BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(END SEMESTER EXAMINATION)

CLASS: BCA   SEMESTER: V
BRANCH: BCA   SESSION: MO/2024

SUBJECT: CA329 DATA ANALYTICS

TIME: 3 Hours   FULL MARKS: 50

INSTRUCTIONS:
1. The question paper contains 5 questions each of 10 marks and total 50 marks.
2. Attempt all questions.
3. The missing data, if any, may be assumed suitably.
4. **Before attempting the question paper, be sure that you have got the correct question paper.**
5. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.

			CO	BL
Q.1(a)	Write three properties of means and compare AM, GM, HM accordingly. For the two data points 10, 20, find the relationship between AM, GM, HM, and arrange them in ascending order.	[5]	CO1,2	BL1,2
Q.1(b)	Briefly explain the measures of dispersion, skewness, and kurtosis with their utility in business decision making. Calculate skewness of the following list of counts: [15, 50, 80, 30, 10, 5].	[5]	CO2	BL3
Q.2(a)	Why data cleaning is required? Briefly Explain the steps involved in data cleaning.	[5]	CO1	BL2,3
Q.2(b)	Define a measure to calculate relationship between two categorical attributes. Give an example to illustrate.	[5]	CO2	BL3
Q.3	What do you mean by Null hypothesis and Alternative hypothesis? Define two-tailed test and one-tailed test. The makers of XYcell batteries claims that the size AA battery lasts on an average of 45 minutes longer than XYcell's main competitor, the ENCell. Two independent random samples of 100 batteries of each kind are selected. Assuming σ1 = 84minutes and σ2 = 67minutes, find the probability that the difference in the average lives of XYcell and ENCell batteries based on samples does not exceed 54 minutes.	[10]	CO2,4	BL 1, 2,4
Q.4(a)	Draw a random graph containing 5 nodes. Define and calculate the terms - {node degree, distance, diameter, effective diameter}. What is role of effective diameter in social network analysis?	[5]	CO2	BL2
Q.4(b)	Define TF, IDF, TF-IDF. Compute TF-IDF for: Case1: {d0 = 'run and run'; d1 = 'run'; d2 = 'field'}, Case2: {d0 = 'runnow'; d1 = 'norun'; d2 = 'runfast'}	[5]	CO3	BL2,
Q.5(a)	Briefly define – global outlier, contextual outlier, collective outlier. How statistical methods of outlier detection preform in detecting these outlier categories?	[5]	CO4	BL1,4
Q.5(b)	Explain a distance-based method for detection of outlier with illustration.	[5]	CO4	BL3,4

::::::21/11/2024 M:::::