R17

Code No: 844AD

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
MCA IV Semester Examinations, April/May - 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) What is a pattern? What are the characteristics of an interesting pattern? [5]
b) Compare and contrast database management system with data warehouse. [5]
c) What characteristics of neural networks make them good classifiers? [5]
d) Give the advantages and disadvantages of partition based methods for clustering. [5]
e) Provide examples for spatial and non-spatial data in database. [5]

PART - B

5 × 10 Marks = 50

2. What is the need of preprocessing of data for mining? Briefly explain various forms of preprocessing. [10]

OR

3. Discuss the major issues pertaining to mining methodology and user interaction in data mining. [10]

4. Illustrate online analytical processing operations. [10]

OR

5. Demonstrate the working of BUC algorithm for data cube computation. [10]

6. What are the limitations of Apriori algorithm? Suggest mechanisms to improve the accuracy of Apriori algorithm. [10]

OR

7. Consider the following data and classify the new sample X= < youth, medium, yes, fair> using Naïve Bayesian classification. [10]

RID	Age	Income	Student	Credit rating	Class: buys computer
1	Youth	High	No	Fair	No
2	Youth	High	No	excellent	No
3	Middle_aged	High	No	Fair	Yes
4	Senior	Medium	No	Fair	Yes
5	Senior	Low	Yes	Fair	Yes
6	Senior	Low	Yes	excellent	No
7	Middle_aged	Low	Yes	excellent	Yes
8	Youth	Medium	No	Fair	No
9	Youth	Low	Yes	Fair	Yes
10	Senior	Medium	Yes	Fair	Yes
11	Youth	Medium	Yes	excellent	Yes
12	Middle_aged	Medium	No	excellent	Yes
13	Middle_aged	High	Yes	Fair	Yes
14	Senior	Medium	No	excellent	No

8.a) What is Є- neighbourhood in density based methods?
b) List the challenges raised by high dimensional data for clustering. [5+5]

OR

9. Write BIRCH algorithm. Does BIRCH follow agglomerative nesting? Justify your answer. [10]

10. What is the importance of sequence mining? Explain Prefix Span algorithm for sequence mining. [10]

OR

11. Demonstrate Latent semantic indexing for text mining with simple example text corpus. [10]

---ooOoo---