R13

Code No: 814BD

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
MCA IV Semester Examinations, January - 2018
DATA WAREHOUSING AND DATA MINING

Time: 3 Hours **Max. Marks: 60**

Note: This question paper contains two parts A and B.
Part A is compulsory which carries 20 marks. Answer all questions in Part A. Part B consists of 5 Units. Answer any one full question from each unit. Each question carries 8 marks and may have a, b, c as sub questions.

PART - A

5 × 4 Marks = 20

1.a) Discuss the data smoothing techniques. [4]
b) Discuss the OLAP query processing. [4]
c) Discuss constraint-based Association mining. [4]
d) How does tree pruning work? [4]
e) Why wavelet transformation useful for clustering? [4]

PART - B

5 × 8 Marks = 40

2. Write the syntax for the following data mining primitives:
a) The kind of knowledge to be mined.
b) Measures of pattern interestingness. [4+4]

OR

3.a) Briefly discuss the data mining functionalities.
b) Briefly discuss the major issues in data mining regarding performance and diverse database types. [4+4]

4.a) Justify the role of data cube aggregation in data reduction process with an example.
b) Differentiate operational database systems and data warehousing. [4+4]

OR

5.a) What is data warehousing? Give their applications.
b) Briefly discuss data warehouse architecture. [4+4]

6. Compare and contrast Apriori algorithm with frequent pattern growth algorithm. Consider a data set apply both algorithms and explain the results. [8]

OR

7.a) Explain how concept hierarchies are used in mining multilevel association rule?
b) Give the classification of association rules in detail. [4+4]

8. a) Discuss the five criteria for the evaluation of classification and prediction methods.
b) Explain how rules can be extracted from training neural networks. [4+4]

OR

9.a) Explain the hold out method for estimating classifier accuracy.
b) Discuss Fuzzy set approach for classification. [4+4]

10.a) Write k-Means and k-Medoids algorithms.
b) Explain COBWEB model. [4+4]

OR

11.a) Explain about Statistical-based outlier detection and Deviation-based outlier detection.
b) Given two objects represented by the tuples(22,1,42,10) and(20,0,36,8)
i) Compute the Manhatten distance between the two objects.
ii) Compute the Euchidean distance between the two objects. [4+4]

---oo0oo---