Code: 15A05602

R15

B.Tech III Year II Semester (R15) Regular Examinations May/June 2018
DATA WAREHOUSING & MINING
(Computer Science & Engineering)

Time: 3 hours

Max. Marks: 70

PART – A
(Compulsory Question)
*****

1   Answer the following: (10 X 02 = 20 Marks)
   (a)   Define data mining. On what kind of data, data mining can be performed?
   (b)   Why do we need to preprocess the data?
   (c)   Bring out the differences between ROLAP and MOLAP server.
   (d)   Define base and apex cuboid with appropriate example.
   (e)   Differentiate between prediction and classification.
   (f)   Define single and multidimensional association rule mining with appropriate example.
   (g)   What are binary variables? Provide appropriate examples.
   (h)   Which method is more robust among K-means and K-medoids clustering? Justify your answer.
   (i)   What is sequential pattern mining?
   (j)   What is spatial data mining?

PART – B
(Answer all five units, 5 X 10 = 50 Marks)

UNIT – I

2   Why do need to transform data? Explain the different approaches for data transformation. For the given data normalize the data 400 using min max normalization and decimal scaling method to the range 0 to 1.

Data: 200, 300, 400, 600, 800, 1000

OR

3   Briefly describe the various data mining functionalities.

UNIT – II

4   (a)   Compare data enterprise, data mart and virtual data warehouse.
   (b)   Suppose that a data warehouse consists of the three dimensions: time, doctor and patient and two measures count and charges. Draw a star schema for the data warehouse.

OR

5   Describe the general optimization techniques for the efficient computation of data cubes.

UNIT – III

6   For the given transactional data, find the frequent item sets using Apriori algorithm. Given minimum support = 2.

Tid	List of item-ids
T1	p1, p2, p5, p8
T2	p2, p4, p7
T3	p2, p3
T4	p1, p2, p4
T5	p1, p3, p9
T6	p2, p3
T7	p1, p3
T8	p1, p2, p3, p5
T9	p1, p2, p3
T10	p10, p11

Given min confidence = 90%, find strong rules with 3 itemset.

OR

7   Describe back propagation algorithm.

UNIT – IV

8   Describe the major clustering methods.

OR

9   (a)   Explain constraints based clustering.

    (b)   Given two objects represented by tuples (22, 1, 42, 10) and (20, 0, 36, 8), compute the distance between the two objects using Euclidean, Manhattan and Minkowski distance (with q = 3).

UNIT – V

10   What is multi-relational data mining? Describe ILP approach for multi-relation classification.

OR

11   Describe multimedia data mining.

*****