Reg. No.										

प्रज्ञानं ब्रह्म

MANIPAL INSTITUTE OF TECHNOLOGY, MANIPAL 576104

(Constituent College of Manipal University)

SIXTH SEMESTER B. Tech (CCE) DEGREE END SEMESTER EXAMINATION, MAY– 2016
SUBJECT: DATA MINING AND PREDICTIVE ANALYSIS(ICT-354)
(REVISED CREDIT SYSTEM)

TIME: 3 HOURS **09/05/2016** **MAX. MARKS: 50**

Instructions to candidates

- Answer any **FIVE FULL** questions.
- Missing data, if any, may be suitably assumed.
- All scripts should be well commented.

1A. For the data set in Table Q.1A , construct FP-Tree and find frequent item sets. Minimum support is 2.

Table Q.1A

Transaction Data Set

TID	Items
1	{a,b}
2	{b,c,d}
3	{a,c,d,e}
4	{a,d,e}
5	{a,b,c}
6	{a,b,c,d}
7	{a}
8	{a,b,c}
9	{a,b,d}
10	{b,c,e}

1B. What are the approaches used for mining multilevel association rules? Explain.
1C. Define DataWarehouse. How do you model data in a data warehouse? Give example. [5+3+2]

2A. What are the approaches used to improve efficiency of Apriori? Explain.
2B. Discuss any three issues of data mining.
2C. Explain different types of attributes with an example for each. [5+3+2]

3A. For the contingency table in Table Q.3A, with expected value, compute support, confidence, lift and chi-square. Find if computergame=>videos is negatively or positively correlated.

Table Q.3A

	Game	Not Game
Video	4000(4500)	3500(3000)
Not Video	2000(1500)	500(1000)

3B. Write the DBSCAN algorithm for clustering and explain.
3C. What is temporal data mining? What are the different types of temporal data? [5+3+2]

4A. Explain the process of web usage mining. What are the different types of data considered in web usage mining?

4B. Write a detailed note on Sequence mining.

4C. Discuss various Spatial mining tasks. [5+3+2]

5A. Cluster the following points into three clusters using K-Medoid Clustering algorithm.
A1(2,10) , A2(2,5) , B1(5,8), B2(7,5) , C1(1,2) , C2(4,9).
Initial representatives are A1,B2,C1.
Find the cost of replacing B2 by B1.

5B. Explain tree pruning methods used in ID3, C4.5 and CART.

5C. Explain any two data mining applications. [5+3+2]

6A. Find root attribute for the decision tree using gini-index as attribute selection measure for the dataset in Table Q.6A.

Table Q.6A

Record Id	Age	income	Repayment of previous loan	Loan sanctioned
1	55	medium	yes	no
2	50	high	yes	yes
3	23	high	Not taken	yes
4	28	medium	Not taken	no
5	25	medium	Not taken	yes
6	45	low	no	no
7	43	low	yes	yes
8	38	high	no	yes
9	36	low	no	no

6B. Find the attributes to be tested in second level(After root level).

6C. Write the basic Page rank algorithm used in web structure mining. [5+3+2]