit II

- 4. How can you perform frequent-set mining using the Apriori algorithm ? Explain in detail by writing the Apriori algorithm and using suitable examples. 15
- 5. How can you create a decision tree ? Explain by writing an algorithm and using a suitable example. Also, discuss the concept of tree pruning using suitable examples. 15

Unit III

- 6. (a) Suppose we wish to represent the relationship between the data using visualization. Discuss the various ways to do this using suitable examples. 7.5
- (b) What are the various problems associated with data analysis ? Explain in detail. 7.5

- 7. (a) How can you manage the data analysis process ? Explain. 7.5
- (b) What are the various data strategies ? Discuss using suitable examples. 7.5

Unit IV

- 8. What are the various types of problems in which tree-based regression may be useful ? How the decision tree be drawn in R ? Explain using suitable examples. 15
- 9. Comment on the need for function in R. How user-defined functions are different from built-in functions ? Describe the various types of built-in functions in R used in data analytics. 15