

Roll No



**PRESIDENCY UNIVERSITY  
BENGALURU**

**SCHOOL OF ENGINEERING  
MID TERM EXAMINATION - OCT 2023**

**Semester :** Semester VII - 2020

**Course Code :** CSE2021

**Course Name :** Sem VII - CSE2021 - Data Mining

**Program :** B. TECH

**Date :** 30-OCT-2023

**Time :** 11:30AM - 1:00PM

**Max Marks :** 60

**Weightage :** 30%

**Instructions:**

- (i) Read all questions carefully and answer accordingly.
- (ii) Question paper consists of 3 parts.
- (iii) Scientific and non-programmable calculator are permitted.
- (iv) Do not write any information on the question paper other than Roll Number.

**PART A**

**ANSWER ALL THE QUESTIONS**

**(5 X 2 = 10M)**

1. List any two application of data mining  
(CO1) [Knowledge]
2. Mention any two issues of data mining  
(CO1) [Knowledge]
3. Define KDD.  
(CO1) [Knowledge]
4. List the reduction techniques.  
(CO2) [Knowledge]
5. Recite types of data.  
(CO2) [Knowledge]

**PART B**

**ANSWER ALL THE QUESTIONS**

**(2 X 15 = 30M)**

6. a. Explain about knowledge discovery in database process.  
b. Explain major issues in data mining.  
(CO1) [Comprehension]
7. a. Demonstrate the handling of missing data and noisy data.  
b. Apply the normalization techniques for the given data.  
200, 400, 800, 1000.  
(CO2) [Comprehension]

## PART C

### ANSWER THE FOLLOWING QUESTION

(1 X 20 = 20M)

8. a. A survey on cars had conducted in 2015 and determine that 48% of car owners have only one car, 40% have two cars and 12% have 3 or more cars. Supposing that you have decided to conduct your own survey and have collected the data below, determine whether your data supports the results of the study.

Use a significance level of 0.05. Also given that out of 180 car owners, 120 had one car and 38 had two cars.

Note:  $DF(1,0.05)=3.841$ ;  $DF(2,0.05)=5.991$ .

- b. Illustrate about similarity, dissimilarity and their properties.

(CO2) [Application]