Tribhuvan University
Institute of Science and Technology
2080
✿

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

Section A

Attempt any **TWO** questions.   [2 × 10 = 20]

1. State Apriori property. Find frequent itemsets and association rules from the transaction database given below using Apriori algorithm. Assume min. support is 50% and min confidence is 75%.   [1+9]

Transaction ID	Items Purchased
1	Bread, Cheese, Egg, Juice
2	Bread, Cheese, Juice
3	Bread, Milk, Yogurt
4	Bread, Juice, Milk
5	Cheese. Juice, Milk

2. How classification differs from regression. Train ID3 classifier using the dataset given below. Then predict class label for the data [Age=Mid, Competition=Yes, Type=HW].   [2+8]

Age	Competition	Type	Profit ((Class Label
Old	Yes	SW	Down
Old	No	SW	Down
Old	No	HW	Down
Mid	Yes	SW	Down
Mid	Yes	HW	Down
Mid	No	HW	Up
Mid	No	SW	Up
New	Yes	SW	Up
New	No	HW	Up
New	No	SW	Up

3. Why the concept of data mart is important? Discuss different data warehouse scheme with examples.   [1+9]

Section B

Attempt any **EIGHT** questions   [8× 5 = 40]

4. How KDD differs from data mining? Explain various stages of KDD with suitable block diagram. [1+4]

5. Discuss different ways of smoothing noisy data along with suitable examples.   [5]

6. How many cuboids are possible from 5-dimensional data? Discuss the concept of full cube and iceberg cube.   [1+4]

7. How K-mediods clustering differs from K-means clustering? Divide the following data points into two clusters using k-mediods algorithm. Show computation upto 3 iterations.   [1+4]

{(70,85),(65,80),(72,88),(75,90), (60,50),(64,55),(62,52),(63,58)}

8. Discuss working of DBSCAN algorithm.   [5]

9. Which algorithm is used for training multi-layer perceptron? Discuss the algorithm in detail.   [5]

10. Explain the OLAP operations with example.   [5]

IOST,TU

11. Discuss the concept of multimedia data mining along with the concept of similarity search.   [5]

12. Write down short notes on   [2×2.5=5]
    a. Support vector Machine
    b. Multi-dimensional Data Model