Student Exam No._____

GANPAT UNIVERSITY

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

Regular Examination November – December 2013

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** Explain star, snowflake and galaxy schemas with the help of suitable example. Also write DMQL for snowflake schema to define cube and its dimensions. **6**

B Explain ROCK algorithm for clustering with the help of an example. **6**

OR

Q.1 **A** What are the major challenges in data mining? **6**

B Explain DBSCAN algorithm for clustering with the help of an example. **6**

Q.2 **A** Explain different approaches to mining multilevel association rules with the help of suitable examples. **6**

B Explain Minkowski distance to measure distance between 2 objects. Also compute distance measure for p=1 and p=2 for the following two objects: X1=(2,3,5) and X2=(4,2,7) **5**

OR

Q.2 **A** Explain various data transformation techniques. **6**

B **5**

	A1	A2	A3	A4	A5	A6
O1	True	True	True	False	False	True
O2	False	True	True	False	True	False

For the above given objects having asymmetric attributes, where True is more significant than False; find a)Dissimilarity measure and b) Jaccard coefficient.

Q.3 **A** Explain various OLAP operations performed on data cube with examples. **6**

B For the given distance matrix apply agglomerative hierarchical clustering using: **6**
a) Single-link b) complete-link
c) Plot the dendograms for the solutions to part a) and b).

	A	B	C	D
A	0	4	5	1
B		0	6	2
C			0	3
D				0

Section – II

Q.4 A Use the k-medoid algorithm to cluster the following 8 objects into three clusters. 8
X1=(2,5), X2=(2,10), X3=(8,4), X4=(5,8), X5=(7,5), X6=(6,4), X7=(4,9), X8=(1,2).
Take initial clusters as X2, X4 and X8 and distance measure as Manhattan distance.
1)Find final three clusters and their medoids formed after 2 iterations.

B What are the various forms of data preprocessing? 4

OR

Q.4 A Use the k-means algorithm to cluster the following 8 objects into three clusters. 8
X1=(2,5), X2=(2,10), X3=(8,4), X4=(5,8), X5=(7,5), X6=(6,4), X7=(4,9), X8=(1,2).
Take initial clusters as X2, X4 and X8 and distance measure as Euclidean distance.
1)Find final three clusters and their centroids formed after 3 iterations.

B Differentiate between OLTP and OLAP. 4

Q.5 A Explain market basket analysis with the help of an example. 6

B What are the various methods to fill missing values in the process of data cleaning? 5

OR

Q.5 A What are the various types of data used in clustering analysis? 6

B Explain Data warehouse and Data mart. What are the various features of data warehouse? 5

Q.6 A Given Minimum Support count=3, find frequent itemsets for the example given below using **FP-Growth** algorithm. 6

Transaction ID	List of items
T1	{g,b,d,e,h,j,n,a}
T2	{b,c,d,g,m,n,p}
T3	{c,g,i,k,p}
T4	{c,d,l,t,a}
T5	{b,g,d,f,m,a,n,o}

B Given minimum support as 40% and minimum confidence as 80%. Find out the frequent itemsets and strong association rules for the example given below using **Apriori** algorithm. 6

TID	ITEMS
100	{I1,I3,I4,I6}
200	{I2,I3,I5,I7}
300	{I1,I2,I3,I5,I8}
400	{I2,I5,I9,I10}
500	{I1,I4}

END OF PAPER