K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

April – May 2024
(B.Tech) Program: Computer Engineering Scheme :II
Examination: LY   Semester: VIII
Course Code: CEDLC8021 and Course Name: Applied Data Science

Date of Exam: 16/5/2024	Duration: 2.5 Hours	Max. Marks: 60

Instructions:
(1) All questions are compulsory.
(2) Draw neat diagrams wherever applicable.
(3) Assume suitable data, if necessary.

		Max. Marks	CO	BT level
Q 1	Solve any six questions out of eight:	12		
i)	Find the arithmetic mean of the following frequency distribution 	x:	1	2	3	4	5	5	7
f:	5	9	12	17	14	10	6		2	CO1	Ap
ii)	Calculate the determinant of following matrix $B = \begin{bmatrix} 1 & 4 & 4 \\ -1 & 3 & -2 \\ -2 & 2 & 3 \end{bmatrix}$	2	CO2	Ap
iii)	Differentiate between estimate of location v/s estimate of variability	2	CO3	U
iv)	List the types of sampling methods.	2	CO4	U
v)	Signify the importance of p-value in hypothesis testing.	2	CO5	U
vi)	State the need of confusion matrix in data science.	2	CO6	U
vii)	Define the central limit theorem.	2	CO4	U
viii)	Compare and contrast between AUC & Lift.	2	CO6	U
Q.2	Solve any four questions out of six.	16		
i)	Eight coins were tossed together and the no of heads resulting was noted. The operation was repeated 256 times and the frequencies (f) that were obtained for different values of x, the number of heads are shown in the following table. Calculate median, quartile, 4th decile & 27th percentile 	x:	0	1	2	3	4	5	6	7	8
f:	1	9	26	59	72	52	29	7	1		4	CO1	Ap
ii)	Calculate the eigenvector and eigen value for following matrix	4	CO2	Ap

Page 1 of 3

	$A = \begin{array}{ccc} -11 & -9 & 14 \\ -6 & -8 & 6 \\ -12 & -22 & 15 \end{array}$			
iii)	What are elements of structural data?	4	CO3	U
iv)	Assume that the life of a packaged magnetic disk exposed to corrosive gases has a Weibull distribution of α=300 hours & β=0.5. Calculate the probability that a)Disk lasts at least 600 hours b)disk failure before 500 hours	4	CO4	Ap
V)	A genetics engineer was attempting to cross a tiger and a cheetah. She predicted a phenotypic outcome of the traits she was observing to be in the following ratio 4 stripes only: 3 spots only: 9 both stripes and spots. When the cross was performed and she counted the individuals she found 50 with stripes only, 41 with spots only and 85 with both. According to the Chi-square test, did she get the predicted outcome?	4	CO5	Ap
Vi)	Write in short about local and global optima	4	CO6	U
Q.3		16		
i)	An analysis of monthly wages paid to the workers of two firms A & B belonging to the same industry gives the following results 		Firm A	Firm B
---	---	---
No of Workers	500	600
Average Monthly wage	Rs 186.00	Rs 175.00

Variance of distribution	81	100 |
		
		1. Which firm A or B has a larger wage bill?
		2. In which firm A or B is there greater variability in individual wages.
		3. calculate the following
		   i) The average monthly wages
		   ii) The variance of the distribution of wages of all the workers in the firm A and B taken together.
			8	CO1	Ap

ii)	Solve the following using binomial distribution
	  a) An irregular six faced die is thrown and the expectation that in 10 throws it will give five even numbers is twice the expectation that it will give four even numbers. How many times in 10,000 sets of 10 throws each. would you expect it to give no even number. (4M)
	  b) In a precision bombing attack there is a 50% chance that anyone bomb will strike the target. Two direct hits are required to destroy the target completely. How many bombs must be dropped to give a 99% chance or better of completely, destroying the target? (4M)
		8	CO4	Ap

Page 2 of 3

iii)	a) Differentiate between chi-square and t-test. (4M) b) Three different kinds of food are tested on three groups of rats for 5 weeks. The objective is to check the difference in mean weight (in grams) of the rats per week. Apply one-way ANOVA using a 0.05 significance level to the following data: (4M)	8	CO5	Ap

Food I	Food II	Food III
8	4	11
12	5	8
19	4	7
8	6	13
6	9	7
11	7	9

Q.4	Solve any two questions out of three.	16		
i)	Calculate the rank and nullity of the following matrix. Describe the importance of rank and nullity in statistics. $$A = \begin{bmatrix} 1 & 4 & 5 & 6 & 9 \\ 3 & -2 & 1 & 4 & -1 \\ -1 & 0 & -1 & -2 & -1 \\ 2 & 3 & 5 & 7 & 8 \end{bmatrix}$$	8	CO2	Ap
ii)	Write a short note on a) Exploring the data distribution b) Exploring the binary and categorical data	8	CO3	U
iii)	Write a short note on a) Constrained optimization b) Unconstrained optimization c) Least square optimization	8	CO6	U