R15

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
B. Tech III Year II Semester Examinations, April - 2018
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 data warehouse.	[2]
b)	Differentiate between ROLAP and HOLAP.	[3]
c)	What is evolution analysis?	[2]
d)	List the various forms of data preprocessing.	[3]
e)	Quote an example for quantitative association rule.	[2]
f)	How to compute confidence for an association rule $X \rightarrow Y$?	[3]
g)	Define classifier accuracy.	[2]
h)	List the characteristics of k-nearest neighbor algorithm.	[3]
i)	What is an outlier?	[2]
j)	What are the weaknesses of hierarchical clustering?	[3]

PART - B

(50 Marks)

2.a)   Compare and contrast operational database systems with data warehouse.
  b)   What is the importance of data marts in data warehouse?   [5+5]

OR

3.a)   With illustrative examples explain various OLAP operations.
  b)   Discuss the characteristics of fact table.   [5+5]

4.a)   Explain data mining as a step in knowledge discovery process.
  b)   Differentiate between data retrieval and data mining.   [5+5]

OR

5.   Demonstrate computation of the following measures for similarity/dissimilarity among
     data:
     a) Cosine measure
     b) Euclidean distance
     c) Manhattan measure.   [10]

6. Find the frequent itemsets and strong association rules for the following transactional database table using Apriori algorithm. Consider the thresholds as support = 30% and confidence = 40%. [10]

TID	ITEM IDs
1	I1,i2,i3,i5
2	I2,i5,i7,i9
3	I1,i3,i5,i7
4	I2,i4,i6,i8
5	I1,i2,i3,i4
6	I2,i3,i4,i5
7	I3,i4,i5,i6
8	I4,i5,i6,i7
9	I5,i6,i7.i8.i9
10	I9.i1.i2.i5
11	I8,i2,i9,i7
12	I5,i6,i3,i2

OR

7.a) Demonstrate construction of FP-tree for the data from Question (6).
  b) What is a closed item set? Is it similar to maximal item set? [5+5]

8.a) Define information gain and explain its importance in decision tree induction.
  b) Give the algorithm for decision tree induction. [5+5]

OR

9.a) State Bayes theorem. How can it be applied for data classification?
  b) With example explain Bayesian belief network. [5+5]

10. What is the main objective of clustering? Give the categorization of clustering approaches. Briefly discuss them. [10]

OR

11.a) Compare k-means with k-medoids algorithms for clustering.
  b) How to evaluate clustering algorithms? [5+5]

---ooOoo---