Code No: 117CD **R13**

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

B. Tech IV Year I Semester Examinations, March - 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) What is a data mart? [2]
b) What is a fact table? [3]
c) What is data mining? [2]
d) List similarity measures. [3]
e) What is maximal frequent itemset? [2]
f) How to compute confidence measure for an association rule? [3]
g) What is classification? [2]
h) Define information gain. [3]
i) What is an outlier? [2]
j) List the demerits of k-means algorithm. [3]

Part-B (50 Marks)

2. What are the various components of data warehouse? Explain their functionality in detail. [10]

OR

3. What is the significance of OLAP in data warehouse? Describe OLAP operations with necessary diagram/example. [10]

4. Explain different data mining tasks for knowledge discovery. [10]

OR

5. What is the need of dimensionality reduction? Explain any two techniques for dimensionality reduction. [10]

6. A database has six transactions. Let min-sup = 50% and min-conf = 75%.

TID	List of items
001	Pencil, sharpener, eraser, color papers
002	Color papers, charts, glue sticks
003	Pencil, glue stick, eraser, pen
004	Oil pastels, poster colours, correction tape
005	Whitener, pen , pencil, charts, glue stick
006	Colour pencils, crayons, eraser, pen

Find all frequent item sets using Apriori algorithm. List all the strong association rules. [10]

OR

7.a) What are the advantages of FP-Growth algorithm? [5+5]

b) Discuss the applications of association analysis.

8. Explain decision tree induction algorithm for classifying data tuples and discuss suitable example. [10]

OR

9.a) What are the characteristics of k-nearest neighbor algorithm? [5+5]

b) How to evaluate the classifier accuracy?

10. What is the goal of clustering? How does partitioning around medoids algorithm achieve this goal? [10]

OR

11.a) Differentiate between AGNES and DIANA algorithms. [5+5]

b) How to access the cluster quality?

--ooOoo—