Roll No											

PRESIDENCY UNIVERSITY
BENGALURU

SCHOOL OF ENGINEERING
MID TERM EXAMINATION - OCT 2023

Semester : Semester III - 2022
Course Code : CSE2021
Course Name : Sem III - CSE2021 - Data Mining
Program : B.TECH

Date : 2-NOV-2023
Time : 9:30AM - 11:00AM
Max Marks : 50
Weightage : 25%

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. Define Classification and typical methods in clasification. (CO1) [Knowledge]
2. Recite the steps from ML and statistics. (CO1) [Knowledge]
3. Define Clustering. (CO1) [Knowledge]
4. Discuss about Attribute subset selection. (CO2) [Knowledge]
5. Define Binning. (CO2) [Knowledge]

PART B

ANSWER ALL THE QUESTIONS **(2 X 10 = 20M)**

6. Describe the steps involved in data mining when viewed as a process of knowledge discovery. (CO1) [Comprehension]

7. a) Compute the similarity between binary attributes using SMC and JC.

p = 1 0 0 0 0 0 0 0 0 0 0 1

q = 0 0 0 0 0 0 1 0 0 1 1 1

b) Define Normalization and types of Normalization with a suitable example.

(CO2) [Comprehension]

PART C

ANSWER THE FOLLOWING QUESTION **(1 X 20 = 20M)**

8. Explain steps in Entropy Based Discretization. Use Entropy Based Discretization to find the best split for the following data. (0,y), (4,y), (12,y), (16,n), (16,n),(18,y), (24,n),(26,n),(28,n). If 'S' has to be permitted into 2 intervals 'S1' and 'S2' using two split points '14' and '21'. Find the Best Split Point.

(CO2) [Application]