Code No: 117CD **R13**

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
B. Tech IV Year I Semester Examinations, November/December - 2017
DATA WAREHOUSING AND DATA MINING
(Computer Science and Engineering)

Time: 3 Hours **Max. Marks: 75**

Note: This question paper contains two parts A and B.
Part A is compulsory which carries 25 marks. Answer all questions in Part A. Part B consists of 5 Units. Answer any one full question from each unit. Each question carries 10 marks and may have a, b, c as sub questions.

PART - A

(25 Marks)

1.a)	Define data warehouse.	[2]
b)	List the Data warehouse Characteristics.	[3]
c)	How can you go about filling in the missing values for this attribute?	[2]
d)	Why is the word data mining a misnomer?	[3]
e)	Give a note on Closed Frequent Item Set.	[2]
f)	Write the FP-graph algorithm.	[3]
g)	How prediction is different from classification?	[2]
h)	What is rule classification?	[3]
i)	Give a note on k means algorithm.	[2]
j)	List the Key Issues in Hierarchical Clustering.	[3]

PART – B

(50 Marks)

2.a) Make a comparisons between the MOLAP and HOLAP.
b) Discuss the star and snowflake schema in detail with suitable example. [5+5]

OR

3.a) Write the difference between designing a data warehouse and an OLAP cube.
b) Give a brief note on ROLAP. [5+5]

4. Explain concept hierarchy generation for the nominal data. [10]

OR

5.a) Describe the Feature Subset Selection.
b) Illustrate the Data Transformation by Normalization. [5+5]

6. Make a comparison of Apriori and ECLAT algorithms for frequent item set mining in transactional databases. Apply these algorithms to the following data:

TID	LIST OF ITEMS
1	Bread, Milk, Sugar, TeaPowder, Cheese, Tomato
2	Onion, Tomato, Chillies, Sugar, Milk
3	Milk, Cake, Biscuits, Cheese, Onion
4	Chillies, Potato, Milk, Cake, Sugar, Bread
5	Bread, Jam, Mik, Butter, Chilles
6	Butter, Cheese, Paneer, Curd, Milk, Biscuits
7	Onion, Paneer, Chilies, Garlic, Milk
8	Bread, Jam, Cake, Biscuits, Tomato

[10]

OR

7. Briefly explain the Partition Algorithms. [10]

8. Discuss K- Nearest neighbor classification-Algorithm and Characteristics. [10]

OR

9. How does the Naïve Bayesian classification works? Explain in detail. [10]

10.a) Give a brief note on PAM Algorithm.

b) What is the drawback of k-means algorithm? How can we modify the algorithm to diminish that problem? [5+5]

OR

11. What are the different clustering methods? Explain in detail. [10]

--ooOoo--