PRESIDENCY UNIVERSITY
BENGALURU

SET B

SCHOOL OF ENGINEERING
END TERM EXAMINATION - JAN 2024

Semester : Semester III - 2022
Course Code : CSE2027
Course Name : Fundamentals of Data Analytics
Program : B.Tech.

Date : 0J-JAN-2024
Time : 9:30AM - 12:30 PM
Max Marks : 100
Weightage : 50%

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 2M = 10M**

1. What is the key difference between Binary and Multiclass Classification?

(CO5) [Knowledge]

2. Write the steps for making the Forecasting in Data Analysis.

(CO4,CO3) [Knowledge]

3. Explain the difference between Observation method and Interview method of Data Collection.

(CO3,CO2) [Knowledge]

4. List and explain the types of Correlation.

(CO3) [Knowledge]

5. Classify the data into nominal, ordinal, interval and ratio scales.
   a) Size of Shirt
   b) Colour of shirt

(CO1) [Knowledge]

PART B

ANSWER ALL THE QUESTIONS   5 X 10M = 50M

6. A psychology student, Sarah, is giving out sleep diaries to her university friends to monitor the number of hours of sleep they have each night. Sarah believes that university students sleep for 6 hours per night on average. Below is the data that they collected. The number of hours of sleep per night for each student was averaged over a one-month monitoring period. Is there any evidence to suggest that Sarah's belief is incorrect? Take 5% level of significance.

Participants	Hours of Sleep per night
1	7.2
2	8.7
3	5.4
4	6.1
5	5.6
6	6.7
7	5.9
8	6.3
9	7
10	4.2

t-test table

df	0.1	0.05	0.025	0.02	0.01	0.005
1	3.078	6.314	12.706	15.895	31.821	63.657
2	1.886	2.920	4.303	4.849	6.965	9.925
3	1.638	2.353	3.182	3.482	4.541	5.841
4	1.533	2.132	2.776	2.999	3.747	4.604
5	1.476	2.015	2.571	2.757	3.365	4.032
6	1.440	1.943	2.447	2.612	3.143	3.707
7	1.415	1.895	2.365	2.517	2.998	3.499
8	1.397	1.860	2.306	2.449	2.896	3.355
9	1.383	1.833	2.262	2.398	2.821	3.250
10	1.372	1.812	2.228	2.359	2.764	3.169
11	1.363	1.796	2.201	2.328	2.718	3.106
12	1.356	1.782	2.179	2.303	2.681	3.055
13	1.350	1.771	2.160	2.282	2.650	3.012
14	1.345	1.761	2.145	2.264	2.624	2.977
15	1.341	1.753	2.131	2.249	2.602	2.947

(CO2,CO5) [Comprehension]

7. If P is the pull required to lift a load W by means of a pulley block, find a linear law of the form P=mW + c connecting P & W, using the following data:

P	12	15	21	25
W	50	70	100	120

here P & W are taken in kg-wt. Compute P when W=150 kg-wt. Also calculate the Mean Squared Error.

(CO4,CO5) [Comprehension]

8. Explain any Five types of charts and graphs for representing different datasets.

(CO4,CO3) [Comprehension]

9. To construct a meaningful sampling distribution, we are investigating the age distribution within a small population(N) of four individuals with ages 18, 20, 22, and 24 years. By gathering information on individual ages, we aim to create a representative sampling distribution that captures the diversity in age perspectives. Your contribution in sharing your age is essential to ensure the accuracy and richness of our sampling distribution, facilitating a better understanding of age dynamics within our limited community.Using the above information, Construct the sample distribution.

(CO3,CO2) [Comprehension]

10. Explain Secondary data collection and its sources. List the Advantages and Disadvantages of using Secondary data. What caution a researcher must ensure before using the secondary data.

(CO3,CO2) [Comprehension]

PART C

ANSWER ALL THE QUESTIONS   2 X 20M = 40M

11. A sample of 25 people is taken. The length of time to prepare dinner is recorded in minutes, as given below:
44.0, 51.9, 49.7, 40.0, 55.5, 33.0, 43.4, 41.3, 45.2, 40.7, 41.1, 49.1
30.9, 45.2, 55.3, 52.1, 55.1, 38.8, 43.1, 39.2, 58.6, 49.8, 43.2, 47.9, 46.6
Is there any evidence that the population mean time to prepare dinner is less than 48 minutes?
Use a level of significance of 0.05.

(CO2) [Application]

12. a)By the method of least squares, find the regression line that best fits the following data:

x	1	3	4	6	8	9	11	14
y	1	2	4	4	5	7	8	9

Also find the MSE, RMSE and MAE of the above data using the best fit regression line.
b) Calculate Accuracy, Precision, Recall and F1-score for the data acquired from a confusion matrix:-
  True Positive(TP):150
  True Negative(TN): 50
  False Positive (FP): 10
   False Negative (FN): 5

(CO5) [Application]