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: 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.	Define Data Discretization.	CO1	[K$_1$]
2.	List and describe the five primitives for specifying a data mining task.	CO1	[K$_1$]
3.	Define concept hierachy. Explain the types of concept hierarchies.	CO1	[K1]
4.	How to generate association rules from frequent item sets?	CO2	[K$_1$]
5.	Summarize few techniques to improve the efficiency of Apriori algorithm.	CO2	[K$_2$]
6.	Why tree pruning is useful in decision tree induction?	CO3	[K1]
7.	Compare pros and cons of eager classification and lazy classification.	CO3	[K$_2$]
8.	Identify different types of data used for cluster analysis?	CO4	[K$_3$]
9.	List the weaknesses of k-means?	CO4	[K$_1$]
10.	Define outliers. How will you determine outliers in the data?	CO4	[K$_3$]

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

11.	a)	Describe the steps involved in data mining when viewed as a process of knowledge discovery.	10	CO1	[K₁]
	b)	In real-world data, tuples with missing values for some attributes are a common occurrence. Discover various methods for handling this problem	6	CO1	[K₃]

12. a) A database has 5 transactions. Let min sup = 60% and min conf = 80%. Find all   12   CO2   [K4]
   frequent item sets using Apriori and FP-growth, respectively. Compare the
   efficiency of the two mining processes.

TID	items_bought
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y }
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, O, K, I ,E}

   b) Outline the Constraint based Association Mining   4   CO2   [K2]

13. a) The following table consists of training data from an employee database. The   10   CO3   [K3]
   data have been generalized. For example, "31 . . . 35" for age represents the age
   range of 31 to 35. For a given row entry, count represents the number of data
   tuples having the values for department, status, age, and salary given in that row.
   Let status be the class label attribute. Construct a decision tree from the given
   data.

department	status	age	salary	count
sales	senior	31...35	46K...50K	30
sales	junior	26...30	26K...30K	40
sales	junior	31...35	31K...35K	40
systems	junior	21...25	46K...50K	20
systems	senior	31...35	66K...70K	5
systems	junior	26...30	46K...50K	3
systems	senior	41...45	66K...70K	3
marketing	senior	36...40	46K...50K	10
marketing	junior	31...35	41K...45K	4
secretary	senior	46...50	36K...40K	4
secretary	junior	26...30	26K...30K	6

b)	Explain the major ideas of naive Bayesian classification.	6	CO3	[K$_2$]

14. a) Explain k-means algorithm with an example. Describe pros and cons of k-means   10   CO4   [K$_2$]
in comparison with the k-medoids algorithm.

    b) Summarize the Data Mining applications.   6   CO4   [K$_1$]

15. a) Illustrate and give examples of each of the following approaches to clustering: 12   CO3   [K$_2$]
partitioning methods,
hierarchical methods,
density-based methods and
grid-based methods.

    b) Given two objects represented by the tuples (22,1,42,10) and (20,0,36,8):   4   CO4   [K$_4$]
Estimate the Euclidean distance between the two objects.
Estimate the Manhattan distance between the two objects.
Estimate the Minkowski distance between the two objects, using h= 3

16. a) Illustrate the three popular schemas for multidimensional databases.   12   CO1   [K2]

    b) Justify, why do we need preprocess data? What are the different forms of pre-
processing?   4   CO1   [K$_4$]

****