Reg. No.
 Image: No.

 IANIPAL INSTITUTE OF TECHNOLOGY



VI SEMESTER B.TECH. (COMPUTER SCIENCE & ENGINEERING) END SEMESTER MAKE UP EXAMINATIONS, JULY 2022

SUBJECT: DATA WAREHOUSE AND DATA MINING [CSE 4060]

REVISED CREDIT SYSTEM

(--/05/2022)

Time: 3 Hours

MAX. MARKS: 50

Instructions to Candidates:

- Answer ALL FIVE questions.
- Missing data may be suitably assumed.

		М	СО	AHEP	BL
		IVI		LO	
1A.	Justify how the diversity of the database is a challenge in Data Mining	4	1	1,3	2
1B.	Compare the advantages and disadvantages of Top Down and Bottom Up construction of a Data Warehouse	4	1	1,3,5	2
1C.	Present a detailed outline of any 4 characteristics of the Fact table in a Star Schema model	2	2	2,3,5	2
2A.	Using an example summarize the general principles and method of application of Type 1 changes to Data Warehouses	5	2	2,3,5	1
2B.	With the help of a diagram justify how Deferred Data Extraction is carried out in Data Warehouses.	3	2	1,4,5	1
2C.	Using suitable examples examine the different types of Data Smoothing techniques in Data Mining	2	2	2,4,6	1
3A.	For the given dataset, assuming minimum support is set to a value of 2, find all frequent itemsets using the Apriori algorithm	5	3	2,3,5	6

MANIPAL INSTITUTE OF TECHNOLOGY MANIPAL

Reg. No.

	ALD BY	A Constituent Institution	of Manipai Oniversi	y						
	TID List of item_IDs									
	T100	T100 I1.12.15								
	T200	T200 I2, I4								
	T300	T300 I2, I3								
	T400 I1, I2, I4									
	T500) I1, I3								
	T600		12, 13							
	T700		11, 13	15						
	1800 T900		11, 12, 13,	15						
	1900		11, 12, 15							
3B.	Justify how the Apriori Property is used in the Apriori algorithm for mining						•	3	2,3,5	6
	frequent itemsets					5				
3C.	Design the FP-Growth algorithm for mining frequent itemsets without						2	3	1,3	1
	candic	late generatio	n.				2			
		U								
4A.	For the following dataset generate the splitting attribute using Information							4	1,3	4
	Gain.									
	Clas	s-Labeled Train	ning Tuple	s from the	AllElectronics	Customer Database				
						<i>C</i>				
	RID	age	income	student	credit_rating	Class: buys_computer				
	1	youth	high	no	fair	no				
	2	youth	high	no	excellent	no				
	3	middle_aged	high	no	fair	yes				
	4	senior	medium	no	fair	yes				
	5	senior	low	yes	fair	yes	5			
	6	senior	low	yes	excellent	no				
	7	middle_aged	low	yes	excellent	yes				
	8	youth	medium	no	fair	no				
	9	youth	low	yes	fair	yes				
	10	senior	medium	yes	fair	yes				
	11	youth	medium	yes	excellent	yes				
	12	middle_aged	medium	no	excellent	yes				
	13	middle_aged	high	yes	fair	yes				
	14	senior	medium	no	excellent	no				
		· · · ·								
4B.	With t	the help of a	diagram ju	istify how	Partitioning t	echnique can help in	3	3	2,4,5	1
	impro	ving the efficie	ency of Ap	riori Appr	oach of mining	g Frequent Itemsets.				
4C.	Preser	nt a detailed s	ummary o	n how the	BOAT algorith	nm handles the	2	4	3,4,5	4
	scalab	ility of Decisio	n Trees							
5A.	Using	sufficient illu	ustration of	outline th	ne working o	f DBSCAN clustering	Δ	5	3,5,6	2
	technique									

	Reg. No.				
A Little	MANIPAL INSTITUTE OF TECH	INC	DL	OGY	7
5B.	With the help of a diagram present a detailed break down of the steps involved in BIRCH algorithm for hierarchical clustering of data.	4	5	4,5	2
5C.	Using proper illustration outline the working of the k-nearest neighbor classifier	2	4	4,5	4