Reg. No.

MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL
A Constituent Institution of Manipal University

VII SEMESTER B.TECH. (COMPUTER SCIENCE ENGINEERING)
END SEMESTER EXAMINATIONS, NOV/DEC 2016

SUBJECT: DATA WAREHOUSING AND DATA MINING [CSE 433]
REVISED CREDIT SYSTEM
(23/11/2016)

Time: 3 Hours   MAX. MARKS: 50

1A.	Describe any two ways in which measures can be categorized in Data Warehousing	**2M**
1B.	Describe how OLAP style analysis can be fused with data mining techniques	**2M**
1C.	Distinguish between ROLAP,MOLAP,HOLAP	**6M**
2A.	Define Exceptions. State why they are used. Explain the measures that are used for indicating exceptions	**3M**
2B.	Consider the following data set: **4, 8, 15, 21, 21, 24, 25, 28, 34** Perform the following operations on the above data set 1. Partition the above data set into equal frequency bins 2. Smoothening by bin means 3. Smoothening by bin boundaries Assume the size of the bins to be 3	**3M**
2C.	Explain the basic procedure of Principal Component Analysis	**4M**
3A.	Explain the Apriori property of generating frequent itemsets .Using Apriori Algorithm generate frequent itemsets for the following data set. Assume Min_sup=2	**5M**

TID	Items
100	1, 3 , 4
200	2, 3, 5
300	1, 2, 3, 5
400	2, 5
500	1, 3, 5

3B.	Describe the phases involved in the execution of Partition Algorithm for generating frequent itemsets	**3M**
3C.	Define Page Rank. With the help of a formula explain how page rank can be calculated	**2M**

4A. Analyze the following data . Using Naïve Bayes classification find out if John will play golf on day with Outlook=Rainy, Temp=Mild, Humidity=Normal ,Windy= True **4M**

Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

4B. Briefly describe why tree pruning is required. Explain what is prepruning and postpruning of trees **3M**

4C. Describe the termination conditions for the recursive partitioning of Decision Tree induction. **3M**

5A. Explain any four measures for evaluating classifier performance. **4M**

5B. Write an algorithm for partitioning using k-means method. **2M**

5C. Explain Density Based Spatial Clustering of Applications with Noise **4M**

6A. With the help of an example explain agglomerative and divisive methods of hierarchical clustering **4M**

6B. Describe Web structure mining **3M**

6C. Find all frequent itemsets in the following database using FP-growth algorithm. Take minimum support count as 3. **3M**

TID	Items
1	E, A, D, B
2	D, A, C, E, B
3	C, A, B, E
4	B, A, D
5	D
6	D, B
7	A, D, E
8	B, C