Reg. No.



A Constituent unit of MAHE. Manipal

VI SEMESTER B.TECH. (COMPUTER SCIENCE & ENGINEERING) END SEMESTER EXAMINATIONS, MAY 2024

SUBJECT: DATA WAREHOUSE AND DATA MINING [CSE 4060]

REVISED CREDIT SYSTEM (--/05/2024)

Time: 3 Hours

MAX. MARKS: 50

Instructions to Candidates:

✤ Answer ALL FIVE questions.

✤ Missing data may be suitably assumed.

Q.No											CO/	AHEP	Blooms		
	Questions								Marks	CLO	LO	Taxonomy			
														Level	
1A.	With the help of examples, summarize the classic problems of machine learning that are related to Data Mining									4	CO1	1,4	6		
1 B .	Present a detailed outline of the types of Meta Data that are related to Data Warehouses. Justify its significance									4	CO2	1,4	4		
1C.	Evaluate the three types of primary keys that are present in fact tables. Justify which one of them is used most commonly.									2	CO2	1,4	6		
2A.	Break down the basic tasks that are involved in Data Transformation								5	CO2	2,4	4			
2B.	Outline any 6 virtues of OLAP in Data Warehouses									3	CO2	1,4	4		
2C.	Suppose that x and y are the first two term-frequency vectors in Table 2.C That is, $x = (5, 0, 3, 0, 2, 0, 0, 2, 0, 0)$ and $y = (3, 0, 2, 0, 1, 1, 0, 1, 0, 1)$. Using cosine similarity evaluate the similarity between x and y Table 2C Document Vector or Term-Frequency Vector												CO2	1,2	6
	Document	team	coach	hockey	baseba	ll socce	r pena	ilty scor	e wii	n los	s season	2			
	Document1	5	0	3	0	2	0	0	2	0	0				
	Document2	3	0	2	0	1	1	0	1	0	1				
	Document3	0	7	0	2	1	0	0	3	0	0				
	Document4	0	1	0	0	1	2	2	0	3	0				
3A.	For the g	iven d all freq	ataset, uent i	assur temset	ning mi s using	nimun the Ap	n supp riori a	ort is s lgorith	et to n	a	value of 2,	5	CO3	1,2	5

	TID List of item_IDs				
	T100 I1, I2, I5				
	T200 I2, I4				
	T'300 I2, I3				
	T400 I1, I2, I4				
	T500 I1, I3				
	T600 I2, I3				
	T700 I1, I3				
	1800 11, 12, 13, 15				
	1900 11, 12, 13				
3R	Formulate the FP-Growth algorithm to discover frequent itemsets without		CO3	2.5	5
50.	candidate generation	3		2,0	5
20	With the help of an example and proper illustration present a detailed		CO3	1 /	1
3 C.	with the help of an example and proper musuation present a detailed	2	005	1,4	4
	outline of now hashing can help in improving the efficiency of Apriori	-			
	algorithm.				
4A.	Consider the following figure showing a multilayer feed-forward neural		CO4	2,5	5
	network. Let the learning rate be 0.9. The initial weight and bias values of				
	the network are given in Table 1. Classify the tuple, $X = (1, 0, 1)$ with a				
	class label of 1 using Backpropagation algorithm. Show all steps in detail				
	fan the first iteration				
	for the first iteration.				
	w ₁₅ 4				
	W46				
		5			
	$x_2(2)$ (6)				
	W25				
	W ₂₄				
	$x_3 \begin{pmatrix} 3 \end{pmatrix}$ w_{25}				
	Initial Input Weight and Bias Values				
	initial input, weight, and bias values				
	x_1 x_2 x_3 w_{14} w_{15} w_{24} w_{25} w_{34} w_{35} w_{46} w_{56} θ_4 θ_5 θ_6				
4B.	Summarize size ordering and rule ordering conflict resolution strategies of	3	CO4	1,4	5
	rule based classification				
4C.	Outline the concept of k-fold cross validation highlighting the importance	2	CO4	1,4	4
	of leave one out and stratified cross validation	-		-	
5A.	Present a detailed overview of the orthogonal aspects using which	1	CO5	1.4	4
	clustering methods can be compared	+		_, .	-
5R	With the help of a figure and detailed explanation break down the working		CO5	14	4
50.	CHAMELEON for elustoring date	4		1,4	-
FC			COF	1 4	
5C.	With the help of a diagram and sufficient explanation summarize the	2	005	1,4	6
	working of k-means clustering	-			