

ROLLNO. _____

ENROLLMENTNO. _____

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING, SCHOOL OF ENGINEERING
SCIENCES & TECHNOLOGY, JAMIA HAMDARD**

SEMESTER EXAMINATION 2023

BTCSEAI DE23

Departmental Elective – II Data Science

TIME:3 HRS MAXIMUM MARKS:75

Note: Do not write anything on paper except Roll No. & Enrollment No. Use of mobile phone is not allowed. Please follow proper numbering while answering the questions Proper steps are required to be specified for all questions

SECTION A

*All questions of Section A are compulsory (7*5)*

S.NO	QUESTIONS	MARKS	CO	BL												
Q1.	The dataset of a pass/fail in an exam for 5 students is given in the table below, using logistic regression optimizer $\log(\text{odds}) = -64 + 2 \times \text{hours}$, calculate probability of pass of a student who studied 33 hours. <table><tr><th>Hours Studied</th><th>Result(1=Pass,0=Fail)</th></tr><tr><td>29</td><td>0</td></tr><tr><td>15</td><td>0</td></tr><tr><td>33</td><td>1</td></tr><tr><td>28</td><td>1</td></tr><tr><td>39</td><td>1</td></tr></table>	Hours Studied	Result(1=Pass,0=Fail)	29	0	15	0	33	1	28	1	39	1	5	CO2	L3
Hours Studied	Result(1=Pass,0=Fail)															
29	0															
15	0															
33	1															
28	1															
39	1															
Q2.	The eigen value and eigen vector of the matrix given below are: $\begin{bmatrix} 1 & 2 & 3 \\ 0 & -4 & 2 \\ 0 & 0 & 7 \end{bmatrix}$ Also justify its importance in data science and analytics.	5	CO3	L2												
Q3.	You're a Data Scientist / Business Analyst working for a new eCommerce company called A&B Co. (similar to Amazon) and you've been asked to prepare a writeup for the Vice President of Sales and the Vice President of Operations that summarizes sales and operations thus far. The summary should include (at a minimum) a summary of current state the business, current customer satisfaction, and a proposal of 2-3 areas where the company can improve. Here are some facts: It's currently September 2018 (e.g., you can ignore all data after September 2018), The company's inception was January 2017 (so you can ignore all data before January 2017), Company is US-based, but launched in Brazil (which is why some information is in Portuguese), You can assume all orders are delivered (so ignore the order state field)	5	CO2	L1												
Q4.	Utilizing data science and analytics case study problem solving approach answer the following:How would you improve bank's existing state-of-the-art credit scoring of borrowers? How will you predict someone can face financial distress in the next couple of years?	5	CO1	L4												
Q5.	A box contains 10 blue balls, 15 green balls and x yellow balls. If the probability of selecting a yellow ball is $\frac{2}{3}$, what is the total number of balls in the box?	5	CO4	L5												
Q6.	Given below the premises, specify whether the Conclusion	5	CO5	L6												

"Taking appropriate medicine, ailment D lasts longer than ailment A" is True or False?

Ailment C lasts twice as long as Ailment D
 Taking Medicine A halves the duration of Ailment B
 Medicine C reduces the duration of Ailment D to five days
 Ailment A and B last for eight days
 Taking Medicine B halves the duration of Ailment A and C
 Ailment B lasts half the duration as Ailment D

Q7.

How would you measure the success of Facebook groups?

5

CO4

L3

SECTION B

Answer any 4 questions in Section B (4*10)

S.NO	QUESTIONS	MARKS	CO	BL																																																							
Q1.	Cluster the following eight points (with (x, y) representing locations) into three clusters: A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9) Initial cluster centers are: A1(2, 10), A4(5, 8) and A7(1, 2). The distance function between two points $a = (x_1, y_1)$ and $b = (x_2, y_2)$ is defined as $P(a, b) = x_2 - x_1 + y_2 - y_1 $. Use K-Means Algorithm to find the three cluster centers after the second iteration.	10	CO3																																																								
Q2.	A company manufactures car batteries with an average life span of 2 or more years. An engineer believes this value to be less. Using 10 samples he measures the average life span to be 1.8 years with a standard deviation of 0.15 (a) State the null and alternative hypothesis. (b) At a 99% confidence level, is there enough evidence to discard the null hypothesis.	10	CO3	L3																																																							
Q3.	Given all the previous patients I've seen (below are their symptoms and diagnosis)... <table border="1"> <thead> <tr> <th>chills</th><th>runny nose</th><th>headache</th><th>fever</th><th>flu?</th></tr> </thead> <tbody> <tr><td>Y</td><td>N</td><td>Mild</td><td>Y</td><td>N</td></tr> <tr><td>Y</td><td>Y</td><td>No</td><td>N</td><td>Y</td></tr> <tr><td>Y</td><td>N</td><td>Strong</td><td>Y</td><td>Y</td></tr> <tr><td>N</td><td>Y</td><td>Mild</td><td>Y</td><td>Y</td></tr> <tr><td>N</td><td>N</td><td>No</td><td>N</td><td>N</td></tr> <tr><td>N</td><td>Y</td><td>Strong</td><td>Y</td><td>Y</td></tr> <tr><td>N</td><td>Y</td><td>Strong</td><td>N</td><td>N</td></tr> <tr><td>Y</td><td>Y</td><td>Mild</td><td>Y</td><td>Y</td></tr> </tbody> </table> Do I believe that a patient with the following symptoms has the flu? <table border="1"> <thead> <tr> <th>chills</th><th>runny nose</th><th>headache</th><th>fever</th><th>flu?</th></tr> </thead> <tbody> <tr> <td>Y</td><td>N</td><td>Mild</td><td>Y</td><td>?</td></tr> </tbody> </table>	chills	runny nose	headache	fever	flu?	Y	N	Mild	Y	N	Y	Y	No	N	Y	Y	N	Strong	Y	Y	N	Y	Mild	Y	Y	N	N	No	N	N	N	Y	Strong	Y	Y	N	Y	Strong	N	N	Y	Y	Mild	Y	Y	chills	runny nose	headache	fever	flu?	Y	N	Mild	Y	?	10	CO1	L1
chills	runny nose	headache	fever	flu?																																																							
Y	N	Mild	Y	N																																																							
Y	Y	No	N	Y																																																							
Y	N	Strong	Y	Y																																																							
N	Y	Mild	Y	Y																																																							
N	N	No	N	N																																																							
N	Y	Strong	Y	Y																																																							
N	Y	Strong	N	N																																																							
Y	Y	Mild	Y	Y																																																							
chills	runny nose	headache	fever	flu?																																																							
Y	N	Mild	Y	?																																																							
Q4.	Suppose a fair six-sided die is rolled once. If the value on the die is 1, 2, or 3, the die is rolled a second time. What is the probability that the sum total of values that turn up is at least 6?	10	CO4	L4																																																							
Q5.	Through an example specify in detail conceptual framework of Data Analysis problems.	10	CO5	L5																																																							
Q6.	Through query analysis disprove the hypothesis: Data scientists who switch jobs more often end up getting promoted faster.	10	CO2	L6																																																							

BLOOMS TAXANOMY LEVELS (BL)

1-REMEMBERING, 2-UNDERSTANDING, 3-APPLYING, 4-ANALYSING,
 5-EVALUATING, 6-CREATING