Reg. No.									

MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL
A Constituent Institution of Manipal University

VI SEMESTER B.TECH. (COMPUTER AND COMMUNICATION ENGINEERING) MAKEUP EXAMINATIONS, JUNE 2017

SUBJECT: DATA MINING AND PREDICTIVE ANALYSIS [ICT 3252]

REVISED CREDIT SYSTEM

(17/06/2017)

Time: 3 Hours

MAX. MARKS: 50

Instructions to Candidates:

- ❖ Answer **ALL** the questions.
- ❖ Write the detailed steps for all the problems.
- ❖ Missing data, if any, may be suitably assumed.

1A. A database has five transactions as shown in Table Q.1A. Let minimum support = 60% and minimum confidence = 50%.

Table Q.1A

TID	Items bought
1	1, 2, 3, 4, 5, 6
2	2, 3, 5, 6, 7, 8
3	1, 2, 5, 8, 9
4	6, 7, 8, 10,11
5	1, 2, 5,8

    (i) Find all frequent itemsets by using FP- Tree growth algorithm.

    (ii) List all the strong association rules of the form {Item1, Item2} → {Item3}.   **5**

1B. Consider the sample database given Table Q.1B. Calculate the support and confidence for the following association rules.

    (i)Milk → {Bread, Butter} (ii){Bread, Butter} → {Egg, Milk} (iii) Milk → {Butter, Egg}

Table Q.1B

TID	Milk	Bread	Butter	Egg
1	1	1	0	0
2	0	0	1	0
3	1	1	0	1
4	1	1	1	1
5	0	1	1	1

3

1C. Explain the difference and similarity between characterization and clustering.   **2**

2A. Cluster the following points into 2 clusters by using k-medoid method with euclidean distance measure. X1(3,1), X2(4,3), X3(6,3), X4(8,3), X5(5,4), X6(7,4), X7(1,5), X8(5,5), X9(3,7), X10(5,7), X11(5,8), X12(4,9). Assume X1 and X12 as the initial cluster centers. Perform 2nd iteration by swapping X1 with X2 and 3rd iteration by swapping X12 with X8. Compare the cost of all the clusters.   **5**

2B. Given the values of variable *age*:18, 22, 25, 42, 28, 43, 33, 35, 56, 28, standardize the