PRESIDENCY UNIVERSITY
BENGALURU

SCHOOL OF ENGINEERING
END TERM EXAMINATION - JAN 2023

Semester : Semester III - 2021

Course Code : CSE2027

Course Name : Sem III - CSE2027 - Fundamentals of Data Analytics

Program : B.Tech. CBC/CBD/CCS/CSD/CDV/CIT/COM/CSE/CSG/ISE/IST/ISD

Date : 11-JAN-2023

Time : 1.00PM - 4.00PM

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.

PART A

ANSWER ALL THE TEN QUESTIONS   **10 X 2 = 20M**

1. What is data analysis?

    (CO1) [Knowledge]

2. What is the use of inferential analysis?

    (CO1) [Knowledge]

3. Define variance and standard deviation with a formula.

    (CO2) [Knowledge]

4. Write down the mathematical formulas for variance and standard deviation.

    (CO2) [Knowledge]

5. What survey tool would be most effective for Hisham to use to find out what instructors think about higher secondary education? What method would he employ to identify the data tool's flaw?

    (CO3) [Knowledge]

6. Without the children's knowledge or consent, Dr. Reddy will watch and document the play behaviour at a nursery school. predict the data gathering method used by Dr. Explain your position.

    (CO3) [Knowledge]

7. How can the effectiveness of a logistic regression classifier be evaluated?

    (CO4) [Knowledge]

8. The number of workshops organised by a school in different areas during the last six years is as follows: Draw a histogram representing the data.

    (CO4) [Knowledge]

9. When is classification preferable than regression?

    (CO5) [Knowledge]

10. What varieties of machine learning are there?

    (CO5) [Knowledge]

ANSWER ALL THE FIVE QUESTIONS   **5 X 10 = 50M**

11. Mr. Samanth works for a business that provides analytics. Various details have been provided to him by his management. He was requested to highlight the variations between the data's structuredness and unstructuredness. To assist Samanth, list and total the same.

(CO1) [Comprehension]

12. Identify the sampling technique and represent it in pictorial format." interested people usually take part by themselves by filling in some sort of survey forms. A good example of this is the you tube survey about "Have you seen any of these ads", which has been recently shown a lot. Here, the researcher who is conducting the survey has no right to choose anyone".

(CO2) [Comprehension]

13. With the introduction of the new teams, a ten-team format was created. This format consists of 74 matches and was introduced as retaining the previous format would result in 94 matches, significantly greater than the 60 matches from the previous season, where teams compete in a double round-robin tournament.

The ten teams are divided into two groups of five. In the group stage, each team plays 14 games: facing the other four teams in their group two times each (one home and one away game), four teams in the other group once, and the remaining team two times. The groups for the tournament were announced on 25 February 2022.

Each team plays the team in the same row and the same column twice, and all others once. For instance, Mumbai Indians will play Chennai Super Kings and the other Group A teams twice but the other teams from Group B (Sunrisers Hyderabad, Royal Challengers Bangalore, Punjab Kings and Gujarat Titans) only once. Similarly, Chennai will play Mumbai and the other Group B teams twice but all other teams from Group A only once. The format used is similar to the one used in 2011, with the only difference being that the teams were drawn according to seedings instead of being drawn randomly. The figures in parenthesis denote the number of titles won by the team.

Group A	Group B
Mumbai Indians (5)	Chennai Super Kings (4)
Kolkata Knight Riders (2)	Sunrisers Hyderabad (1)
Rajasthan Royals (1)	Royal Challengers Bangalore
Delhi Capitals	Punjab Kings
Lucknow Super Giants	Gujarat Titans

(CO3) [Comprehension]

Develop the schedule.

14. We investigate the impact of COVID-19 through screening and surveillance methods adopted in India, as well as the potential health system, social, political, and economic consequences. The research was done in a chronological manner, and data was collected between 30 January 2020 till 12 June 2020. Initial containment measures, including point of entry screenings and testing protocols, appeared insufficient. However, testing capacity was gradually expanded after the commencement of a nation-wide lockdown. Modeling predictions have shown varying results on the emergence of cases depending on the infectiousness of asymptomatic individuals, with a peak predicted in mid-July having over two million cases. The country also faces risks of the economic plunge by losing approximately 4% of its gross domestic product, due to containment measures and reduction in goods importation. The low public health expenditure combined with a lack of infrastructure and low fiscal response implies several challenges to scale up the COVID-19 response and management. Therefore, an emergency preparedness and response plan is essential to integrate into the health system of India.

Develop the Questionnaire.

(CO4) [Comprehension]

15. Solve the prediction of y non linear model with least square method for the following observation

x	y
3	4
6	5
9	7
8	6
10	8
11	10
12	12
13	14
13.5	16
14	18

(CO5) [Comprehension]

PART C

ANSWER ALL THE TWO QUESTIONS   2 X 15 = 30M

16. The population of all verbal KCET scores are known to have a standard deviation of 8.5. The UW Psychology department hopes to receive applicants with a verbal KCET scores over 210. This year, the mean verbal KCET scores for the 42 applicants was 212.79. Using a value of = 0.05 is this new mean significantly greater than the desired mean of 210?

Z	Area between mean and Z	Area beyond Z
....	....	....
1.62	0.4474	0.0526
1.63	0.4484	0.0516
1.64	0.4495	0.0505
1.65	0.4505	0.0495
1.66	0.4515	0.0485

Z	Area between Mean and Z	Area beyond Z
....		
2.11	0.4826	0.0174
2.12	0.4830	0.0170
2.13	0.4834	0.0166
2.14	0.4838	0.0162
2.15	0.4842	0.0158

(CO2) [Application]

17. We want to determine if on average, boys score 15 marks more than girls in the exam. We do not have the information related to variance (or standard deviation) for girls' scores or boys' scores. To perform a t-test. we randomly collect the data of 10 girls and boys with their marks. We choose our α value (significance level) to be 0.05 as the criteria for Hypothesis Testing.
Girls: 589, 604, 629, 612, 621,624,607, 610, 598, 594
Boys: 628, 645, 649, 636, 632, 651, 627, 625.619, 609

(CO5,CO2) [Application]

*****