

Nov – Dec 2024		
B.Tech. (All Branches - Honors) Scheme II		
Regular Examination: TY Semester: V		
Course Code: HDSC501 and Course Name: Mathematics for Data Science		
Date of Exam: 5/12/2024	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.

Q. No.	Question	Max. Marks	CO	BT level																																
Q 1	Solve any two questions out of three: (05 marks each)	10																																		
a)	Find the dimension and basis for the four fundamental subspaces for $A = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{bmatrix}$		CO1	U																																
b)	A survey samples 100 students to determine their average study time per week. The sample mean is 15 hours, and the population standard deviation is 4 hours. Construct a 95% confidence interval for the population mean. ($z=1.96$).		CO2	AP																																
c)	Compare t-distribution and normal distribution in terms of application.		CO2	U																																
Q 2	Solve any two questions out of three: (05 marks each)	10																																		
a)	The heights and weights of 10 individuals are given. Create a scatter plot, calculate the correlation coefficient, and interpret the relationship. <table><tr><th>Individual</th><th>Height (in cm)</th><th>Weight (in kg)</th></tr><tr><td>1</td><td>150</td><td>50</td></tr><tr><td>2</td><td>160</td><td>55</td></tr><tr><td>3</td><td>155</td><td>53</td></tr><tr><td>4</td><td>170</td><td>65</td></tr><tr><td>5</td><td>165</td><td>60</td></tr><tr><td>6</td><td>175</td><td>68</td></tr><tr><td>7</td><td>180</td><td>72</td></tr><tr><td>8</td><td>158</td><td>54</td></tr><tr><td>9</td><td>162</td><td>58</td></tr><tr><td>10</td><td>168</td><td>62</td></tr></table>		Individual	Height (in cm)	Weight (in kg)	1	150	50	2	160	55	3	155	53	4	170	65	5	165	60	6	175	68	7	180	72	8	158	54	9	162	58	10	168	62	CO3
Individual	Height (in cm)	Weight (in kg)																																		
1	150	50																																		
2	160	55																																		
3	155	53																																		
4	170	65																																		
5	165	60																																		
6	175	68																																		
7	180	72																																		
8	158	54																																		
9	162	58																																		
10	168	62																																		

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

Nov – Dec 2024	
B.Tech. (All Branches - Honors) Scheme II	
Regular Examination: TY Semester: V	
Course Code: HDSC501 and Course Name: Mathematics for Data Science	
Date of Exam: 5/12/2024	Duration: 02.5 Hours Max. Marks: 60

b)	Compare the effectiveness of exponential graphs to linear graphs to model population growth. Also discuss which is more appropriate and why?		CO3	AN
c)	Discuss the limitations of linear dimensionality reduction techniques in handling complex datasets with non-linear relationships. Also, discuss how non-linear techniques address these limitations?		CO6	AP
Q.3	Solve any two questions out of three. (10 marks each)	20		
a)	Veena spends her vacations in random order visiting 4 cities (A,B, C and D). What is the probability that she will go to i) A before B ii) A before B and B before C.		CO2	U
b)	Consider a dataset of 50 students with their test scores to construct a frequency distribution table and graph. Analyze the central tendency and variability of the scores. Also, compare the graph to a box plot for current scenario. (Consider the marks out of 100) 45, 67, 72, 88, 53, 47, 90, 56, 61, 73, 82, 95, 54, 68, 75, 49, 85, 92, 58, 62, 80, 87, 66, 70, 78, 64, 77, 91, 55, 59, 84, 76, 65, 48, 74, 93, 50, 52, 81, 89, 57, 46, 79, 63, 71, 83, 60, 69, 51, 94		CO3	AP
c)	Define feature engineering and its significance. Discuss various techniques to handle dataset with mixed types of variables.		CO4	U
Q.4	Solve any two questions out of three. (10 marks each)	20		
a)	Discuss EDA and its importance in a machine-learning environment. Explain how you would decide the appropriate machine-learning model for a given dataset.		CO4	AP
b)	Find the root of $f(x) = x^2 - 2$ using the Newton-Raphson Method. ($\epsilon = 0.00001$)		CO5	AN
c)	Solve the equation $f(x) = x^3 - x - 2 = 0$ using the Bisection Method (use function method instead of derivative method).		CO5	AP
