TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2076 Ashwin

Exam.		Back	
Level	BE	**Full Marks**	80
Programme	BEX, BCT	**Pass Marks**	32
Year / Part	IV / 1	**Time**	3 hrs.

Subject: - Data Mining (Elective I) (CT72502)

✓ Candidates are required to give their answers in their own words as far as practicable.
✓ Attempt __All__ questions.
✓ The figures in the margin indicate __Full Marks__.
✓ Assume suitable data if necessary.

1. What are the fundamental differences between Data Mining and Data Warehousing? Describe the steps of KDD for data mining. [3+7]

2. What do you mean by dimensional data? What are base & apex cuboid? Slicing & Dicing? Roll Down and Roll UP operations? Give example. [2+3+3+3]

3. How do you measure the accuracy of classifiers? How do you select best root attribute in decision tree? Explain. [4+6]

4. What are prior and posterior probabilities? Explain the algorithmic steps of Bayesian classifier and write its strengths. [3+7]

5. For the transactions given below, consider confidence=60% and minimum support=30%. Identify large itemsets (L-Itemset) at L=3 with possible associations using A-priori algorithm and generate F-List using FP-Growth algorithm. [12]

Transactions	Items description
T1	A, B, C, T, M, P, D, K
T2	A, B, T, P, D, K
T3	B, C, T, D, M, A, P
T4	A, C, T, M, D,
T5	A, C, D, K, M
T6	B, C, T

6. How DBSCAN algorithm works? How do we avoid the issues of DBSCAN? [8+2]

7. Explain web mining taxonomy. [8]

8. Write short notes on (__Any Three__) [3+3+3]

   a. Data smoothing techniques

   b. Clustering and its application in anomaly detection

   c. AprioriALL: Sequential pattern mining algorithm

   d. Various similarity measures between data tuples.

***

Exam.			Back	
Level	BE		Full Marks	80
Programme	BCT, BEX		Pass Marks	32
Year / Part	IV / I		Time	3 hrs.

Subject: - Data Mining *(Elective I) (CT72502)*

✓ Candidates are required to give their answers in their own words as far as practicable.
✓ Attempt **All** questions.
✓ The figures in the margin indicate **Full Marks**.
✓ Assume suitable data if necessary.

1. How is data warehouse different from a database? How are they similar? [2+2]

2. Discuss issues to consider during Data Integration. Describe OLAP and operations on OLAP with suitable example. [5+5]

3. Explain Naïve Bayesian classification with suitable example. [8]

4. The confusion matrix for a classifier is given as follows: [10]

		Predicted Class	
		Class 1	Class 2
Actual Class	Class 1	21	6
	Class 2	7	41

   Calculate: Accuracy, Sensitivity, Specificity and Precision.

5. Why association analysis is required in data mining? Explain Apriori principle with example. [2+6]

6. What are the advantages of FP growth method? Explain FP growth algorithm. [2+6]

7. Explain K-means clustering with limitation. Generate two clusters from following dataset using K-means clustering. [4+6]

A	B
1	2
2.5	4.5
4	6
3.5	4
4	5.5
3	6

8. What are outliers? Explain an algorithm that can be used to generate density based clusters. [8]

9. Why anamoly detection is important? Explain distance based method for anamoly detection. [2+6]

10. Explain Web mining and Multimedia mining. [6]

***

Exam.		**Regular**	
Level	BE	Full Marks	80
Programme	BEX, BCT	Pass Marks	32
Year / Part	IV / I	Time	3 hrs.

Subject: - Data Mining *(Elective I) (CT72502)*

✓ Candidates are required to give their answers in their own words as far as practicable.
✓ Attempt __All__ questions.
✓ The figures in the margin indicate __Full Marks__.
✓ Assume suitable data if necessary.

1. What is data warehouse and data mart? Describe Snowflake scheme with example. [2+4]

2. What are the approaches to handle missing data? Describe OLAP and operations on OLAP with suitable example. Differentiate between OLAP and OLTP. [2+5+3]

3. Draw clear block diagram depicting different stages in classification. Explain the inverse relation between precision and recall. Given the confusion matrix, determine accuracy, sensitivity and precision of the classifier model. [2+3+5]

Predicted Actual	Positive	Negative
Positve	142	40
Negative	98	720

4. Explain decision tree with the concept of Naive base classification with appropriate example. [10]

5. Why association analysis is required in data mining? Explain apriori principle with example. [2+6]

6. How does FP growth approach overcomes the disadvantages of Apriori algorithm. For the transaction data given in table generate FP-Tree. [2+8]

Transaction ID	Item set
T1	Camera, Laptop, Pen drive
T2	Laptop, Pen drive
T3	Laptop, Mobile, Earphone
T4	Earphone, Mobile
T5	Camera, Earphone
T6	Laptop, Mobile, Earphone

7. Describe the difference between Hierarchical and partitioning clustering. How K-means clustering is applied? Verify using example. [2+8]

8. What do you mean by anomaly detection and why is it important? Describe distance based approaches for anomaly detection. [4+3]

9. Write short notes on: (any three) [3×3]
   i) Issues in clustering
   ii) Multimedia mining
   iii) Time series data mining
   iv) Web mining

***