R15

Code No: 126VW

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
B. Tech III Year II Semester Examinations, May - 2019
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)	What is the role of meta data repository in data warehouse?	[2]
b)	Give an example for snow-flake schema.	[3]
c)	What is data cleaning?	[2]
d)	How binning can handle noisy data?	[3]
e)	Define confidence of an association rule.	[2]
f)	Give an example for maximal frequent itemsets.	[3]
g)	Define gain ratio of an attribute.	[2]
h)	How effective are Bayesian classifiers?	[3]
i)	What are the objectives of clustering?	[2]
j)	What is the need of outlier detection? List two applications of it.	[3]

PART - B

(50 Marks)

2. Compare and contrast online transaction processing with online analytical processing. [10]

OR

3. With necessary diagrams and examples of data cubes explain various OLAP operations. [10]

4. Describe the various phases in knowledge discovery process with a neat diagram. [10]

OR

5.a) What is the curse of dimensionality? How to reduce it?
b) What are the typical methods for data discretization? [5+5]

6.a) How are association rules generated from frequent itemsets? Illustrate.
b) Discuss the limitations of apriori algorithm. [5+5]

OR

7. Apply FP-Growth algorithm to the following transactional data to find frequent itemsets. List all frequent itemsets with their support count. [10]

TID	List of Item IDs
1	I1,i3,i5,i7
2	I2,i4,i6,i8
3	I1,i3,i5,i7
4	I9,i7,i5,i1
5	I2,i4,i6,i7
6	I1,i2,i3,i4
7	I3,i4,i5,i6
8	I7,i8,i6,i1
9	I8,i5,i3,i2
10	I1,i3,i4,i6

8. Explain Decision tree induction algorithm for classification. Discuss the usage of information gain in this. [10]

OR

9.a) What is meant by sensitivity and specificity?

b) Differentiate between eager learners and lazy learners. [5+5]

10. Discuss the similarity measures and distance measures frequently used in clustering the data. [10]

OR

11.a) What are the advantages of PAM algorithm over k-means algorithm?

b) How to evaluate a clustering algorithm? [5+5]

---ooOoo---