TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2075 Chaitra

Exam.	Regular / Back		
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. Explain Data Warehouse architecture with its analytical processing. [8]
2. Why data preprocessing is necessary? Explain the methods for data preprocessing to maintain data quality. [4+4]
3. Define Decision Tree Classifier with Gini-Index with suitable example. How can you handle overfitting in Decision Tree? [6+4]
4. What do you mean by frequent Pattern growth, draw FP-tree with given tabular data. [4+4]

TID	Items
01	f, a, c, d, g, i, m, p
02	a, b, c, f, l, m, o
03	b, f, h, j, o, w
04	b, c, k, s, p
05	A, f, c, e, l, p, m, n

5. How ANN works? Explain with Algorithm. [8]
6. What is the application of clustering in data mining? Explain K-means clustering with example. [2+6]
7. How DBSCAN clustering is used for handling noise in data? [8]
8. What is outlier? Explain the distance base approaches for the anomaly detection. [5]
9. What are the challenges of web mining? Explain about time series data mining with an example. [5]
10. Write short notes on: (Any three) [4+4+4]
    a) Market Basket Analysis
    b) Visual Data Mining
    c) OLAP and OLTP
    d) Data Normalization

***

For given support 60% & confidence 60% then with support be

Total Scanned with CamScanner

36 B TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2073 Chaitra

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. How is data warehouse different from RDBMS? Also list the similarities. [2+2]
2. What is data pre-processing? Explain data sampling and dimensionality reduction in data pre-processing with suitable example. [2+4+4]
3. How data in most real application becomes Asymmetric. Explain the difference between symmetric and asymmetric data. [5]
4. What is ID3 algorithm? Calculate TPR, FPR and Accuracy for given confusion matrix. [2+6]

	Predicted +	Predicted -
Predicted +	100	40
Predicted -	60	300

5. Explain Apriori algorithm in market basket analysis? Derive association rule from the following market basket transactions with 50% of minimum support and confidence respectively. [3+7]

Transaction	Itemsets
1	A, B, C
2	A, C
3	A, D
4	B, E, F

6. What is the use of FP-Growth method in market basket analysis? Explain FP-Growth method with a suitable example. [10]
7. How clustering differ from classification? Given the one-dimensional points {5, 12, 18, 24, 30, 42, 48} with initial centroids {5, 12, 18}, create three clusters by K-Means algorithm and calculate SSE for this clustering result. [4+8]
8. Explain Sequential Pattern and Sub-graph Pattern with suitable example. [4+4]
9. What is anomaly detection? Explain the issues associated with anomaly detection. [2+3]
10. Write short notes on: (Any two) [2×4]
    a) Time series data mining
    b) Overfitting and ROC
    c) www mining

***

26B TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2074 Chaitra

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]

Actual \ Predicted	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

***

37B

TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2073 Shrawan

Exam.	New Back (2066 & Later Batch)		
Level	BE	Full Marks	80
Programme	BE, BCT	Pass Marks	32
Year / Part	IV / I	Time	3 hrs.

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

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

1. "The world is data rich but information is poor". Justify with your own words. [8]
2. What are the measuring elements of data Quality? Explain different data transformation by normalization methods with an example. [2+6]
3. What is a decision tree and how information gain is used for attribute selection? Explain with example. [8]
4. Explain ROC. Using the following data, calculate TPR, FPR, precision for given confusion matrix. [1+3+6]

	A	B
A	20	5
B	10	40

Classify, A = Yes, B = No

5. What is FP Tree? How FP-growth algorithm eliminate the problem of Apriori algorithm? Construct the FP tree and find association rules for the following transaction database using FG- Growth algorithm. Support = 30% and confidence = 75%. [10]

Transaction ID	Items
1	P,R,S
2	R,S,T
3	P,Q,R
4	P,R,S,T
5	P,S,T
6	P,Q,T
7	Q,S,T
8	Q,R,T

6. What are Categorical data? What are the possible issues arrives when using Categorical data? How can you handle such issues? [2+3+3]
7. What is the application of clustering in data mining? Explain the k-means algorithm with example. [8]
8. What is anamoly detection? Explain distance based method for anamoly detection. [8]
9. Write short notes on: [4×3]
   i) Data transformation
   ii) Web mining
   iii) OLAP

***

34B TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2072 Chaitra

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

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

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

1. What is data mining? Explain all the steps of knowledge discovery. [2+6]
2. How do you perform analysis of multidimensional data? Explain with the concept of OLAP. [10]
3. Predict Class label using naive Bayesian classifier for X = (age = youth, income = medium, student = yes, credit-rating = fair) using the following data set. [10]

RID	Age	Income	Student	Credit-rating	Class Buy computer
1	Youth	High	No	Fair	No
2	Youth	High	No	Excellent	No
3	Middle-age	High	No	Fair-	Yes
4	Senior	Medium	No	Fair	Yes
5	Senior	Low	Yes	Fair	Yes
6	Senior	Low	Yes	Excellent	No
7	Middle-age	Low	Yes	Excellent	Yes
8	Youth	Medium	No	Fair	No
9	Youth	Low	Yes	Fair	Yes
10	Senior	Medium	Yes	Fair	Yes
11	Youth	Medium	Yes	Excellent	Yes
12	Middle-age	Medium	No	Excellent	Yes
13	Middle-age	High	Yes	Fair	Yes
14	Senior	Medium	No	Excellent	No

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

		actual class	
		class1	class2
predicted class	class1	21	6
	class2	7	41

calculate a. accuracy
b. sensitivity
c. specificity
d. precision
e. recall

5. What is the importance of SUPPORT and COFIDENCE during association analysis? Explain <u>FP-Growth</u> method with example.
6. What are the types of clustering methods? Explain DBSCAN method of clustering with an example.
7. What is the use of Apriori Algorithm in market basket analysis? Explain with suitable example.
8. Write short notes on:
   i) Time series Data mining
   ii) Issues in anomaly/Fraud detection
   iii) Categorical data and relational [4]

27C TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2072 Kartik

Exam.	[illegible]		
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 a data mining? Explain general steps in brief. [4]
2. Why data preprocessing is required in the data mining? Explain some of approaches of data clearing. [5+5]
3. Write about Hunt's Algorithm for Decision Tree induction. Explain the test conditions that can be used for different attribute types. [10]
4. What is an ANN classifier? Explain its general consideration that required for the classifier. [2+6]
5. What is an association analysis? Explain its importance in market-basket analysis. [2+5]
6. What is a Frequent item set? Explain FP growth method with example. [1+8]
7. What is a cluster analysis? How it is different from classification? [5]
8. Explain a DBSCAN algorithm with example. [5]
9. What is an Anomaly detection? Discuss its importance in security. [5]
10. Explain Time series data mining in brief. [6]
11. Write short notes on: [3×3]
    i) Data transformation
    ii) Sequential pattern
    iii) Cluster evaluation

***

24C TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2071 Chaitra

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.*
- *All questions carry equal marks.*
- *Assume suitable data if necessary.*

1. What is a Data Mining? Explain its application.
2. Explain the properties that a Distance Metric needs to support with respect to Minkowski's distance.
3. What is a decision tree? Explain Gini Index with suitable example.
4. Explain a Bayes classifier. In what cases can Naive Bayes and Bayesian Belief Network be used?
5. Why is a clustering an unsupervised learning? How can hierarchical clusters be generated using Bisecting K-means algorithm?
6. Explain the different measures of cluster validity.
7. How does Apriori Algorithm optimize the brute force approach for frequent item set generation?
8. What is an Anomaly Detection? Explain few distance based approaches that can be used for Anomaly Detection.

***

TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2071 Shawan

Exam.	New Back (2068 and Later)	Full Marks	80
Level	BE	Pass Marks	32
Programme	BEX, BCT	Time	3 hrs.
Year / Part	IV / I		

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

- ✓ Candidates are required to give their answers in their own words as far as practicable.
- ✓ *Attempt All questions.*
- ✓ *All questions carry equal marks.*
- ✓ *Assume suitable data if necessary.*

1. What is data mining? Explain different data types of attributes in a dataset.
2. How can principle component analysis be used for dimentionality reduction?
3. Why is classification a super vised learning method? Explain different impurity measures used in decision tree classifier.
4. Explain Naive Bayes classifier. How can over fitting problem be solved in case of classification?
5. Explain FP-growth algorithm in detail.
6. What are association rules? How can spriori algorithm be used to generate association rules.
7. What is contiguous cluster? Explain an algorithm that can be used to generate contiguous clusters.
8. Explain K-means clustering with limitation Use k-means clustering to cluster the following dataset.

A	B
1.0	1.0
1.5	2.0
3.0	4.0
5.0	7.0
3.5	5.0
4.5	5.0
3.5	4.5

9. How can Nearest-Neighbor algorithm be used for anomaly defection?
10. Write short notes on:
    a) Time-series data mining
    b) Data warehouse on data mining

INSTITUTE OF ENGINEERING
Examination Control Division
2070 Chaitra

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

Subject: - Data Mining *(Elective I) (CT725)*

- ✓ 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. a) What is "curve of Dimensionality"? How can it be avoided? [5]
   b) Discuss the impact of noisy data in data mining? [5]
2. Explain rule based classifier? How can CN2 Algorithm be used for rule based classification? Define "Accuracy" and "Laplace" measures used for rule evaluation. [9]
3. An input sequence "A A B B B A A A B B" was used for classification. The Classifier 'X' predicted the sequences as: "A A B B B A A A B B" where as the Classifier 'Y' predicted the sequences as: "A A A A B B A A A B". Develop the corresponding confusion matrix for the classifiers and find their corresponding. [10]
   i) Accuracy
   ii) Precision
   iii) True Positive Rate
   iv) False Positive Rate
4. Explain Apriori algorithm. Use Apriori to generate frequent item sets with support of 50% for the following transaction database. [10]

TID	Items
1	ACD
2	BD
3	ABCE
4	BDF

5. Why is pattern evaluation important in association rule mining? Explain with example the statistical based measures used for measuring interestingness of association rules. [8]
6. What is a density based cluster. Explain an algorithm that can be used to generate density based clusters. [8]
7. What is Hierarchical Clustering? Differentiate between agglomerative and divisive approach of hierarchical clustering. Augment your answer with appropriate illustrative examples. [10]
8. [illegible] [15]