Roll No											

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

PRESIDENCY UNIVERSITY
BENGALURU

SCHOOL OF ENGINEERING
MID TERM EXAMINATION - APR 2023

Semester : Semester IV - 2021
Course Code : CSE2021
Course Name : Sem IV - CSE2021 - Data Mining
Program : B.Tech - All Programs

Date : 17-APR-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 **(10 X 1 = 10M)**

1. The initial steps concerned in the process of knowledge discovery is:
a) Data Cleaning (CO1) [Knowledge]
b) Data Integration
c) Data Reduction
d) Data Transformation

2. Two fundamental goals of Data Mining are _________ and ___________.
a) Analysis and Description (CO1) [Knowledge]
b) Prediction and Description
c) Data cleaning and organizing the data
d) Data cleaning and summarization

3. To remove noise and inconsistent data is called __________
a) Data Integration (CO1) [Knowledge]
b) Data Cleaning
c) Data Transformation
d) Data Reduction

4. ___________ is not a data mining functionality?

a) Clustering and Analysis (CO1) [Knowledge]

b) Selection and interpretation

c) Classification and regression

d) Characterization and Discrimination

5. A sequence of patterns that occur frequently is known as?

a) Frequent Item Set (CO1) [Knowledge]

b) Frequent Subsequence

c) Frequent Sub Structure

d) All of the above

6. The classification or mapping of a class using a predefined class or group is called.

a) Data Sub Structure (CO2) [Knowledge]

b) Data Set

c) Data Discrimination

d) Data Characterisation

7. In Binning, we first sort data and partition into (equal-frequency) bins and then which of the following is not a valid step.

a) smooth by bin boundaries (CO2) [Knowledge]

b) smooth by bin median

c) smooth by bin means

d) smooth by bin values

8. Multiple numbers of data sources get combined in which step of the Knowledge Discovery?

a) Data Transformation (CO2) [Knowledge]

b) Data Selection

c) Data Integration

d) Data Cleaning

9. Which of the following is NOT a common binning strategy?

a) Equiwidth binning (CO2) [Knowledge]

b) Equidepth binning

c) Homogeneity – based binning

d) Equilength binning

10. Let x1 = (1, 2) and x2 = (3, 5) represent two objects . Calculate The Euclidean distance between the two objects.

a) 3.61 (CO2) [Knowledge]

b) 4.2

c) 3.22

d) 3.11

PART B

ANSWER ALL THE QUESTIONS **(4 X 5 = 20M)**

11. Describe in detail about the stages of KDD process with neat diagram?

(CO1) [Comprehension]

12. Discuss in detail about mining methodology as an issue in Data Mining?

(CO1) [Comprehension]

13. What is Noisy data in Data Cleaning? How do you handle noisy data on the given set of datasets [5, 10, 11, 13, 15, 35, 50, 55, 72, 92, 204, 215] . Apply Binning techniques by considering it as 3 Equal width bins.

(CO2) [Comprehension]

14. Suppose that a group of 1500 people was surveyed. The gender of each person was noted. Each person was polled as to whether his or her preferred type of reading material was fiction or nonfiction. Thus, we have two attributes, gender and preferred reading. The data is as follows:-1. male who preferred fiction=250 2. male who preferred non-fiction=50 3. female who preferred fiction=200 4. female who preferred non-fiction=1000. Find the correlation for these nominal attributes for the given value ?

(CO2) [Comprehension]

PART C

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

15. Compare Jaccard Coefficient with Simple Matching Coefficient (SMC). Also Find the SMC and Jaccard Coefficient for the data given below.

M	1	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	1

(CO2) [Application]

16. Perform Entropy Based Discretization on the Given a set of samples “s”, S=(0,Y)(4,Y)(12,Y)(16,N)(16,N)(18,Y)(24,N)(26,N)(28,N). If S has to be partitioned into 2 intervals S1 & S2 using 2 possible split points 14 & 21. Find the best Split among them?

(CO2) [Application]