M.Sc.(DS).II/04.25.001 Reg.No.

M.Sc. COMPUTER SCIENCE DEGREE WITH SPECIALIZATION IN DATA SCIENCE SECOND SEMESTER EXAMINATION, APRIL 2025

22-359-0201 NETWORKS AND DATA COMMUNICATIONS

(Regular)

Write any FIVE questions.
(Each Question Carries 10 Mark)

Time : 3 Hours **Maximum Marks :50**

No		QUESTIONS	MARKS	CO	BL	PL
1	a	Encode the sequence 1111 1001 1100 0101 using NRZ-I and Manchester methods. Draw the encodings.	5	CO1	L3	1.7.1
	b	Explain the Medium Access procedure in CSMA/CA.	5	CO1	L2	1.7.1
		OR				
2	a	List and explain any five factors that influence data rates in network cables.	5	CO1	L2	1.7.1
	b	What is slotted ALOHA? How does it differ from pure ALOHA?	5	CO1	L2	1.7.1
3	a	Show how bit errors are detected in Cyclic Redundancy Check scheme if MSB of the following data sequence gets modified as 1 in transit. Use 1011 as the divisor. 0011 1001 1011	5	CO2	L3	1.7.1
	b	Draw the frame format of 802.3. Write down the role of each of the fields in the frame.	5	CO2	L3	1.7.1
		OR				
4	a	Illustrate checksum calculation using the following 16-bit sequences. Also, explain the error checking mechanism using checksums. 1110 0110 0110 0001 0101 0101 0101 0100 1000 1101 0000 1100	5	CO2	L3	1.7.1

	b	Can we recover the original data from the erroneous data in the two dimensional parity illustration shown below. Give reasons. 1 0 0 1 | 1 0 1 0 1 | 0 1 1 0 0 | 0 1 1 0 1 | 1 0 1 1 1 |	5	CO2	L3	1.7.1
5	a	Create five subnets of equal size using 226.13.16.200/26. Give the subnet id and host id ranges of each of the subnets.	5	CO3	L3	1.7.1
	b	Discuss the role of SNMP in communication networks.	5	CO3	L2	1.7.1
		OR				
6	a	What is Network Address Translation (NAT) ? Explain the mechanism of network address translation. What is the advantage with NAT? Why is NAT criticized?	5	CO3	L2	1.7.1
	b	Explain network management with ICMP.	5	CO3	L2	1.7.1
7	a	Explain how sequence numbers, acknowledgements and timeouts help in reliable data transfer.	5	CO4	L2	1.7.1
	b	List and explain the flags in TCP segments.	5	CO4	L3	1.7.1
		OR				
8	a	Explain the congestion control mechanism of TCP.	5	CO4	L3	1.7.1
	b	Compare the TCP dump given below with the standard TCP segment format and find (in decimal) the values of destination port, receive window size and header length. 01 bb 97 3c 57 b2 e6 0f 85 5f a3 c4 80 10 11 ed cc d0 00 00 01 01 08 0a 27 49 9e a1 ad e0 34 43	5	CO4	L3	1.7.1

9	a	Explain the server hierarchy of DNS. What is the role of each of these servers?	5	CO5	L3	1.7.1
	b	Explain how load balancing is facilitated by DNS.	5	CO5	L3	1.7.1
		OR				
10		Build a client-server chat application in Python. Make suitable assumptions regarding the communication.	10	CO5	L3	1.7.1

****

M.Sc.(DS).II/04.25.002 Reg.No. | | | | | | | | |

B

M.Sc. COMPUTER SCIENCE WITH SPECIALIZATION IN DATA SCIENCE SECOND SEMESTER EXAMINATION APRIL 2025

22-359-0203 R FOR DATA ANALYTICS

(Regular)

Write any FIVE questions.
(Each Question Carries 10 Mark)

Time-3 Hours **Maximum Marks :50**

Q.No		QUESTIONS	MARKS	CO	BL	PL
1.	a.	With an example explain which loop statement in R can be used to iterate through a set of statements a large number of times without checking for any particular conditions? How can we exit from such a looping structure?	3	CO1	L2	2.5.3
	b.	Explain the concept of factors data structures in R?	2	CO1	L3	1.7.1
	c.	Write an R program to create a dataframe as shown below and perform the following operations.	5			

Student_ID	Name	Age	Grade_Level
1.	Elsa	17	11
2.	Hima	16	10
3.	Minu	18	12
4.	Sonam	17	11

i. Connect to MySQL database.
ii. Save the dataframe as a table in MySql
iii. From the table select only those rows in which the Grade_level=12 and display the result in R studio
iv. Insert a new row into the table from R studio
v. Update the row which contains Age=18 to 20

OR

2.	a.	Write the R code for performing the following operations. i. Create a list named student_list with the following information: • Student IDs: 200, 202, 204 • student Names: Alice, Bob, Charlie a list courses_enrolled with three vector element as given below I. First vector values: Maths,Physics II. Second vector value: Chemistry,Biology III. Third vector value: History,Literature ii. List how many courses Bob is enrolled into.	2	CO1	L3	1.
	b.	With suitable examples explain the different data types and data structures used in R language.	8	CO1	L2	1.6.1
3.	a.	Describe the different ways in which scatterplots can be created using R language.	3	CO4	L2	1.6.1
	b.	Write the R code for creating a boxplot of the different attributes of the iris dataset and find the five number summary of the dataset given below; (8,67,69,78,79,89,99,156) Identify the outliers if any.	7	CO1	L3	1.7.1
		OR				
4.	a.	Explain with an example how area charts are created in R including the essential parameters needed for its construction?	3	CO4	L2	1.6.1
	b.	Consider the below given time series with nine periods of data: 34, 38, 46, 41, 43, 48, 51, 50, 56. Use exponential smoothing method to forecast the value for period 10. Assume $\alpha = 0.2$	7	CO1	L3	1.7.1
5.	a.	Explain in detail the KNN algorithm.	5	CO2	L2	1.6.1
	b.	Write the R code for implementing KNN algorithm including the necessary packages and functions?	5	CO2	L3	2.5.3
		OR				
6.	a.	Explain how a Naïve Bayes classifier can be implemented in R language	4	CO2	L2	1.6.1
	b.	Given a training dataset as shown below, using Naive Bayesian algorithm. Find whether a data object with color=Red, Type=SUV, origin= Domestic have the class label as YES OR NO.	6	CO5	L3	1.7.1

Example No	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

7.	a.	Explain how Apriori rules can be generated from any Transaction dataset in R language. Write the R statement for applying a constraint on the rules generated using Apriori algorithm	4	CO3	L3	1.7.1
	b.	Using Apriori algorithm find the frequent itemset from the below given transaction dataset. Assume minimum support count as 2. Also find the possible association rules that can be generated from any one of the frequent item sets take confidence threshold as 40%	6	CO3	L3	1.7.1

Tid	Items Bought
1	Book, Table, Ball
2	File,Table, Bag
3	Book,File,Table, Bag
4	File, Bag
5	Laptop

OR

8.	a.	Consider the below given text and write the R code to perform the following operations on the text Varun works in ABC university //@ Kreeshma works at DEF University. They both works as Bigdata Developers. Bigdata Developers work on large datasets. Big data developers architect and construct scalable data systems using distributed computing technologies empowering organizations to harness the potential of vast datasets for insights and decision making.	7	CO2	L3	1.7.1

		i. Perform preprocessing to remove special characters, convert lower case to upper case, remove English common stop words, remove punctuations, eliminate extra white spaces and perform stemming. ii. Find the term document matrix iii. Find frequent terms with low frequency =3 iv. Draw the word cloud v. Find the terms correlated to the term 'data' with correlation limit =0.2				
	b.	With an example explain the use of TF-IDF in text mining.	3	CO2	L2	1.6.1
9.	a.	Explain how the accurate value for K can be selected in K means algorithm?	3	CO2	L2	1.6.1
	b.	For the given data, compute two clusters using K-means algorithm for clustering where initial cluster centers are (1.0, 1.0) and (5.0, 7.0). Execute for two iterations.	7	CO2	L3	1.7.1

Record	A	B
R1	1	1
R2	1	2
R3	3	4
R4	5	7
R5	3	5
R6	4	5
R7	3	4

OR

10.	a.	Explain the steps in performing agglomerative hierarchal clustering.	4	CO2	L2	1.6.1
	b.	Using single linkage/ MIN agglomerative clustering algorithm find the clusters of the data objects in the below given dataset.	6	CO2	L3	1.7.1

RECORD	X	Y
P1	0.07	0.83
P2	0.85	0.14
P3	0.66	0.89
P4	0.49	0.64

****

M.Sc.(DS).II/04.25.003 Reg.No. | | | | | | | | |

C

M.Sc. COMPUTER SCIENCE WITH SPECIALIZATION IN DATA SCIENCE SECOND SEMESTER EXAMINATION APRIL 2025

22-359-0204 MACHINE LEARNING

(Regular)

Answer any FIVE questions.
(Each Question Carries 10 Mark)

Time-3 Hours **Maximum Marks :50**

Q. No		QUESTIONS	MARKS	CO	BL	PI
1.	a.	Explain the different types of learning models with examples.	6	CO1	L2	1.7.1
	b.	Discuss the importance of training versus testing in machine learning.	4	CO1	L2	1.7.1
		OR				
2.	a.	Explain various data transformation strategies with examples.	5	CO1	L2	1.7.1
	b.	Describe how data normalization helps in improving machine learning model performance.	5	CO1	L2	1.7.1
3.	a.			CO2	L3	2.6.4

Transaction ID	Items Bought
T1	Milk, Bread, Butter
T2	Bread, Butter
T3	Milk, Bread
T4	Milk, Butter
T5	Bread, Butter
T6	Milk, Bread, Butter
T7	Milk, Bread

Using the Apriori algorithm, perform the following steps:

a) List all the frequent itemsets using a minimum support count of 3. — 6

b) Generate strong association rules with minimum confidence of 70% from the frequent itemsets found. — 4

OR

4.		Describe the FP-Growth algorithm for mining frequent patterns. How does FP-Growth improve upon the Apriori algorithm? Illustrate with an example.	10	CO2	L3	1.7
5.		Compare SVM and logistic regression for classification. Include mathematical formulations and scenarios where each excels.	10	CO4	L2	1.6.1
		OR				
6			10	CO3	L3	2.8.4

Day	Weather	Temperature	Traffic	Time	GoOut
D1	Sunny	Hot	Heavy	Morning	No
D2	Sunny	Hot	Light	Afternoon	No
D3	Overcast	Hot	Light	Morning	Yes
D4	Rainy	Mild	Heavy	Morning	Yes
D5	Rainy	Cool	Light	Evening	Yes
D6	Rainy	Cool	Light	Morning	No
D7	Overcast	Cool	Light	Evening	Yes
D8	Sunny	Mild	Heavy	Morning	No
D9	Sunny	Cool	Light	Evening	Yes
D10	Rainy	Mild	Light	Evening	Yes
D11	Sunny	Mild	Light	Evening	Yes
D12	Overcast	Mild	Heavy	Afternoon	Yes
D13	Overcast	Hot	Light	Morning	Yes
D14	Rainy	Mild	Heavy	Evening	No

Given the dataset shown above, use the ID3 algorithm to construct a decision tree that predicts whether a person will GoOut or not, based on the following attributes: Weather (Sunny, Overcast, Rainy), Temperature (Hot, Mild, Cool), Traffic (Heavy, Light) and Time (Morning, Afternoon, Evening).

7.	a.	Explain the k-Means clustering algorithm.	4	CO3	L2	1.7.1
	b.	Compare hierarchical clustering methods with density-based clustering methods.	6	CO3	L2	1.7.1

		OR				
8	a.	Implement the k-means clustering algorithm manually for the dataset: {2, 4, 10, 12, 3, 20, 30, 11} with k=2 and initial centroids at 4 and 12. Show iterations until convergence	6	CO3	L3	1.6.1
	b.	Explain how hierarchical clustering differs from partitioning methods. When would you prefer BIRCH over k-means?	4	CO3	L3	1.6.1
9.	a.	Explain the architecture of a Multilayer Perceptron with a diagram.	4	CO5	L2	1.7.1
	b.	Describe the backpropagation algorithm for training neural networks.	6	CO5	L2	1.7.1
		OR				
10.	a.	Explain different activation functions used in neural networks with their derivatives.	5	CO5	L3	1.7.1
	b.	You are tasked with building a classification model to predict whether a person has diabetes using the Pima Indians Diabetes dataset. The dataset contains features like Glucose, BMI, Age, Insulin, and the target variable Outcome (1 = diabetic, 0 = non-diabetic). Write the Python code snippet (not full program) to do the following: i. Split the dataset into training and test sets. ii. Create and train an MLPClassifier with one hidden layer of 10 neurons. iii. Predict and print the accuracy of the model on the test data.	5	CO5	L2	2.8.4

****

M.Sc.(DS).II/04.25.004 Reg.No. | | | | | | | | |

A

M.Sc. COMPUTER SCIENCE WITH SPECIALIZATION IN DATA SCIENCE SECOND SEMESTER EXAMINATION APRIL 2025

22-359-0212 DIGITAL IMAGE PROCESSING

(Regular)

Answer any FIVE questions.
(Each Question Carries 10 Mark)

Time-3 Hours **Maximum Marks :50**

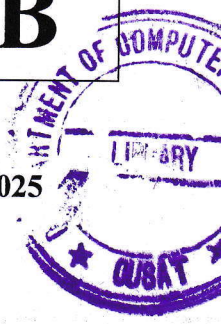
Q. Nos		QUESTIONS	MARKS	CO	BL	PI
1	a.	Explain the concepts of sampling and quantization. How do changes in sampling rate and quantization levels affect the quality of a digital image? Illustrate your answer with suitable examples.	6	CO1	L2	1.4.1
	b.	Given a 3×3 kernel and a small 5×5 image matrix, demonstrate both convolution and correlation operations step by step. Highlight the differences in the results.	4			
		OR				
2	a.	Explain the steps involved in performing histogram equalization on a grayscale image. How can histogram equalization enhance the visibility of features in low-contrast medical images?	6	CO1	L2	1.4.1
	b.	Compare gradient-based filters with the Laplacian filter. In what scenarios would you prefer one over the other?	4			
3	a.	What is the purpose of applying the Discrete Fourier Transform (DFT) to an image, and how does it help in understanding the frequency characteristics of the image? Explain with an example.	6	CO2	L2	1.4.1
	b.	Describe the steps involved in applying homomorphic filtering to an image. Why is a logarithmic transformation used before processing in the frequency domain?.	4			
		OR				

4	a.	Explain the steps involved in applying a frequency domain filter to an image using DFT. What is the significance of shifting the zero-frequency component to the center of the spectrum before applying a frequency filter?	6	CO2	L2	1.4.1
	b.	What is an Ideal low-pass filter? How does the cut-off frequency in a low-pass filter affect the degree of blurring in the output image? Explain with an example.	4			
5	a.	What is salt-and-pepper noise? How does the density of salt-and-pepper noise impact the performance of filtering techniques like the median filter? Explain with examples.	6	CO3	L2	1.4.1
	b.	In what situations would you choose a max filter or min filter over a median filter in noise reduction tasks? Explain with examples.	4			
		OR				
6	a.	Explain Adaptive Median Filter. Give the advantages of Adaptive Median Filter over traditional Median Filter.	6	CO3	L2	1.4.1
	b.	What is Gaussian noise? Explain a method for reducing Gaussian noise in an image.	4			
7.	a.	Explain in detail the Canny Edge Detector.	7	CO4	L2	1.4.1
	b.	Differentiate between Region Growing and Region Splitting and Merging Segmentation.	3		L2	
		OR				
8.	a.	Explain segmentation using K-Means clustering.	7	CO4	L2	1.4.1
	b.	Give the significance of Image Gradient in detection of edges.	3			
9.		Explain in detail the Huffman Coding with example.	10	CO5	L3	1.4.1
		OR				
10.		What are the disadvantages of Golomb codes? Compute the Golomb code, $G_2(5)$.	10	CO5	L3	1.4.1

****

M.Sc.(DS).II/04.25.005 Reg.No. | | | | | | | | |

B



M.Sc. COMPUTER SCIENCE WITH SPECIALIZATION IN DATA SCIENCE SECOND SEMESTER EXAMINATION, APRIL 2025

22-359-0202 DATA BASE MANAGEMENT SYSTEMS

(Regular)

Answer any FIVE questions.

(Each Question Carries 10 Mark)

Time : 3 Hours **Maximum Marks : 50**

Q.No		QUESTIONS	MARKS	CO	BL	PL
1	a.	With the help of a diagram, illustrate how the three-schema architecture separates user views from physical storage.	3	CO1	L2	1.7.1
	b.	In a hospital management system, the hospital maintains data about patients and the doctors they are assigned to. Each patient is admitted under a specific doctor, while a doctor can treat multiple patients. The system records details about each patient such as their name, age, and disease, and maintains doctor details like name and specialization. Design an Entity-Relationship (E-R) diagram to represent the above scenario. Your answer should include the following: 1. Identify and draw the entity sets for patients and doctors, including all relevant attributes. 2. Show the relationship between the entities, clearly indicating the mapping cardinality. 3. Indicate the type of attributes (e.g., simple, composite, derived) where applicable. 4. Mention the key attributes for each entity.	7	CO1	L3	1.7.1
		OR				
2	a.	An online bookstore stores data about books, authors, and publishers. A book can have multiple authors, but each author may write many books. Each book is published by only one publisher. The bookstore stores the following details: • Books: BookID, title, price, genre • Authors: AuthorID, name, email • Publishers: PublisherID, name, contactNumber Construct the E-R diagram showing all entities and relationships. Identify the attribute types and mapping cardinalities.	6	CO1	L3	1.7.1

	b.	Explain the types of attributes in the E-R model. Differentiate between simple, composite, derived, and multivalued attributes with examples.	**4**	**CO1**	**L2**	**1.7.1**
3	**a.**	Explain the various types of functional dependencies.	**5**	**CO2**	**L2**	**1.7.1**
	b.	Illustrate how nested queries are used with the IN, EXISTS, and ANY operators by providing suitable examples for each.	**5**	**CO2**	**L2**	**1.7.1**
		OR				
4	**a.**	You are tasked with designing a database for a university course registration system. The table, REGISTRATION *(StudentID, StudentName, CourseID, CourseName, InstructorName, Instructor Phone)* has the following dependencies. • StudentID → StudentName • CourseID → CourseName • CourseID → InstructorName, InstructorPhone 1. Identify the candidate keys. 2. Determine which normal form the table satisfies. 3. Normalize the table to 3NF with explanation.	**5**	**CO2**	**L3**	**1.7.1**
	b.	Consider a relation schema R(A,B, C, D) with functional dependencies A→BC,C→AB,D→A. Check whether it is in 3NF or not?	**5**	**CO2**	**L3**	**1.7.1**
5	**a.**	A ticket reservation system handles multiple transactions such as booking tickets, cancelling reservations, checking seat availability, and updating passenger details. These transactions frequently access and modify shared data like seat availability and booking records. During peak hours or special events, many users try to book tickets at the same time. 1. Identify the potential concurrency problems that may arise in this ticket reservation system. 2. Suggest an appropriate concurrency control strategy to ensure data consistency and prevent booking conflicts during high-traffic periods.	**8**	**CO3**	**L3**	**1.7.1**

	b.	Define ACID properties.	**2**	**CO3**	**L2**	**1.7.1**
		OR				
6	**a.**	Consider the following schedule S involving three transactions T1, T2, and T3: T1: R(A) W(A) C T2: R(A) W(A) C T3: R(B) W(B) C (R-Read, W-Write, C- Commit) 1. Draw the precedence graph for the given schedule. 2. Is the schedule conflict serializable? Justify your answer. 3. Is the schedule recoverable? If not, explain why and how it can be made recoverable.	**7**	**CO3**	**L3**	**1.7.1**
	b.	Differentiate between shared lock and exclusive lock.	**3**	**CO3**	**L2**	**1.7.1**
7		What is clustering in file organization? How does it improve performance? Describe the benefits and limitations of clustering records in files.	**10**	**CO4**	**L2**	**1.7.1**
		OR				
8		Discuss how records are organized in files. Compare heap files, sequential files, and hash files in terms of structure, performance, and use cases.	**10**	**CO4**	**L2**	**1.7.1**
9	**a.**	Assume that an "employees" collection contains the following fields: name, salary, status, and department. Perform the following operations: a) Update the collection to add a bonus field for all employees whose salary is greater than 50,000. b) Delete all employees from the collection whose status is "resigned".	**7**	**CO5**	**L3**	**1.7.1**

	b.	What is NoSQL? List any two key features that distinguish NoSQL databases from traditional relational databases.	**3**	**CO5**	**L3**	**1.7.1**
		OR				
10	**a.**	Assume that the following documents are in the collection Book. • { title: "Mongo Basics", author: "John Smith", price: 300, stock: 10 } • { title: "Advanced Mongo", author: "John Smith", price: 450, stock: 0 } • { title: "Mongo Guide", author: "Jane Doe", price: 350, stock: 0 } Write and explain MongoDB queries to perform the following on a "books" collection: a) Update the price of all books by a specific author. b) Delete all books that are out of stock.	**6**	**CO5**	**L3**	**1.7.1**
	b.	How are documents deleted in MongoDB? Explain the usage of deleteOne() and deleteMany() methods with proper syntax and examples. Also mention how deletion based on a condition is performed.	**4**	**CO5**	**L3**	**1.7.1**

****