Tribhuvan University
Institute of Science and Technology
2081
✡

Bachelor Level / Forth Year / Seventh Semester / Science
Computer Science and Information Technology (CSC 410)
(Data Warehousing and Data Mining)
(NEW COURSE)

Full Marks :- 60
Pass Marks :- 24
Time :- 3 Hrs

Candidates are required to give their answers in their own words as for as practicable.
The figures in the margin indicate full marks.

Section A

Attempt any TWO questions.

[2 × 10 = 20]

1. When do we prefer trim mean for statistical description of data? Justify with an example. Describe about multi-dimensional data model and conceptual modeling of data warehouse.

[3 + 7]

2. How do you generate strong association rules? From the following dataset find the frequent item set using FP growth algorithm using 3 as minimum support.

[2 + 8]

Transaction ID	Items
T1	{K, E, M, O, Y}
T2	{K, E, O, Y}
T3	{K, E, M}
T4	{K, M, Y}
T5	{K, E, O}

3. Define over fitting and under fitting. Train the decision tree classifier using the ID3 algorithm based on the following training data.

[2 + 8]

TID	Age	Car Type	Class
1	≤30	Family	High
2	≤30	Sports	High
3	>30	Sports	High
4	>30	Family	Low
5	>30	Truck	Low
6	≤30	Family	High

Section B

Attempt any EIGHT questions

[8 × 5 = 40]

4. Describe any two methods for handling noisy data.

[5]

5. Using k-means++ algorithm and Euclidean distance, find the initial 3 cluster centroids from A1 = (3, 11), A2 = (3, 6), A3 = (9, 5), A4 = (6, 9), A6 = (7, 5), A7 = (2, 3), A8 = (5, 10). Choose (3, 11) as one of the initial centroid.

[5]

6. Explain the general strategies for cube computation.

[5]

7. Distinguish between data characterization and data discrimination. What are the challenges of multimedia mining?

[2 + 3]

IOST,TU

8. Define graph mining. Discuss about conflict between the theory of balance and theory of status.

$[2+3]$

9. What is support vector? How do you evaluate the accuracy of a classifier? Describe. $[2+3]$

10. Differentiate between k-means and k-medoids clustering algorithm.

$[5]$

11. List any two OLAP operations with example. How do you compute rule coverage and rule accuracy? $[2.5+2.5]$

12. Define link mining. What are the roles of epsilon and MinPts in DBSCAN. $[2+3]$