

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



**BENGALURU**  
**School of Computer Science and Engineering**  
**Mid - Term Examinations - November 2024**

**Semester:** V

**Date:** 07-11-2024

**Course Code:** CSE 2021

**Time:** 02:00pm – 03:30pm

**Course Name:** DATA MINING

**Max Marks:** 50

**Program:** B. Tech (ISE)

**Weight age:** 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**

- |          |   |            |    |     |
|----------|---|------------|----|-----|
| <b>1</b> | Define the Preprocessing with an example.                               | 2<br>Marks | L1 | C01 |
| <b>2</b> | List the reduction techniques.  | 2<br>Marks | L1 | C01 |
| <b>3</b> | Explain about Data cube.  | 2<br>Marks | L1 | C01 |
| <b>4</b> | Derive the relationship between covariance and correlation coefficient. | 2<br>Marks | L1 | C02 |
| <b>5</b> | Define Data Integration.  | 2<br>Marks | L1 | C02 |

---

**Part B**

**Answer ALL Questions. Each question carries 10 marks.**

**4QX10M=40M**

- |           |  |          |    |     |
|-----------|--|----------|----|-----|
| <b>6</b>  | Explain about knowledge discovery in database process with a neat diagram. | 10 Marks | L1 | C01 |
| <b>or</b> |  |          |    |     |
| <b>7</b>  | Explain functionalities of data mining in detail with suitable example.    | 10 Marks | L1 | C01 |

- 8 Explain major issues of Data Mining with respect to society. 10 Marks L2 CO1
- or**
- 9 Explain Technologies used in data mining with detail architecture. 10 Marks L1 CO1
- 10 Explain types of attributes and their properties with suitable example. 10 Marks L2 CO2
- or**
- 11 Explain Data reduction, define Attribute Subset Selection on {A1, A2, A3, A4, A5, A6} given data sets. 10 Marks L2 CO2
- 12 Apply entropy based discretization on the given set  $S = (4,y),(0,y),(16,n),(12,y),(16,n),(18,y),(26,n),(24,n),(28,n)$ . If S has partitioned in to 2 intervals S1 & S2 with 2 possible split points 14 & 21. Find the Best split point. 10 Marks L3 CO2
- or**
- 13 Suppose that the minimum and maximum values for the attribute income are \$12,000 and \$98,000, respectively. We would like to map income to the range [0.0,1.0]. By min-max normalization, a value of \$73,600 for income is transformed find the normalization. 10 Marks L3 CO2
- i. Min-Max  
ii. Z-Score  
iii. Decimal Scaling