

HARCOURT BUTLER TECHNICAL UNIVERSITY, KANPUR

B.Tech(CS/IT)

End Semester Examination

Odd Semester (V), 2022-23

ECS-359: DATA SCIENCE

Time: 2:30 Hours

Max. Marks: 50

Note: 1. Attempt all questions. All questions carry marks as shown against them. Use the fact that normal curve is symmetrical to calculate areas for negative Z score values.

Please mention all the Course Outcomes (CO) in statement form

1. Understand the core concepts and methods in data science.
2. Understand the issues and challenges in data collection, storage and management.
3. Understand and Apply various techniques for data analysis.
4. Understand various data visualization techniques.
5. Understand Learn Python programming tools for data science.

		Related CO	Marks
1.	Attempt all parts:		
	a) Explain any two examples of applications of Data Science in real life.	CO1	(05)
	b) What is Z-score? How is it significant?	CO1	(05)
2.	Attempt all parts:		
	a) Explain 2 types of errors in Hypothesis testing.	CO2	(05)
	b) What is dirty data? How can we clean it?	CO2	(05)
3.	Attempt all parts:		
	a) 85% of students in JEE Mains prefer CSE branch. If we sample 80 students, what is the probability that more than 60 will opt for CSE?	CO3	(05)
	b) A report stated that the mean cost of oil change is Rs. 53. You don't think this is correct; hence you sample 36 random values and find the mean value to be Rs. 49, with std. deviation 12 in the past. Is there enough evidence to support your claim for a level of significance = 0.05?	CO3	(05)
4.	Attempt all parts:		
	a) What is exploratory analysis? How can scatterplots help in understanding correlation for regression?	CO4	(04)
	b) Explain with the help of diagrams, clearly showing the acceptance and fail to accept regions for null hypothesis: • Two tail test • Left tail test	CO4	(06)

5 Answer all parts:

- a) Given `myvar = 'hello'`, how would you return `myvar` in uppercase?
- b) Consider the string `NAME="ABCDE"`, what is the result of `NAME.find("B")`
- c) What is the result of '`str(1) +str(2)`'?
- d) Consider the dataframe `df`, how would you access the element in the 2nd row and 1st column?
- e) What function will you use to load a `.csv` file in Pandas?
- f) What is the output after the following code is run?

```
X=np.array([[1,0,1],[2,2,32]])
Out=X[0:2,2]
Out
```
- g) If we have 10 columns and 100 samples, how large is the output of `df.corr()`?
- h) Which method provides summary statistics of a dataframe `df`?
- i) What is the minimum possible value of Pearson coefficient?
- j) What does the vertical axis of scatterplot represent: Dependent or Independent variable?

HARCOURT BUTLER TECHNICAL UNIVERSITY, KANPUR

III B.Tech. (CSE/IT)

END SEMESTER EXAMINATION

ODD SEMESTER (V), 2022-23

ECS 355: Design & Analysis of Algorithms

Time: 2:30 Hours

Max. Marks: 50

Note: Attempt all questions. All questions carry marks, as shown against them.

Course Outcomes (CO)

1. Understand and apply mathematical preliminaries to the analysis and design stages of different types of algorithms. (Understand, Apply)
2. Analyze worst-case time complexity of various algorithms using asymptotic methods. (Analyze)
3. Understand and apply the divide-and-conquer paradigm and synthesize divide-and-conquer algorithms on problems of Sorting, Searching, finding MST etc. (Understand, Apply)
4. Describe the greedy paradigm and explain when an algorithmic design situation calls for it. For a given problem develop the greedy algorithms. (Apply, Analyze)
5. Apply the dynamic-programming paradigm to model engineering problems using graph and write the corresponding algorithm to solve the problems. (Apply)
6. Explain the ways to analyze randomized and approximation algorithms (Apply, Analyze)

Related Course Outcome (CO)	Marks
-----------------------------	-------

Q. No. 1: Attempt all parts.

- | | | |
|--|----|---|
| (a) Discuss the need and characteristics of analysis of an algorithm? | 01 | 2 |
| (b) Differentiate between Big-oh and omega notation with example. | 01 | 2 |
| (c) Describe the Algorithm Analysis of Binary Search. | 02 | 2 |
| (d) Using master theorem find asymptotic bound for the recurrence-
$T(n)=2T(n/2) + n^3$ | 02 | 2 |

Q. No.2: Attempt all parts.

- | | | |
|--|----|---|
| (a) Create a B-Tree of order 5 from the list of elements [30,20,35,95,15,60,55,25,5,65,70,10,40,50,80,45]. | 03 | 4 |
| (b) Discuss the Amortized analysis with an example. | 03 | 4 |

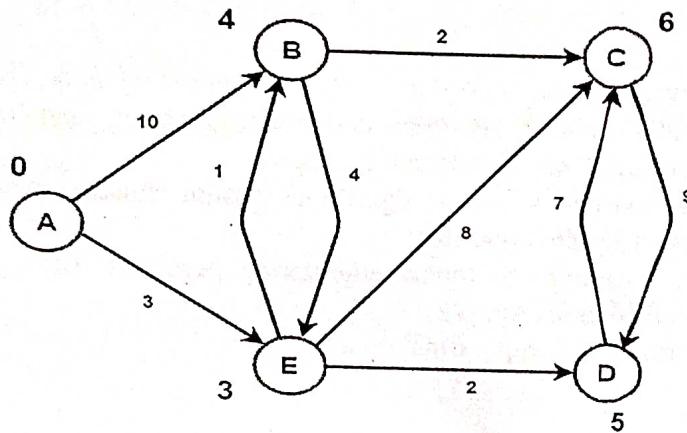
Q. No. 3: Attempt all parts.

- | | | |
|---|----|---|
| (a) Define Greedy approach of problem solving with the help of any suitable standard example. | 04 | 4 |
| (b) A thief enters a house for robbing it. He can carry a maximal weight of 5 kg into his bag. There are 4 items in the house with the following weights and values. What items should thief take if he either takes the item completely or leaves it completely? | 05 | 4 |

Item	Weight (kg)	Value (\$)
Mirror	2	3
Silver nugget	3	4
Painting	4	5
Vase	5	6

Q. No. 4: Attempt all parts.

- (a) Explain Prim's Minimum cost spanning tree algorithm with suitable example. **04** **4**
 (b) Find the single-source shortest path considering the source vertex as A. **05** **4**



Q. No. 5: Attempt all parts.

- (a) Write Short notes on the following: **06** **04**
 (i) PRAM Algorithms
 (ii) Approximation Algorithms
 (b) Explain NP -Hard and NP-Complete problems with the help of suitable examples **06** **04**

Q. No. 6: Attempt all parts.

- (a) Explain Merge sort algorithm and derive the time complexity of this algorithm. **03** **05**
 OR

Write the algorithm to find the K-th minimum element from a list of elements.

- (b) Explain Heap sort algorithm and derive the time complexity of this algorithm **03** **05**

HARCOURT BUTLER TECHNICAL UNIVERSITY, KANPUR

B.Tech (CS/IT)

End Semester Examination

Odd Semester (V), 2022-23

ECS-353: Database Management System

Time: 2:30 Hours

Max. Marks: 50

Note: *I. Attempt all questions. All questions carry marks, as shown against them.*

Course Outcomes (CO)

1. Understand and Develop Entity Relationship (ER) and Relational Models for a given application.
2. Develop and manipulate relational database using Structured Query Language and relational languages.
3. Develop a normalized database for a given application by incorporating various constraints like integrity and value constraints.
4. Understand and apply transaction processing concepts and convert schedules to serializable schedules.
5. Illustrate different concurrency control mechanisms to preserve data consistency in a multi-user environment.

Related Course Outcome (CO)	Marks
-----------------------------	-------

Q. No. 1: Attempt both questions.

- (a) What are different types of anomalies associated with database? Define constraints and its types in DBMS. CO1 (04)
- (b) Draw an E-R diagram for hospital with a set of patients and a set of medical doctors, with each patient a log of the various conducted tests is also associated. CO1 (04)

OR

Discuss three level of abstraction or schemas architecture of DBMS.

Q. No. 2: Attempt both questions.

- (a) Write the difference between cross join, natural join, left outer join and right outer join with suitable example. CO2 (04)
- (b) Define the terms aggregation and generalisation with example. CO2 (04)

Q. No. 3: Attempt both questions.

- (a) Why do we normalise database? Explain BCNF and 3NF with examples. CO3 (04)
- (b) What do you mean by decomposition of a relation? Consider the relational scheme R(A,B,C,D,E,F) and FDs $A \rightarrow BC$, $C \rightarrow A$, $D \rightarrow E$, $F \rightarrow A$, $E \rightarrow D$. Is the decomposition of R into R1 (A, C, D), R2 (B, C, D) and R3 (E, F, D) lossless?

Q. No. 4: Attempt both questions.

- (a) What do you mean by serializability? Discuss the conflict and view serializability CO4 (04)

with example.

- (b) What is deadlock? What are necessary conditions for it? How it can be detected and recovered? CO4 (04)

Q. No. 5: Attempt both questions.

- (a) What do you mean by time stamping protocols for concurrency control? Discuss multi-version scheme of concurrency control also. CO5 (04)
- (b) Explain two phase locking protocol with suitable example. CO4 (04)

Q. No. 6: Attempt all questions.

- (a) Consider the following relation. The Primary key is Roll no, ISBN, Student (Roll No, Name, Branch), Book (ISBN, Title, Author, Publisher) Issue (Rollno, ISBN, date_of_issue). Write the query in SQL of the following
- List the Roll Number and Name of All CSE Branch Student.
 - Find the name of students who have issued a book of publication 'BPB'.
 - List the title and author of all books which are issued by a student name started with a.
 - List the title of all books issued on or before 20/09/2012.
 - List the name of student who will read the book of author named 'Sanjeev'.

- (b) Write SQL query for the given table. CO2 (05)

Given the relational schema:

ENROLL (S#, C#, Section), S# is the student number, TEACH (Prof, C#, Section), C# is the course number. ADVISE (Prof, S#), Prof is Thesis advisor of S#. PRE_REQUEST(C#, pre-C#), pre-C# is prerequisite course. GRADE (S#, C#, grade, year), STUDENT (S#, Sname), Sname is the student name.

Give SQL queries for the following:

- List of students taking courses with Smith or Jones.
- List all students taking at least one course that their advisor teaches.
- List those professors who teach more than one section of the same course.
- List all student number and course number.
- List the student number and course number who got grade A.

HARCOURT BUTLER TECHNICAL UNIVERSITY
End-Semester Examination (odd semester 2022-23)
III-B. Tech (IT, CS, ME, EE, ET, CE)
Operations Research (BMA-351)

Time: 2.30 Hrs.

Max. Marks: 50

Note: 1. Attempt all questions.
2. All questions carry marks, as shown against them.

1. (a) Give a general mathematical formulation of L.P. problem and describe the CO1 degeneracy in L.P. problem. 5

Use penalty (Big-M) method to solve the following LPP.

$$\text{Max } z = 5x + 4y$$

$$\text{Subject to: } 10x + 4y \geq 15$$

$$5x + 8y \geq 10,$$

$$\text{and } x, y \geq 0$$

- (b) An advertising agency wishes to reach two types of audiences- customers with annual income greater than one lakh rupees (target audience A) and customers with annual income of less than one lakh rupees (target audience B). The total advertising budget is Rs. 2,00,000. One programme of TV advertising costs Rs. 50,000; one programme of radio advertising costs Rs. 20,000. For contract reasons, at least three programmes ought to be on TV and the number of radio programmes must be limited to 5. Surveys indicate that a single TV programme reaches 4,50,000 prospective customers in target audience A and 50,000 in target audience B. One radio programme reaches 20,000 prospective customers in target audience A and 80,000 in target audience B. 2.5

Use graphical method to determine the media mix so as to maximize the total reach.

- (c) Write the computational procedure of simplex method to solve an LPP and give a flow chart of the simplex algorithm. 2.5

OR

Explain the duality theory of the Linear Programming and prove that the dual of the dual of a primal is primal itself.

2. (a) A manufacturer wants to ship 22 loads of his product as shown below. The matrix CO2 gives the kilometers from sources of supply to the destinations. Find the initial basic feasible solution of the following T.P. by using Vogel's Method and check the optimality of the solution by MODI method. 5
The shipping cost is Rs. 10 per load per kilometer. What shipping schedule should be used in order to minimize the total transportation cost?

Destination → Source ↓	D ₁	D ₂	D ₃	D ₄	D ₅	Capacity
S ₁	5	8	6	6	3	8
S ₂	4	7	7	6	5	5
S ₃	8	4	6	6	4	9
Requirement	4	4	5	4	8	

- (b) Describe the mathematical formulation of an Assignment Problem.

2.5

A manager has four salesmen and four sales districts. He estimates that the profit per day from each salesman in each district would be as follows (in Rs.). Find the assignment of salesman to district that will result in maximum profit.

Districts → Salesman ↓	D ₁	D ₂	D ₃	D ₄
M ₁	33	21	35	23
M ₂	29	23	37	31
M ₃	31	31	33	29
M ₄	27	25	35	37

2.5

- (c) Explain all Integer Programming Problem. Describe Gomory's Cutting Plane method to solve the IPP and its use.

OR

Use Branch-and-Bound technique to solve the integer programming problem.

$$\text{Max } z = 7x_1 + 9x_2$$

$$\begin{aligned} \text{Subject to: } & -x_1 + 3x_2 \leq 6 \\ & 7x_1 + x_2 \leq 35, \\ & \text{and } 0 \leq x_1, x_2 \leq 7 \end{aligned}$$

3. (a) What is PERT? Define optimistic time, pessimistic time, most likely time and CO3 explain how you will estimate the expected time to complete the activity in PERT technique. Differentiate between CPM and PERT.

5

An established company has decided to add a new product to its line. It will buy the product from a manufacturing concern, package it, and sell it to a number of distributors that have been selected on a geographical basis. Market research has already indicated the volume expected and the size of sales force required. The steps

shown in the following table are to be planned.

Activity	Description	Predecessors	Duration(days)
A	Organize sales office	-	6
B	Hire salesmen	A	4
C	Train salesmen	B	7
D	Select advertising agency	A	2
E	Plan advertising campaign	D	4
F	Conduct advertising campaign	E	10
G	Design package	-	2
H	Setup packaging facilities	G	10
I	Package initial stocks	J, H	6
J	Order stock from manufacturer	-	13
K	Select distributors	A	9
L	Sell to distributors	C, K	3
M	Ship stocks to distributors	I, L	5

Draw the network diagram and calculate the expected project completion time.
Find the critical path of the project.

Calculate EST, LST, EFT, LFT and floats for each activity of the project.

- (b) What do you understand by sequencing problem? Describe the Johnson's algorithm to solve the problem of processing n jobs through two machines. 5
 Find the optimal sequence of jobs which minimizes the total elapsed time based on the following information and find total elapsed time. The mode of operation being ABC for each job.

Job	A	B	C
1	3	3	5
2	8	4	8
3	7	2	10
4	5	1	7
5	2	5	6

4. (a) Describe the replacement problem and explain the various factors which are responsible to replace the equipment although it may be running. CO4 5

Determine the optimal replacement policy for the data given below:

- (i) Group replacement cost Rs. 20 per unit.
- (ii) Cost of individual replacement of failure is Rs. 90 per unit.

- (iii) Total number of units in a system are 1000.
 (iv) Mortality data of units to be used in the system.

Interval of time period(hours)	Probability of failure
0-200	0.00
201-400	0.06
401-600	0.30
601-800	0.48
801-1000	0.16

- (b) What is inventory? Describe advantages and disadvantages of increased inventory.
 Derive the optimum economic lot size (EOQ) formula for the deterministic model with shortage (fixed time model), infinite rate of production and uniform demand rate.

5

A factory requires 1,500 units of an item per month, each costing Rs. 27. The cost per order is Rs. 150 and the inventory carrying charges work out to 20% of the average inventory. Find the economic order quantity and the number of orders per year. Would you accept a 2% discount on a minimum supply quantity of 1,200 units? Compare the total costs in both the cases.

5. (a) Use the principle of optimality to find the minimum value of

CO5 5

$$\begin{aligned} \text{Min } z &= y_1^2 + y_2^2 + \dots + y_n^2 \\ \text{when } y_1, y_2, \dots, y_n &= c \\ \text{and } y_j &> 0; j = 1, 2, \dots, n \end{aligned}$$

- (b) Explain the concept of dynamic programming problem and state the principle of optimality in dynamic programming. Discuss the relation between linear programming and dynamic programming.

5

Date of showing evaluated answer books: 10.1.2023

No. of Printed Pages: 02

Roll No.

HARCOURT BUTLER TECHNICAL UNIVERSITY, KANPUR

B. Tech (CSE / IT)

END SEMESTER EXAMINATION

ODD SEMESTER (V), 2022-23

ECS 351: COMPUTER NETWORKS**Time: 2:30 Hours****Max. Marks: 50**

- Note: 1. Attempt all questions. All questions carry marks, as shown against them.
2. Q.No. 6 is from the lab component of the subject.

Please mention all the Course Outcomes (CO) in statement form

1. Explain the functions of the different layer of the OSI Protocol. (Understand)
2. Design of wide-area networks (WANs), local area networks (LANs) and Wireless LANs (WLANS) based on available network devices. (Apply, Analyze)
3. Develop network programming application for a given problem related to TCP/IP protocol stack. (Apply, Analyse)
4. Understand and analyze different routing algorithms. (Understand, Analyze)
5. Understand the use of IP addressing schemes as per IPV4 and IPV6. (Understand)
6. Modify the existing protocols of TCP/IP protocol stack for performance improvement. (Apply, Analyze)

		Related Course Outco me (CO)	Marks
Q. No. 1:	Attempts all parts		
(a)	Explain the functions of Data link layer, Network layer, Transport and Application layer of TCP/IP protocol suite.	01	04
(b)	Compare and contrast circuit switching and packet switching techniques. What is the number of cable required for n devices connected in mesh, ring, bus and star topology?	01	04
Q. No.2:	Attempts all parts		
(a)	Explain ALOHA and Pure ALOHA in detail. Why is the channel throughput doubled in slotted ALOHA compared to pure ALOHA?	02	04
(b)	What do you mean by flow control in data link layer? Describe GO back N ARQ protocol.	02	04
Q. No. 3:	Attempts all parts		
(a)	Discuss the Distance vector routing algorithm with an example and explain the count to infinity problem in detail.	04	04
(b)	Explain the following terms with example (a) Multicast addressing (b) Unicast addressing. A block of IP addresses is granted to a small organization. We know that one of the addresses is 205.16.37.39/28. What are the first and last addresses in the block?	05	04

(OR)

What are the main causes of congestion? Describe congestion control techniques used at network layer?

- Q. No. 4:** Attempts all parts
- (a) Explain the term QoS (Quality of services) primitives looked at transport layer and Discuss three way handshake techniques in detail. **05** **04**
- (b) Compare and contrast UDP and TCP. How the performance of TCP protocol can be improved? **05** **04**
- Q. No. 5:** Attempts all parts
- (a) What is domain name system? How does it work? Explain typical resolution process in DNS in detail. **06** **04**
- (b) Elaborate about the working of TELNET and SMTP (Simple Mail Transfer Protocol) protocols. **06** **04**
- Q. No. 6:** Attempts all parts
- (a) Discuss cyclic redundancy code (CRC) and write a program in C. **03** **05**
- (b) Explain the method and write program in C for it :
i. Character stuffing
ii. Bit stuffing **03** **05**



HARCOURT BUTLER TECHNICAL UNIVERSITY, KANPUR
B.Tech.

End Semester Examination
Odd Semester (V), 2022-23

Theory of Automata and Formal Languages (ECS-357)

Time: 2:30 Hours

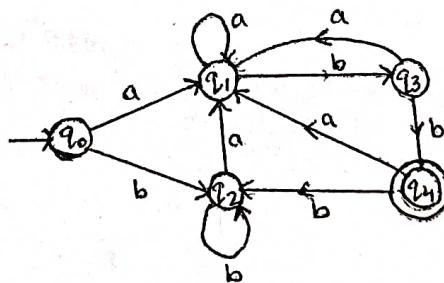
Max. Marks: 50

Note: Attempt all questions. All questions carry marks as shown against them.

-
1. Describe the capabilities and limitations of abstract machines including finite automata, pushdown automata, Turing machines, and their associated languages.
 2. Construct finite automata, pushdown automata, and Turing machines for the given grammar and vice versa.
 3. Show that a language is not regular / not context-free using a pumping lemma.
 4. Outline the characteristics of NP-Complete problems in the context of Turing machines.
-

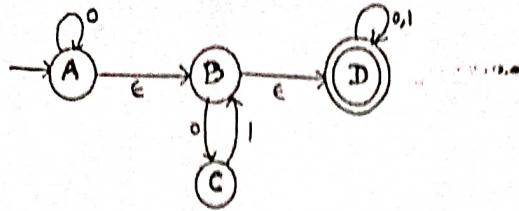
Q. No.1. Attempt both questions.

- (a) Construct the DFA which accepts a set of all strings over $\Sigma = \{a, b\}$, where CO1 (05)
(i) String start and end with the same symbol
(ii) String start and end with the different symbol
- (b) Construct the minimal DFA corresponding to the following machine CO1 (05)



Q. No.2. Attempt both questions.

- (a) Construct the mealy machine which calculates residue mod-4 for each binary string treated as a binary integer. CO2 (05)
- (b) Construct the NFA corresponding to the following ϵ -NFA. CO2 (05)



OR

Convert the given MEALY machine to the corresponding MOORE machine

CO2 (05)

	$a = 0$	$a = 1$
$\rightarrow q_1$	$q_3, 0$	$q_2, 0$
q_2	$q_1, 1$	$q_4, 0$
q_3	$q_2, 1$	$q_1, 1$
q_4	$q_4, 1$	$q_3, 0$

Q. No.3. Attempt both questions.

(a) Consider the following grammar G

$$S \rightarrow ABAC$$

$$A \rightarrow a \text{ or } \epsilon$$

$$B \rightarrow b \text{ or } \epsilon$$

$$C \rightarrow c$$

remove the ϵ -production in the above CFG.

CO3 (2.5)

(b) Convert the following grammar into GNF form

CO3 (2.5)

$$S \rightarrow AB/BC$$

$$A \rightarrow aB/bA/a$$

$$B \rightarrow bB/cC/b$$

$$C \rightarrow c$$

(c) What do you understand by derivation tree and explain an ambiguity in the grammar with an example.

CO3 (05)

Q. No.4. Attempt both questions.

(a) State the pumping lemma for CFL and prove that the family of CFL is not closed under intersection and complementation property.

CO3 (05)

(b) Construct a PDA for $L = \{a^n b^n c^{n+m} \mid n, m \geq 1\}$

CO4 (05)

Q. No.5. Attempt both questions.

(a) Design a Turing machine for even palindromes over $\Sigma \cup \{\epsilon\} \{a, b\}$.

CO4 (05)

OR

Design a Turing machine that accepts the number of a's = number of b's over the $\Sigma \cup \{\epsilon\} \{a, b\}$.

(b) Define recursive and recursively enumerable languages with examples. Also, explain the Universal Turing Machine in detail.

CO4 (05)