Enrol. No. | S | R | H | U | | | | | | | | | Regn. No. | D | D | | | | | | | | | |

END SEMESTER EXAMINATION, JUNE- 2024

B. Sc. (Hons.) in Data Science Semester – II

Course Name: Discrete Mathematics **Course Code: BDS121**

Time allotted: 03 Hours **Max. Marks: 100**

Note: Read all the instructions carefully.

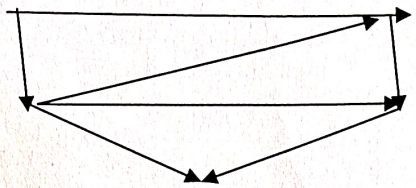
Q.1. Objectives Type **(10 X 2=20)**

(Very short/ Short Answer, MCQ, One-liners, Fill in the Blank, True/ False) Attempt all.

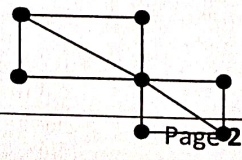
S. No.	Question	CO	PO
a.	List all the subsets of the set $A = \{1,2,3,4\}$	1	1
b.	Write the negation of following statement: If I am ill, then I cannot go to University.	3	2
c.	Let $A = \{1, 2, 3\}$, $B = \{8,9\}$ and $R = \{(1,8),(2,8),(1,9),(3,9)\}$ be a relation from A to B. Determine the complement of relation R.	2	1
d.	$\lvert A \cup B \cup C \rvert =$ _____	1	1
e.	Define a partial order relation.	2	1
f.	Define an Euler path.	5	1
g.	What is a pendent vertex.	5	1
h.	In how many ways a president and a vice president can be chosen from a committee of 25 members.	4	2
i.	How many permutations of the letters of the word BANANA are possible.	4	1
j.	Consider the sets $A = \{1,2,4,6,8\}$, $B = \{2,4,5,9\}$ and $C = \{x : x \text{ is a positive integer and } x^2 \le 16\}$. calculate $A \cap (B \cup C)$.	1	1

Q.2. Short notes / Short Answer Type (Attempt any five out of Seven) **(5 X 4=20)**

a.	In a group of 25 students, 12 have taken Mathematics, 8 have taken Mathematics but not Biology. Find the number of students who have taken Mathematics and Biology and those who have taken Biology but not Mathematics.	1	2
b.	Apply mathematical induction to show the following $1 + 2^n < 3^n$ for $n \ge 2$.	1	2
c.	Let P and Q be the relations on set $A = \{1,2,3,4\}$ defined by	2	1

	$P = \{(1,2),(2,2),(2,3),(2,4),(3,2),(4,2),(4,3)\}$ and $Q = \{(2,2),(2,3),(3,2),(3,3),(3,4),(4,1),(4,2)\}$. Then, calculate P0P.		
d.	Find the adjacency matrix of the following graph.  V_1	5	2
e.	Given $A = \{1,2,3,4\}$, B = {x, y, z}and Let R be a relation from A to B defined as: $R = \{(1,x),(2,z),(3,x),(3,y),(3,z)\}$ (i) Determine the matrix of relation, (ii) find the inverse relation R^{-1},	2	1
f.	Write the converse, inverse and contrapositive of the following statement If I work hard, then I will able to get good job.	3	1
g.	Suppose that an urn contains 15 balls, of which eight are red and seven are black. In how many ways can five balls be chosen so that at least two are red.	4	2

Q.3. Structured Question (Attempt any three out of four) **(3 X 10=30)**

a.	**(i)** Show algebraically that: $(A \cup B)^c = A^c \cap B^c$. (ii) Prove that if n^2 is odd then n is odd, n being an integer.	1 3	2 2
b.	Suppose that 75% of all farmers in a district fertilize their fields, 60% apply herbicides and 35% apply insecticides. In addition, 20% apply none of these, 30% apply all three, 56% apply herbicides and fertilizer, and 33% apply insecticides and fertilizer. Find **(8)** (i) What percentage apply herbicides and insecticides. (II) What percentage apply herbicides and insecticides but not fertilizer. (III) Fill all the regions of Venn diagram. (IV) State Pigeonhole principal **(2)**	1 4	2 1
c	(i) Define the followings in a graph (I)Degree of a vertex (II) pendent edge (III) adjacent vertices (IV) isolated vertex (V) Connected graph. **(5)** (ii) Determine whether the following graph has a Hamiltonian circuit, a Hamiltonian path but no Hamiltonian circuit, or neither. Give reasons for your answer.  **(5)**	5	1

d.	(i)Solve the following recurrence relation with initial condition $a_0 = 2$ and $a_1 = 1$, $a_{n+2} - 2a_{n+1} + a_n = 2^n$. **(7)** (ii) Solve the following recurrence relation $a_n - 2a_{n-1} = 7n^2$. **(8)**	4	2

Q.4. Long Answer / Essay type (Attempt any two out of three) **(2 X 15=30)**

a	(i) Check the validity of following argument. **(7)** If my plumbing plans do not meet the construction code, then I cannot build my house. If I hire a licensed contractor, then my plumbing plans will meet the construction code. I hire a licensed contractor. ∴ I can build my house. (ii) Using mathematical induction, show that $n(n+1)(n+2)$ is a multiple of 6, n being positive integer. **(8)**	3 1	2 2
b.	(i) Without using truth table check the validity of the following argument. $p \rightarrow q,\ q \rightarrow r,\ r \rightarrow s, \sim s,\ p \vee t \vdash t$. **(7)** (ii) Let R be a relation on the set R of real numbers defined as $_aR_b$ if and only if $1 + ab > 0$. Check whether this relation is an equivalence relation or is a partial order relation. **(8)**	3 2	2 2
c.	(i) A panel has to select a president, vice president and secretary from the committee of five members namely Ankit, Arjit, Sonu, Monu and Nonu. **(8)** (I) In how many selections include Sonu and Monu, (II) How many selections are there in whicheither Nonu is a secretary or he isexcluded, (III) How many selections are there in which Ankit is president. (ii) Define one-one and onto function. If R is the set of real numbers, then show that the function $f : R \rightarrow R$ defined by $f(x) = 5x^3 - 1$, is an invertible function. **(7)**	4 3	2 2

***

Enrol. No. | S | R | H | U | | | | | | | | | Regn. No. | D | D | | | | | | | | | |

END SEMESTER EXAMINATION, JUNE- 2024

B. Sc. (Hons.) in Data Science Semester - II

Course Name: Object Oriented Programming using C++ **Course Code: BDS123**

Time allotted: 03 Hours **Max. Marks: 100**

Note: Read all the instructions carefully.

Q.1. Objectives Type

(Very short/ Short Answer, MCQ, One-liners, Fill in the Blank, True/ False)Attempt all. (10×2=20)

		CO	PO
a.	What is Polymorphism in C++	1	1,2,3,12
b.	How we identify attribute in C++	1	1,2,3,12
c.	List type of access derivates in C++	2	1,2,3,12
d.	When we use Constructors	2	1,2,3,12
e.	What is private Access	3	1,2,3,12
f.	Through diagram explain multilevel Inheritance	3	1,2,3,12
g.	List different file mode in C++	4	1,2,3,12
h.	What is Virtual Function	4	1,2,3,12
i.	Write syntax for command line arguments	5	1,2,3,12
j.	File open , File close to operations in C++(true/false)	5	1,2,3,12

Q.2. Short notes / Short Answer Type (Attempt any five out of Seven) (5×4=20)

		CO	PO
a.	Write C++ Program to generate Fibonacci sequence of 10 elements.	1	1,2,3,12
b.	WAP in C++ to demonstrate the usage of constructor and destructor in a class.	2	1,2,3,12
c.	WAP in C++ to Demonstrate the concept of abstract class.	2	1,2,3,12
d.	Write a C++ program to demonstrate function overloading.	3	1,2,3,12
e.	Elaborate the concepts of runtime Polymorphism	3	1,2,3,12
f.	What is the role of Friend function? Explain with example.	4	1,2,3,12
g.	Write C++ function to count the number of spaces in a file MY.TXT.	5	1,2,3,12

Q.3.Structured Question (Attempt any three out of four) **(3×10=30)**

		CO	PO
a	i) Write a program in C++ to find the sum of the first 100 odd numbers. ii) List the advantage and disadvantages of OOPS	1	1,2,3,12
b	i) How we can implement the concept of Encapsulation, Information hiding in C++ Explain with program. ii) What is dynamic memory allocation? Explain with Example	2	1,2,3,12
c	i) C++ program to read and print shape information using inheritance. ii) Elaborate the concept hybrid inheritance explain with Block Diagram and code also .	3	1,2,3,12
d	i) Write a user-defined function in C++ to read the content from a text file OUTPUT.TXT, count and display the number of alphabets present in it ii) Elaborate on Generic class , Write a proper code with explanation.	5	1,2,3,12

Q.4.Long Answer / Essay type (Attempt any two out of three) **(2×15=30)**

		CO	PO
a.	Discuss the types of polymorphism supported in C++. Compare and contrast compile-time (static) polymorphism and runtime (dynamic) polymorphism, highlighting their respective use cases and implementations. Write a program to demonstrate the concepts runtime (dynamic) polymorphism.	3	1,2,3,12
b.	Model a banking system using inheritance and abstract classes in C++. Design a base class Account with virtual functions for deposit, withdrawal, and displaying account information. Derive classes **SavingsAccount** and **CheckingAccount** from Account, each with specific attributes and methods. Make Account an abstract class by declaring at least one pure virtual function. Implement constructors, destructor, and methods for deposit, withdrawal, and displaying information.	2	1,2,3,12
c.	i) Explain the difference between reading and writing binary files compared to text files in C++. Provide an example demonstrating reading and writing binary files. ii) Explain the concept of file handling in C++ and discuss its importance in programming.	5	1,2,3,12

****

Enrol. No. | S | R | H | U | | | | | | | | | Regn. No. | D | D | | | | | | | | | |

END SEMESTER EXAMINATION, JUNE- 2024

B. Sc. (Hons.) in Data Science Semester – II

Course Name: Probability and Distribution Theory **Course Code: BDS122**

Time allotted: 03 Hours **Max. Marks: 100**

Note: Read all the instructions carefully.

Q.1. Objectives Type

(Very short/ Short Answer, MCQ, One-liners, Fill in the Blank, True/ False) Attempt all. (10×2=20)

a.	Write down sample space if an unbiased die is rolled twice.	(CO 1, PO 1)
b.	Write down mean and variance of Poisson distribution.	(CO 4, PO 2)
c.	If A, B, C are three arbitrary events. Find expression for 'all three events occur'.	(CO 2, PO 1)
d.	Addition theorem of probability for any two events A and B.	(CO 2, PO 2)
e.	Formula for marginal probability mass function of random variable X from f(X,Y).	(CO 5, PO 2)
f.	E(aX+b)=....., where E is expectation and a and b are constants.	(CO 3, PO 1)
g.	Write down probability mass function of Binomial distribution.	(CO 4, PO 2)
h.	Addition theorem of probability.	(CO 2, PO 1)
i.	Explain distribution function of a random variable X.	(CO 5, PO 2)
j.	Formula for finding expectation of a discrete random variable X.	(CO 3, PO 2)

Q.2. Short notes / Short Answer Type (Attempt any five out of Seven) (5×4=20)

a.	If P(A)=0.35, P(B)=0.42 and P(A∩B) =0.28. Find P(A/B). Are A and B two independent events?	(CO 1, PO 1)
b.	Derive mean of Poisson distribution.	(CO 4, PO 2)
c.	An urn contains 4 white, 3 red, and 5 black balls. If 2 balls are drawn at random, find the probability that 'one ball is of each colour'.	(CO 2, PO 1)
d.	Explain the concept of joint distribution function of random variables X and Y with an example.	(CO 3, PO 2)
e.	The joint density function of two random variables X and Y is given below. f(X,Y)={2-x-y; 0≤x≤1, 0≤y≤1, and 0, otherwise, Find conditional density functions of X and Y.	(CO 5, PO 1)
f.	In an examination test, 25% of the students have failed in English, 18% in Mathematics, and 10% in both subject. A student is selected at random. What is the probability that the student has failed in English if it is known that she has failed in Mathematics?	(CO 1, PO 1)
g.	Binomial distribution.	(CO 4, PO 2)

P.T.O./-

Q.3. Structured Question (Attempt any three out of four) **(3×10=30)**

a. Explain Uniform distribution and hence derive its mean and variance. **(CO 5, PO 2)**

b. A box contains 6 red, 4 white, and 5 black balls. A person draws 4 balls from the box at random. Find the probability that among the balls drawn there is at least one ball of each colour. **(CO 2, PO 1)**

c. Let X be a random variable with the following probability distribution: **(CO 2, PO 2)**

x	-3	6	9
P(x)	1/6	1/2	1/3

Find E(X), $E(X^2)$, V(X), and $E(6X+8)^2$

d. Two random variables X and Y have the following joint density function: **(CO 3, PO 1)**
$f(X,Y)=\{2-x-y;\ 0\le x\le 1,\ 0\le y\le 1$, and 0, otherwise,
Find
(i) Marginal probability density function of X and Y.
(ii) E(x) and E(Y)
(iii) V(X) and V(Y)

Q.4. Long Answer / Essay type (Attempt any two out of three) **(2×15=30)**

a. (i), A letter is known to have come either from TATANAGAR or from CALCUTTA. On the envelop just two letters TA are visible. What is the probability that the letter come from CALCUTTA? **(CO 1, PO 1)**

(ii) Explain Cumulative Distribution Function (CDF) **(CO 3, PO 2)**

b. (i), Three coins are tossed. Using Binomial distribution, find the probability that there are more than one head. **(CO 4, PO 2)**

(ii) Of every 100 mosquitoes selected, it is found in average, 25 were in grade A and 10 were in grade B. Use the addition rule of probability that the selected mosquitoes will have either grade A or grade B. **(CO 1, PO 1)**

c. (i), From a city population, the probability of selecting (a) a male or a smoker is 7/10, (b) a male smoker is 2/5, and (c) a male, if a smoker is already selected is 2/3. Find the probability of selecting a male. **(CO 3, PO 1)**

(ii) For the adjoining bivariate probability distribution of X and Y, given below: **(CO 4, PO 2)**

Y → / X ↓	1	2	3	4	5	6
0	0	0	1/32	2/32	2/32	3/32
1	1/16	1/16	1/8	1/8	1/8	1/8
2	1/32	1/32	1/64	1/64	0	2/64

Find P (X<2, Y≤3).

***

14/6/24

Enrol. No. | S | R | H | U | | | | | | | | | Regn. No. | D | D | | | | | | | | | |

END SEMESTER EXAMINATION, JUNE- 2024

B. Sc. (Hons.) in Data Science Semester – II

Course Name: Disaster Management **Course Code: DMT126**

Time allotted: 03 Hours **Max. Marks: 100**

Note: Read all the instructions carefully.

Q.1. Attempt all. **(10×2=20)**

QN	Questions	CO	PO
a.	Give two examples of pandemics.	1	12
b.	Give two examples of traditional hazards.	1	1
c.	Define focus of earthquake.	1	12
d.	Flood is an example of hydroclimatic disaster. (True/False)	1	1
e.	Define rehabilitation.	1	6
f.	Name any two phases of disaster management cycle.	3	10
g.	Write one negative effect of urbanization.	2	12
h.	Development and disasters are interrelated. (True/False)	2	12
i.	Write full form of NDRF.	1	1
j.	Tsunameter is used for _____.	1	1

Q.2. Attempt any five out of Seven. **(5×4=20)**

QN	Questions	CO	PO
a.	Differentiate between flash flood and riverine flood.	1	1
b.	Write the causes of global warming.	2	12
c.	Brief the negative aspects of development.	2	1
d.	Draw DRR cycle.	3	10
e.	How development decreases the vulnerability for future disasters?	3	1
f.	Write the functions of NDMA in disaster management.	1	1
g.	Brief the causes of acid rain.	1	1

Q.3. Attempt any three out of four. **(3×10=30)**

QN	Questions	CO	PO
a.	Explain the causes of cyclone formation or atmospheric disturbances.	1	1
b.	Brief the negative effects big hydropower dams.	2	12
c.	Describe different preventive measures for landslides.	3	1
d.	Brief the national framework for disaster management in India.	3	1

Q.4. Attempt any two out of three. **(2×15=30)**

QN	Questions	CO	PO
a.	Discuss the need for and importance of sustainable development in disaster risk reduction.	2	6
b.	Describe the structural and non-structural mitigation measures for disasters.	3	6
c.	Explain the process of preparedness and communication at family and community level.	4	9,10

***

12/6/24

Enrol. No. | S | R | H | U | | | | | | | | | Regn. No. | D | D | | | | | | | | | |

END SEMESTER EXAMINATION, JUNE- 2024

B. Sc. (Hons.) in Data Science Semester – II

Course Name: Data Warehousing and Mining **Course Code: BDS125**

Time allotted: 03 Hours **Max. Marks: 100**

Note: Read all the instructions carefully.

Q.1. Objectives Type

(Very short/ Short Answer, MCQ, One-liners, Fill in the Blank, True/ False) Attempt all. (10×2=20)

		CO	PO
a.	What do you understand by data mining	1	1
b.	What do you understand by information gain and entropy?	2	2
c.	What is the need for developing a data warehouse?	3	3
d.	What is DBSCAN method?	2	2
e.	What are independent data marts?	4	1
f.	Describe the role of metadata in a data warehouse system.	2	2
g.	What do you understand by data warehouse?	4	2
h.	What do you understand by cluster analysis?	3	2
i.	Describe the process of data cleaning.	5	1
j.	There are some factors that should be considered while designing of a data warehouse. Explain.	5	2

Q.2. Short notes / Short Answer Type (Attempt any five out of Seven) (5×4=20)

a.	Elaborate some key components of data warehouse.	5	2
b.	What does OLAP stand for, and how does it differ from OLTP?	4	4
c.	Role of association rule mining to the field of data mining?	3	3
d.	Write characteristics of OLAP system.	2	2
e.	Describe the importance of multidimensional view in OLAP.	1	2
f.	What is the main objective of classification in data mining?	2	2
g.	What are data marts? Explain with the help of an example.	3	4

Q.3. Structured Question (Attempt any three out of four) (3×10=30)

a.	(i) Describe the conceptual modeling of data warehouses using star schema. (ii) What is the difference between star schema and snowflake schema? (iii) What are fact constellations, and how do they relate to data warehousing?	3	4
b.	(i) Differentiate between ROLAP, MOLAP, and HOLAP OLAP server architectures. (ii) Explain the term "slice and dice" in the context of OLAP. (iii) Define data reduction in the context of data mining.	2	2
c.	(i) How does data preprocessing contribute to the effectiveness of data mining algorithms? (ii) Explain the concept of OLAP cube partitioning and its benefits (iii) Explain the concept of support and confidence measures in association rule mining.	4	4
d.	(i) Draw the structure of a decision tree along with its components. (ii) Discuss the concept and role of the Gini index in decision tree split algorithms. (iii) Name basic clustering methods along with their general characteristics.	5	2

Q.4. Long Answer / Essay type (Attempt any two out of three) (2×15=30)

a	(i) How does the Apriori algorithm work? Explain with the help of an example. (ii) What is dynamic itemset counting (DIC), and how does it address the scalability issues of traditional association rule mining algorithms?	4	4
b	(i) Describe the FP-Growth algorithm for mining frequent patterns with the help of an example. (ii) Why FP-Growth algorithm is more effective as compare to Apriori algorithm? Explain.	3	3
c	(i) Explain the process of tree induction in decision tree algorithms with the help of an example. (ii)What is market basket analysis? Explain with help of suitable example.	2	2

***

ol. No. | S | R | H | U | | | | | | | | | Regn. No. | D | D | | | | | | | | | |

END SEMESTER EXAMINATION, JUNE- 2024

BCA Semester – II/ B. Sc. (Hons.) in Data Science Semester - II

Course Name: Data Structures **Course Code: BC122T/BDS124**

Time allotted: 03 Hours **Max. Marks: 100**

Note: Read all the instructions carefully.

Q.1. Attempt all. **(10×2=20)**

S.No	Questions	CO	PO
a.	Define stack.	1	1
b.	What is circular queue?	2	2
c.	What is the difference between stack and queue?	3	2
d.	Mention the different ways of tree traversal.	2	1
e.	Give the node structure of singly linked list.	1	1
f.	Define Data structure.	2	1
g.	Mention the graph traversal methods.	2	1
h.	Define space and time complexities of an algorithm.	2	1
i.	What is recursion?	3	1
j.	What is dynamic memory allocation?	2	1

Q.2. Attempt any five out of Seven. **(5×4=20)**

S.No	Questions	CO	PO
a.	What are queue and its applications?	1	1
b.	Explain in detail about ordered list, sparse matrix and Vector.	1	1
c.	Write a program to calculate the factorial of a number using recursion.	1	1
d.	What is recursion? Explain the significance of recursion.	1	1
e.	Create function in C which checks full or empty for a linear queue. Also find out the number of elements in a queue.	3	1
f.	What is priority queue? How it is implemented?	2	3
g.	Write an algorithm to delete a given item from a two way list.	3	1

Q.3. Attempt any three out of four. **(3×10=30)**

S.No	Questions	CO	PO
a.	Write an algorithm to evaluate postfix expression. Convert the following infix expression into equivalent postfix expression (a+b)*c-d+e/(f-g)	1	1
b.	Define a tree. Explain the different types of binary tree. What is predecessor and successor of a node in a binary search tree?	2	2
c.	Given the following input(4322,1334,1471,9679,1989,6171,6173,4199) and the hash function x mod 10, what is the final hash table.	3	2
d.	Compare BFS and DFS traversal methods for graph with suitable example.	3	3

Q.4. Attempt any two out of three. **(2×15=30)**

S.No	Questions	CO	PO
a.	i) Write an algorithm to insert an item into a one way list. ii) What are the various operations possible on doubly linked list? Explain with suitable example.	1	1
b.	i) What is hash table? What is the purpose of using hash table? How the use of hash table is more beneficial as compared to its contemporary techniques. ii) What are the various applications of Stack? Explain with an example.	2	2
c.	i) Explain circular queue with example and write an algorithm to insert and delete an element from a circular queue. ii) Apply bubble sort on the following elements: {21, 11, 5, 78, 49, 54, 72, 88}	3	3

***

Roll No. | S | R | H | U | | | | | | | | | Regn. No. | D | D | | | | | | | | | |

END SEMESTER EXAMINATION, JUNE- 2024

B. Sc. (Hons.) in Data Science Semester - IV

Course Name: Estimation and Inferential Statistics **Course Code: BDS241**

Time allotted: 03 Hours **Max. Marks: 100**

Note: Read all the instructions carefully.

Q.1. Objectives Type **(10 X 2=20)**

(Very short/ Short Answer, MCQ, One-liners, Fill in the Blank, True/ False) Attempt all.

S. No.	Question	CO	PO
a.	The probability of committing Type I error is called _____.	3	1
b.	In point estimation the error of estimate never depends on the size of sample. (True/False)	1	1
c.	Write the confidence interval for the mean of a population when large sample is taken with known population variance.	2	1
d.	In the test concerning goodness of fit the number of degrees of freedom is_____.	5	1
e.	Write the criteria for rejecting the null hypothesis $p = p_0$ against the alternative hypothesis $p < p_0$ in the testing of hypotheses concerning one proportion.	3	1
f.	In the interval estimation of difference between means with small samples the population variance is estimated by_____	4	1
g.	Write the Cramer- Rao inequality.	1	1
h.	The value of point estimator is called _____.	1	1
i.	In a certain testing of hypothesis concerning differences of two means if samples are of sizes $n_1 = 12$ and $n_2 = 8$, then the test statistics to be used is _____.	4	1
j.	Define efficient estimator in point estimation.	1	1

Q.2. Short notes / Short Answer Type (Attempt any five out of Seven) **(5 X 4=20)**

a.	Show that the sample mean is an unbiased estimator of population mean.	1	1
b.	In six determinations of the melting point of tin, a chemist obtained a mean of 232.26 degrees Celsius with a standard deviation of 0.14 degrees. If he uses this mean to estimate the actual melting point of tin, what can the chemist assert with 98% confidence about the maximum error.	2	2
c.	A random sample of size n =100 is taken from a population with $\sigma = 5.1$. Given that the sample mean is 21.6, construct a 95% confidence interval for the population mean.	2	2

P.T.O./-

d.	In a sample survey conducted in a large city, 136 of 400 persons answered yest to the question whether their city's public transportation is adequate. With 99% confidence, what can we say about the maximum error if $\frac{x}{n} = 0.34$ is used an as estimate of the corresponding true proportion.	1	
e.	A sample survey at a super market showed that 204 of 300 shoppers regularly use cents-off coupons. Construct a 95% confidence interval for the corresponding true proportion.	2	2
f.	Explain the Mann Whitney U test or the rank sum test.	5	1
g.	It is desired to determine whether there is less variability in the silver planting done by company 1 than in that done by company 2. If independent random samples of size 12 of two companies work yield $S_1 = 0.035$ mil and $S_2 = 0.062$ mil, test the null hypothesis $\sigma_1^2 = \sigma_2^2$ against the alternative hypothesis $\sigma_1^2 < \sigma_2^2$ at 0.05 level of significance.	4	2

Q.3. Structured Question (Attempt any three out of four) **(3 X 10=30)**

a.	Show that $\bar{X}$ is a minimum variance unbiased estimator of the mean μ of a normal population.	1	2
b.	**(i)** An experiment obtained the following observations 0.59, 0.56, 0.49, 0.55, 0.65, 0.55, 0.51, 0.60, 0.56, 0.47, 0.58, 0.61, 0.54, 0.68, 0.56, 0.50, 0.57 and 0.53. Use sign test at 0.05 level of significance to test the null hypothesis $\bar{\mu} = 0.55$ against the alternative hypothesis $\bar{\mu} \neq 0.55$. **(ii)** The following are the average weekly losses of worker-hours due to accidents in 10 industrial plants before and after certain safety program was put into operation: Before: 45, 73, 46, 124, 33, 57, 83, 34, 26, 17 After: 36, 60, 44, 119, 35, 51, 77, 29, 24, 11 Use sign test at 0.05 level of significance to test whether the safety program is effective.	5	2
c.	The following random samples are measurements of the heat-producing capacity of specimens of coal from two mines: Mine 1: 8260, 8130, 8350, 8070, 8340 Mine 2: 7950, 7890, 7900, 8140, 7920, 7840. Use the 0.01 level of significance to test whether the difference between the means of these two samples is significant.	4	2
d.	A random sample from a company's very extensive files shows that orders for a certain piece of machinery were filled, respectively, in 10, 12, 19, 14, 15, 18, 11 and 13 days. Use the level of significance α=0.01 to test the claim that on the average such orders are filled in 10.5 days. Choose the alternative hypothesis so that rejection of the null hypothesis μ=10.5 implies that it takes longer than indicated. Assume normality.	4	2

Long Answer / Essay type (Attempt any two out of three) **(2 X 15=30)**

	To determine whether there really is a relationship between an employee's performance in the company's training program and his or her ultimate success in the job, the company takes a sample of 400 cases from its very extensive files and obtain the results shown in the following table: Performance in training program Success in job (employer's rating) Use the 0.01 level of significance to test the null hypothesis that performance in the training program and success in the job are independent.	5	1
b.	The following are the numbers of minutes it took random samples of 15 men and women to complete a written test given for the renewal of their driver's licenses: Men: 9.9, 7.4, 8.9, 9.1, 7.7, 9.7, 11.8, 9.2, 10.0, 10.2, 9.5, 10.8, 8.0, 11.0, 7.5 Women: 8.6, 10.9, 9.8, 10.7, 9.4, 10.3, 7.3, 11.5, 7.6, 9.3, 8.8, 9.6 Use the U test, at 0.05 level of significance to decide whether to accept the null hypothesis $\mu_1 = \mu_2$ or the alternative hypothesis $\mu_1 \neq \mu_2$, where μ_1 and μ_2 are the average amounts of time it takes men and women to complete the test.	5	2
c.	**(i)**Twelve randomly selected tree of one variety have a mean height of 13.8 feet with a standard deviation of 1.2 feet, and 15 randomly selected trees of another variety have mean height of 12.9 feet with a standard deviation of 1.5 feet. Assuming that the samples were selected from normal populations with equal variances, construct a 95% confidence interval for the difference between the true average height of the two kinds of trees. **(9)** **(ii)** In a random sample, the weights of 24 Black Angus steers of a certain age have a standard deviation of 238 pounds. Assuming that the weights constitute a random sample from a normal population, test the null hypothesis $\sigma = 250$ pounds again the two sided alternative hypothesis $\sigma \neq 250$ at the 0.01 level of significance. **(6)**	4 3	2 2

Success in job (employer's rating)	Poor	23	60	29
	Average	28	79	60
	Very good	9	49	63

***

Roll No. | S | R | H | U | | | | | | | | | | Regn. No. | D | D | | | | | | | | | | |

END SEMESTER EXAMINATION, JUNE- 2024

B. Tech. (Hons.) CSE in AI/DS & ML Semester - IV

Course Name: Probability and Statistics **Course Code: BS401T**

Time allotted: 03 Hours **Max. Marks: 100**

Note: Read all the instructions carefully.

Q.1. Attempt all. **(10×2=20)**

S. No.	Question	CO	PO
a.	In an experiment of rolling a pair of dice, describe the event that the sum of points rolled with the pair of dice is 8.	1	1
b.	The four possible mutually exclusive outcomes A, B, C and D of an experiment have followi assignments of probability. Check whether this assignment is permissible. $P(A) = 0.17$, $P(B) = 0.33$, $P(C) = 0.26$, $P(D) = 0.34$.	1	1
c.	If we do sampling with replacement then which of the following distribution will be applicable (i) Binomial (ii) Poisson (iii) Hypergeometric (iv) none of these.	2	1
d.	If X_1 has a mean 5 and variance 9 while X_2 has a mean -2 and variance 5, and two are independent find $E(3X_1 + 2X_2 - 5)$.	3	1
e.	Find $z_{0.05}$.	2	2
f.	With standard notations if $S_{xy} = 10.17744$ and $S_{xx} = 0.68088$, find the least square estimate of intercept.	5	1
g.	Write the model for simple linear regression.	5	1
h.	Define type II error.	4	1
i.	In a certain testing of hypothesis concerning differences of two means if samples are of sizes $n_1 = 42$ and $n_2 = 35$, then the test statistics to be used is	4	1
j.	Define unbiased estimator in point estimation.	3	1

Q.2. Attempt any five out of Seven. **(5×4=20)**

		CO	PO
a.	If a random variable X has the probability density $f(x) = \begin{cases} 2.e^{-2x}, x > 0 \\ 0, x \leq 0. \end{cases}$ Then Find the value probabilities that it will take on a value (i) between 1 and 3, (ii) greater than 0.50.	2	2
b.	When we sample from an infinite population, what happens to the standard error of the mean if sample size is (i) increased from 50 to 200, (ii) decreased from 225 to 25.	1	1

c.	The number of customers who visit a car showroom on a given day is a random variable with $\mu=18$ and $\sigma=2.5$. with what probability can we assert that there will be between 8 and 28 customers.	2	2
d.	In six determinations of the melting point of tin, a chemist obtained a mean of 232.26 degrees Celsius with a standard deviation of 0.14 degrees. If he uses this mean to estimate the actual melting point of tin, what can the chemist assert with 98% confidence about the maximum error.	3	2
e.	If x = 36 of n = 100 persons interviewed are familiar with the tax incentives for installing certain device, construct a 95% confidence interval for the corresponding true proportion.	3	2
f.	With standard notations if $S_{xy} = 510.40$, $S_{xx} = 142{,}000$, $S_{yy} = 5.13745$ $\sum_{i-1}^{n} x_i = 2{,}050$, $\sum_{i-1}^{n} y_i = 9.35$ and $n = 10$. Fit a straight line and use it to find the value of dependent variable when independent variable is 190.	5	2
g.	Two random samples have the variances 15, 750 and 10,920 respectively. Use 0.02 level of significance to test whether it is reasonable to assume that the variances of the two populations samples are same.	4	2

Q.3. Attempt any three out of four. **(3×10=30)**

		CO	PO
a.	(i) Three machines A, B, and C produce respectively 50%, 30% and 20% of the total number of items of a factory. The percentage of defective output of these machines are 3%, 4% and 5% respectively. If an item is selected at random, find the probability that: (1) The item is defective; (2) The chosen defective item is from machine B. (5) (ii) Find the probabilities of getting (5) (1) eight heads in a row with a balanced coin; (2) three 3's and then a 4 or a 5 in four rolls of a balanced die; (3) five multiple-choice questions answered correctly, if for each question the probability of answering it correctly is 1/3.	1	2
b.	(i) A quality control engineer inspects a random sample of 3 batteries from each lot of 24 car batteries. If such a lot contains 6 batteries with slight defects, what are the probabilities that the inspector's sample will contain (5) (1) None of the batteries with defects; (2) only one of the batteries with defects; (3) at least two of the batteries with defects. (ii) Find the variance of probability distribution given by $f(x) = \frac{1}{n}$ for x = 1, 2, 3,....n. (5)	2	2
c.	For the simple linear regression calculate the following for the given data a. the estimates of slope b. the estimates of intercept $\sum_{i-1}^{n} x_i = 43,\ \sum_{i-1}^{n} y_i = 572,\ \sum_{i-1}^{n} x_i^2 = 157.42,\ \sum_{i-1}^{n} y_i^2 = 23530,\ \sum_{i-1}^{n} x_i\, y_i = 1697.80$ and $n = 14$.	5	2
d.	The specification for a certain kind of ribbon call for a mean breaking strength of 185 pounds. If five pieces randomly selected from different rolls have breaking strength of 171.6, 191.8, 178.3, 184.9, and 189.1 pounds against the alternative hypothesis $\mu<185$ pounds at the 0.05 level of significance.	4	2

Attempt any two out of three. **(2×15=30)** **CO** **PO**

		CO	PO
a.	(i) From a box containing 3 blue, 2 red, and 3 green marbles, two marbles are selected at random. **(1)** Find the joint probability distribution of $X_1 =$ the number of blue marbles, and $X_2 =$ the number of red marbles. **(2)** Find the marginal probability distribution of X_1. **(3)** Find the conditional probability distribution X_1 given $X_2 . = 0$. **(9)**	2	2
	(ii) A manufacturer of submersible pumps claims that at most 40% of the pumps require repairs with in first 5 years of operation. If a random sample of 150 of these pumps include 49 which require repairs within the first 5 years, test the null hypothesis $p = 0.40$ against the alternative hypothesis $p < 0.40$ at the 0.01 level of significance. **(6)**	4	2
b.	(i) If the probability density of a random variable is given by $f(x) = \begin{cases} x & for\ 0 < x < 1 \\ 2 - x & for\ 1 \le x < 2 \\ 0 & elsewhere \end{cases}$. Find the corresponding distribution function and use it to find the probability that the random variable will take on a value greater than 1.8. **(8)**	2	2
	(ii) A random variable has a normal distribution with $\sigma = 10$. If the probability is 0.8212 that it will take on a value less than 82.5, what is the probability that it will take on a value greater than 58.3. **(7)**		
c.	(i) Two randomly selected samples of sizes 5 and 6 have a means 8230 and 7940 respectively. Their variances are 15, 750 and 10,920 respectively. Assuming that the samples were selected from normal populations with equal variances, use 0.01 level of significance to test whether the difference between the means of these two samples is significant. **(7)**	3	2
	(ii) Fit a simple linear regression model to the following data taking x as independent and y as dependent variable **(8)**	5	1

x	27	45	41	19	35	39	19	49	15	31
y	57	64	80	46	62	72	52	77	57	68

***

06/06/24

Enrol. No. | S | R | H | U | | | | | | | | | Regn. No. | D | D | | | | | | | | | |

END SEMESTER EXAMINATION, JUNE- 2024

B. Sc. (Hons.) in Data Science Semester - IV

Course Name: Operating System **Course Code: BDS244**

Time allotted: 03 Hours **Max. Marks: 100**

Note: Read all the instructions carefully.

Q.1. Attempt all. **(10×2=20)**

a.	Define system call.	CO1	PO2
b.	List out the different OS functions.	CO1	PO1
c.	List out CPU scheduling algorithms.	CO2	PO2
d.	What is Preemptive and Non Preemptive process Scheduling?	CO5	PO3
e.	What is Virtual memory?	CO3	PO2
f.	Define Deadlock.	CO3	PO1
g.	What is starvation?	CO4	PO1
h.	List out the conditions of deadlock.	CO4	PO2
i.	What is belady anomaly and which allocation algorithm suffer by it?	CO2	PO5
j.	Define semaphore.	CO2	PO2

Q.2. Attempt any five questions. Each carries equal marks **(5X4=20)**

a.	Explain multiprogramming and multitasking system.	CO1	PO1
b.	Explain functions and system calls.	CO5	PO2
c.	Compare UNIX operating system and Windows operating system.	CO5	PO5
d.	Explain any two disk scheduling algorithms.	CO4	PO4
e.	Explain any two conditions of deadlock.	CO2	PO2
f.	Discuss the following: (i) Race Condition (ii) Critical Section	CO3	PO3
g.	Explain the cooperating process and producer consumer problem.	CO1	PO2

P.T.O./-

Q3. Attempt any three questions. Each carries equal marks. **(3X10=30)**

a.	Explain Operating system components and functions of OS in detail.	CO2	PO2
b.	Explain file access methods in detail with suitable examples.	CO4	PO3
c.	Explain CPU scheduling algorithms. Consider a reference string: 3, 7, 5, 1, 7, 5, 1, 2, 7, 2. The number of frames in the memory is 3. Find out the number of page faults in MRU, LRU and FIFO Page Replacement Algorithm.	CO3	PO5
d.	Define deadlock. Why it occurs? What are the four conditions of deadlock? Explain deadlock detection and recovery mechanism with suitable example.	CO2	PO4

Q4. Attempt any two questions. Each carries equal marks. **(2X15=30)**

a. Suppose that a disk drive has 300 cylinders numbered from 0 to 299. The drive is currently serving a request at cylinder 113. The queue of pending request in order is 186, 170, 213, 174, 48, 150, 12, 250, 30 starting from the current head position. What is the total distance moved by head using following disk scheduling algorithm

i. FCFS and C Look
ii. C-Scan and Scan

CO4 | PO5

b. Consider the following processes

PROCESS	BURST TIME	Arrival Time
P1	8	0
P2	4	0
P3	6	1
P4	1	1

Draw GANTT CHART to show the execution of the following scheduling and Calculate average Waiting Time.
a) FCFS
b) SJF (Preemptive and NonPreemptive)
c) Round Robin (Time quantum= 2 ms)

CO3 | PO4

c. Using Banker's algorithm, answer the following questions:

PROCESS	Allocation				Max				Available			
	A	B	C	D	A	B	C	D	A	B	C	D
P0	0	1	1	0	0	2	1	0	1	5	2	0
P1	1	2	3	1	1	6	5	2				
P2	1	3	6	5	2	3	6	6				
P3	0	6	3	2	0	6	5	2				
P4	0	0	1	4	0	6	5	6				

a. What is the content of the matrix Need?
b. Is the system in a safe state?
c. If a request from process P1 arrives for (0,4,2,0), can the request be granted immediately?
d. How many resources of type A, B, C, D are there?

CO2 | PO2

***

Enrol. No. | S | R | H | U | | | | | | | | | 8/6/24 Regn. No. | D | D | | | | | | | | | |

END SEMESTER EXAMINATION, JUNE- 2024

B. Sc. (Hons.) in Data Science Semester – IV

Course Name: Data Warehousing and Mining **Course Code: BDS243**

Time allotted: 03 Hours **Max. Marks: 100**

Note: Read all the instructions carefully.

Section (A)

Q1. Attempt **all** questions. All carry equal marks. **(10 x 02 = 20)**

a. Distinguish between Object-Oriented Database and Object Relational Database.
b. What is private Key
c. Why we need KDD?
d. Explain online analytical processing.
e. What is the role of Meta Data in Mining?
f. Define Data integration.
g. How are data marts different from data warehousing?
h. What is star schema?
i. List the characteristics of a data warehouse
j. What is Unsupervised learning?

Section (B)

Q2. Attempt **any 5** questions. All carry equal marks **(5 x 7 = 35)**

a. Show with an example and diagrammatic illustrations the following OLAP operations in a multidimensional data model: (i) Roll-up (5) (ii) Drill-down (5) (iii) Slice and Dice
b. Define: Generalization, Normalization, Min-Max and Z-score.
c. Solve Numerical – Using K means clustering algorithm form two clusters for given data.

ID	HEIGHT	WEIGHT
1	185	72
2	170	56
3	168	60
4	179	68
5	182	72
6	188	77
7	180	71
8	180	70
9	183	84
10	180	88

d. How will you define Multidimensional data model and Aggregation?
e. A database has 4 transactions, shown below. Assuming a minimum level of support min_sup= 60% and a minimum level of confidence min_conf=80%. Fine all frequent item sets using the Apriori algorithm. Also list strong association rule.

TID	Date	Items_Bought
T100	17/07/23	{K, A, D, B}
T200	10/07/23	{D, A, C, E, B}
T300	15/07/23	{C, A, B, E}
T400	1/07/23	{B,A,D}

f. What is Data preprocessing and data cleaning, explain the ways to handle missing data during cleaning.
g. What are the different components of a data warehouse? Explain with the help of a diagram.

P.T.O./-

Section (C)

Attempt **any 3** questions. All carry equal marks. **(03x15 = 45)**

Q3.

a. Consider whether a dataset based on which we will determine whether to play football or not(using decision tree). The independent variables are Outlook, Temperature, Humidity, and Wind. Find the entropy of the class variable. Also calculate IG, Entropy and total gain of attribute Outlook. And Temperature only.

Outlook	Temperature	Humidity	Wind	Play Football(yes/no)
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	Yes
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

b. What is on-line transaction processing (OLTP)? Describe the evolution of OLTP. What are the critical features of OLTP system.

Q4.

a. Perform Agglomerative clustering on the following data and plot a dendrogram using Complete Linkage Approach. The given data indicates the distance between elements

Item	A	B	C	D	E
A	0				
B	9	0			
C	3	7	0		
D	6	5	9	0	
E	11	10	2	8	0

b. What is Data Mining? Describe the evolution of Data Mining.

Q5.

a. Write short notes on:

(i) DBSCAN

(ii) Schema

b. What is agglomerative clustering?

Q6.

a. Write seven basic applications of Data mining.

b. What is ETL? Explain the steps in ETL

Q7.

a. Explain the operation of data-cube with suitable examples

b. Write the algorithm for k-nearest neighbor classification

***

Enrol. No. | S | R | H | U | | | | | | | | | Regn. No. | D | D | | | | | | | | | |

END SEMESTER EXAMINATION, JUNE- 2024

B. Sc. (Hons.) in Data Science Semester - IV

Course Name: Applied Data Science using Python **Course Code: BDS243**

Time allotted: 03 Hours **Max. Marks: 100**

Note: Read all the instructions carefully.

Q.1. Objectives Type

(Very short) Attempt all. **(10×2=20)**

a.	What is meant by List in python?	CO1	PO1
b.	What are tuples in Python?	CO1	PO1
c.	Define lambda function with examples	CO2	PO1,4
d.	Extract the second row of a given 3x3 NumPy array as array = np.array([[1, 2, 3], [4, 5, 6], [7, 8, 9]])	CO3	PO1,2
e.	Slice a NumPy array to extract every alternate element. array = np.array([1, 2, 3, 4, 5, 6, 7, 8, 9])	CO3	PO1,2,4
f.	Differentiate between a pandas Series and a DataFrame.	CO4	PO1,2,4,5
g.	Explain the structure of a pandas DataFrame.	CO4	PO1,2,
h.	Briefly explain the purpose of Scikit-learn in Data Science.	CO5	PO1,2
i.	Name two machine learning algorithms commonly implemented in Scikit-learn.	CO5	PO1
j.	How do you drop rows with missing values from a DataFrame in pandas?	CO2	PO1

Q.2. Short notes Type (Attempt any five out of Seven) **(5×4=20)**

a.	Write a Python code snippet to demonstrate dictionary manipulation.	CO1	PO1,2
b.	Discuss the usage of map() function in Python with an example.	CO2	PO1,2
c.	Describe the process of transposing arrays and swapping axes in NumPy.	CO3	PO1,2,3
d.	Discuss the differences between Series and DataFrame in Pandas.	CO4	PO1
e.	How do you import and export data in Python using libraries such as SciKit Learn and Matplotlib?`	CO5	PO1,2
f.	Explain the process of importing and exporting data in Python using libraries such as Pandas	CO4	PO1,2
g.	Generate a random NumPy array of shape (3, 4) containing integers between 1 and 10. Calculate the mean, median, and standard deviation of the array, and print the results.	CO3	PO1,2

Q.3. Structured Question (Attempt any three out of four) **(3×10=30)**

a.	Given two NumPy arrays, array1 = np.array([1, 2, 3]) and array2 = np.array([4, 5, 6]), perform element-wise addition, subtraction, multiplication, and division between the arrays. Print the results for each operation.	CO3	PO1,2
b.	Describe the purpose of universal functions (ufuncs) in NumPy and provide examples of commonly used mathematical functions available in NumPy's ufunc library.	CO2	PO1,2,3
c.	Create a Python program that simulates a simple dictionary for storing contact information. The program should allow users to add new contacts, delete existing contacts, and display the entire contact list. Use dictionaries to represent contacts, with names as keys and contact details (e.g., phone numbers, email addresses) as values.	CO1	PO1,2,3

P.T.O./-

d.

ProductID	Quantity	Revenue	CustomerID	Region
Product3	NaN	2739.948652	8504	Region3
Product3	NaN	1679.149019	2571	Region3
Product3	NaN	882.511665	3199	Region3
Product2	NaN	1511.998497	8698	Region1
Product3	47.0	4605.004648	6212	Region2
Product2	NaN	562.870815	1181	Region1
Product1	18.0	288.524524	8069	Region3
Product1	31.0	740.447113	5023	Region1
Product3	36.0	2509.334012	5387	Region2
Product2	44.0	3657.869004	3569	Region3

CO4 | PO1,2,3

i. Define Missing Values?
ii. Write a code to Fill missing values by mean of Quantity Attribute.
iii. Write the output of print(dataset.isnull().count())
iv. Write a code to replace Region1 with Dehradun, Region2 with Rishikesh and Region3 with Haridwar.
v. Calculate descriptive statistics (mean, median and standard deviation) for Quantity and Revenue.

Q.4. Long Answer type (Attempt any two out of three) **(2×15=30)**

a.	**(i)** Discuss the use of Matplot package **(ii)** Develop line chart for the following data. years = [1950, 1960, 1970, 1980, 1990, 2000, 2010] gdp = [300.2, 543.3, 1075.9, 2862.5, 5979.6, 10289.7, 14958.3].	CO5	PO1,2,3
b.	Discuss the Following with example **(i)** List **(ii)** Tuple **(iii)** Dictionary	CO1	PO1,2,4
c.	**(i)** Describe how indexing, selection, and filtering work in Pandas DataFrames. **(ii)** Explain the concept of the Series data structure in the Pandas library. Discuss its characteristics, including index labels and data alignment, and provide examples of creating and manipulating Series objects.	CO4	PO1,2

****

13/6/24

Enrol. No. | S | R | H | U | | | | | | | | | Regn. No. | D | D | | | | | | | | | |

END SEMESTER EXAMINATION, JUNE- 2024

B. Sc. (Hons.) in Data Science Semester – IV

Course Name: Design and Analysis of Algorithms **Course Code: BDS245**

Time allotted: 03 Hours **Max. Marks: 100**

Note: Read all the instructions carefully.

Q.1. Attempt all. **(10×2=20)**

QN	Questions	CO	PO
a.	Define the basic principles of backtracking.	1	2
b.	What is pseudo-code? Explain with an example.	2	1
c.	Define the Minimum Spanning tree problem.	3	2
d.	How to calculate the efficiency of Dijkstra's algorithm?	2	2
e.	Write the running time of 0/1 knapsack problem.	2	1
f.	Compare NP-hard and NP-completeness?	2	2
g.	Explain the principles of FIFO branch and bound?	2	1
h.	Explain any one application back tracking with example?	3	2
i.	Describe job scheduling with deadlines?	2	2
j.	Define Knapsack Problem.	2	1

Q.2. Attempt any five out of Seven. **(5×4=20)**

QN	Questions	CO	PO
a.	Explain Big Omega and Theta notations? Explain what are they used for?	1	2
b.	What is an optimal Huffman code for the following set of frequencies? a : 1, b:1, c:2, d:3, e:5, f:8,g:13,h : 21	2	1
c.	Write a short note on time-space trade-off.	3	2
d.	Explain Kruskal's Algorithm?	2	2
e.	Outline the Dynamic Programming approach to solve the Optimal Binary Search Tree problem and analyze it time complexity.	2	1
f.	State Hamiltonian Circuit Problem.	2	2
g.	Compare Greedy method and Dynamic Programming.	2	1

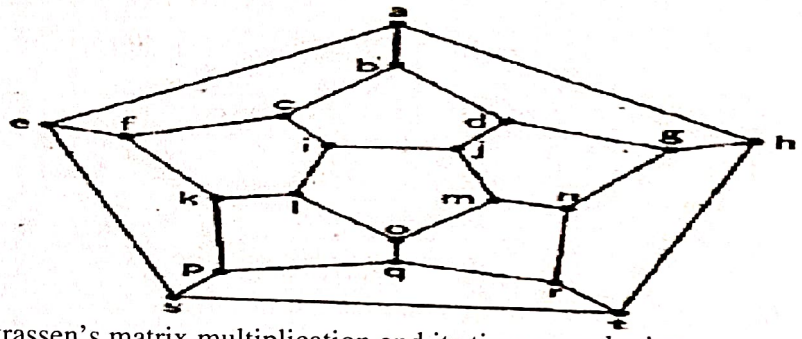
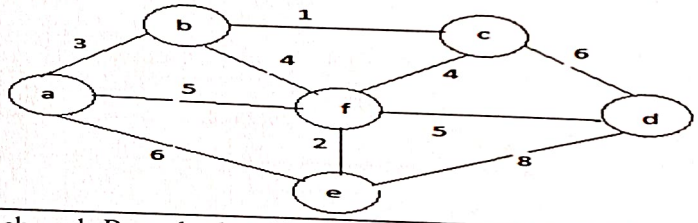
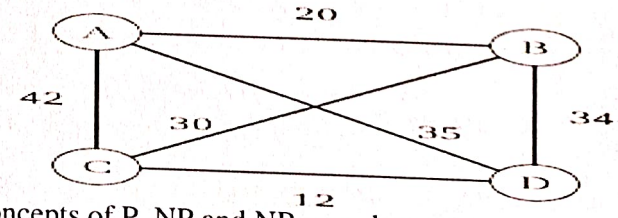
Q.3. Attempt any three out of four. **(3×10=30)**

QN	Questions	CO	PO
a.	**I.** Write and explain bellman ford Generation Algorithm? **II.** Solve the following instance of knapsack problem by branch-and-bound algorithm.	2	2

ITEM	WEIGHT	VALUE
1	4	$ 40
2	7	$42
3	5	$25
4	3	$12

The knapsack capacity W = 10

P.T.O./-

b.	I.	Find a Hamiltonian circuit or disprove its existence in the graph given below. 	2	
	II.	Explain Strassen's matrix multiplication and its time complexity		
c.	I.	Write an algorithm for quick sort and write its time complexity with example list are 5,3,1,9,8,2,4,7	3	2
	II.	Discuss about the algorithm and pseudocode to find the Minimum Spanning Tree using Prim's Algorithm. Find the Minimum Spanning Tree for the graph shown below. 		
d.	I.	Apply Branch and Bound algorithm to solve the travelling sales man problem for 	3	2
	II.	Explain the concepts of P, NP and NP completeness.		
	III.	Illustrate Merge sort algorithm and discuss its time complexity.		

Q.4. Attempt any two out of three. **(2×15=30)**

QN		Questions	CO	PO
a.	I.	Explain General method of Greedy method. Find the greedy solution for following job sequencing with deadlines problem n = 7, (p1,p2,p3,p4,p5,p6,p7) = (3,5,20,18,1,6,30), (d1,d2,d3,d4,…,d7) = (1,3,4,3,2,1,2)	2	1
	II.	Find an optimal solution to the knapsack instance n=7 objects and the capacity of knapsack m=15. The profits and weights of the objects are (P1,P2,P3, P4, P5, P6, P7)= (10, 5,15,7,6,18,3) (W1,W2,W3,W4,W5,W6,W7)= (2,3,5,7,1,4,1).		
b.	I.	Show the result of inserting the following items in an initially empty B- tree of order 5. (25,31,37,76,5,60,38,08,30,15,35,17,23,53,27,43,65,48)and Explain and compare best and worst time complexity of Quick Sort.	3	2
	II.			
	III.	What is a Spanning tree? Explain Prim's Minimum cost spanning tree algorithm with suitable example and also find the time complexity		
c.	I.	Insert the following keys into an initially empty red black tree. (4,7,12,15,3,5,14,18,16).	3	2
	II.	Explain the process of rebalancing an AVL tree after performing an insertion or deletion operation with algorithms.		

***

Enrol. No. | S | R | H | U | | | | | | | | | Regn. No. | D | D | | | | | | | | | |

END SEMESTER EXAMINATION, JUNE- 2024

B. Tech. (CSE) Semester – VI/ B. Sc. (Hons.) in Data Science Semester - IV

Course Name: Artificial Intelligence/ Introduction to Artificial Intelligence

Course Code: CCS361/BDS242

Time allotted: 03 Hours

Max. Marks: 100

Note: Read all the instructions carefully.

Q.1. Objectives Type **(10 X 2=20)**

(Very short/ Short Answer, MCQ, One-liners, Fill in the Blank, True/ False) Attempt all.

		CO	PO
a.	What is a heuristic function in AI search algorithms?	1	1
b.	What is depth-first search, and when is it useful?	1	2
c.	Parsing mean.....	2	1
d.	Regular grammar is Stronger than CFG. True/False	2	1
e.	Prolog is an OOP type of language. True/False	3	2
f.	Zero order logic is also known as.....	3	1
g.	Why Fuzzy sets are required?	4	2
h.	Fuzzy sets are used to represent Boolean logic. True/False	4	2
i.	What is speech recognition?	5	1
j.	Write one application of machine learning	5	1

Q.2. Short notes / Short Answer Type (Attempt any five out of Seven) **(5 X 4=20)**

a.	How an 8-Puzzle problem can be solved using informed search. Show with an example.	1	1
b.	Explain Bottom-up Parsing with the help of an example	2	2
c.	What are the advantages and disadvantage of logical representation of knowledge?	3	1
d.	How union operation is performed in Fuzzy set? Show with the help of an example.	3	1
e.	What is a pattern recognition? Where it can be used?	5	2
f.	Write the differences between top-down parsing and bottom-up parsing.	2	2
g.	What are the problems associated with Hill-climbing algorithm? How they can be solved?	1	3

P.T.O./-

Q.3. Structured Question (Attempt any three out of four) **(3 X 10=30)**

a.	(i) Write three application of pattern recognition. (ii) Define Training and testing data? (iii) How many Phases are required in Pattern Recognition System?	5	2
b.	(i) Write the two differences between Informed and Uninformed search (ii) Explain the Heuristic function (iii) Who is considered as father of AI?	1	1
c.	(i) Why first order logic is required? (ii) Elaborate the Case grammar? (iii) How forward chaining is used?	3	2
d.	(i) What is Inference engine? (ii) Differentiate between digital controller and fuzzy controller (iii) Name any two-membership function	3	2

Q.4. Long Answer / Essay type (Attempt any two out of three) **(2 X 15=30)**

a.	(i) Compare DFS with BFS. (ii) What is minimax algorithm? (iii) Why pruning is required in it	1	2
b.	(i), Why Fillmore grammar is required? (ii) Explain about Transition nets in detail.	3	2
c.	(i) Write two advantages and disadvantages of Fuzzy logic.? (ii) Explain the use of fuzzy logic taking the example of a real-life scenario?	4	1

****

Enrol. No. | S | R | H | U | | | | | | | | | Regn. No. | D | D | | | | | | | | | |

END SEMESTER EXAMINATION, JUNE- 2024

B. Sc. (Hons.) in Data Science Semester - VI

Course Name: Data Visualization Techniques **Course Code: DDS362**

Time allotted: 03 Hours **Max. Marks: 100**

Note: Read all the instructions carefully.

Q.1. Attempt all. **(10×2=20)**

QN	Questions	CO	PO
a.	Why is Syntactic Processing Important in NLP?	CO3	PO1
b.	Define Non-spatial data.	CO3	PO1
c.	How do we use shape of an object in visualization?	CO1	PO1
d.	What is color hue?	CO1	PO1
e.	Define Proximity in data visualization.	CO1	PO1 ,3
f.	What are the disadvantages of data visualization?	CO1, 4	PO1
g.	Why it's important to avoid information overload?	CO3	PO1
h.	Define Iconic memory in perception.	CO2	PO1
i.	What do mean by continuous Time Series Data?	CO3	PO2
j.	What do you mean by poor chart design?	CO2	PO1

Q.2. Attempt any five out of Seven. **(5×4=20)**

QN	Questions	CO	PO
a.	How to deal with the missing data in data cleaning process? Explain	CO2	PO1 ,2
b.	Explain about any four Keys factors of Data Visualization.	CO4	PO1
c.	Write about following charting Primitives: 1) Data Points 2) Bar Charts	CO2	PO1
d.	What are Infographics? Explain in detail.	CO4	PO1
e.	Explain about relating data table to a chart in details.	CO2	PO1
f.	What are the key properties of a data set?	CO1	PO1
g.	Discuss about GIS Data Visualization.	CO5	PO1

P.T.O./-

Q.3. Attempt any three out of four. **(3×10=30)**

QN	Questions	CO	PO
a.	Explain the following terms – **(5+5)**		
	(i) What is the relation between data analysis and visualization? Explain.	CO1, 2	PO1
	(ii) How to choose the right visualization for your data? Explain.	CO1, 4	PO1 ,2,3
b.	Attempt all. **(5+5)**		
	(i) Briefly explain about the line plot and scatter plot.	CO1	PO1
		CO2	PO1
	(ii) What are two advantages and two drawbacks of the tree map method of visualization?	CO3	PO1
c.	Discuss about different types of data visualization. **(10)**	CO1	PO1, 2,3,1 2
d.	Attempt all. **(5+5)**		
	(i) What are the main processes that can be applied on vector data?	CO5	PO1 ,2
	(ii) What is meant by spatial data or Geographic data?	CO5	PO1

Q.4. Attempt any two out of three. **(2×15=30)**

QN	Questions	CO	PO
a.	(i) Define Data Visualization. What are the advantages of data visualization in comparison to old methods?	CO1, 4	PO1
	(ii) How do you visualize graph-based visualization, Explain with suitable example?	CO4	PO1
b.	(i) What is a Scatterplot? What kind of data is it used to represent? Explain.	CO1	PO1 ,2,3
	(ii) What perceptual properties are best for preattentive processing?	CO2	PO1 ,3
c.	(i) Briefly Explain about real time issues in data gathering and preparation with example.	CO2	PO1
	(ii) What are the major GIS functions? Explain.	CO5	PO2 ,3

***

Enrol. No. | S | R | H | U | | | | | | | | | Regn. No. | D | D | | | | | | | | | |

END SEMESTER EXAMINATION, JUNE- 2024

B. Tech. (CSE) Semester – VIII/B. Sc. (Hons.) in Data Science Semester – VI

Course Name: Entrepreneurship Development **Course Code: OCS483/GDS362**

Time allotted: 03 Hours **Max. Marks: 100**

Note: Read all the instructions carefully.

Q.1	VERY SHORT ANSWER QUESTIONS (write a short note)– *Attempt All.* (10x2 = 20 Marks)	
a.	Creativity	CO1
b.	Slack	CO1
c.	Sole Proprietorship	CO1
d.	Critical activities	CO1
e.	Industry attractiveness	CO1
f.	Product focus	CO1
g.	Trend analysis	CO1
h.	Problem solving	CO1
i.	Trends	CO1
j.	Innovation	CO1

Q.2	SHORT ANSWER QUESTIONS – *Attempt any Five.* (5x4 = 20 Marks)	
a.	What are major ways of identifying opportunities for new business idea?	CO1
b.	Elaborate on characteristics of successful entrepreneurs.	CO2
c.	How does entrepreneurs impact society in general?	CO2
d.	How entrepreneur perform competitive analysis?	CO2
e.	For following project cashflow given determine the NPV for the discount rate of 10%.	CO3

Option	Investment (Rs.)	Year 1	Year 2	Year 3	Year 4
A	350,000	50,000	90,000	120,000	60,000

f.	What are various government policies that assist for entrepreneur development?	CO5
g.	What are major factors of failure of entrepreneurship?	CO5

Q.3	STRUCTURED QUESTIONS – *Attempt any Three.* (3x10 = 30 Marks)	
a.	What various business models an entrepreneurs can apply for new business?	CO4
b.	Explain the term entrepreneurship and elaborate more on entrepreneurship development process.	CO2
c.	What are various legal ways of forming new business?	CO3
d.	Mr. Harsh has started a pilot project of renting bicycles in area near university. He had developed miniature cycle tracking and prepaid device to control the usage of bicycle. From last six months the feedback from user is positive and Mr. Harsh is willing to expand his business at other nearby university but is worried about funds for expansion. Assist him regarding what are various ways of arranging finance for his enterprise.	CO3

P.T.O./-

Q.4	Case Analysis/ Situation Based QUESTIONS – *Attempt any Two.* (2x15 = 30 Marks)	
a	**Automated Order-picking system at Coleman Inc.** R. C. Coleman distributes a variety of food products that are sold through grocery store and supermarket outlets. The company receives orders directly from the individual outlets, with a typical order requesting the delivery of several cases of anywhere from 20 to 50 different products. Under the company's current warehouse operation, warehouse clerks dispatch order- picking personnel to fill each order and have the goods moved to the warehouse shipping area. Because of the high labor costs and relatively low productivity of hand order-picking, management has decided to automate the warehouse operation by installing a computer-controlled order-picking system, along with a conveyor system for moving goods from storage to the warehouse shipping area. First the team has to determine **equipment needs** according to the space available and material nature to be handled. This activity might take on an average 1 week to complete through a consultant. Depending on equipment needed the **proposals** will obtained from various vendors. The proposal gathering will take 2 weeks. After proposal evaluation and **negotiation** process vendor will be **selected** for the ordering system. Due to negotiation process the vendor selection might take next 3 weeks to complete. As per vendor analysis the **order** will be placed for the **computer systems and equipment** and its will take 4 weeks to receive the system. After vendor selection, vendor simultaneously works on the **design of the new warehouse layout** (1 week) design of computer interface (2 weeks). After designing the warehouse, the complete **warehouse will be rearranged** and remodeled to equip the automated ordering system. This will take around 5 weeks. As soon as system is received, completion of warehouse design, and computer interface design, the vendor starts working on **installing and synchronizing the complete system** (3 weeks). The testing department of vendor then **conduct testing** on system (1 weeks). After successful testing, the **operators are trained** on various aspects and modules of the system (2 weeks) and then system is handed over to company. **I.** Develop precedence relationship table with duration. (Assign Alphabets to the activities). **II.** Draw a network diagram for this project. Identify all the paths through the network diagram. **III.** Also develop the Earliest and Latest schedule for the activities. Determine the critical path and project completion time. **IV.** Write a small report on schedules of the activities, when to be started earliest, by how much the activity can be delayed.	CO 2
b	Miss. Neha has an idea for a unique android application for tourist which will provide detail weather reports and current situation of most visited places. Explain how should she perform the fesibility analysis of her idea	CO4
c	Two graduates form SRHU are planning to implement their idea for a unique android application for tourist which will provide detail weather reports and current situation of most visited places in Uttarakhand. Develop detailed Business plan for their idea.	CO5

***

Enrol. No. | S | R | H | U | | | | | | | | | Regn. No. | D | D | | | | | | | | | |

END SEMESTER EXAMINATION, JUNE- 2024

B. Sc. (Hons.) in Data Science Semester – VI

Course Name: Introduction to Deep Learning **Course Code: BDS361**

Time allotted: 03 Hours **Max. Marks: 100**

Note: Read all the instructions carefully.

Q.1. Attempt all. **(10×2=20)**

S.No	Questions	CO	PO
a.	What is deep learning?	1	1
b.	What is an artificial neural network (ANN)?	4	4
c.	Define "gradient descent."	1	2
d.	What is a convolutional neural network (CNN)?	2	4
e.	Explain the term "activation function."	1	3
f.	What is the role of regularization in deep learning?	2	2
g.	What is the vanishing gradient problem?	1	3
h.	What is backpropagation?	4	1
i.	Define "dropout" in deep learning.	1	12
j.	What is transfer learning?	3	1

Q.2. Attempt any five out of Seven. **(5×4=20)**

S.No	Questions	CO	PO
a.	What is the role of activation functions in deep learning, and how do they contribute to model non-linearity?	1	1
b.	Explain the concept of "vanishing gradients" in deep learning, and what strategies can be used to mitigate this issue?	2	2
c.	Describe the concept of "transfer learning" in deep learning and provide examples of its applications.	1	3
d.	What are the advantages and disadvantages of using convolutional neural networks (CNNs) compared to fully connected neural networks in image recognition tasks?	5	2
e.	Discuss the importance of hyperparameter tuning in deep learning models and some common techniques used for this purpose.	5	1
f.	How do autoencoders differ from other types of neural networks, and what are their applications in unsupervised learning?	4	3
g.	Explain the concept of "dropout" regularization in deep learning and how it helps prevent overfitting during training.	1	2

P.T.O./-

Q.3. Attempt any three out of four. **(3×10=30)**

S.No	Questions	CO	PO
a.	Explore the challenges and advancements in continual learning or lifelong learning in deep learning, including methods for adapting models to new tasks, concepts, and environments over time, and discuss their potential applications in dynamic and evolving scenarios.	1	1
b.	Discuss the principles and applications of graph neural networks (GNNs) in deep learning, including their ability to model relationships and dependencies in graph-structured data such as social networks, molecular structures, and knowledge graphs.	1	5
c.	Evaluate the role of meta-reinforcement learning in enabling agents to learn how to learn, including its applications in fast adaptation to novel tasks, environments, and reward structures, and discuss its potential for building flexible and adaptive autonomous systems.	2	12
d.	Describe the principles and applications of deep learning in medical imaging, including the development of deep convolutional neural networks (CNNs) for tasks such as disease diagnosis, medical image segmentation, and treatment planning, and discuss the challenges and opportunities in translating research advancements into clinical practice.	5	3

Q.4. Attempt any two out of three. **(2×15=30)**

S.No	Questions	CO	PO
a.	Explore the intersection of deep learning and robotics, including the development of deep reinforcement learning algorithms for robot control, manipulation, and navigation tasks, and discuss the challenges and opportunities in deploying deep learning models in real-world robotic systems.	1	1
b.	Discuss the principles and applications of deep learning in natural language understanding (NLU), including the development of transformer-based models such as BERT and GPT, and their applications in tasks such as language translation, sentiment analysis, and conversational AI.	1	2
c.	Evaluate the role of unsupervised learning in deep generative modeling, including the development of variational autoencoders (VAEs), generative adversarial networks (GANs), and flow-based models for tasks such as image generation, style transfer, and data augmentation, and discuss the challenges and opportunities in designing and training generative models.	5	3

***

Enrol. No. | S | R | H | U | | | | | | | | | Regn. No. | D | D | | | | | | | | | | |

END SEMESTER EXAMINATION, JUNE- 2024

B. Sc. (Hons.) in Data Science Semester - VI

Course Name: Business Intelligence Tools and Applications **Course Code: BDS362**

Time allotted: 03 Hours **Max. Marks: 100**

Note: Read all the instructions carefully.

Q.1. Attempt all. **(10×2=20)**

QN	Questions	CO	PO
a.	What is the difference between dimensions and measures?	2	1
b.	What is the use of characterization?	2	1
c.	Why do we use training data set in classification?	2	1
d.	What do you mean by classification and prediction?	2	1
e.	What do you mean by raw data?	1	1
f.	What is the role of data transformation in data pre-processing?	3	1
g.	Explain the term data discretization.	3	1
h.	What is the need of BI?	1	1
i.	What do you mean by analytics process?	4	1
j.	Describe meta data in your words.	5	1

Q.2. Attempt any five out of Seven. **(5×4=20)**

QN	Questions	CO	PO
a.	How do you remove the redundancy? Explain in your words.	3	1,2
b.	What are the advantages of data warehouse development?	2	1,2
c.	What do you mean by association rule mining?	2	1,2, 3
d.	What is the concept of classification? How is it different from clustering?	2	1,2
e.	Explain data, information and knowledge. Also explain the role of mathematical models in BI.	1	1,2, 3,4
f.	Differentiate operational intelligence and business intelligence.	5	1,2, 3
g.	Draw the Business Intelligence Architecture and explain its components.	1	1,2, 3

P.T.O./-

Q.3. Attempt any three out of four. **(3×10=30)**

QN	Questions	CO	PO
a.	Define the following terms: • **Star Schema** • **Fact Constellations**	2	1,2
b.	Differentiate ROLAP and MOLAP with the help of architecture.	2	1,2, 3
c.	Explain multitier architecture of data warehouse.	2,3	1,2, 3
d.	What do you mean by data analytics process? Explain the following phases of analytics process: • Data Requirements Specification • Data Collection • Data Processing • Communication	4	1,2, 3

Q.4. Attempt any two out of three. **(2×15=30)**

QN	Questions	CO	PO
a.	i. What is the difference between Online Analytical Processing and Online Transactional Processing? ii. What is the need of data mining in today's world? Explain in your words.	2,3	1,2
b.	Explain the following terms: • Business Process Management • Complex Event Processing	5	1,2
c.	Describe the term by business activity monitoring? How does BAM work? Explain with a suitable diagram.	5	1,2, 3,4

***