

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

May-June 2024		
(B. Tech) Program: Computer Engineering Scheme: II		
Regular Examination: TY Semester: VI		
Course Code: HDSC601 and Course Name: Statistical Learning in Data Science		
Date of Exam: 24/5/24	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
<b>Q 1</b>	<b>Solve any six questions out of eight:</b>	<b>12</b>		
i)	What is the difference between supervised and unsupervised learning?	2	CO1	U
ii)	Describe the difference between parametric and non-parametric models.	2	CO2	U
iii)	Discuss the differences between classification and regression problems.	2	CO4	U
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 bias-variance-trade-off?	2	CO1	R
vii)	What is the ROC curve used for?	2	CO4	R
viii)	Describe the components of a time series.	2	CO5	R
<b>Q.2</b>	<b>Solve any four questions out of six.</b>	<b>16</b>		
i)	What is the purpose of probability distributions in statistical learning?	4	CO2	R
ii)	Describe discrete probability distributions in detail.	4	CO2	R
iii)	What is sampling, describing different sampling techniques?	4	CO3	R
iv)	Explain the difference between null and alternative hypotheses.	4	CO4	U

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

May-June 2024		
(B. Tech) Program: Computer Engineering Scheme: II		
Regular Examination: TY Semester: VI		
Course Code: HDSC601 and Course Name: Statistical Learning in Data Science		
Date of Exam: 24/5/24	Duration: 02.5 Hours	Max. Marks: 60

v)	How does dimensionality reduction help with high-dimensional data?	4	CO5	U
vi)	What are autoregressive (AR), moving average (MA) models?	4	CO5	R
<b>Q.3</b>	<b>Solve any two questions out of three.</b>	<b>16</b>		
i)	Explain the bias-variance trade-off in the context of model complexity.	8	CO1	R
ii)	Explain the concept of Bayes' theorem and its significance.	8	CO2	R
iii)	Out of 800 families with 4 children each, how many families would be expected to have (i) 2 boys and 2 girls (ii) at least 1 boy, (iii) at most 2 girls and (iv) children of both sexes. Assume equal probabilities for boys and girls. Considering each child as a trial, $n = 4$ . Assuming that the birth of a boy is a success, $p = \frac{1}{2}$ and $q = \frac{1}{2}$ . Let $X$ denote the number of successes (boys).	8	CO2	Ap
<b>Q.4</b>	<b>Solve any two questions out of three.</b>	<b>16</b>		
i)	Define conditional probability and provide an example.	8	CO2	R
ii)	What is sampling bias, and how can it affect the results of a study? Explain how stratified random sampling differs from quota sampling.	8	CO3	R
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? $H_0 : p = P$ i.e. hospital is not efficient. $H_1 : p < P$ . Test it by one and two tailed test.	8	CO4	Ap

\*\*\*\*\*