

JAMIA HAMDARD
School of Engineering Sciences and Technology
Department of Computer Science & Engineering

Even Semester Examination - 2023

Subject: Data Analytics with Python

Subject Code: BTCSE MOOCS-2 / BTCSE MOOCS-2 Lateral Entry

TIME: 3 Hours

MAX. MARKS: 75

Note: 1) Immediately write your Roll No. and Enrolment No. after receiving the Question Paper.

2) Do not write anything on the Question Paper.

3) Scientific Calculator is allowed without Memory Feature.

4) Please follow proper numbering while answering the questions.

SECTION – A (35 Marks)

❖ **Attempt ALL questions. Each question is of 5 marks.**

Q. No.	Questions	CO*	BL*
1.	Write the differences between data science and machine learning in terms of definition, role and scope.	CO1	L1, L2
2.	Explain the concept of general MLE method for estimating the parameters of a probability distribution.	CO4	L3, L4
3.	Differentiate between list, tuple, sets and dictionary in python programming with examples for each.	CO2	L4, L6
4.	Differentiate between supervised learning and unsupervised learning with examples.	CO1	L1, L2
5.	Compare K-means clustering with Hierarchical Clustering Techniques.	CO2	L2, L3
6.	What is precision and recall? How they are used in ROC curve?	CO2	L2, L3
7.	The National Weather Service keeps records of rainfall in valleys. Records indicate that in a certain valley, the annual rainfall has a mean of 70 inches and a standard deviation of 12 inches. Suppose the rainfalls are sampled during randomly picked years, and $\bar{x}$ is the mean amount of rain in these years. For samples of size 36, determine the mean and standard deviation of $\bar{x}$.	CO3	L4, L5, L6

SECTION – B (40 Marks)

❖ **Attempt any FOUR questions. Each question is of 10 marks.**

Q. No.	Questions	CO*	BL*
1.	A patient takes a lab test and the result comes back negative. It is known that the test returns a correct positive result in only 96% of the cases and a correct negative result in only 95% of the cases. Furthermore, only 0.008 % of the entire population has this disease. 1. What is the probability that this patient has cancer? 2. What is the probability that he does not have cancer? 3. What is the diagnosis?	CO5	L5, L6

2.	Using the following data set, perform an oneway analysis of variance using $\alpha=0.05$. Write up the results in APA format. <table><tr><th>GROUP 1</th><th>GROUP 2</th><th>GROUP 3</th></tr><tr><td>51</td><td>23</td><td>56</td></tr><tr><td>45</td><td>43</td><td>76</td></tr><tr><td>33</td><td>23</td><td>74</td></tr><tr><td>45</td><td>43</td><td>87</td></tr><tr><td>67</td><td>45</td><td>56</td></tr></table>	GROUP 1	GROUP 2	GROUP 3	51	23	56	45	43	76	33	23	74	45	43	87	67	45	56	CO4	L3, L4, L5						
GROUP 1	GROUP 2	GROUP 3																									
51	23	56																									
45	43	76																									
33	23	74																									
45	43	87																									
67	45	56																									
3.	Create the relationship model for the data given in table below and also predict Y for $X_1=4.0$ and $X_2=5.0$ respectively. <table><tr><th>Observations</th><th>X1</th><th>X2</th><th>Y</th></tr><tr><td>1</td><td>5.1</td><td>4.9</td><td>10.4</td></tr><tr><td>2</td><td>6.2</td><td>3.8</td><td>10.2</td></tr><tr><td>3</td><td>7.3</td><td>2.7</td><td>10.1</td></tr><tr><td>4</td><td>8.4</td><td>1.6</td><td>10.3</td></tr><tr><td>5</td><td>9.5</td><td>0.5</td><td>10.5</td></tr></table>	Observations	X1	X2	Y	1	5.1	4.9	10.4	2	6.2	3.8	10.2	3	7.3	2.7	10.1	4	8.4	1.6	10.3	5	9.5	0.5	10.5	CO5	L1, L5, L6
Observations	X1	X2	Y																								
1	5.1	4.9	10.4																								
2	6.2	3.8	10.2																								
3	7.3	2.7	10.1																								
4	8.4	1.6	10.3																								
5	9.5	0.5	10.5																								
4.	What do you mean by the Logistic Regression? What are the assumptions made in Logistic Regression? Why is Logistic Regression termed as Regression and not classification?	CO3	L1, L3, L4																								
5.	What is CART? Explain in detail with advantages and limitations of the CART model and also write the pseudo code of the CART algorithm.	CO4	L3, L5																								
6.	Explain the following terms: (a) Hypothesis Testing (b) Type I and Type II error (c) P-value (d) Critical Region (e) Difference Between a T-Test and Z-Test	CO3	L2, L3, L4																								

CO* - Course Outcomes from Syllabus, BL* - Bloom's Taxonomy Levels (1: Remembering, 2: Understanding, 3: Applying, 4: Analyzing, 5: Evaluating, 6: Creating)

END