GANPAT UNIVERSITY

B. Tech. Semester: VII (CE/IT) Engineering

Regular Examination November - December 2014

2CE703/2IT703 - DATA MINING & DATA WARE HOUSING

Time:	Ho Ho	urs]								[Total Marks:	70
Instrue	tion	1 Attempt all Questions. 2 Figures to the right ind 3 Each Section should be	icate	full m	arks (of the	que	estion.			
					Secti						
Q.1	A	Why preprocessing is repreprocessing.	equire					Explain	n various	forms of data	6
	В	Define the terms core of density connected in DBS	bject. CAN	dire algor	ectly o	densi	ty r	eachable	e, density xamples.	reachable and	6
					0	R			YA		
Q.1	A	What is supervised and ur	super	vised	learn	ing?	Brie	fly expl	ain BIRC	H algorithm.	6
	B	What are the major issues				kiasi.		Const.		an bort A	6
Q.2 A	A	How Correlation Analysis can be helpful in data mining. Following readings were observed according to the survey done in a city of 2000 persons on the interest on 2 issues. State the correlation between issues ABC & XYZ using Chi-Square analysis for the given data:								6	
				XYZ		-	int	erested	in XYZ	Total	
		ABC		200			-	00		700	
		Not interested in ABC		1000			3(00		1300	
		Total		1200			8(00		2000	
	B	Explain various schemas u write DMQL to create cub	sed ir e & it	data s dim	ensio	ns for	ng v r sta	with the	help of ex	camples. Also	5
					Ol						
Q.2	A	For the given distance ma a) Single-link c) Plot the dendogram for the	b) (comp	lete-li	nk			ical cluste	ering using:	6
						T.					
			-	P	Q	R	S	T			
		n. Topien	P	0	-						
			Q	3	0		-				
		A A CONTRACTOR	R	1	2	0	0				
		S In Dim with	T	5	6	7	8	0			
			1	13	110	17	10	0			

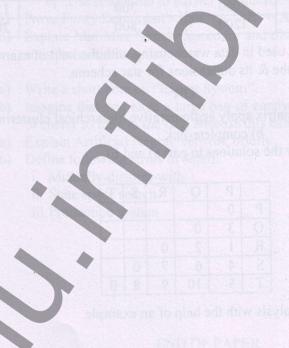
B Explain market basket analysis with the help of an example.

Q.3 A Using Naïve Bayesian Classification algorithm, predict whether a child can play in the condition X=(outlook=sunny, temperature=<=75, Windy=true) for the dataset given below.

outlook	temperature	windy	play
sunny	>75	FALSE	No
Sunny	>75	TRUE	No
Overcast	>75	FALSE	Yes
Rainy	<=75	FALSE	Yes
Rainy	<=75	FALSE	Yes
Rainy	<=75	TRUE	No
Overcast	<=75	TRUE	Yes
Sunny	<=75	FALSE	No
Sunny	<=75	FALSE	Yes
rainy	<=75	FALSE	Yes
Sunny	<=75	TRUE	Yes
Overcast	<=75	TRUE	Yes
Overcast	>75	FALSE	Yes
rainy	<=75	TRUE	No

B Find the Jaccard's coefficient between all the objects for the given below attributes. Also state which two objects are likely to have similar properties?

	Δ1	TA2	A3	A4	A5	A6
Object-X	True	True	True	False	False	True
Object-Y	False	True	True	False	True	False
Object-Z	False	True	False	True	False	True



Section - II

- Q.4 A Use the k-medoid algorithm to cluster the following 8 objects into three clusters.

 P1=(2,5), P2=(5,4), P3=(2,4), P4=(7,5), P5=(3,4), P6=(6,4), P7=(2,1), P8=(0,2).

 Take initial clusters as P1, P6 and P8 and distance measure as Manhattan distance.
 - 1) Find final three clusters and their medoids formed after 2 iterations.
 - B Explain various data transformation techniques in data mining.

6

6

OR

- Q.4 A Use the k-means algorithm to cluster the following 8 objects into three clusters. P1=(2,5), P2=(5,4), P3=(2,4), P4=(7,5), P5=(3,4), P6=(6,4), P7=(2,1), P8=(0,2). Take initial clusters as P1, P6 and P8 and distance measure as Manhattan distance.
 - 1) Find final three clusters and their centroids formed after 2 iterations.
 - B Explain various OLAP operations in the multidimensional data model with the help of suitable examples.
- Q.5 A Find frequent itemsets using FP-growth algorithm for the given dataset. Consider Minimum Support Count as 50%.

Transaction ID	List of items
T101	{12,15,16}
T102	{12,13,15}
T103	{12,13,14,15,16}
T104	{11,13}
T105	{11,12,13,15}
T106	{11,14,15,16}

B Explain 3-Tier data warehouse architecture.

5

OR

Q.5 A Find frequent itemsets using Apriori algorithm for the given dataset. Also find strong association rules for the highest frequent itemset. Consider Minimum Support Count as 50% & Minimum Confidence as 80%.

Transaction ID	List of items
T101	{12,15}
T102	{12,13,15}
T103	{12,13,14,15}
T104	{11,13}
T105	{11,12,13,15}
T106	{11,14,15}

B Explain various data reduction strategies used in data mining.

5

Q.6 A Discuss any three methods to improve Apriori algorithm.

- 6
- B Explain Data warehouse and Data mart. Explain various features of data warehouse.

6

END OF PAPER