

Tribhuvan University
Institute of Science and Technology
 2082
 ☆

Bachelor Level / Fourth 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 far as practicable.
 The figures in the margin indicate full marks.

Section A

Attempt any **TWO** questions.

[2 × 10 = 20]

1. Define strong association rule. What are the limitations of APriori algorithm? Create a FP tree from the following data set. [2 + 2 + 6]

TID	List of Items
T1	{A, B, C}
T2	{B, C, D}
T3	{C, D}
T4	{B, D}
T5	{A, C}
T6	{A, C, D}

2. What is the role of Laplace smoothing? Create a decision tree from the following data set using ID3 as attribute selection approach. [2 + 8]

Object	A1	A2	Class
1	T	T	C1
2	T	T	C1
3	T	F	C2
4	F	F	C1
5	F	T	C2
6	F	T	C2

3. Consider the data set (6,3), (7,2), (4,8), (2,2), (0,2), (9,0). Taking $k = 3$, show the result after first iteration using k – means algorithm. For choosing initial centroid, use k – means++ by taking (6, 3) as initial cluster center. [10]

Section B

Attempt any **EIGHT** questions

[8 × 5 = 40]

4. Explain about data mining primitives. [5]
5. Define support vector. Write the algorithm for back propagation for classification. [1 + 4]
6. What is data mart? Why do we need multidimensional data model? [2 + 3]
7. Describe the different types of data object and attribute types. [5]
8. What is data cube? List the different variations of cube materializations. [1 + 4]

9. What is the concept behind beam search? Discuss about theory of balance and status. [2 + 3]
10. Explain about web content, web usage and web structure mining. [5]
11. Given the following distance matrix, find the core points and outliers using DBSCAN. Take Eps = 2.5 and MinPts = 3. [5]

Data Points	A	B	C	D	E	F	G	H
A	0	1.41	2.83	4.24	5.66	5.83	6.40	5.83
B		0	1.41	2.82	4.24	4.47	5.00	4.47
C			0	1.41	2.82	3.16	3.60	3.16
D				0	1.41	2.00	2.24	2.00
E					0	1.41	1.00	1.41
F						0	1.00	2.82
G							0	2.24
H								0

12. List the components of data warehouse. Discuss about the trust propagation on social network. [2 + 3]