R17

Code No: 844AD

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
MCA IV Semester Examinations, December - 2019
DATA WAREHOUSING AND DATAMINING

Time: 3hrs **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

5 × 5 Marks = 25

1.a) Describe three challenges to data mining regarding data mining methodology and user interaction issues. [5]
b) Illustrate the OLAP operations in the Multidimensional Data Model. [5]
c) How to develop a scalable SVM algorithm for efficient classification in large datasets? [5]
d) Discuss about Outlier detection using deviation based approach. [5]
e) Differentiate structured, unstructured and semi-structured data. [5]

PART - B

5 × 10 Marks = 50

2.a) What is Data Mining? Explain the steps involved in KDD with neat sketch.
b) Describe the classification of Data Mining System. [6+4]

OR

3.a) Distinguish the methods used for integration of Data Mining System with data warehousing system.
b) Describe various methods involved in Data Cleaning process. [5+5]

4.a) Describe the similarities and the differences of star schema, and snowflake schemathen analyze their advantages and disadvantages from each other.
b) Discuss about the algorithm that computes closed iceberg cubes efficiently. [5+5]

OR

5.a) A data cube C, has n dimensions, and each dimension has exactly p distinct values in the base cuboid. Assume that there are no concept hierarchies associated with the dimensions.
(i) What is the maximum number of cells possible in the base cuboid?
(ii) What is the minimum number of cells possible in the base cuboid?
(iii) What is the maximum number of cells possible in the data cube, C? (including both base cells and aggregate cells)
(iv) What is the minimum number of cells possible in the data cube, C?
b) Discuss about bitmap indexing. [6+4]

6. A database has five transactions. Let min sup = 60% and min conf = 80%.

TID	items bought
T100	M, O, N, K, E, Y
T200	D, O, N, K, E, Y
T300	M, A, K, E
T400	M, U, C, K, Y
T500	C, O, O, K, I ,E

Find all frequent item sets using Apriori and FP-growth, respectively. Compare the efficiency of the two mining processes. [10]

OR

7.a) Discuss about Multilayer Feed-Forward Neural Network.
b) Describe Back Propagation Algorithm for classification and explain with an example. [6+4]

8.a) Differentiate AGNES and DIANA clustering algorithms.
b) Explain about k-means algorithm with an example. [5+5]

OR

9. Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8):
a) Compute the Euclidean distance between the two objects.
b) Compute the Manhattan distance between the two objects.
c) Compute the Minkowski distance between the two objects, using q = 3. [10]

10.a) What is time-series data? Explain in brief about mining Time-series data.
b) Enumerate the steps in Multimedia Data mining. [5+5]

OR

11. Design aWeb document classification method that can take index classification standard to classify a set of Web documents automatically. [10]

---ooOoo---