Register Number:……………..

KUMARAGURU
college of technology
character is life

B.TECH. DEGREE EXAMINATIONS: APRIL / MAY 2023

(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: Implement 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. Distinguish between Classification Vs Clustering techniques. CO1 [K_2]
2. Outline OLAP. CO1 [K_2]
3. List some of the applications of Knowledge Representation methods. CO4 [K_2]
4. Illustrate different types of constraints that can be applied in association mining. CO2 [K_2]
5. Summarize classification and prediction. CO3 [K_1]
6. Define Lazy Learners. CO4 [K_1]
7. List the steps involved in the cluster analysis. CO4 [K_1]
8. Define interestingness constraint. CO2 [K_1]
9. State confidence and support count. CO2 [K_1]
10. Develop a rule-based classifier model for student grading system. CO3 [K_2]

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

11. a) Explain the various stages of the Data Mining Process. 10 CO1 [K_2]

b) The following table shows the heights of sample of Eight fathers and their oldest adult sons. Find correlation coefficient and show that the heights of father and son are positively or negatively correlated. 6 CO1 [K_3]

x	y
165	167
166	168
167	165
167	168
168	172
169	172
170	169
172	171

12. a) Consider the 3D view of the home appliance sales data in the table below and perform the following operations: 6 CO1 [K_3]

	LOCATION=”CHENNAI”			LOCATION=”TRICHY”			LOCATION=”SALEM”		
TIME	**ITEMS**			**ITEMS**			**ITEMS**		
	TV	FRIDGE	GRINDER	TV	FRIDGE	GRINDER	TV	FRIDGE	GRINDER
Q1	77	25	45	71	12	11	81	64	51
Q2	23	91	53	37	45	35	52	58	63
Q3	15	19	28	55	68	67	16	31	41
Q4	17	18	75	52	15	89	45	63	15

i) Roll-up (or) Aggregation of cube for Salem location.

ii) Slice (or) Selection of the cube on quadrant Q1.

b) Develop a python code for the following preprocessing function 10 CO2 [K_3]

i. Forwardfill

ii. Backwardfill

iii. Interpolation

iv. Mean

v. Median

13. Assume that below table shows the set of items purchased in a fruit stall during different transactions showing the sequence of its evolution. 16 CO2 [K_3]

Tid	**Items Bought**
100	Apple, Banana, Cherries, Lemon
200	Apple, Banana, Lemon
300	Apple, Mango, Orange
400	Apple, Lemon, Mango
500	Banana, Lemon, Mango

i) Develop the frequent patterns from the conditional FP-Tree
ii) Find frequent itemset in single iteration with minimum support count of 2
iii) List any four association rule with minimum confidence = 50%

14. a) Demonstrate logistic regression with neat sketch 8 CO3 [K$_3$]

b) Examine the Bayesian classification. 8 CO3 [K$_2$]

15. a) Apply KNN classification on the following dataset and predict the quality of paper_5 having Acid Durability = 3 and Strength = 7 for K= 3 (Nearest Neighbor).The data from a survey and objective testing with two attributes (Acid durability and Strength) can be used to classify whether the quality of the sample paper is good or bad. The below table shows four training samples 8 CO3 [K$_3$]

Sample Paper	**Acid Durability**	**Strength**	**Quality**
Paper_1	7	7	Bad
Paper_2	7	4	Bad
Paper_3	3	4	Good
Paper_4	1	4	Good

Now consider a new sample paper called Paper_5 that passes laboratory tests with Acid Durability = 3 and Strength = 7. Without any expensive survey find out the quality of this new paper by using the KNN classifier.

b) Consider the following data point and cluster the given points using density-based clustering algorithm 8 CO4 [K$_2$]

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) Demonstrate the iteration process in agglomerative clustering. 8 CO4 [K_3]

b) Apply centroid based clustering technique to cluster the following five points (with x, y representing location) into two clusters: 8 CO4 [K_3]

A1 (2, 2)

A2 (3, 2)

A3 (1, 1)

A4 (3, 1)

A5 (1.5, 0.5)

Let initial interation1 centroid points as A1, A3.

****