Reg. No.										

प्रज्ञानं ब्रह्म

Manipal
INSPIRED BY LIFE

MANIPAL INSTITUTE OF TECHNOLOGY, MANIPAL 576104

(Constituent College of Manipal University)

KNOWLEDGE IS POWER
MANIPAL INSTITUTE OF TECHNOLOGY

SIXTH SEMESTER B-TECH. (CCE) DEGREE MAKEUP EXAMINATION JUNE/JULY 2016
SUBJECT: DATA MINING AND PREDICTIVE ANALYSIS– ICT 354
(REVISED CREDIT SYSTEM)

TIME: 3 HOURS **01/07/2016** **MAX. MARKS: 50**

Instructions to candidates
Answer any **FIVE FULL** questions.
Missing data, if any, may be suitably assumed.

1A. Apply Apriori partitioning algorithm on the dataset given in Table Q.1A. Number of partitions=3, global support=3. Generate association rule by considering confidence=60%.

Table Q. 1A

T_ID	Itemsets
T1	1,2,3,4,
T2	1,2,4
T3	1,2
T4	2,3,4
T5	2,3
T6	3,4

1B. Write and explain Generalised Sequential Pattern algorithm with an example.
1C. Write a note on data mining applications in finance.

[5+3+2]

2A. Given Min_Sup=3, apply FP Tree algorithm on the dataset given in Table Q.2A to obtain the frequent patterns.

Table Q.2A

T-ID	Itemsets
T1	1,2,4
T2	1,3,4,5
T3	1,4,5,6
T4	2,5,6
T5	1,2,4,5,6

2B. Obtain average link hierarchical cluster for the following data.
A=(0.4,0.4,0.5) B=(0.1,0.8,0.1) C(0.3,0.3,0.4) D=(0.1,0.1,0.8) E=(0.4,0.2,0.4) F=(0.1,0.4,0.5) G=(0.7,0.2,0.1) H=(0.5,0.4,0.1)
2C. The pincer search algorithm finds only maximal frequent sets. Justify.

[5+3+2]

3A. Apply Dynamic Itemset Counting algorithm for the dataset given in Table Q.3A

Table Q. 3A

T_ID	Itemsets
1	Burger, Coke, Juice
2	Juice, Potato chips
3	Coke, Burger
4	Juice, Groundnuts
5	Coke, Groundnuts

0123 Mrs Swathi B.P Swathi.B.P

0 4,5,6. Mrs Manjula Shenoy.K

3B. Explain with an example how the following approaches improves the efficiency of Apriori algorithm:

i. Hash based technique
ii. Transaction reduction.

3C. Explain majority voting with an example.

[5+3+2]

4A. Obtain decision tree for the dataset given in Table Q.4A

Table Q.4A

Wings Maint.	Engs	Nose	Intake	Fuselage	Class
Mid	1	Hat	Nose	Cigar	Foreign
Mid	2	Hat	Nose	Sluck	Foreign
Low	1	Snule	Nose	Sluck	Foreign
High	3	Point	Body	Thick	Domestic
High	4	Point	Body	Thick	Domestic

4B. What is the advantage of gain ratio over Information Gain? Explain with an example.

4C. What type of pruning is used in CART? Explain.

[5+3+2]

5A. Apply PAGERANK algorithm for the following web pages:
A→B, A→D, A→C, B→C, B→A, C→D, D→C, D→A.
List out the real time applications of spatial and temporal data mining.

5B. K-means algorithm is sensitive to outliers. Justify.

5C. Explain KDD process with an example for each step.

[5+3+2]

6A. List out the difference between symmetric and asymmetric binary variables. Find the dissimilarity matrix for the dataset given in Table Q.6A.

Table Q.6A

Car (Nominal)	Range (Ordinal)	Quality [0-bad, 1-Good] Asymmetric binary
1	Red	Senior
2	Green	Junior
3	Blue	Mid
4	Green	Senior

6B. Give an example to show that items in a strong association rule may actually be negatively correlated.

6C. List the mathematical properties satisfied by Euclidean and Manhattan distance.

[5+3+2]

****