

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



**PRESIDENCY
UNIVERSITY
BENGALURU**

**School of Information Science
Mid - Term Examinations - November 2024**

Semester: III

Date: 07-11-2024

Course Code: CSA2007

Time: 09.30am to 11.00am

Course Name: Data Mining

Max Marks: 50

Program: BCA

Weightage: 25%

Instructions:

(i) Read all questions carefully and answer accordingly.

(ii) Do not write anything on the question paper other than roll number.

Part A

Answer ALL the Questions. Each question carries 2marks.

5Qx2M=10M

- | | | | | |
|---|---|---------|----|-----|
| 1 | Define Data Mining. | 2 Marks | L1 | CO1 |
| 2 | Explain about Classification with an example. | 2 Marks | L1 | CO1 |
| 3 | List the types of attributes. | 2 Marks | L1 | CO2 |
| 4 | Explain Pure and Impure bin with an example. | 2 Marks | L2 | CO2 |
| 5 | Distinguish between discrete and continuous attributes. | 2 Marks | L2 | CO2 |

Part B

Answer ALL Questions. Each question carries 10 marks.

4QX10M=40M

- | | | | | |
|---|---|---------|----|-----|
| 6 | Explain major challenges of data mining regarding mining methodology and user interactions. | 10Marks | L2 | CO1 |
|---|---|---------|----|-----|

Or

- | | | | | |
|---|--|---------|----|-----|
| 7 | Describe in details the stages of KDD Process with a neat diagram. | 10Marks | L2 | CO1 |
| 8 | Elucidate a brief note on Data mining Functionalities with an example. | 10Marks | L2 | CO1 |

Or

- 9 a. Outline advantages of data mining. 5Marks L2 C01
 b. Discuss about confluence of multiple disciplines in Data Mining. 5Marks L2 C01

- 10 a. Explain Attribute subset selection methods with an example. 5Marks L2 C02
 b. Elucidate about quality measures of data preprocessing. 5Marks L2 C02

Or

- 11 Compare Jaccard and Simple Matching coefficient for the data given below: 10Marks L2 C02

A	1	1	0	0	0	0	0	0	0	0
B	0	1	0	0	0	0	1	0	0	1

- 12 Normalize the following group of data: 200 , 300 , 400 , 600, 1000 using 10Marks L3 C02
- i. Min-Max
 ii. Z-Score
 iii. Decimal Scaling

Or

- 13 Apply entropy based discretization on the given set S= 10Marks L3 C02
 (4,y),(0,y),(16,n),(12,y),(16,n),(18,y),(26,n),(24,n),(28,n). If S has partitioned in to 2 intervals S1 & S2 with 2 possible split points 14 & 21. Find the Best split point.