R13

Code No: 126EW

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
B. Tech III Year II Semester Examinations, May - 2016
DATA WAREHOUSING AND DATA MINING
(Information Technology)

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 concept hierarchy.	[2]
b)	What is meta data repository?	[3]
c)	What are the challenges of KDD?	[2]
d)	What is data mining?	[3]
e)	Define association rule.	[2]
f)	Define maximal frequent itemset.	[3]
g)	Why is tree pruning useful in decision tree induction?	[2]
h)	Define Bayesian belief network.	[3]
i)	What is clustering?	[2]
j)	How does Chameleon work?	[3]

PART - B

(50 Marks)

2.a) Explain any two of the schemas for multidimensional databases.
b) Describe Fully Addictive, Semi-Addictive, Non Addictive Measures. [5+5]

OR

3.a) What are OLAP operations in the multidimensional data model? Explain.
b) Describe efficient computation of data cubes. [5+5]

4.a) Discuss about dimensionality reduction.
b) Explain in detail about data cleaning. [5+5]

OR

5.a) List and describe the five *primitives* for specifying a data mining task.
b) In real-world data, tuples with *missing values* for some attributes are a common occurrence. Describe various methods for handling this problem. [5+5]

6. Write the the Apriori algorithm for discovering frequent item sets for mining Boolean association rules. [10]

OR

7.a) How can we mine closed frequent item sets? Explain.
b) Write the FP-growth algorithm. [5+5]

8. Compare the advantages and disadvantages of *eager* classification (e.g., decision tree, Bayesian, neural network) versus *lazy* classification (e.g., *k*- earest neighbor, casebased reasoning). [10]

OR

9.a) What are the measures for selecting the Best Split? Explain.

b) What are the general approaches for classification problems? Explain. [5+5]

10.a) Write and explain about the k-medoids algorithm.

b) Describe distance based outlier detection. [5+5]

OR

11. Briefly describe the following approaches to clustering: partitioning methods, hierarchical methods, density-based methods, grid-based methods, model-based methods, methods for high-dimensional data, and constraint-based methods. Give examples in each case. [10]

–--ooOoo---