KUMARAGURU
college of technology
character is life

B.TECH DEGREE EXAMINATIONS: DEC 2022

(Regulation 2018)

Fourth Semester

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

U18AII4203: Data Mining & Modeling

COURSE OUTCOMES

CO1: Understand about data mining basics, issues and the working principle of classification technique

CO2: Explain the basic concepts of Association Rule Mining and evaluate the working of various Association Rule Mining algorithms

CO3: Implement classification and prediction techniques

CO4: Analyze the working of different clustering algorithms

Time: Three Hours   **Maximum Marks: 100**

Answer all the Questions: -
PART A (10 x 2 = 20 Marks)
(Answer not more than 40 words)

1. Compare and contrast KDD and Data mining	CO1	[K2]
2. List the various forms of data pre-processing	CO1	[K1]
3. Build a code to forward fill the null value in the dataset	CO2	[K3]
4. Justify how confidence calculation plays a major role in association rule	CO2	[K2]
5. Differentiate operational database systems with data warehouse	CO3	[K2]
6. Define lazy learner. Give an example	CO3	[K2]
7. Describe data normalization methods	CO2	[K2]
8. Explain grid based clustering	CO3	[K2]
9. List any four tools which are extensively used in data mining	CO4	[K1]
10. Discuss the weaknesses of DBSCAN algorithm	CO4	[K2]

Answer any FIVE Questions:-
PART B (5 x 16 = 80 Marks)
(Answer not more than 400 words)

11. a)	Illustrate various steps in KDD processing with sketch	8	CO1	[K2]
b)	Explain various slicing method in data mining	8	CO1	[K2]

12. a) Suppose we have the following dataset that has various transactions, and from this dataset, we need to find the frequent itemsets and generate the association rules using the Apriori algorithm: 16 CO2 [K₃]

TID	ITEMSETS
T1	A, B
T2	B, D
T3	B, C
T4	A, B, D
T5	A, C
T6	B, C
T7	A, C
T8	A, B, C, E
T9	A, B, C

Given: Minimum Support= 2, Minimum Confidence= 50%

13. a) Write a python code for the following pre-processing function 16 CO2 [K₃]
   i. Backwardfill
   ii. Interpolation
   iii. Mean
   iv. Median

14. a) Apply bin by border method and clean the given dataset A=[6, 10, 11, 13, 15, 35, 51, 55, 72, 93, 205, 216] 6 CO3 [K₃]

   b) Explain agglomerative and divisive hierarchical clustering algorithm with an example 10 CO3 [K₂]

15. a) Consider the following data point and cluster the given points using density-based clustering 16 CO4 [K₃]

Points	X	Y
P1	2	10
P2	2	5
P3	8	4
P4	5	8
P5	7	5
P6	6	4
P7	1	2
P8	4	9

Where eps=2 and minPts=3

16. a) Apply centroid based clustering technique to cluster the following eight points (with x, y representing location) into three clusters: A1 (2, 10), A2 (2, 5), A3 (8, 4), A4 (5, 8), A5 (7, 5), A6 (6, 4), A7 (1, 2), & A8 (4, 9). Assume that the initial centroids are A1, A4, A7. 16 CO4 [K₃]

****