B. Tech. Sem - VIII (Computer Engg.) (2014 COURSE) (CBCS) :
SUMMER - 2019
SUBJECT: DATA MINING AND KNOWLEDGE DISCOVERY

Day: Thursday   Time: 02.30 PM TO 05.30 PM

Date: 30/05/2019   Max Marks: 60

S-2019-2891

N.B.:
1) All questions are **COMPULSORY**.
2) Use of non programmable calculator is **ALLOWED**.
3) Figures to the right indicate **FULL** marks.
4) Assume suitable data, if necessary.

Q.1   What is data preprocessing? List and explain different tasks/methods **(10)** involved in data preprocessing?

OR

Q.1   How data mining systems can be classified? Describe the different data **(10)** mining functionalities with example.

Q.2   Explain OLAP operations in multidimensional data models. Also **(10)** differentiate between ROLAP, MOLAP and HOLAP.

OR

Q.2   What is OLAP server? Describe the OLAP server architecture in detail. Also **(10)** differentiate between OLTP and OLAP.

Q.3   What are the major weaknesses of the Apriori algorithm? Explain the various **(10)** techniques to improve the efficiency of the Apriori algorithm.

OR

Q.3   A database has 5 transactions. Let minimum support = 60% and Minimum **(10)** confidence = 80%.

Transaction	Items Bought
T10	{ M, O, N, K, E, Y}
T20	{ D, O, N, K, E, Y}
T30	{ M, A, K, E}
T40	{ M, U, C, K, Y}
T50	{ C, O, O, K, I, E}

Find all frequent itemsets using Apriori and FP growth respectively. Compare the efficiency of the two mining process.

P.T.O.

Q.4   Explain the different attribute selection methods? Why is tree pruning useful **(10)** in decision tree induction? Also explain the drawbacks of using separate set of samples to evaluate pruning.

OR

Q.4   What is decision tree classifier? Explain in detail the phases of learning and **(10)** classification in the working of any classifier.

Q.5   Explain in detail the KDD process using the KDD system architecture.   **(10)**

OR

Q.5   Explain in detail the integration of KDD system with Database/Data **(10)** warehouse system.

Q.6   What is clustering? What are the different types of clusters? Explain in detail **(10)** the k-means and k-medoid clustering algorithm.

OR

Q.6   Using a single and complete link agglomerative clustering to group the data **(10)** described by the following distance matrix. Show the dendograms.

	A	B	C	D
A	0	1	4	5
B		0	2	6
C			0	3
D				0

*   *   *   *   *