## 2024

## **DATA MINING**

Full Marks: 100

Time: Three hours

## The figures in the margin indicate full marks for the questions.

Answer any five questions.

1.	Answer the following questions:				
	a)	Match the following "Pre-Processing approach" with their "tasks":		1x6=6	
		<b>Pre-Processing approach</b>	Tasks		
		Data Cleaning,	Data Compression, To fill the missing values,		
		Data Transformation,	To handle high dimensionality,		
		Data Reduction Concept hierarchy,			
			To resolve inconsistencies, Normalization		
	b)	True or False:	9474	1x8=8	
			to equipment malfunction.		
	(ii) Binning is used to handle noisy data.				
		<ul> <li>(iii) K-means handles outlier well.</li> <li>(iv) AGNES is a top down approach.</li> <li>(v) In DIANA iteratively clusters are merged together.</li> </ul>			
		(vi) STING is a grid based clustering approach.			
		(vii) There will an arbitrary sha	ape of cluster in density based clustering.		
		(viii) The more detailed to less	detailed data is retrieve in drill down operation.		
	c)	Whether or not each of the following activities is a data mining task (Yes/No)			
		(i) Dividing the customers	Dividing the customers of a company according to their gender.		
		(ii) Finding the name of page	atients having blood pressure in database.		
		(iii) Handwritten Digit Rec	git Recognition.		
		(iv) Fraudulent and Abusiv	e Data.		
		(v) Anomaly detection.			
		(vi) Predicting the Covid-1	9 behavior.		
2.	a)	What is the need of data clear	ning and data reduction in data-preprocessing?	4	
	b)	Apply the Min-max normalization by setting $min = 0$ and $max = 1$ on the		8	
		following group of data value	s 60, 100, 200, 300, 400, and 500.	Ü	
	c)	Find out the three clusters using the k-medoids algorithm for the given data			
		objects {4, 8, 12, 18, 3, 22, 32, 10, 24}. (Hint k=3 and keep continue till the			
		no change in the formed clusters)			
3.	a)	What are the weakness and st	rength of Neural Network?	4	
	b)	How a multi-layer neural nety	vork works?	4	

c) Apply the Bayesian classification for predicting the buys_comp of the given test sample data, x= ((3140, HIGH, Y, EXCELLENT)    Age	
X = ((3140, HIGH, Y, EXCELLENT)   Age	
Age Income Student Credit_rating Class:Buys_comp (=30 HIGH N Excellent N N   N   N   N   N   N   N   N   N	
A   A   A   A   A   A   A   A   A   A	
Color   Colo	1
3140   NIGH   N   FAIR   Y    -40   MEDIUM   N   FAIR   Y    -40   LOW   Y   FAIR   Y    -40   LOW   Y   EXCELLENT   N    -40   LOW   Y   EXCELLENT   Y    -40   LOW   Y   EXCELLENT   Y    -40   MEDIUM   N   FAIR   N    -40   MEDIUM   Y   FAIR   Y    -40   MEDIUM   N   EXCELLENT   Y    -41   3140   HIGH   Y   FAIR   Y    -42   MEDIUM   N   EXCELLENT   N    -43   Medium   N   EXCELLENT   N    -44   MEDIUM   N   EXCELLENT   N    -45   MEDIUM   N   EXCELLENT   N    -46   MEDIUM   N   EXCELLENT   N    -47   MEDIUM   N   EXCELLENT   N    -48   Medium   N   EXCELLENT   N    -49   MEDIUM   N   EXCELLENT   N    -40   MEDIUM   N   EXCELLENT   N    -	1
A   What do you mean by Association rules?	as a
4. a) What do you mean by Association rules?  b) What are the types of Association rules?  c) Describe the frequent itemsets.  d) A database has five transactions. Let minimum support is 50%.  Find all frequent item sets using Apriori algorithm.  TID Item sets  T100 {1, 2, 3, 4}  T200 {1, 2, 3, 4, 5}  T400 {3, 4, 5, 6}   Write short notes on the following (any four):  a) Decision Tree  b) Data Warehousing  c) Dendogram  A caso we feat the fair of	
4. a) What do you mean by Association rules?  b) What are the types of Association rules?  c) Describe the frequent itemsets.  d) A database has five transactions. Let minimum support is 50%.  Find all frequent item sets using Apriori algorithm.  TID Item sets  T100 {1, 2, 3, 4}  T200 {1, 2, 3, 4, 5}  T300 {2, 3, 4, 5}  T400 {3, 4, 5, 6}   Write short notes on the following (any four):  a) Decision Tree  b) Data Warehousing  c) Dendogram  d) Data Cube	
Si40   LOW   Y   EXCELLENT   Y	12
4. a) What do you mean by Association rules?  b) What are the types of Association rules?  c) Describe the frequent itemsets.  d) A database has five transactions. Let minimum support is 50%. Find all frequent item sets using Apriori algorithm.  TID Item sets  T100 {1, 2, 3, 4}  T200 {1, 2, 3, 5, 6}  T300 {2, 3, 4, 5}  T400 {3, 4, 5, 6}  T400 {3, 4, 5, 6}   Write short notes on the following (any four):  a) Decision Tree  b) Data Warehousing  c) Dendogram  Decident itemsets  C) Dendogram  Data Cube	
C=30	
A company   Part   Pa	
3140 MEDIUM N EXCELLENT Y 3140 HIGH Y FAIR Y >40 MEDIUM N EXCELLENT N  4. a) What do you mean by Association rules? b) What are the types of Association rules? c) Describe the frequent itemsets. d) A database has five transactions. Let minimum support is 50%. Find all frequent item sets using Apriori algorithm.  TID Item sets T100 {1, 2, 3, 4} T200 {1, 2, 3, 5, 6,} T300 {2, 3, 4, 5} T400 {3, 4, 5, 6}   5. Write short notes on the following (any four):  a) Decision Tree  b) Data Warehousing  c) Dendogram  d) Data Cube	
4. a) What do you mean by Association rules? b) What are the types of Association rules? c) Describe the frequent itemsets. d) A database has five transactions. Let minimum support is 50%. Find all frequent item sets using Apriori algorithm.  TID Item sets  T100 {1, 2, 3, 4}  T200 {1, 2, 3, 5, 6,}  T300 {2, 3, 4, 5}  T400 {3, 4, 5, 6}   5. Write short notes on the following (any four):  a) Decision Tree  b) Data Warehousing  c) Dendogram  d) Data Cube	
4. a) What do you mean by Association rules? b) What are the types of Association rules? c) Describe the frequent itemsets. d) A database has five transactions. Let minimum support is 50%. Find all frequent item sets using Apriori algorithm.  TID Item sets  T100 {1, 2, 3, 4}  T200 {1, 2, 3, 5, 6,}  T300 {2, 3, 4, 5}  T400 {3, 4, 5, 6}   5. Write short notes on the following (any four):  a) Decision Tree  b) Data Warehousing  c) Dendogram  d) Data Cube	
4. a) What do you mean by Association rules? b) What are the types of Association rules? c) Describe the frequent itemsets. d) A database has five transactions. Let minimum support is 50%. Find all frequent item sets using Apriori algorithm.  TID Item sets T100 {1, 2, 3, 4} T200 {1, 2, 3, 5, 6,} T300 {2, 3, 4, 5} T400 {3, 4, 5, 6}   5. Write short notes on the following (any four): a) Decision Tree b) Data Warehousing c) Dendogram	
b) What are the types of Association rules?  c) Describe the frequent itemsets.  d) A database has five transactions. Let minimum support is 50%. Find all frequent item sets using Apriori algorithm.  TID Item sets  T100 {1, 2, 3, 4}  T200 {1, 2, 3, 5, 6, }  T300 {2, 3, 4, 5}  T400 {3, 4, 5, 6}   5. Write short notes on the following (any four):  a) Decision Tree  b) Data Warehousing  c) Dendogram  d) Data Cube	
b) What are the types of Association rules?  c) Describe the frequent itemsets.  d) A database has five transactions. Let minimum support is 50%. Find all frequent item sets using Apriori algorithm.  TID Item sets  T100 {1, 2, 3, 4}  T200 {1, 2, 3, 5, 6, }  T300 {2, 3, 4, 5}  T400 {3, 4, 5, 6}   5. Write short notes on the following (any four):  a) Decision Tree  b) Data Warehousing  c) Dendogram  d) Data Cube	
b) What are the types of Association rules?  c) Describe the frequent itemsets.  d) A database has five transactions. Let minimum support is 50%. Find all frequent item sets using Apriori algorithm.  TID Item sets  T100 {1, 2, 3, 4}  T200 {1, 2, 3, 5, 6, }  T300 {2, 3, 4, 5}  T400 {3, 4, 5, 6}   5. Write short notes on the following (any four):  a) Decision Tree  b) Data Warehousing  c) Dendogram  d) Data Cube	
c) Describe the frequent itemsets.  d) A database has five transactions. Let minimum support is 50%. Find all frequent item sets using Apriori algorithm.  TID Item sets  T100 {1, 2, 3, 4}  T200 {1, 2, 3, 5, 6,}  T300 {2, 3, 4, 5}  T400 {3, 4, 5, 6}   5. Write short notes on the following (any four):  a) Decision Tree  b) Data Warehousing  c) Dendogram  d) Data Cube	3
d) A database has five transactions. Let minimum support is 50%.  Find all frequent item sets using Apriori algorithm.  TID Item sets  T100 {1, 2, 3, 4}  T200 {1, 2, 3, 5, 6,}  T300 {2, 3, 4, 5}  T400 {3, 4, 5, 6}    Write short notes on the following (any four):  a) Decision Tree  b) Data Warehousing  c) Dendogram  d) Data Cube	3
d) A database has five transactions. Let minimum support is 50%.  Find all frequent item sets using Apriori algorithm.  TID Item sets  T100 {1, 2, 3, 4}  T200 {1, 2, 3, 5, 6,}  T300 {2, 3, 4, 5}  T400 {3, 4, 5, 6}    Write short notes on the following (any four):  a) Decision Tree  b) Data Warehousing  c) Dendogram  d) Data Cube	2
Find all frequent item sets using Apriori algorithm.    TID	
TID Item sets  T100 {1, 2, 3, 4}  T200 {1, 2, 3, 5, 6,}  T300 {2, 3, 4, 5}  T400 {3, 4, 5, 6}   5. Write short notes on the following (any four):  a) Decision Tree  b) Data Warehousing  c) Dendogram  d) Data Cube	12
T100 {1, 2, 3, 4} T200 {1, 2, 3, 5, 6, } T300 {2, 3, 4, 5} T400 {3, 4, 5, 6}   S. Write short notes on the following (any four):  a) Decision Tree  b) Data Warehousing  c) Dendogram  d) Data Cube	
T100 {1, 2, 3, 4} T200 {1, 2, 3, 5, 6, } T300 {2, 3, 4, 5} T400 {3, 4, 5, 6}   S. Write short notes on the following (any four):  a) Decision Tree  b) Data Warehousing  c) Dendogram  d) Data Cube	
T200 {1, 2, 3, 5, 6, } T300 {2, 3, 4, 5} T400 {3, 4, 5, 6}   T400 [4]  T200 {1, 2, 3, 5, 6, }  T300 {2, 3, 4, 5}  T400 [4]  T400 [5]  T400 [6]  T400 [6]  T300 [2]  T400 [6]  T300 [2]  T400 [6]  T400 [6]  T300 [6]  T400 [6]  T300 [6]  T400 [6]  T300 [6]  T400 [6]  T300 [6]  T300 [6]  T400 [6]  T300 [6]  T400 [6]  T300 [6]  T400 [6]  T4	
T300 {2, 3, 4, 5 } T400 {3, 4, 5, 6 }   Write short notes on the following (any four):  a) Decision Tree  b) Data Warehousing  c) Dendogram  d) Data Cube	
5. Write short notes on the following (any four):  a) Decision Tree  b) Data Warehousing  c) Dendogram  d) Data Cube	
5. Write short notes on the following (any four):  a) Decision Tree  b) Data Warehousing  c) Dendogram  d) Data Cube	
5. Write short notes on the following (any four):  a) Decision Tree  b) Data Warehousing  c) Dendogram  d) Data Cube	
<ul> <li>a) Decision Tree</li> <li>b) Data Warehousing</li> <li>c) Dendogram</li> <li>d) Data Cube</li> </ul>	
<ul> <li>a) Decision Tree</li> <li>b) Data Warehousing</li> <li>c) Dendogram</li> <li>d) Data Cube</li> </ul>	
<ul> <li>a) Decision Tree</li> <li>b) Data Warehousing</li> <li>c) Dendogram</li> <li>d) Data Cube</li> </ul>	
b) Data Warehousing c) Dendogram d) Data Cube	4x5=20
b) Data Warehousing c) Dendogram d) Data Cube	
c) Dendogram d) Data Cube	
c) Dendogram d) Data Cube	
d) Data Cube	
d) Data Cube	
e) OLAP operations	
c) OLAI operations	
6 Differentiate between the City is a	
6. Differentiate between the following (any four):	4x5 = 20
a) Classification and Clustering	
a) Classification and Clustering	
b) Data Matrix and Dissimilarity matrix	
b) But Hatin and Bissinnarity matrix	
c) Support and Confidence (ARs)	
. ,	
d) OLAP and OLTP	
e) Binary variables and Nominal variables	