TRIBHUVAN UNIVERSITY

FACULTY OF MANAGEMENT

Office of the Dean

Full Marks: 40

Pass Marks: 18

Time: 2 hrs.

April 2021

BIM / Eighth Semester / IT 308: Data Mining and Data Warehousing

Candidates are required to answer all the questions in their own words as far as practicable.

Group "A"

1. **Brief Answer Questions:** *[10 × 1 = 10]*

   i. How data mining differ with data warehousing?

   ii. What are the main parameters to be considered in capacity planning?

   iii. What is the role of damping factor in search engine?

   iv. How do you compare two classifier?

   v. What are the complexities in multimedia datamining?

   vi. List any two types of noise in data.

   vii. Mention any two limitations of k-means algorithm.

   viii. Give any two applications of time series data mining.

   ix. What are the possible limitations of data mining?

   x. What is the role of minimum support?

Group "B"

Exercise Problems: *[5 × 4 = 20]*

2. Why data preprocessing is necessary before mining? Explain some techniques to clean the data.

3. Explain the Link Analysis page ranking algorithm.

4. Consider the following transaction data sets.

TID	Items
1	E, A, D, B
2	D, A, C, E, B
3	C, A, B, E
4	B, A, D
5	D
6	D, B
7	A, D, E
8	B, C

Now, construct the FP Tree.

5. What are the roles of MinPts and epsilon in DBSCAN algorithm? Explain.

6. Consider the 14 training datasets with 9 positive and 5 negative classes. Suppose one of the attributes is Wind, which have values Weak and Strong. There are 8 occurrences of weak winds and 6 occurrences of strong winds. For the weak winds, 6 are positive and 2 are negative. For the strong winds, 3 are positive and 3 are negative. Calculate the information gain of wind.

Group "C"

[2 × 5 = 10]

Comprehensive Questions:

7. What are the needs of multidimensional data model? Describe about web usage mining and web content mining.

8. Explain ETL process. Distinguish between OLAP and OLTP.

✿✿✿