

Tribhuvan University  
**Institute of Science and Technology**  
 2075  
 ☆

Bachelor Level / Forth Year / Eighth Semester / Science  
**Computer Science and Information Technology-(CSc. 451)**  
**(Data warehousing and Data mining)**

Full Marks: 60  
 Pass Marks: 24  
 Time: 3 hours.

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

**Group A**

**Attempt any two questions.**

[2×10=20]

1. List some issues of multimedia mining. Describe how back propagation is used in classification. [3 + 7]
2. Describe how bitmap and join indexing are used to represent OLAP data. Explain the different components of data warehouse. [4 + 6]
3. Give any two types of association rules with examples. Trace the results of using the Apriori algorithm on the grocery store example with support threshold 2 and confidence threshold 60%. Show the candidate and frequent itemsets for each database scan. Enumerate all the final frequent itemsets. Also indicate the association rules that are generated. [4 + 6]

Transaction_ID	Items
T1	HotDogs, Buns, Ketchup
T2	HotDogs, Buns
T3	HotDogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	HotDogs, Coke, Chips

**Group B**

**Attempt any eight questions.**

[8×5=40]

**Question No. 13 is compulsory.**

4. What is the purpose of cluster analysis in data mining? Explain. [5]
5. How does KDD differ with data mining? Describe the stages of data mining. [1 + 4]
6. Explain OLAP operations with examples. [5]
7. Explain the primitives of data mining query language. [5]
8. How different schema are used to model data warehouse? Explain. [5]  
 - snowflake  
 - star  
 - Data cube  
 - Fact constellation
9. Describe the significances of pre computation of data cube. [5]

10. How text mining can be used in social networking? Give your own opinion. [5]
11. Describe any five applications of data mining. — [5]
12. How ID3 algorithm can be used as attribute selection while building decision tree? Explain. [5]
13. Write short notes on (Any TWO): [2.5 + 2.5]
- a. Outlier Analysis
  - b. Data warehouse back end tools *ETL*
  - c. Virtual data warehouse
  - d. Market basket analysis