USN												CMRIT

Sub:	DataMining and DataWarehousing				Sub Code:	17CS651/ 18CS561	Branch:	CSE
Date:	23/06/2021	Duration:	90 min's	Max Marks:	50	Sem / Sec:	VI- 15,17,18 Scheme	OBE

	Answer any FOUR FULL questions from Q1 to Q5, Attempt any 1 question from Q6,Q7.	MARKS	CO	RBT
1	a. Consider the following vectors. Find a) Simple Matching Coefficient b) Jaccard Coefficient c) Hamming Distance. i)X: 0101010001 Y: 0100011000 ii)X: 1000000000 Y: 0000001001 b. What is data mining? what are the applications of data mining.	[6] [2+2]	CO1	L1
2.	a. State and explain various Data mining tasks. b. Explain Knowledge data discovery KDD with a neat diagram.	[04] [06]	CO1	L2
3.	a. Describe the difference between ROLAP & MOLAP. b. Explain data Preprocessing steps.	[05] [05]	CO1	L3
4	a. What is association analysis? Define support and confidence with example. b. State Apriori Principle. Differentiate between Apriori and FP Growth methods of generating frequent item sets.	[05] [2+3]	CO2	L2, L3
5	Explain FP growth algorithm for discovering frequent item sets. What are its limitation.	[07+03]	CO1	L3
6.	A database has five transactions. Let min-sup = 60% and min-conf = 80% TID ITEM T1 {M, O, N, K, E, Y} T2 {D, O, N, K, E, Y} T3 {M, A, K, E} T4 {M, U, C, K, Y} T5 {C, O, O, K, I, E} Find all frequent item sets and strong association rules using apriori Algorithm.	[10]	CO1	L3
7.	Consider the following transaction data set: TID ITEM T100 I1, I2, I5 T200 I2, I4 T300 I2, I3 T400 I1, I2, I4 T500 I1, I3 T600 I2, I3 T700 I1, I3 T800 I1, I2, I3, I5 T900 I1, I2, I3 Construct FP Tree. Generate List of frequent item set ordered by their corresponding suffixes.	[10]	CO1	L2