Student Exam No.________________

GANPAT UNIVERSITY

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

Regular Examination November – December 2014

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

Time: 3 Hours] [Total Marks: 70

Instruction: 1 Attempt all Questions.
2 Figures to the right indicate full marks of the question.
3 Each Section should be written in separate answer book.

Section – I

Q.1 A Why preprocessing is required in data mining? Explain various forms of data preprocessing. 6

B Define the terms core object, directly density reachable, density reachable and density connected in DBSCAN algorithm with the help of examples. 6

OR

Q.1 A What is supervised and unsupervised learning? Briefly explain BIRCH algorithm. 6

B What are the major issues in data mining? 6

Q.2 A How Correlation Analysis can be helpful in data mining. 6
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:

	XYZ	Not interested in XYZ	Total
ABC	200	500	700
Not interested in ABC	1000	300	1300
Total	1200	800	2000

B Explain various schemas used in data warehousing with the help of examples. Also write DMQL to create cube & its dimensions for star schema. 5

OR

Q.2 A For the given distance matrix apply agglomerative hierarchical clustering using: 6
a) Single-link b) complete-link
c) Plot the dendogram for the solutions to part a) and b).

	P	Q	R	S	T
P	0				
Q	3	0			
R	1	2	0		
S	4	6	7	0	
T	5	10	9	8	0

B Explain market basket analysis with the help of an example. 5

gnu.inflibnet.ac.in

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. 6

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? 6

	A1	A2	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

gnu.inflibnet.ac.in

Section – II

Q.4 A Use the k-medoid algorithm to cluster the following 8 objects into three clusters. 6
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

OR

Q.4 A Use the k-means algorithm to cluster the following 8 objects into three clusters. 6
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. 6

Q.5 A Find frequent itemsets using **FP-growth** algorithm for the given dataset. Consider Minimum Support Count as 50%. 6

Transaction ID	List of items
T101	{I2,I5,I6}
T102	{I2,I3,I5}
T103	{I2,I3,I4,I5,I6}
T104	{I1,I3}
T105	{I1,I2,I3,I5}
T106	{I1,I4,I5,I6}

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%. 6

Transaction ID	List of items
T101	{I2,I5}
T102	{I2,I3,I5}
T103	{I2,I3,I4,I5}
T104	{I1,I3}
T105	{I1,I2,I3,I5}
T106	{I1,I4,I5}

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

gnuinfinet.ac.in

16