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

July / Aug 2024
(B. Tech) Program: [illegible], Scheme: II
Regular Examination: TY Semester: VI
Course Code: HDSC601 and Course Name: Statistical Learning in Data Science
Date of Exam: 07/08/24 Duration: 02.5 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)	What is the difference between supervised and unsupervised learning?	2	CO1	R
ii)	What is cross-validation, and why is it important in machine learning?	2	CO4	R
iii)	What is the purpose of probability distributions in statistical learning?	2	CO2	R
iv)	List the different types of sampling techniques.	2	CO3	R
v)	Define population mean and sample mean.	2	CO3	R
vi)	What is the difference between precision and recall	2	CO4	R
vii)	What is clustering?	2	CO3	R
viii)	Describe the components of a time series.	2	CO5	R
Q.2	**Solve any four questions out of six.**	**16**		
i)	What is the purpose of probability distributions in statistical learning?	4	CO2	R
ii)	Describe common probability distributions such as the normal, binomial, and Poisson distributions.	4	CO2	R
iii)	What is sampling, and why is it important in statistics?	4	CO3	R
iv)	Explain the difference between null and alternative hypotheses.	4	CO4	R

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

July / Aug 2024
(B. Tech) Program: [illegible] Scheme: II
Regular Examination: TY Semester: VI
Course Code: HDSC601 and Course Name: Statistical Learning in Data Science
Date of Exam: 07/08/24 Duration: 02.5 Hours Max. Marks: 60

v)	What is the difference between linear and nonlinear regression models?	4	CO5	R
vi)	What are autoregressive (AR), moving average (MA) models?	4	CO5	R
Q.3	**Solve any two questions out of three.**	**16**		
i)	Explain the bias-variance trade-off in the context of model complexity.	8	CO2	R
ii)	What is the purpose of the coefficient of determination (R-squared) in regression analysis?	8	CO4	R
iii)	An irregular 6 faced dice is such that the probability that it gives 3 even numbers in 5 throws is twice the probability that it gives 2 even numbers in 5 throws. How many sets of exactly 5 trials can be expected to give no even number out of 2500 sets ? Let the probability of getting an even number with the unfair dice be p. Let X denote the number of even numbers obtained in 5 trials.	8	CO2	Ap
Q.4	**Solve any two questions out of three.**	**16**		
i)	Define conditional probability and provide an example.	8	CO2	R
ii)	Explain how stratified random sampling differs from quota sampling.	8	CO3	U
iii)	The fatality rate of thyroid patients is believed to be 17.26 %. In a certain year 640 patients suffering from thyroid were treated in a metropolitan hospital and only 63 patients died. Can you consider the hospital efficient ? H0 : p = P i.e. hospital is not efficient. H1: p < P. Test it by one and two tailed test.	8	CO6	Ap

****