Tribhuvan University
Institute of Science and Technology
2074

Bachelor Level i Forth Year /Eighth Semester/Science — Full Marks: 60
Computer Science and Information Technology-(CSc. 451) — Pass Marks: 24
(Data warehousing and Data mining) — Time: 3 hours.

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

Group A

Attempt any two questions. [2x 10=20]

1. You are given the transaction data shown below from a fast food restaurant. There are 9 distinct transactions (orderl --> order 9). There is total 5 meal (M1 --> M5) involved in the transactions.

Meal Item	List of Item IDs	Meal Item	List of Item IDs
Order 1	MI, M2, M5	Order 6	M2, M3
Order 2	M2, M4	Order 7	Ml, M3
Order 3	M2, M3	Order 8	Ml, M2, M3, M5
Order 4	Ml, M2, M4	Order 9	MI, M2, M3
Order 5	Ml, M3		

Minimum Support= 2, Minimum Confidence = 0.7
Apply the Apriori algorithm to the dataset to identify frequent k-itetnset and **find all strong** association rules. **[10]**

2. Why do we need to preprocess the data before running the algorithm? What are the processes for this? Explain. Give some examples of noise that must be removed in data while extracting the pattern. [2+5+3]
3. List the two steps used in classification approach with its issues. Is this right decision to use neural network always as a classifier? Give your opinion. Discuss the working mechanism of back propagation classification algorithm. [2+3+5]

Group B

Attempt any eight questions. [8x5=40]
Question No. 13 is compulsory.

4. List and describes the five primitives for specifying a data mining task. [5]
5. Describe the types of data used in data mining. [5]
6. Explain the similarities and dissimilarities between operational database and data warehouse. [5]
7. List the types of OLAP operations with examples. [5]
8. Illustrate the strength and weakness of k-means in comparison with k-medoids algorithm. [5]
9. Why data cube computation is essential task in data mining? Describe the general strategies. [5]
10. Describe the different components of a data warehouse.[5]
11. Define dimension table and fact table. What makes the necessity of multidimensional data model?[2+3]
12. Discuss the approach behind Bayesian classification. Why smoothing technique is necessary in Bayesian classification? [3+2]
13. Write short notes. (Any Two): [2.5×2=5]
    a. Concept hierarchy
    b. Data Mining Query Language
    c. Text Mining
    d. ROLAP vs MOLAP