

Code: 20CS4501A

**III B.Tech - I Semester – Regular / Supplementary Examinations
NOVEMBER 2023**

**DATA SCIENCE
(COMPUTER SCIENCE & ENGINEERING)**

Duration: 3 hours

Max. Marks: 70

Note: 1. This paper contains questions from 5 units of Syllabus. Each unit carries 14 marks and have an internal choice of Questions.

2. All parts of Question must be answered in one place.

BL – Blooms Level

CO – Course Outcome

			BL	CO	Max. Marks
UNIT-I					
1	a)	Elaborate the phases of Data Science with example.	L2	CO1	7 M
	b)	What is Exploratory data analysis? Explain how the process of data analysis carried out for data preparation.	L2	CO1	7 M
OR					
2	a)	Explain the business understanding in engineering aspects of Data Science.	L2	CO1	7 M
	b)	Illustrate the process of Data modeling in detail with examples.	L2	CO1	7 M
UNIT-II					
3	a)	Describe data cleaning with example.	L3	CO2	7 M
	b)	State and explain the following: i) Aggregation ii) Normalization	L3	CO2	7 M

OR

4	a)	Explain how to measure the quality of data, during data preprocessing.	L3	CO2	7 M
	b)	Define data reduction and explain the usage of Wavelet Transforms in data reduction.	L3	CO2	7 M

UNIT-III

5	a)	Explain the mean and variance of Binomial distribution and its properties.	L2	CO1	7 M
	b)	Say you have two subsets of a dataset for which you know their means and standard deviations. How do you calculate the blended mean and standard deviation of the total dataset? Can you extend it to K subsets?	L3	CO3	7 M

OR

6	a)	Suppose you were interviewed for a technical role. 50% of the people who sat for the first interview received the call for second interview. 95% of the people who got a call for second interview felt good about their first interview. 75% of people who did not receive a second call, also felt good about their first interview. If you felt good after your first interview, what is the probability that you will receive a second interview call?	L3	CO3	7 M
	b)	Differentiate Systematic Sampling and Cluster sampling with example.	L3	CO3	7 M

UNIT-IV

7	<p>a) A company manufactures an electronic device to be used in a very wide temperature range. The company knows that increased temperature shortens the life time of the device, and a study is therefore performed in which the life time is determined as a function of temperature. The following data is found:</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td style="padding: 2px;">Temperature in Celcius (t)</td> <td style="padding: 2px;">10</td> <td style="padding: 2px;">20</td> <td style="padding: 2px;">30</td> <td style="padding: 2px;">40</td> <td style="padding: 2px;">50</td> <td style="padding: 2px;">60</td> <td style="padding: 2px;">70</td> <td style="padding: 2px;">80</td> <td style="padding: 2px;">90</td> </tr> <tr> <td style="padding: 2px;">Life time in hours (y)</td> <td style="padding: 2px;">420</td> <td style="padding: 2px;">365</td> <td style="padding: 2px;">285</td> <td style="padding: 2px;">220</td> <td style="padding: 2px;">176</td> <td style="padding: 2px;">117</td> <td style="padding: 2px;">69</td> <td style="padding: 2px;">34</td> <td style="padding: 2px;">5</td> </tr> </table> <p>Calculate the 95% confidence interval for the slope in the usual linear regression model, which expresses the life time as a linear function of the temperature.</p>	Temperature in Celcius (t)	10	20	30	40	50	60	70	80	90	Life time in hours (y)	420	365	285	220	176	117	69	34	5	L3	CO4	7 M
Temperature in Celcius (t)	10	20	30	40	50	60	70	80	90															
Life time in hours (y)	420	365	285	220	176	117	69	34	5															
	<p>b) Give a Detailed explanation on Least Square method.</p>	L3	CO4	7 M																				

OR

8	<p>a) Explain Linear Discriminative analysis with an example.</p>	L3	CO4	7 M
	<p>b) Using the logistic regression model with probability cutoff = 0.75, classify the following 6 customers as "Owner" or "Nonowner": if $p \geq 0.75$ then the case as a "Owner". Present the results in a classification matrix.</p>	L3	CO4	7 M

Customer#	Income	Lot_size	Ownership
1	60.0	18.4	Owner
2	64.8	21.6	Owner
3	84.0	17.6	Nonowner
4	59.4	16.0	Nonowner
5	108.0	17.6	Owner
6	75	19.6	Nonowner

UNIT-V

9	a)	How the Bias – variance and model complexity is measured during model assessment.	L3	CO4	7 M
	b)	What is the holdout approach? What is the limitation of this approach? Name four alternative approaches for it.	L3	CO4	7 M

OR

10	a)	Illustrate the process of optimism of the training error rate in model selection and assessment.	L3	CO4	7 M
	b)	Explain concept of cross validation with example.	L3	CO4	7 M