PRESIDENCY UNIVERSITY
BENGALURU

SET - B

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

SCHOOL OF ENGINEERING
END TERM EXAMINATION - JUN 2023

Semester : Semester IV - B.Tech CSE - 2021
Course Code : CSE2027
Course Name : Sem IV - CSE2027 - Fundamentals of Data Analytics=
Program : CAI,CEI&CST

Date : 10-JUN-2023
Time : 9:30 AM - 12:30PM
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
(10 X 2 = 20M)

1. Find the median of 4,1,3,8,2,6,7,5.

   (CO2) [Knowledge]

2. Define correlation with an example.

   (CO3) [Knowledge]

3. Explain any two types of interview methods.

   (CO3) [Knowledge]

4. What is classification in context of machine learning?

   (CO5) [Knowledge]

5. Define Data Analysis.

   (CO1) [Knowledge]

6. _____ is about sampling items from the population at regular predefined intervals.

   (CO2) [Knowledge]

7. Write the formula for Accuracy and Precision with respect to Confusion matrix.

   (CO5) [Knowledge]

8. What is line graph?

   (CO4) [Knowledge]

9. List the common use cases where data visualization is popularly used.

   (CO4) [Knowledge]

10. There is a case study where a data analysis team is working in investigating why the company has got slow shipment in certain regions.
What type of analysis is this case study related to ?

(CO1) [Knowledge]

PART B

ANSWER ALL THE QUESTIONS   (5 X 10 = 50M)

11. Explain different types of Observation methods for collecting primary data.

(CO3) [Comprehension]

12. What is Support Vector Machine? How is it helpful in differentiating the high dimensional data in two or more classes?

(CO5) [Comprehension]

13. In a class 10 pupils took a Science test and an English test. Their scores are listed in the following table: Pupil A B C D E F G H I J
English Score 2 10 18 4 9 7 18 19 3 10
Science Score 18 12 6 3 11 20 4 17 7 2.
Using this data table draw the scatter plot and describe the correlation between two scores

(CO4) [Comprehension]

14. Table shows a random sample of 200 cyclists and the routes they prefer. Let M=males and H=hilly path.
Out of the males, what is the probability that the cyclist prefers a hilly path?
Are the events "being male" and "preferring the hilly path" independent events?

Table1

Gender	Lake Path	Hilly Path	Wooded path	Total
Female	45	38	27	110
Male	26	52	12	90
Total	71	90	39	200

(CO2) [Comprehension]

15. Explain different types of digital data with neat diagram.Also indicate the growth trend for types of digital data.

(CO1) [Comprehension]

PART C

ANSWER ALL THE QUESTIONS   (2 X 15 = 30M)

16. Mr Hansal has started a company that has developed a machine learning algorithm which is supposed to predict the student's percentage in a given semester for the program of study. It is known that the standard deviation of percentage in the university is 15. The algorithm is tested on 36 students and obtain a mean percentage of 97.65. Using an alpha value of 0.05, is this percentage signicantly different than the population mean of 100?

Z	Area between mean and Z	Area beyond Z
....	....	....
1.94	0.4738	0.0262
1.95	0.4744	0.0256
1.96	0.4750	0.0250
1.97	0.4756	0.0244
1.98	0.4761	0.0239
......	....	....

Z	Area between mean and Z	Area beyond z
....	.....	.....
-0.96	-0.3315	0.8315
-0.95	-0.3289	0.8289
-0.94	-0.3264	0.8264
-0.93	-0.3238	0.8238
-0.92	-0.3212	0.8212
.....	....	....

(CO2) [Application]

17. a. Find the mode, median and mean of the given data
6,4,7,8,5,9,3,7,2,10
b. Transform the given data into the range of [-1, 1] using minmax normalization technique
Data: 13, 9, 26, 14, 8, 11, 21, 25, 23

(CO1) [Application]