

Enrollment Number: _____

Roll Number: _____

JAMIA HAMDARD
Department of Computer Science & Engineering
School of Engineering Sciences & Technology
New Delhi – 110062

Examination: Odd Semester Examination 2024

Sub: Data Science

Sub. Code: BTCSE – DED51 / BTCSE – DED51 (Backlog)

Max. Time: 3 hours

Max. Marks: 75

Note: Follow proper numbering for answering the questions.

SECTION – A

*Attempt all questions.**7 x 5 = 35 Marks*

SL. NO	Questions	Marks	CO	BL																																				
1.	Give your views on the statement “Data Science is an evolutionary extension of statistics capable of dealing with the massive amounts of data produced today”.	5	CO1	L5																																				
2.	Discuss the relationship between Big Data and Data science and thus write their applications.	5	CO1	L2																																				
3.	What is Big Data? Discuss the characteristics of big data.	5	CO2	L1																																				
4.	Explain the various techniques used to handle large volumes of data.	5	CO2	L2																																				
5.	Given below is a distance matrix for 5 Data Points. <table><tr><td></td><td>A</td><td>B</td><td>C</td><td>D</td><td>E</td></tr><tr><td>A</td><td>0</td><td></td><td></td><td></td><td></td></tr><tr><td>B</td><td>4</td><td>0</td><td></td><td></td><td></td></tr><tr><td>C</td><td>3</td><td>8</td><td>0</td><td></td><td></td></tr><tr><td>D</td><td>6</td><td>3</td><td>9</td><td>0</td><td></td></tr><tr><td>E</td><td>7</td><td>10</td><td>2</td><td>8</td><td>0</td></tr></table> Perform Hierarchical Clustering on these data points.		A	B	C	D	E	A	0					B	4	0				C	3	8	0			D	6	3	9	0		E	7	10	2	8	0	5	CO3	L3
	A	B	C	D	E																																			
A	0																																							
B	4	0																																						
C	3	8	0																																					
D	6	3	9	0																																				
E	7	10	2	8	0																																			
6.	What is Deep Learning? Discuss the significance of Deep Learning in Data Science.	5	CO4	L4																																				
7.	Describe 5 Cs of data science ethics.	5	CO5	L2																																				

SECTION B

*Attempt any four out of six questions.**4 x 10 = 40 Marks*

SL. NO	Questions	Marks	CO	BL
1.	Suppose you are a data scientist working on a project to detect fraud transactions over credit cards. Illustrate the stepwise process for credit card fraud detection.	10	CO1	L6

2.	Differentiate the following: i. Linear Regression vs Logistics Regression ii. Supervised Learning vs Reinforcement Learning iii. Classification vs Clustering.	10	CO3	L4																					
3.	The following data shows the sugar content of a fruit (SUGAR) for different number of days after picking (DAYS) <table border="1"><tr><td>DAYS</td><td>0</td><td>1</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td></tr><tr><td>SUGAR</td><td>7.9</td><td>12.0</td><td>9.5</td><td>11.3</td><td>11.8</td><td>11.3</td><td>4.2</td><td>0.4</td></tr></table> i. Find the regression equation between DAYS and SUGAR. ii. Use the model to find the Sugar content after 2 days of picking. iii. Also, check the fitness of the model.	DAYS	0	1	3	4	5	6	7	8	SUGAR	7.9	12.0	9.5	11.3	11.8	11.3	4.2	0.4	10	CO3	L3			
DAYS	0	1	3	4	5	6	7	8																	
SUGAR	7.9	12.0	9.5	11.3	11.8	11.3	4.2	0.4																	
4.	Use K-means clustering, and distance given by formula D to cluster the following 10 points into 2 clusters. Show the calculation performed at each iteration. Assume that the initial clusters are A and B and perform K-means clustering for 2 iterations. $D(X, Y) = x_2 - x_1 + y_2 - y_1 $ Where X is (x1, y1) and Y is (x2, y2). <table border="1"><tr><td>Point</td><td>X</td><td>Y</td></tr><tr><td>A</td><td>185</td><td>72</td></tr><tr><td>B</td><td>170</td><td>56</td></tr><tr><td>C</td><td>168</td><td>60</td></tr><tr><td>D</td><td>179</td><td>68</td></tr><tr><td>E</td><td>182</td><td>72</td></tr><tr><td>F</td><td>188</td><td>77</td></tr></table>	Point	X	Y	A	185	72	B	170	56	C	168	60	D	179	68	E	182	72	F	188	77	10	CO3	L3
Point	X	Y																							
A	185	72																							
B	170	56																							
C	168	60																							
D	179	68																							
E	182	72																							
F	188	77																							
5.	What are the reasons for moving towards deep learning? Write some applications of Deep Learning.	10	CO4	L2																					
6.	Write a short note on i. Data Visualization tools ii. Data visualization options	10	CO5	L2																					

BL – Bloom's Taxonomy Levels (1 – Remembering, 2 – Understanding, 3 – Applying, 4 – Analyzing, 5 – Evaluating, 6 – Creating)
