



PRESIDENCY UNIVERSITY
BENGALURU
SCHOOL OF ENGINEERING

Roll No.																			
----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

TEST 2

Winter Semester: 2021 - 2022

Course Code: BCA211

Course Name: DATA MINING

Program & Sem: BCA & IV Sem

Date: 1st June 2022

Time: 10.00 AM TO 11.00 AM

Max Marks: 30M

Weightage: 15%

Instruction:

(i) *Read the all questions carefully and answer accordingly.*

Part A [Memory Recall Questions]

Answer ALL the Questions. Each question carries TWO marks.

(4Qx2M=8M)

1. Define Data Mining. [2M] (CO1) [Knowledge]
2. Briefly explain on category of Data Mining Tasks. [2M] (CO1) [Knowledge]
3. Describe Data Warehouse and its advantages. [2M] (CO1) [Knowledge]
4. Explain the types of Data Sets. [2M] (CO2) [Knowledge]

Part B [Thought Provoking Questions]

Answer both Questions. Each question carries SIX marks.

(2Qx6M=12M)

5. Describe in detail the challenges on Data Mining [6M] (CO1) [Comprehension]
6. For the following vectors, X and Y, Calculate the Similarity and Dissimilarity measures. [6M] (CO2) [Comprehension]
 - i) $x = (2, 2, 2, 2)$, $y = (1, 1, 1, 1)$ cosine, Euclidean
 - ii) $x = (1, 0, 1, 0)$, $y = (0, 1, 0, 1)$ cosine, SMC, Euclidean

Part C [Problem Solving Questions]

Answer the Question. The question carries TEN marks.

(1Qx10M=10M)

7. a. Use the Min-Max normalization to normalize, by setting min=5, max=10. the following group of data: 35, 55, 70, 75. Calculate is the normalized data for all the given numbers. [4M] (CO2) [Application]
- b. Use the Z-score normalization to normalize the following group of data: 20, 25, 40, 45. Calculate is the normalized data for all the given numbers. [6M] (CO2) [Application]



Roll No.																			
----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

**PRESIDENCY UNIVERSITY
BENGALURU**

SCHOOL OF ENGINEERING

TEST – 2

Winter Semester: 2021 - 22
Course Code: BCA 211
Course Name: DATA MINING
Program & Sem: BCA & IV

Date: 2nd June 2022
Time: 10.00 AM TO 11.00 AM
Max Marks: 30
Weightage: 15%

Instructions:

(ii) Read the all questions carefully and answer accordingly.

Part A [Memory Recall Questions]

Answer ALL Questions. Each Question Carries TWO Marks (3Qx2M=6 Marks)

- 7. Explain the reasons for missing data. (CO2) [Knowledge]
- 8. Define the following: a) Support. B) Confidence (CO3) [Knowledge]
- 9. Define Association Rules and how are they represented. (CO3) [Knowledge]

Part B [Thought Provoking Questions]

Answer ALL Questions. Each Question Carries SEVEN Marks (2Qx7M=14 Marks)

- 10. Data Transformation of data using Discretization need to be done on the given data. Identify the types of discretization techniques for the data and elaborate on the same. (CO2) [Comprehension]
- 11. Given a set of samples $S = (4, Y), (26, N), (0, Y), (18, Y), (28, N), (12, N), (16, N), (24, N), (16, N)$, Partition S into two intervals S1 & S2 with 21 as the split point. (CO2) [Comprehension]

Part C [Problem Solving Questions]

Answer ALL Questions. Each Question Carries TEN Marks

(1Qx10M=10 Marks)

6. Using Apriori algorithm, formulate and validate the association rules for the transaction data given below considering Minimum Support = 30% and Minimum Confidence = 60%.

(CO3) [Application]

Transactions List

1	Milk	Egg	Bread	Butter
2	Milk	Butter	Egg	Ketchup
3	Bread	Butter	Ketchup	
4	Milk	Bread	Butter	
5	Bread	Butter	Cookies	
6	Milk	Bread	Butter	Cookies
7	Milk	Cookies		
8	Milk	Bread	Butter	
9	Bread	Butter	Egg	Cookies
10	Milk	Butter	Bread	
11	Milk	Bread	Butter	
12	Milk	Bread	Cookies	Ketchup



Roll No																			
---------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

**PRESIDENCY UNIVERSITY
BENGALURU
SCHOOL OF INFORMATION SCIENCE
END TERM EXAMINATION**

Winter Semester: 2021 - 22

Date: 28th June 2022

Course Code: BCA 211

Time: 01:00 PM to 04:00 PM

Course Name: DATA MINING

Max Marks: 100

Instructions:

Program & Sem / Qs & Hs / Sr Read questions carefully and answer accordingly.

Weightage: 50%

Part A [Memory Recall Questions]

Answer all the Questions. Each question carries TWO marks.

(15Qx 2M= 30M)

- 1 a. Which of the following Data transformation normalizes data using division by 10 power k?
A. Decimal scaling B. Min-max C. Z score D. None of these (C.O.1) [Knowledge]
- b. Data selection can be defined as (C.O.1) [Knowledge]
A. The actual discovery phase of a knowledge discovery process
B. The stage of selecting the right data for a KDD process.
C. A subject-oriented integrated time variant non-volatile collection of data in support of management.
D. None of these.
- c. Which of the following is an example of sequence data? (C.O.1) [Knowledge]
A. weather forecast B. genomic data C. data matrix D. market basket data
- d. Which of the following is NOT an example of ordinal attributes? (C.O.1) [Knowledge]
A. Ordered numbers B. Movie ratings C. Military ranks D. Zip codes
- e. Find min-max normalization by setting min=0 and max=1 for the following data:
200,300, 400, 600, 1000 (C.O.1) [Knowledge]
A. Original data 200,300, 400, 600, 1000 Normalized data 0, 0.125, 0.25, 0.5, 1
B. Original data 200,300, 400, 600, 1000 Normalized data 0.125, 0, 0.25, 1, 0.5
C. Original data 200,300, 400, 600, 1000 Normalized data 1, 0.25, 0.5, 0, 0.25
D. Original data 200,300, 400, 600, 1000 Normalized data 0, 0.25, 0.5, 1, 0.25
- f. If the data about seismic activity in japan is given and the magnitude of the next earthquake has to be predicted. Which of the following is suitable? (C.O.1) [Knowledge]
A. Classification B. Unsupervised Learning
C. Supervised Learning D. Dimensionality Reduction
- g. _____ is an essential process where intelligent methods are applied to extract data patterns. (C.O.2) [Knowledge]
A. Data transformation B. Data mining C. Data selection D. Text mining
- h. Height and width are example for which type of data? (C.O.2) [Knowledge]
A. Finite B. Discrete C. Continuous D. None of the above
- i. Let μ (mean) = 54,000, σ (standard deviation) = 16,000, then Z-score normalization value for 45200 is: (C.O.3) [Knowledge]
A. 0.55 B. 0.54 C. -0.55 D. 0.56

- j. Given $X=(2,3)$ and $Y=(3,4)$, Euclidean distance between the two points
 A. 1.414 B. 1.44 C. 1.325 D. 1.432 (C.O.3) [Knowledge]
- k. In Context of Data Mining, describe what is NOISE? (C.O.3) [Knowledge]
- l. If dangerous fires are rare (1%) but smoke is fairly common (10%) due to barbecues, and 90% of dangerous fires make smoke then what is the Probability of dangerous Fire when there is Smoke
 (C.O.4) [Knowledge]
- m. Explain Decision Trees. (C.O.4) [Knowledge]
- n. Write any 4 operations performed in Rapid Miner tool. (C.O.5) [Knowledge]
- o. Explain how Rapid Miner tool is used in Data Mining. (C.O.5) [Knowledge]

Part B [Thought Provoking Questions]

Answer all the Questions. Each question carries TEN marks. (4Qx10M=40M)

2. Describe in detail the challenges on Data Mining (CO1) [Comprehension]
3. Data Transformation of data using Discretization need to be done on some data. Identify the types of discretization techniques for the data and elaborate on the same. (CO2) [Comprehension]
4. Describe k-Means clustering with a scenario along with its pros and cons. (CO4) [Comprehension]
5. a. Use the Min-Max normalization to normalize, by setting min=5, max=10 the following group of data: 35, 55, 70, 75. Calculate is the normalized data for all the given numbers. (5M) (CO2) [Comprehension]
- b. You are planning a picnic today, but the morning is cloudy
- 50% of all rainy days start off cloudy!
 - But cloudy mornings are common (about 40% of days start cloudy)
 - And this is usually a dry month (only 3 of 30 days tend to be rainy, or 10%)
- What is the chance of rain during the day? (5M) (CO4) [Comprehension]

Part C [Problem Solving Questions]

Answer both the Questions. Each question carries FIFTEEN marks. (2Qx15M=30M)

6. Using Apriori algorithm, formulate and validate the association rules for the transaction data given below considering Minimum Support = 30% and Minimum Confidence = 60%. (CO3) [Application]

Transactions List

1	Milk	Egg	Bread	Butter
2	Milk	Butter	Egg	Ketchup
3	Bread	Butter	Ketchup	
4	Milk	Bread	Butter	
5	Bread	Butter	Cookies	
6	Milk	Bread	Butter	Cookies
7	Milk	Cookies		
8	Milk	Bread	Butter	
9	Bread	Butter	Egg	Cookies
10	Milk	Butter	Bread	
11	Milk	Bread	Butter	
12	Milk	Bread	Cookies	Ketchup

7. Given a set of samples:
 $S = (18, Y), (28, N), (0, Y), (24, N), (4, Y), (26, N), (16, N), (12, N), (16, N)$
 Partition S into two intervals S1 & S2 with 14 & 21 as the split points. Find the Information Gain and find the best split. (CO2) [Application]