



END Semester Examination

Programme: B.Tech

Semester: VI

Course Code: CT-21012

Course Name: Data Science

Branch: Computer Engineering

Academic Year: 2023-24

Duration: 03 Hrs.

Max Marks: 60

Student PRN No.:

--	--	--	--	--	--	--	--

Instructions:

1. Figures to the right indicate the full marks.
2. Mobile phones and programmable calculators are strictly prohibited.
3. Writing anything on question paper is not allowed.
4. Exchange/Sharing of stationery, calculator etc. not allowed.
5. Write your PRN Number on Question Paper.

Marks CO PO

Q 1 A Categories the following task into Data Science / Non Data Science task. Justify your answer. 4 1,3 1,3,
5,12

- i) Predicting the future stock price of a company using historical records as well as trending news of stock.
- ii) Extracting the frequencies of a sound wave
- iii) Prediction of graduation year of a student from MIS number of a student
- iv) Monitoring physical data for athletes and accordingly providing required suggestions.

B Classify the given dataset into their most appropriate categories: graph dataset, time series dataset, spatio-temporal dataset. 2

- i) Sales of a particular product of a company.
- ii) Dataset giving details of worldwide pollution incident.
- iii) Dataset having pages on world wide web.
- iv) Dataset showing details of various forest fires in different parts of the world.

C Consider an dataset storing employee details like employee id, name, designation, salary for analysis. Give 2 sample examples in given dataset to demonstrate one noisy and other an outlier example for salary attribute. 2

Q 2 A Calculate dissimilarity measure between the sample no. 1 & 3 for the dataset given below. 4 2 1,2,
3,4,
5,12

Sample No.	Job(categorical)	Score(numeric)	Grade (ordinal)
1	testing	88	Excellent
2	developer	75	Above average
3	analyst	60	Average
4	marketing	45	Poor
5	testing	53	Below average



COLLEGE OF ENGINEERING PUNE
(An Autonomous Institute of Government of Maharashtra.)

- B Assume that you are building a decision tree classifier for some dataset having ten features. Draw a flowchart to explain stepwise forward selection method for attribute subset selection. Stopping criteria for adding attributes to classifier is: stop if accuracy of classifier is above 90%.

4

8 4 1,2,
4,5

- Q3 A The daily expenditure of 32 families is given below:

25	37	23	26	30	40	25	26
39	32	21	26	19	27	32	23
18	26	34	18	31	35	21	33
33	9	16	32	35	42	15	24

- i) Calculate Q1, Q2, Q3 for the given dataset.
- ii) Calculate IQR range.
- iii) Draw box-plot with whiskers for obtained data.
- iv) Analyze if any outliers in the given dataset.
- v) Construct a frequency distribution table with proper column heading using width of 20 and use it to calculate mean, median, mode for given data.
- vi) Use information calculated in v) to comment on skewness / symmetricness of given data.

- B In an anti malaria campaign in a certain area, quinine was administered to 812 persons out of a total population of 3248. The number of Malaria (fever) cases is shown below. Use Chi-square test to comment on usefulness of quinine in checking malaria at 5% level.

4

NOTE: Distribution table is given for reference at the end of question paper.

Treatment	Fever	No Fever
Quinine	20	792
No Quinine	220	2216

- Q4 A
- i) Write a generalized hypothesis for a Logistic Regression Classifier which is using single feature to predict the class of a sample. (NOTE: use 0_0 's and x_1 's).
 - ii) Assume that optimized '0' parameters for a dataset are $0_0 = 64$ and $0_1 = 2$. Calculate decision boundary for Class = 1 and decision boundary for Class = 0.
 - iii) What is a least value of a feature so that the chance of sample belonging to Class 1 is 95% or more.
 - iv) Find the class of a sample if sample is having feature value 30.

8 3 1,5

- B Compare between Linear Regression and Logistic Regression.

4

- Q5 A
- Draw a neat diagram to explain Data Stream Management System. Explain different types of queries that are used in streaming system.

6 5 1,2,

5

- B Discuss different issues in handling streaming data.

3

- Q6 A
- Suppose we want to find a pair of documents having 0.8 similarity and we do so by LSH with 20 bands and 6 rows each.
- i) If two documents have Jaccard Similarity 0.8, what is the chance that they will be true positive?
 - ii) If two documents have Jaccard Similarity 0.4, what is the chance that they will be true negative?

4 3 1,5

TID	Items
1	P, R, S
2	Q, R, T
3	P, Q, R, T
4	Q, T
5	P, Q, R, T

Consider the dataset given above, assume minsupport = 60%.

- i) Calculate the support of below given itemsets and categories them into frequent, closed frequent or maximal frequent itemsets and justify it: {P}, {Q}, {R}, {T}, {P, R}, {Q, R}, {Q, T}, {R, T}, {Q, R, T}.
- ii) Use diagram to show relation between frequent, closed frequent and maximal frequent itemsets.
- iii) Calculate the confident of all rules that can be generated using the apriori algorithm for the itemset {Q, R, T}

Distribution Table for Q3(B)

DF	P	0.995	0.975	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.002	0.001
1	0.00003930.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828		
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816	
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266	
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467	
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515	
6	0.676	1.237	8.558	10.545	12.592	14.449	15.033	16.812	18.548	20.791	22.458	
7	0.989	1.690	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322	
8	1.344	2.180	11.030	13.362	15.507	17.535	18.168	20.090	21.955	24.352	26.124	
9	1.735	2.700	12.242	14.584	16.919	19.023	19.679	21.666	23.589	26.056	27.877	
10	2.156	3.247	13.442	15.987	18.307	20.483	21.161	23.209	25.188	27.722	29.588	
11	2.603	3.816	14.631	17.275	19.675	21.920	22.618	24.725	26.757	29.354	31.264	
12	3.074	4.404	15.812	18.549	21.026	23.337	24.054	26.217	28.300	30.957	32.909	
13	3.565	5.009	16.985	19.312	22.362	24.736	25.472	27.688	29.819	32.535	34.528	
14	4.075	5.629	18.151	21.064	23.685	26.119	26.873	29.141	31.319	34.091	36.123	
15	4.601	6.262	19.311	22.307	24.996	27.488	28.259	30.578	32.801	35.628	37.697	
16	5.142	6.908	20.465	23.542	26.296	28.845	29.633	32.000	34.267	37.146	39.252	
17	5.697	7.564	21.615	24.769	27.587	30.191	30.995	33.409	35.718	38.648	40.790	
18	6.265	8.231	22.760	25.989	28.869	31.526	32.346	34.805	37.156	40.136	42.312	
19	6.844	8.907	23.900	27.204	30.144	32.852	33.687	36.191	38.582	41.610	43.820	
20	7.434	9.591	25.038	28.412	31.410	34.170	35.020	37.566	39.997	43.072	45.315	
21	8.034	10.283	26.171	29.515	32.671	35.479	36.343	38.932	41.401	44.522	46.797	
22	8.643	10.982	27.301	30.313	33.924	36.781	37.659	40.289	42.796	45.962	48.268	
23	9.260	11.689	28.429	32.007	35.172	38.076	38.968	41.638	44.181	47.391	49.728	
24	9.886	12.401	29.553	33.196	36.415	39.364	40.270	42.980	45.559	48.812	51.179	
25	10.520	13.120	30.675	34.382	37.652	40.646	41.566	44.314	46.928	50.223	52.620	



College of Engineering Pune
(An Autonomous Institute of Government of Maharashtra.)

END Semester Examination

Programme: T. Y. B. Tech

Semester: VI

Course Code: [IOC-21001]

Course Name: PLC & It's Applications

Branch: All Branches

Academic Year: 2023-24

Duration: 3 Hrs

Max Marks: 60

Student MIS No.

--	--	--	--	--	--	--	--	--	--

Instructions:

1. Figures to the right indicate the full marks.
2. Mobile phones and programmable calculators are strictly prohibited.
3. Writing anything on question paper is not allowed.
4. Exchange/Sharing of stationery, calculator etc. not allowed.

Q.1	Answer the following questions (Any FIVE)	Marks
	1. Explain logical AND & OR operation using ladder logic. 2. Write typical specification and features of Industrial grade PLC. 3. Explain about bits associated with timer instruction of PLC? 4. Explain the importance and purpose of Input Module, Output Module, CPU of the PLC 5. Explain different types of input devices and outputs devices typically get connected to the PLC. 6. Explain the structure of Ladder Diagram as programming language of PLC, also draw a symbol for Normally OPEN, Normally CLOSED, Relay COIL. 7. Explain seven step sequence for development and implementation of logic for a PLC based application.	20 CO1
Q.2	A temperature control system consists of four thermostats. The system operates three heating units. Thermostats are set at 55-, 60-, 65- and 70-degree C. Below 55 degree C, three heaters are to be on. A temperature between 55 to 60 degree C causes two heaters to be on. For 60 to 65 degree C, one heater is to be on. List down inputs and outputs and sequence of operation. List inputs and outputs, Design the wiring diagram for the controller connections, assign the inputs and outputs and develop the ladder diagram which will accomplish the task.	15 CO3



COLLEGE OF ENGINEERING PUNE
(An Autonomous Institute of Government of Maharashtra.)

END Semester Examination

Programme: T.Y.B. Tech

Semester: VI

Course Code: AS(HS)-21007

Course Name: Engineering Economics

Branch: E Group (All divisions)

Academic Year: 2023-24

Duration: 2hrs

Max Marks: 60

Student PRN No.

--	--	--	--	--	--	--	--	--	--

Instructions:

- Figures to the right indicate the full marks.
- Mobile phones and programmable calculators are strictly prohibited.
- Writing anything on question paper is not allowed.
- Exchange/Sharing of stationery, calculator etc. not allowed.
- Write your PRN Number on Question Paper.

		Each Question carries 12 Marks.	Marks	CO	PO
		Answer the following Questions. (Solve any 5 Questions)			
Q 1.	a.	Write short statements with explanations on Veblen Good-			
	b.	Cross Price elasticity	12		
Q 2.	a.	Explain the term Demand in economics and state its determinants.			
	b.	If the consumer expects an increase in petrol prices in May, will this impact the prices in April? If yes, how?	12		
Q 3.		Define Cost-Push factors of inflation. Critically examine the statement that CPI is the best indicator of inflation without bias.	12		
Q 4.		Use appropriate diagrams to elucidate the difference between binding and non-binding price floors. Give an example of a price floor.	12		
Q 5.		Write short notes on: a. Frictional Unemployment b. Economic development c. Seasonal Unemployment d. Headline Inflation	12		



COLLEGE OF ENGINEERING PUNE

(An Autonomous Institute of Government of Maharashtra.)

Q6.	Minimum wage laws tend to hurt some segments of the labourers. Explain using appropriate figures.	12		
Q7.	Amongst Agriculture, Service, & Manufacturing, which sector should India focus on to ensure sustainable growth? State your reasons clearly.	12		



COLLEGE OF ENGINEERING, PUNE

(An Autonomous Institute of Government of Maharashtra.)

END Semester Examination

(CT-21014) Design and Analysis of Algorithms

Programme: T.Y. BTech

Semester: VI

Branch: Computer Engineering

Max Marks: 60

Academic Year: 2023-24

Date : 28-04-2024

Duration: 3 Hours

Student MIS No.

Instructions

- Figures to the right indicate the full marks.
- Mobile phones and programmable calculators are strictly prohibited.
- Writing anything on question paper is not allowed.
- Exchange/Sharing of stationery, calculator etc. not allowed.
- Write your MIS Number on Question Paper
- Assume suitable data if necessary.
- Specify question no. clearly on answer-sheet and write answers in sequence.

Q.I

- A. (a) Using recurrence tree method compute the time complexity of the following recurrence relation and assume constant time for small value of n. [8]

$$T(n) = 2T(n/10) + T(9n/10) + n$$

6

- (b) Compute the time complexity of the following recurrence relation using back substitution method :-

$$T(n) = \begin{cases} 2T\sqrt{n} + \log n & n > 2 \\ 2 & 0 < n \leq 2 \end{cases}$$

- B. Compute the time complexity of the following recurrence relation using master method by specifying the respective case: [4]

3

(a) $T(n) = 2T(n/2) + n/\log n$

(b) $T(n) = 8T(n/2) + n^2$

- C. A data file of 100,000 characters contains only the characters a-f with the following frequencies. If we use variable length code then how much saving in number of bits can be achieved using Huffman code compared to the fixed length code? Draw the tree and show the calculation for both the methods. [4]

3

character	a	b	c	D	e	f
Frequencies (in thousands)	45	13	12	16	9	5

Q.II

- A. Given 4 items where each item has some weight (w) and profit (p) associated with it. [6]

6

Weights (w): { 3, 4, 6, 5 }

Profits (p): { 2, 3, 1, 4 }

The weight of the knapsack is 8 kg. The task is to put the items into the bag such that the sum of profits associated with them is the maximum possible. Use dynamic programming tabulation method to solve the 0/1 knapsack problem.

OR

- A. The Hamiltonian cycle in an undirected graph $G = (V, E)$ traverses every vertex exactly once. Prove that the Hamiltonian cycle problem is NP complete.
- B. Using dynamic programming find the optimum parenthesis for matrix chain multiplication given below. Compute m and s table for storing the cost and indexes achieved. Calculation for each matrix combination to be shown clearly and result of cost and indexes to be shown in table.

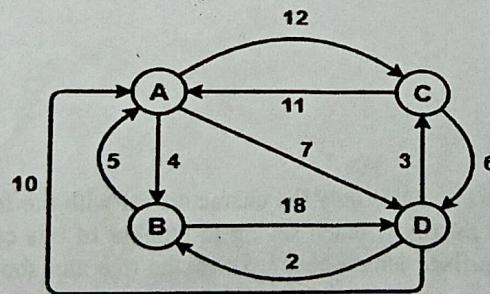
Matrix	A1	A2	A3	A4	A5
Dimension	4 x 10	10 x 3	3 x 12	12 x 20	20 x 7

Q.III

- A Let there are 4 workers and 4 tasks and each task is assigned to any workers exactly once. Also task is assigned in such a way that the cost is to be minimized. Use branch and bound method to assign the job to each workers. Show the state space tree for job allocation with lower bound calculation.

Workers	Task 1	Task 2	Task 3	Task 4
A	9	2	7	2
B	6	4	3	3
C	5	8	1	1
D	7	6	9	4

- B Solve travelling salesman problem for the following graph using branch and bound method in which city set = { A, B, C, D } and A is the home location of the salesman. Also show the recursive steps used for each path chosen and its state space tree.



- C Find a subset of a given set $A = \{5, 10, 12, 13, 15, 18\}$ contains 6 positive integers whose sum is equal to a given positive integer $d=30$. Use backtracking approach and draw the state-space tree showing all possible solution for subset sum problem.

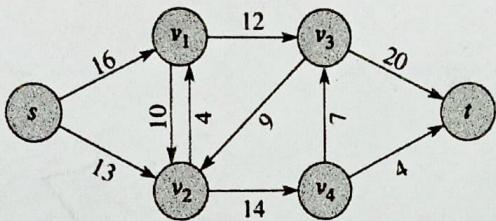
Q.IV

- A Compute a prefix function or LPS for the given pattern
Pattern P: a b a b a c a

Text T : b a c b a b a b a c a a b

[6]
6

- B** Given a graph which represents a flow network where every edge has a capacity. Also, given two vertices source 's' and sink 't' in the graph, find the maximum possible flow from s to t using Ford-Fulkerson Algorithm with the following constraints:
- Flow on an edge doesn't exceed the given capacity of the edge.
 - Incoming flow is equal to outgoing flow for every vertex except s and t.





COLLEGE OF ENGINEERING, PUNE

(An Autonomous Institute of Government of Maharashtra.)

END Semester Examination

Programme: B.Tech

Semester: VI

Course Code: CT-21015

Course Name: Software Engineering Mini Project Stage- II

Branch: T.Y. B.Tech(Computer)

Academic Year: 2023-24

Duration: 2hrs

Max Marks: 40

Student PRN No.

--	--	--	--	--	--	--	--	--	--

Instructions:

1. Questions are to be attempted strictly in the given order only.
2. Figures to the right indicate the full marks.
3. Mobile phones and programmable calculators are strictly prohibited.
4. Writing anything on question paper is not allowed.
5. Exchange/Sharing of stationery, calculator etc. not allowed.
6. Write your PRN Number on Question Paper.

		Marks	CO	PO
Q 1	a	5	1,4 ,5	a,d c,d
	b	5	1,2 ,3	c,d, e
Q2	a	4	3,4	c,d
	b	6	1,2 c,f	a,b, c,f

Considering above requirements, as a software engineer suggest best solution to design and develop the Totalware system by answering the questions below with proper justification.

a. Process Model b. Type of system
c. Architectural style d. Security aspects for keeping client data secure

b Consider a "dice game" in which a software system simulates a player rolling two dice at a time. If the sum of the rolls is equal to nine, the player wins;



COLLEGE OF ENGINEERING, PUNE

(An Autonomous Institute of Government of Maharashtra.)

i. Define Use Case(s) and corresponding Use Case Diagram(s) for the dice game.

ii. Construct a Class Diagram for the dice game. Clearly mention attributes and methods for each class specified in the class diagram. Further, specify various class relationships among different classes.

Provide proper justification and explanation for your design.

Q 3 a Consider a software project with following important functional units:

5 4,5 b,d
 ,g

No of user inputs=30,

No of user outputs=40,

No of user Inquiries=45,

No of internal logical files=08,

No of external interface files=05.

Assuming all complexity adjustment factors are significant and the weighting factors are average. Compute the FP of the software project.

b Explain any two indirect measures of software metrics significant to the online commercial software system designed to handle payment transactions during peak festival time.

5 3,4 a,d,
 ,5 g

Q 4 a Use the Basic COCOMO model to estimate efforts and duration of an embedded software development project with size of 600 KLOC. How many workers should be hired for this project? If the project must be completed within 15 months, how many additional staff should be hired? 178,518 = 340

5 4,5 a,d,
 g

b For the RMS software system in Q.1b, developed using C programming language, indicate and justify the role of structural testing and functional testing in deployment readiness of RMS software to the client with the help of test cases.

5 1 a,f



COLLEGE OF ENGINEERING BANGALORE

(Autonomous Institute of Government of Karnataka)

END Semester Examination

Programme: B.Tech

Semester: VI

Course Code: (CT(HO)-21003)

Course Name: Big Data Analytics

Branch: Computer Science and Engineering

Academic Year: 2023-24

Duration: 3 hr

Max Marks: 60

Student PRN No.

--	--	--	--	--	--	--	--	--	--	--	--

Instructions:

1. Figures to the right indicate the full marks.
2. Mobile phones and programmable calculators are strictly prohibited.
3. Writing anything on question paper is not allowed.
4. Exchange/Sharing of stationery, calculator etc. not allowed.
5. Write your PRN Number on Question Paper.

			Marks	CO	PO
Q 1	a	Explain HDFS architecture in detail?	5	2	1,12
	b	Explain shuffle and sort in Mapreduce?	4	2,3	1,3, 12
	c	Explain Anatomy of Mapreduce Job Run?	8	2,3	1,3,4,12
Q 2	a	Explain different types of Schedulers in Hadoop?	6	2	1,12
	b	Explain Lambda architecture of big data pipeline?	5	3	1,2, 12
	c	Explain Apache Spark Ecosystem Components?	6	3	1, 12
Q 3	a	What are RDD operations and differentiate RDD and Dataframe in Spark ?	8	3	1,2, 5, 12
	b	What is ETL Process?	6	4	1,12
	c	Explain Pig data type and data models in detail ?	6	4	1,12
Q 4	a	Explain architecture of Hive and its Components?	6	5	1,2, 12