| Reg. No. | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|

# MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

*A Constituent Institution of Manipal University*

**VI SEMESTER B.TECH. (INFORMATION TECHNOLOGY)**

**MAKEUP EXAMINATIONS, JUNE 2017**

SUBJECT: DATA WAREHOUSING AND DATA MINING [ICT 3202]

**REVISED CREDIT SYSTEM**
**( 17 /06/2017)**

Time: 3 Hours MAX. MARKS:

**Instructions to Candidates:**

- Answer **ALL** the questions.
- Missing data if any, may be suitably assumed.

**1A.** Find the value of the correlation coefficient from Table Q.1A and discuss the relationship between X and Y. Explain the three issues to be considered in the data integration step. 5

Table Q.1A

| | | | | | |
|---|---|---|---|---|---|
| X | 28 | 33 | 29 | 31 | 24 |
| Y | 10 | 25 | 1 | 5 | 18 |

**1B.** Normalize the data points 10,20,30,40 using: 3
a) min-max normalization by setting min=0 and max=1
b) z-score normalization
c) decimal scaling

**1C.** Find the variance of the following data points: 6, 7, 10, 11, 11, 13, 16, 18, 25. 2

**2A.** Write the k-medoid clustering algorithm. Consider the below mentioned data points. Let P1(2,2) be the cluster 1 medoid and P5(3,4) be the cluster 2 medoid. Apply k-medoid clustering algorithm and calculate the total cost for the given data-set. 5
P1 (2, 2), P2 (1, 14), P3 (10, 7), P4 (1, 11), P5 (3, 4), P6 (11, 8), P7 (4, 3), P8 (12, 9)

**2B.** Briefly list the steps of CLARA clustering algorithm. Discuss its strength and weaknesses with respect to the PAM algorithm. 3

**2C.** Compare web structure mining and web usage mining. 2

**3A.** Discuss the three-Tier architecture of a Data warehouse with a neat diagram. Explain the four types of OLAP servers. 5

**3B.** Find all the frequent item-sets from the transaction data-set given below by applying Apriori algorithm with minimum support =50%. Indicate all the steps. 3
T1: {8, 9,7,10} T2: {7, 10,11,8} T3: {11,12,8} T4:{7, 10}

**3C.** List the various ways of handling missing values in the data cleaning step of knowledge discovery from data. 2

**4A.** Construct a FP-tree and find the frequent patterns for the following data points. 5
T1:{ A,B,D,E} T2:{B,C,E} T3:{A,B,D,E} T4:{A,B,C,E} T5:{A,B,C,D,E}
T6:{B,C,D}. Assume min_sup=3.

**4B.** State the various data structures used in Dynamic Itemset Counting algorithm and 3

explain its roles. Justify how Dynamic Itemset Counting algorithm is an improvement over Apriori algorithm.

4C. Discuss any two OLAP operations by giving an example for each. 2

5A. Define null invariant correlation measure with an example. Compute the four correlation measures for the contingency Table Q.5A. 5

Table Q.5A

| | Computer | $\overline{Computer}$ |
|---|---|---|
| Printer | 80 | 18 |
| $\overline{Printer}$ | 40 | 42 |

5B. Draw the box plot for the following data points: 3
27, 28, 30, 42, 45, 50, 50, 61, 62, 64, 120

5C. Compute the dissimilarity matrix for the data given in Table Q.5C. 2

Table Q.5C

| Object ID | Grades (Ordinal) |
|---|---|
| 1 | A+ |
| 2 | A |
| 3 | A+ |
| 4 | B |