Computer Science & Engineering

Model Question Papers

For Undergraduate Program

The model question papers are suggestive blueprints. The primary aim of these question papers is to bring clarity about the process of connecting questions to performance indicators and hence to course outcomes. Further, these question papers demonstrate how bloom's taxonomy can be used to understand the quality of question papers and their effectiveness in assessing higher order abilities. The structure of question papers, number of questions, choices given, time given for examination etc., can vary based on the practices of the University or college.

Table of Contents

Course Name: Discrete Mathematical Structures

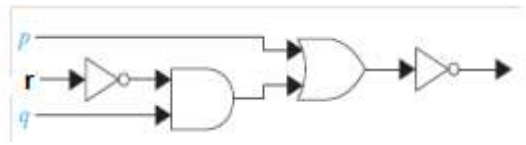
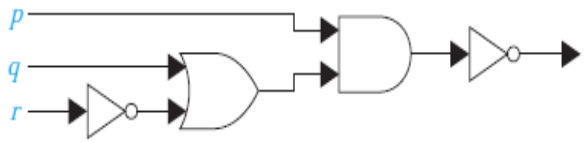
Course Outcomes (CO):

1. Apply logic and rules of inference to draw a conclusion from a set of premises in a finite sequence of steps.
2. Apply principles of sets operations and functions.
3. Apply various operations on sets and represent them using Venn diagram.
4. Use the fundamental counting principles to determine the number of outcomes for a specified problem.
5. Develop the recurrence relation for the given problems
6. Discuss and differentiate the types of functions, relations and groups.

Model Question Paper
Total Duration (H: M): 3:00
Course: Discrete Mathematical Structures
Maximum Marks: 100

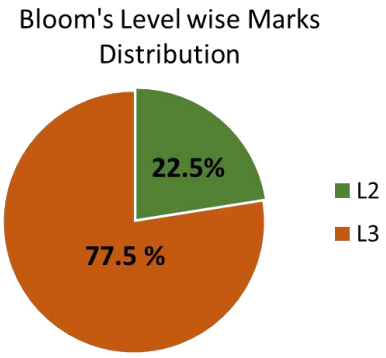
Note: Answer Any two questions from UNIT I, UNIT II and one question from UNIT III

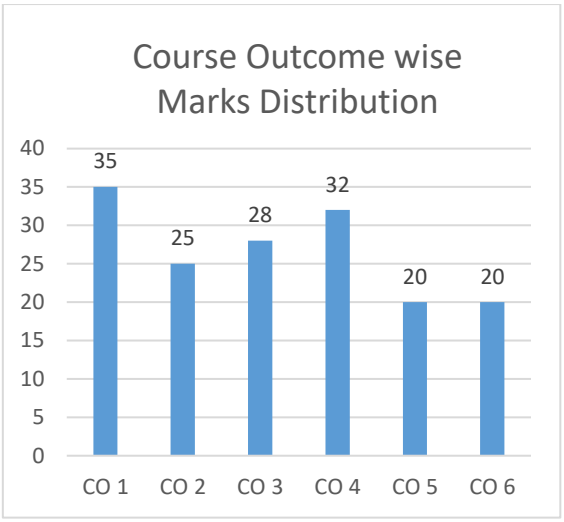
Q.No	Questions	Marks	CO	BL	PI
	UNIT I				
1a	In asynchronous transfer mode (ATM), data are organized into cells of 53 bytes. Identify the range (number of ATM cells transmitted) for the domain (minutes) set M={1, 2, 3, 4, 5, 6} if connection that transmits data at the rate of i) 128 kilobits per second ii) 300 kilobits per second iii) 1 megabit per second	10	CO2	L3	1.1.1
1b	Write the propositions for the following English statements. To use the wireless network in the airport you must pay the daily fee unless you are a subscriber to the service. Express your answer in terms of w: You can use the wireless network in the airport. d: You pay the daily fee. and s: You are a subscriber to the service.	5	CO 1	L3	1.1.1
1c	Let p,q and r be the propositions P: You have attended cultural audition. q: You miss the first minor exam. r: You will not get the make-up exam. Express each of these propositions as an English sentence i) $(p \rightarrow \neg r) \vee (q \rightarrow \neg r)$ ii) $(p \wedge q) \vee (\neg q \wedge r)$ iii) $\neg q \leftrightarrow r$	5	CO1	L2	1.1.1

Q.No	Questions	Marks	CO	BL	PI
2a	Let A, B, and C be sets. Show that $\overline{A \cup (B \cap C)} = (\bar{C} \cup \bar{B}) \cap \bar{A}$	5	CO 2	L2	1.1.1
2b	Consider the following system specifications using the propositions "The message is scanned for viruses" or "The message was sent from an unknown system" "When a message is not sent from an unknown system it is not scanned for viruses." "The message is scanned for viruses" Is the specification consistent? Justify your answer	5	CO 1	L3	1.1.1
2c	Consider the combinatorial circuit shown in below figure and answer the following.  (a)  (b) 1. Find the output of combinatorial circuits (a) and (b). 2. Write the simplified form of negation of the output. 3. Assume appropriate p, q and r and express the output in English sentence.	10	CO1	L3	1.1.1
3a	Let f, g, h be functions from $\mathbf{R} \rightarrow \mathbf{R}$ where $f(x)=x^2$, $g(x)=x+5$ and $h(x)=\sqrt{x^2+2}$. Determine ((h o g) o f) (x).	5	CO 2	L2	1.1.1
3b	Identify which of the following propositional statements are tautology using laws of equivalence. i) $[p \vee q \vee (\neg p \wedge \neg q \wedge r)] \longleftrightarrow (p \vee q \vee r)$ ii) $\neg (p \rightarrow q) \rightarrow \neg q$	10	CO 1	L3	1.1.1
3c	State whether the following statements are true or false i) Every infinite sets are countable ii) Every relation is not necessarily function iii) What time is it? is a proposition iv) Every bijective functions are inverse functions v) $(f \circ g)(a) = f(g(a))$.	5	CO 2	L3	1.1.1
	UNIT II				
4a	Suppose that at some future time every telephone in the world is assigned a number that contains a country code 1 to 3 digits long, that is, of the form X, XX, or XXX, followed by a 10-digit telephone	6	CO4	L3	1.1.1

Q.No	Questions	Marks	CO	BL	PI
	number of the form NXX-NXX-XXXX. How many different telephone numbers would be available worldwide under this numbering plan?				
b	How many positive integers between 100 and 999 inclusive i) are divisible by 7? ii) are not divisible by 4? iii) are divisible by 3 and 4? iv) are divisible by 3 or 4? v) are divisible by 3 but not by 4 and 7?	8	CO4	L3	1.1.1
c	For the relations R_1={(a,b), (a,c), (b,d), (d,d)} and R_2={(a,a), (a,d), (b,a), (b,b), (c,e), (d,d), } on sets {a,b,c,d,e} to {a,b,c,d,e} determine $R_2 \circ R_1$. Represent the output relation using directed graph.	6	CO3	L2	1.1.1
5a	Consider the following relation R={(1,1),(1,2),(1,3),(1,4), (2,2),(2,3),(2,4),(3,3),(3,4),(4,4)} defined over the set S={1,2,3,4} i) Is (S,R) is a Poset? Justify your answer. ii) Is (S,R) Linearly ordered? Justify your answer. iii) Is (S,R) Well-ordered? Justify your answer. iv) Identify the minimal, maximal, greatest and least elements v) Identify the lower bound and upper bound for the set {3}and also find the least upper bound and greatest lower bound.	10	CO3	L3	1.1.1
b	In how many possible orders a student can answer 5 questions in the SEE exams considering the following conditions i) There are 3 units UNIT1, UNIT2 and UNIT3 consisting of 3, 3 and 2 questions respectively. ii) Student has to answer 2 questions from UNIT 1 , 2 questions from UNIT 2 and one from UINIT 3	6	CO4	L2	1.1.1
c	In order to conduct the SEE examination, In how many ways seating arrangement can be made for 240 CS students and 240 EC students such that CS and EC students should sit alternatively.	4	CO4	L3	1.1.1
6a	School of Computer Science and Engineering is planning to create a Computer network lab of 15 computers. In how many ways every computer is connected to every other computer for each of the following assumptions. i) Every computer is implicitly connected to itself ii) Every computer is explicitly connected to itself iii) Every connection is one-way communication iv) Every connection is two-way communication	8	CO4	L3	1.1.1

Q.No	Questions	Marks	CO	BL	PI
b	Let R be the relation on the set of people with doctorates such that (a, b) ∈ R if and only if '***a'*** was the thesis advisor of '***b'***. When is an ordered pair (a, b) in R^2? When is an ordered pair (a, b) in R^n, when n is a positive integer? (Assume that every person with a doctorate has a thesis advisor.)	8	CO3	L3	1.1.1
c	Let R1 and R2 be the "congruent modulo 3" and the "congruent modulo 4" relations, respectively, on the set of integers. That is, R1 = {(a, b) | a ≡ b (mod 3)} and R2 = {(a, b) | a ≡ b (mod 4)}. Find i) R1 U R2. ii) R1 ∩ R2. iii) R1 − R2. iv) R2 − R1.	4	CO3	L2	1.1.1
	UNIT III				
7a	A vending machine dispensing books of stamps accepts only one-dollar coins, $1 bills, and $5 bills. a) Find a recurrence relation for the number of ways to deposit n dollars in the vending machine, where the order in which the coins and bills are deposited matters. b) What are the initial conditions? c) How many ways are there to deposit $10 for a book of stamps?	6	CO5	L3	1.1.2
b	Solve these recurrence relations together with the initial conditions given. i. $a_n = 2a_{n-1}$ for $n \geq 1$, $a_0 = 3$ ii. $a_n = a_{n-1}$ for $n \geq 1$, $a_0 = 2$	6	CO5	L2	1.1.2
c	a) Find a recurrence relation for the number of steps needed to solve the Tower of Hanoi puzzle. b) Show how this recurrence relation can be solved using iteration.	8	CO5	L3	1.1.2
8a	i) Check whether the binary operation * is commutative and associative on the seta) On Z, where a*b is ab b) on Z+, where a*b is a+b+2 ii) Prove or disprove the binary operation on Z+ of a*b = GCD(a,b) has the idempotent property.	8M	CO6	L3	1.1.1
b	Check whether set Z with the binary operation of subtraction is a semi group.	6M	CO6	L2	1.1.1
c	Define – i) Group ii) Rings iii) Fields give one example for each with domain as set of positive integers.	6M	CO6	L2	1.1.1





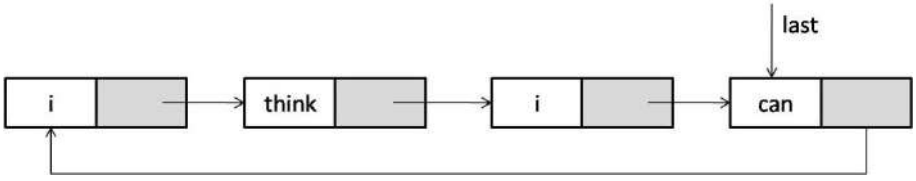
BL – Bloom's Taxonomy Levels (1- Remembering, 2- Understanding, 3 – Applying, 4 – Analysing, 5 – Evaluating, 6 - Creating)
CO – Course Outcomes
PO – Program Outcomes; PI Code – Performance Indicator Code

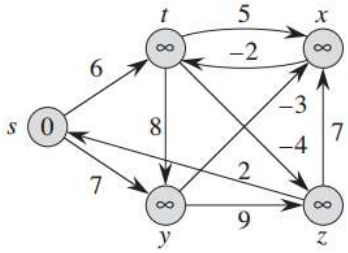
Course Name: Data Structures and Algorithms

Course Outcomes (CO):

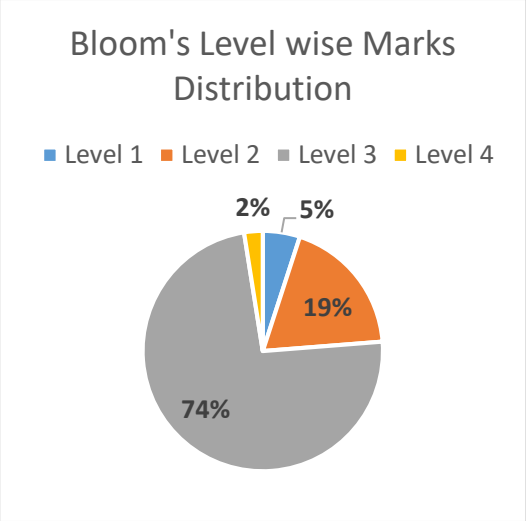
1. Discuss the C language features and analyze the differences between recursive and iterative programming structures
2. Analyze the role of data structures in structuring and manipulating data and implement them using array or list representation
3. Discuss the properties, operations, applications, strengths and weaknesses of the different data structures and their effect on algorithms
4. Analyze, interpret and compare various sorting, searching and graph algorithms and perform efficiency analysis
5. Discuss the file structures and storage management for efficient access of data

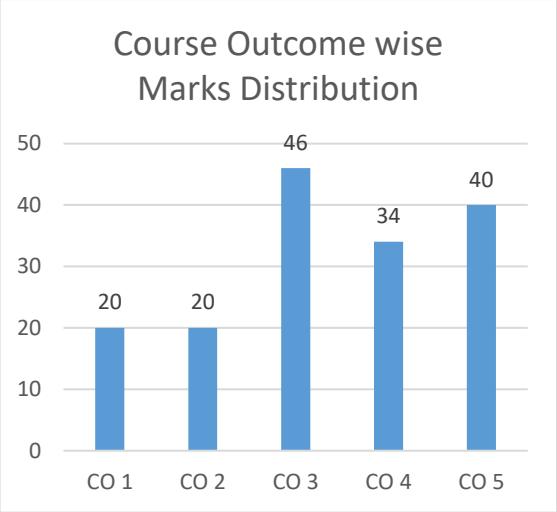
Model Question Paper
Total Duration (H:M): 3:00
Course: Data Structures and Algorithms
Maximum Marks: 100

Q.No	Questions	Marks	CO	BL	PI
1(a)	Suppose we wish to search a linked list of length n, where each element contains a key k long with a hash function h(k). Each key is a long character string. How might we take advantage of the hash values when searching the list for an element with a given key?	4	CO3	L3	1.4.1
1(b)	With the help of suitable code snippets, Prove That: "Queue is NO more exactly a First In First Out data structure"	6	CO2	L2	1.4.1
1(c)	Using state space tree prove that: • There is no solution for a 2 queen problem • There are multiple solutions for a 4 queen problem	10	CO1	L2	1.4.1
2(a)	Differentiate between Structures and Unions with suitable code snippets.	4	CO1	L3	1.4.1
2(b)	A linear probing has a hash function of the form: h(k, i) = (h'(k) + i) mod m and a quadratic probing has a hash function of the form: h(k, i) = (h'(k) + c 1 i + c 2 i 2) mod m. Linear probing suffers from a problem known as primary clustering and quadratic probing from secondary clustering. Discuss.	6	CO3	L3	1.4.1
2(c)	Consider the circular list given below with string data:  Write a function which will display the output in following fashion: *i think i can* *think i can*	10	CO2	L3	1.4.1

Q.No	Questions	Marks	CO	BL	PI
	i can *can* At each line, the function should display data from all the nodes present. After printing each line, an appropriate node has to be deleted. After printing the last line, "last" pointer should be holding the NULL value.				
3(a)	Write a program to print the nth node from end from a singly linked list.	4	CO2	L3	1.4.1
3(b)	Complete the function described below: Function Name: summon Input Params: base address of string Return Type: base address of summoned string Description: A magician wants to generate summoning charms. For input string "firebolt", the function should produce "summon firebolt". Do not use any inbuilt string handling functions.	6	CO1	L3	1.4.1
3(c)	Write the modules to implement the following using Stack data structure: • Check if the given string is palindrome • Sort the given set of integers	10	CO3	L3	1.4.1
4(a)	You have been invited to a post-exam party. i) You walk in and shake everyone's hand. As the number of attendees N increases, what is the order of growth to shake everyone's hand? Justify. ii) You meet everyone else and during each meeting, you talk about everyone else in the room. To what efficiency class does this belong to? Justify.	4	CO3	L4	1.1.2
4(b)	Create a AVL Tree for: 50, 60, 80, 30, 20, 40, 70 Can you perform the three tree traversals on AVL tree? Justify your answer.	6	CO3	L3	1.4.1
4(c)	Apply Bellman-Ford Algorithm on the given graph.  How is Bellman-Ford different from Dijkstra's Algorithm? To what design technique does the algorithm belong to? Explain.	10	CO4	L3	1.4.1
5(a)	Bring out the differences between Prim's and Kruskal's algorithm. Also compare with respect to efficiency analysis.	4	CO4	L2	1.4.1
5(b)	Write a algorithm for given below description: ALGORITHM CountLeafNodes(T) // Recursively counts the number of leaf nodes in the tree T	6	CO3	L3	1.4.1

Q.No	Questions	Marks	CO	BL	PI
5(c)	Apply Quick Sort on the following: Q U I C K S O R T Write the efficiency analysis of quick sort (Best, Worst, and Average).	10	CO4	L3	1.1.2
6(a)	Write the algorithm design technique for the given below algorithms/problems: i) N-Queen's Problem ii) Binary Search iii) Insertion Sort iv) AVL Trees v) Heap Sort vi) Hashing vii) Boyer-Moore viii) Breadth First Search	4	CO4	L2	1.4.1
6(b)	A DNA sequence consists of a text on the alphabet {A, C, G, T} and the gene or gene segment is the pattern. For the pattern for chromosome-10: TCCTATTCTT construct the following tables: i) π- table ii) Bad Symbol Shift Table	6	CO4	L3	1.4.1
6(c)	Write a function to delete a node from a Binary Search Tree. Suitably comment the code explaining each of the cases.	10	CO3	L3	1.4.1
7(a)	What are indexed sequential files?	4	CO5	L1	1.4.1
7(b)	Explain fseek( ) API with help of a C suitable program. Explain each of the parameter it takes in detail.	6	CO5	L3	1.4.1
7(c)	A file consists of binary data. Write a program to read and count the number of 0's and 1's in it. Write the individual count in a separate file. Also find the size of the file.	10	CO5	L3	1.4.1
8(a)	What do you mean by storage release?	4	CO5	L1	1.4.1
8(b)	Differentiate between the fixed block and variable block storage management.	6	CO5	L2	1.4.1
8(c)	With a help of a suitable program explain the concept of Bit Maps and how they can be used as an efficient storage means.	10	CO5	L3	1.4.1





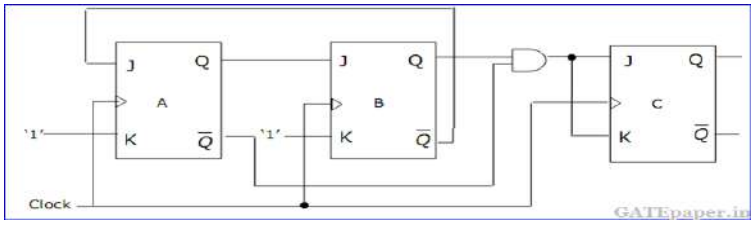
BL – Bloom's Taxonomy Levels (1- Remembering, 2- Understanding, 3 – Applying, 4 – Analysing, 5 – Evaluating, 6 - Creating)
CO – Course Outcomes
PO – Program Outcomes; PI Code – Performance Indicator Code

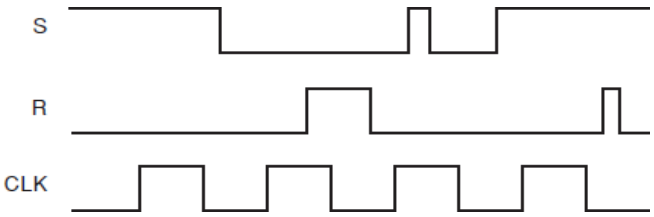
Course Name: Computer Organization & Architecture

Course Outcomes (COs):

At the end of the course the student should be able to:

1. Design combinational and sequential circuits using digital components.
2. Analyze different types of communication between processor and peripherals.
3. Design memory units for given specifications.
4. Describe the architectures of various processors.
5. Analyze the performance of pipelined architecture in a processor.

Model Question Paper for End Semester Examination					
Course Code:		**Course Title: Computer Organization and Architecture**			
Duration: 3 hrs		**Max. Marks: 100**			
Note: Answer five questions; any two full questions from each unit-I and unit-II and one full question from unit-III					
Unit-I					
Q.No	Questions	Marks	CO	PI Code	B L
1 a	Implement the Boolean function F(A,B,C,D)=BC+ABC̀+A'C'D using only a single 8X1 Multiplexer where signals A,B,C,D and D' represent the inputs.	6	CO1	1.4.4	L2
b	The circuit diagram of a synchronous counter is shown in the given figure. Determine the sequence of states of the counter assuming that the initial state is "000". Give your answer in a tabular form showing the present state QA, QB, QC, J-K inputs (JA, KA, JB, KB, JC, KC) and the next state QA+, QB+, QC+. From the table, determine the modulus of the counter. 	6	CO1	1.4.4	L3
c	Consider the instruction Load R2,(R3). What control signals are generated in the datapath during the execution of this instruction. Justify your answer with proper reasoning.	8	CO2	1.4.4	L3
2 a	Realize a Boolean expression f(w,x,y,z)=∑m(1,2,6,7,8,9,11,13,14,15) using Multiplexer Tree structure. The first level should consist of two 4: 1 MUX with variables **w** and **z** on their select lines S1 and S0 respectively and second level should consist of single 2:1 MUX with variable **y** on the select lines	6	CO1	1.4.4	L2

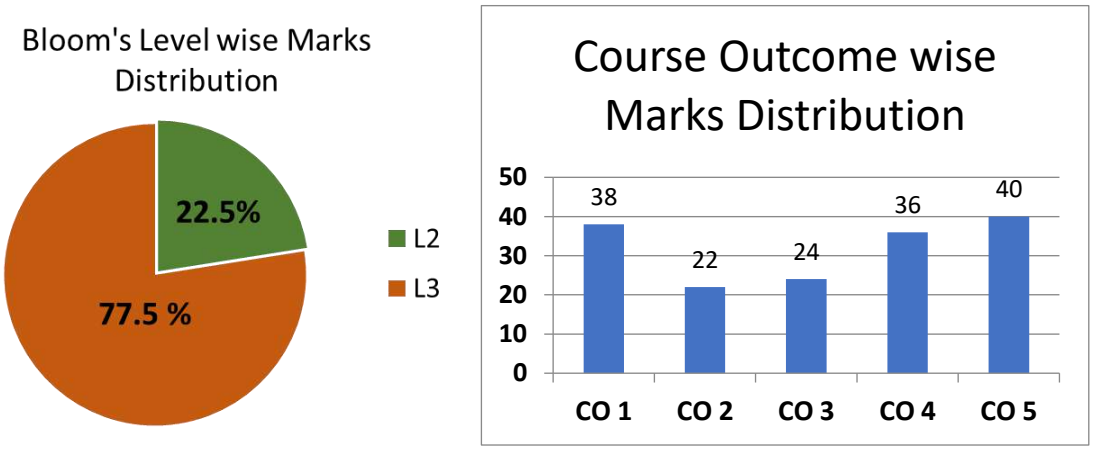
<table>
<tr><td>b</td><td>i. Explain the working of Gated SR latch.
ii. What is forbidden state?
iii. Determine the output waveform for a gated SR latch
  </td><td>6</td><td>CO1</td><td>1.4.4</td><td>L3</td></tr>
<tr><td>c</td><td>i. Load R4,(R3)
ii. Store R4,(R3)
For the above instructions draw the timing diagrams for synchronous bus operation</td><td>8</td><td>CO2</td><td>1.4.4</td><td>L3</td></tr>
<tr><td>3 a</td><td>Explain Hardwired Control unit.</td><td>6</td><td>CO2</td><td>1.4.4</td><td>L2</td></tr>
<tr><td>b</td><td>Design , Simulate and implement a 4-bit Universal Shift Register for following operation
<table><tr><th>Sl.no</th><th>Selection Bits</th><th>Operation</th></tr><tr><td>1.</td><td>00</td><td>Circular Shift Right</td></tr><tr><td>2.</td><td>01</td><td>Parallel Load</td></tr><tr><td>3.</td><td>10</td><td>Circular Shift Left</td></tr><tr><td>4.</td><td>11</td><td>Hold</td></tr></table></td><td>6</td><td>CO1</td><td>1.4.4</td><td>L3</td></tr>
<tr><td>c</td><td>There is a queue for exchange of old 500 rupees currency notes after demonetization, in the post office. But a single customer can exchange only upto Rs. 4000/- per day. What components are necessary to design a system which alerts the user on reaching the maximum count? Also design the system to automate the counting process.</td><td>8</td><td>CO1</td><td>2.1.2</td><td>L3</td></tr>
<tr><td colspan="6">Unit-II</td></tr>
<tr><td>4a</td><td>Explain the operational model of SIMD computers.</td><td>6</td><td>CO4</td><td>1.4.4</td><td>L2</td></tr>
<tr><td>b</td><td>A 40 MHz processor was used to execute a benchmark program with the following instructions and clock cycle counts:</td><td>6</td><td>CO4</td><td>1.4.4</td><td>L3</td></tr>
<tr><td></td><td><table><tr><th>Instruction type</th><th>Instruction count</th><th>Clock cycle count</th></tr><tr><td>Integer arithmetic</td><td>45000</td><td>1</td></tr><tr><td>Data transfer</td><td>32000</td><td>2</td></tr><tr><td>Floating point</td><td>15000</td><td>2</td></tr><tr><td>Control transfer</td><td>8000</td><td>2</td></tr></table>
Determine the effective CPI, MIPS rate and execution time for this program.</td><td></td><td></td><td></td><td></td></tr>
</table>

<table>
<tr><td>c</td><td>Consider a main memory built with SDRAM chips. Data are transferred in burst lengths of 8. Assume that 32 bits of data are transferred in parallel. If a 400-MHz clock is used, how much time does it take to transfer:
(a) 32 bytes of data
(b) 64 bytes of data
What is the latency in each case?</td><td>8</td><td>CO3</td><td>1.4.4</td><td>L3</td></tr>
<tr><td>5a</td><td>Explain the pipelining process in VLIW processors.</td><td>6</td><td>CO4</td><td>1.4.4</td><td>L2</td></tr>
<tr><td>b</td><td>The execution times (in seconds) of four programs on three computers are given below:
<table><tr><td rowspan="2">Program</td><td colspan="3">Execution Time (in seconds)</td></tr><tr><td>Computer A</td><td>Computer B</td><td>Computer C</td></tr><tr><td>Program 1</td><td>1</td><td>10</td><td>20</td></tr><tr><td>Program 2</td><td>1000</td><td>100</td><td>20</td></tr><tr><td>Program 3</td><td>500</td><td>1000</td><td>50</td></tr><tr><td>Program 4</td><td>100</td><td>800</td><td>100</td></tr></table>Assume that 100,000,000 instructions were executed in each of the four programs. Calculate the MIPS rating of each program on each of the three machines. Based on these ratings, can you draw a clear conclusion regarding the relative performance of the three computers? Rank them statistically.</td><td>6</td><td>CO4</td><td>1.4.4</td><td>L3</td></tr>
<tr><td>c</td><td>With neat figure show how a very small memory chip consisting of 16 words of 8 bits each is organized (16 X 8 organization). What is Double-Data-Rate SDRAM?</td><td>8</td><td>CO3</td><td>1.4.4</td><td>L3</td></tr>
<tr><td>6a</td><td>With a neat diagram of computer architecture explain the components involved in modern computer architecture.</td><td>6</td><td>CO4</td><td>1.4.4</td><td>L2</td></tr>
<tr><td>b</td><td>How does instruction set, compiler technology, CPU implementation and control and cache and memory hierarchy affect the CPU performance and justify the effects in terms of program length, clock rate and effective CPI</td><td>6</td><td>CO4</td><td>1.4.4</td><td>L3</td></tr>
<tr><td>c</td><td>Design a 8M x 32 size memory chip using memory chips of
i. 512K x 16
ii. 256K x 8
Draw the diagrams showing appropriate connections for each of the above cases.</td><td>8</td><td>CO3</td><td>1.4.4</td><td>L3</td></tr>
</table>

	Unit III				
7a	With a neat diagram explain 5-stage pipeline organization. Consider the following instructions at the given addresses in the memory: 1000 ADD R4, R3, R2 1004 OR R7, R6, R5 1008 SUBTRACT R8, R7, R4 Initially, registers R2 and R3 contain 4 and 8, respectively. Registers R5 and R6 contain 128 and 2 respectively. Assume that pipeline provides forwarding paths to the ALU from registers RY and RZ. The first instruction is fetched in clock cycle 1, and the remaining instructions are fetched in successive cycles. Draw a pipelined execution of these instructions assuming that processor is using operand forwarding. Describe the contents of registers, RY, and RZ in the pipeline during cycles 4 to 7.	10	CO5	1.4.4	L3

b	Explain with an example the different types of hazards. The following instructions are executed in pipelined architecture.
	SUBTRACT R4,R2,R3
	BRANCH NEXT
	OR R1, R2,#5
	MUL R3,R4,R2
	NEXT: LOAD R5,[R0]
	ADD R6,R5,R2
	Identify hazards and suggest hardware or software approach to minimize the hazards.
		10	CO5	1.4.4	L3

8a	Assume that 25% of dynamic count of instructions executed for a program is branch instructions. There are pipeline stalls due to data dependencies; static branch prediction is used with non-taken assumption. i. Determine execution time for two cases: when 30% of branches are taken, and when 80% of branches are taken. ii. Determine the speedup for one case relative to other. Express the speedup as percentage relative to 1.	10	CO5	1.4.4	L3
b	Explain the role of Dispatch unit to resolve Deadlock in scalar processor.	10	CO5	1.4.4	L2



BL – Bloom's Taxonomy Levels (1- Remembering, 2- Understanding, 3 – Applying, 4 – Analysing, 5 – Evaluating, 6 - Creating)
CO – Course Outcomes
PO – Program Outcomes; PI Code – Performance Indicator Code

Competency addressed in the Course and corresponding Performance Indicators

Competency	Performance Indicators
1.4: Demonstrate competence in computer science engineering knowledge	**1.4.4** Apply machine dependent/independent features to build system modules.
2.1: Demonstrate an ability to identify and characterize an engineering problem.	**2.1.2**: Identify processes, modules, variables, and parameters of computer based system to solve the problems.

Eg: 1.2.3: Represents Program Outcome „1", Competency „2" and Performance Indicators „3".

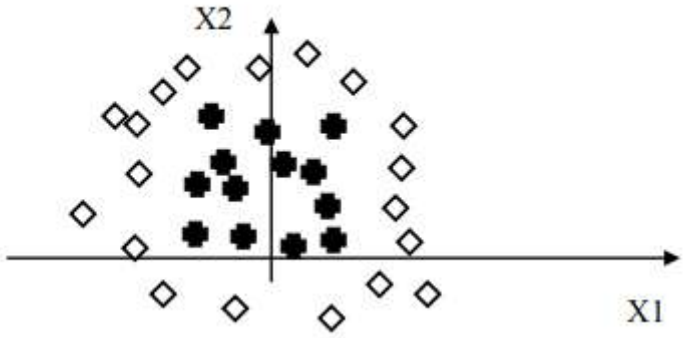
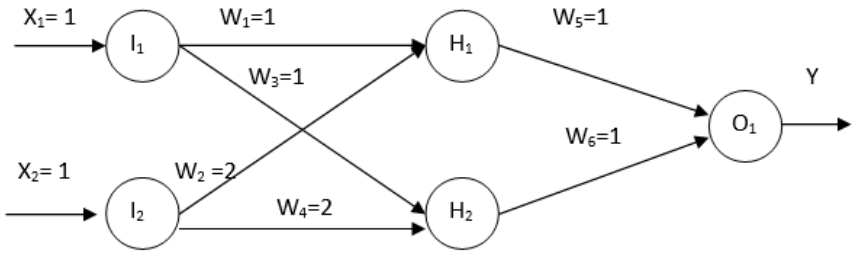
Course Name: Machine Learning

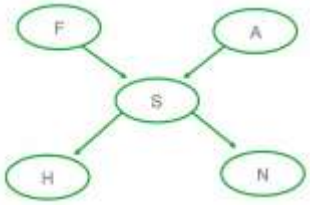
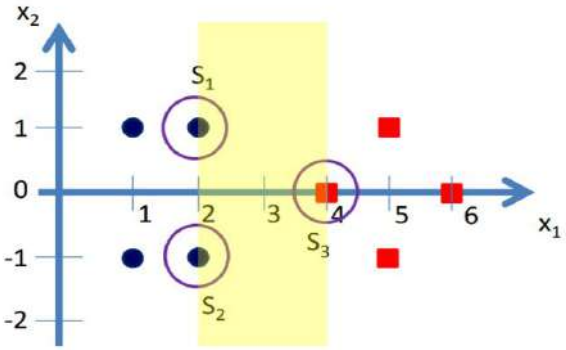
Course Outcomes (CO):

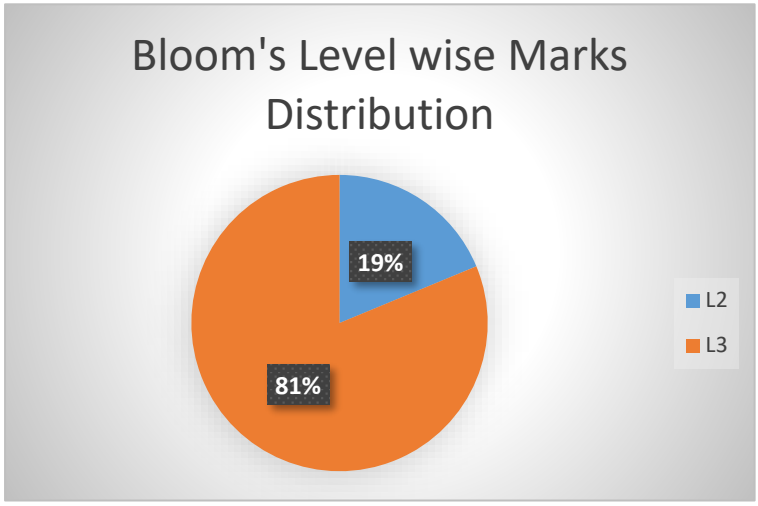
1. Interpret and apply machine learning concepts.
2. Develop a machine learning model to extract knowledge from given data.
3. Apply a suitable supervised/un-supervised learning algorithm to solve a given problem.
4. Develop an autonomous system using reinforcement learning.
5. Evaluate various machine learning algorithms and build a solution for real-world applications.

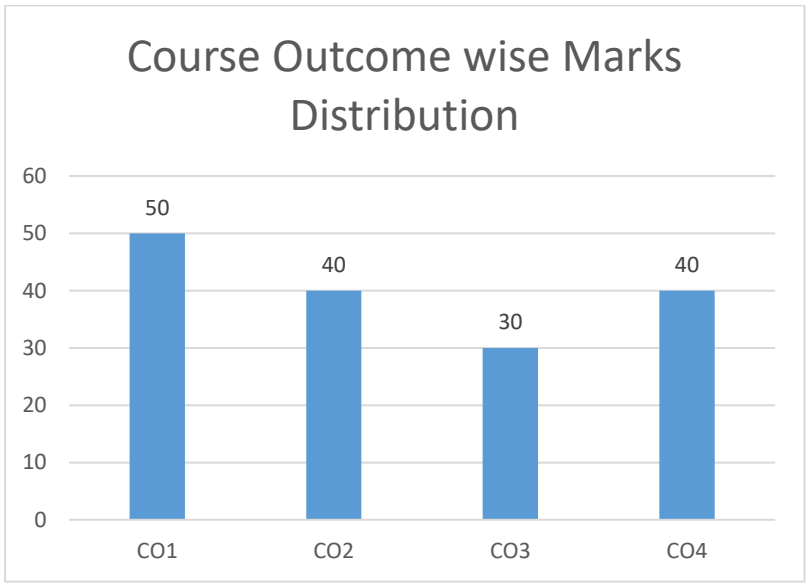
Model Question Paper
Total Duration (H:M): 3:00
Course: Machine Learning
Maximum Marks: 100

Q.No	Questions	Marks	CO	BL	PI
1a	A dealer has a warehouse that stores a variety of fruits and vegetables. When fruit is brought to the warehouse, various types of fruit may be mixed together. The dealer wants a model that will sort the fruit according to type. Justify with reasons how machine learning model is efficient compared to feature based classification technique.	10	CO1	L2	1.4.1
1b	Suppose you are only allowed to use binary logistic classifiers to solve a multi-class classification problem. Given a training set with 2 classes, this classifier can learn a model, which can then be used to classify a new test point to one of the 2 classes in the training set. You are now given a 6 class problem along with its training set, and have to use more than one binary logistic classifier to solve the problem, as mentioned before. Propose the following scheme - you will first train a binary logistic classifier for every pair of classes. Now, for a new test point, you will run it through each of these models, and the class which wins the maximum number of pairwise contests, is the predicted label for the test point. How many binary logistic classifiers will you need to solve the problem using your proposed scheme?	10	CO2	L3	2.1.3
2a	Describe the two error functions that are used for neural networks. Suppose we are training a neural network for binary classification, justify the type of error function which is suitable to solve the problem.	10	CO1	L3	1.4.1
2b	Why it is necessary to estimate the accuracy of hypothesis. Explain procedure to estimate difference in error between two learning methods.	10	CO3	L3	1.4.1
3a	Explain the effect of following factors in achieving global minima with gradient descent algorithm.	10	CO1	L3	2.1.3

Q.No	Questions	Marks	CO	BL	PI
	• Epochs • Learning rate • Weights of hypothesis (Θ_0 and Θ_1)				
3b	Calculate the cost value of linear regression for the following dataset. X=[1,2,3,4,5] Y=[3,6,7,11,15] Slope of hypothesis (Θ_1) = 7 Constant/ intercept (Θ_0)=5 Epoch=3 Learning rate (α) = 0.5	10	CO2	L3	2.1.3
4a	Determine the Principal Components for the given 2-Dimensional dataset. (1, 2), (2, 4), (3, 6).	10	CO3	L3	2.1.3
4b	Suppose that we want to build a neural network that classifies two dimensional data (i.e., X = [x1, x2]) into two classes: diamonds and crosses. We have a set of training data that is plotted as follows:  Draw a network that can solve this classification problem. Justify your choice of the number of nodes and the architecture. Draw the decision boundary that your network can find on the diagram.	10	CO1	L3	1.4.1
5a	Consider the following Neural Network with alpha = 0.5, eta=0.24, desired output = 1 and sigmoid activation function. i. Perform one forward pass and calculate the error. ii. Calculate the updated weights for w5 and w6 using back-propagation. 	10	CO3	L3	2.1.3

Q.No	Questions	Marks	CO	BL	PI
5b	Consider the following graphical model, which defines a joint probability distribution over five Boolean variables. Apply Expectation Maximization to train this Bayesian network, given training data in which the variables F, S, H and N are fully observed, and where the variable A is sometimes unobserved.  a. What are the conditional probability distributions associated with each of the five random variables in this network? b. What variables are in the Markov blanket for variable A? c. Given that all variables in the Markov blanket of A are observed, it should be possible to compute the distribution over A based on only these variables. Simplify your expression from part (2) so that it uses only the variables in the Markov blanket of A.	10	CO1	L3	1.4.1
6a	Apply SVM algorithm for the data-points and find dimension of hyper plane to classify the data-points for the figure. (assume bias =1) 	10	CO2	L3	2.1.3
6b	What is linearly in separable problem? Design a two layer network of perceptron to implement XOR and AND gates.	10	CO2	L2	1.4.1
7a	Develop a Q learning task for recommendation system of an online shopping website. What will be the environment of the system? Write the cost function and value function for the system.	10	CO4	L3	2.1.3
7b	Explain how an agent can take action to move from one state to other state with the help of rewards.	10	CO4	L3	1.4.1
8a	Explain reinforcement learning in detail along with the various elements involved in forming the concept. Also define what is meant by partially observed state.	10	CO4	L3	2.1.3
8b	Explain how reinforcement learning problem differs from other function approximation.	10	CO4	L2	1.4.1





BL – Bloom's Taxonomy Levels (1- Remembering, 2- Understanding, 3 – Applying, 4 – Analysing, 5 – Evaluating, 6 - Creating)
CO – Course Outcomes
PO – Program Outcomes; PI Code – Performance Indicator Code

Course Name: Operating System

Course Outcomes (COs):

At the end of the course the student should be able to:

1. Explain the core structure and functionality of operating system.
2. Evaluate and analyze various algorithms in process management.
3. Describe different deadlock prevention, avoidance and memory management algorithms.
4. Analyze protection and security aspects of mobile and network operating systems.
5. Demonstrate scheduling and memory management algorithms.

<table>
<tr><td colspan="6">Model Question Paper for End Semester Examination</td></tr>
<tr><td colspan="3">Course Code:</td><td colspan="3">Course Title: Operating System</td></tr>
<tr><td colspan="3">Duration: 3 hrs</td><td colspan="3">Max. Marks: 100</td></tr>
<tr><td colspan="6">Note: Answer five questions; any two full questions from each unit-I and unit-II and one full question from unit-III</td></tr>
<tr><td colspan="6">Unit-I</td></tr>
<tr><td>Q.No</td><td>Questions</td><td>Marks</td><td>CO</td><td>PI Code</td><td>B L</td></tr>
<tr><td>1 a</td><td>Classify and justify the following applications as Batch oriented or Interactive.
Generating monthly bank statement, Word processing, Generating personal tax returns.</td><td>06</td><td>CO1</td><td>1.4.1</td><td>L2</td></tr>
<tr><td>b</td><td>Discus implementation of mutual exclusion with semaphores. Explain the role of wait( ) and signal( ) functions in this regard.</td><td>04</td><td>CO1</td><td>1.4.1</td><td>L2</td></tr>
<tr><td>c</td><td>Consider the following processes with their arrival time and burst time as given below:
<table>
<tr><td>Process</td><td>Arrival Time</td><td>Burst Time</td></tr>
<tr><td>P1</td><td>0</td><td>12</td></tr>
<tr><td>P2</td><td>2</td><td>4</td></tr>
<tr><td>P3</td><td>3</td><td>6</td></tr>
<tr><td>P4</td><td>8</td><td>5</td></tr>
</table>
Identify appropriate data structure to implement following scheduling algorithms for the above scenario. Which one among them gives optimized scheduling with respect to waiting time?
a. FCFS
b. Shortest remaining time first (SJF preemptive)
c. Round Robin (time quantum = 4 units)</td><td>10</td><td>CO2</td><td>1.4.1</td><td>L3</td></tr>
</table>

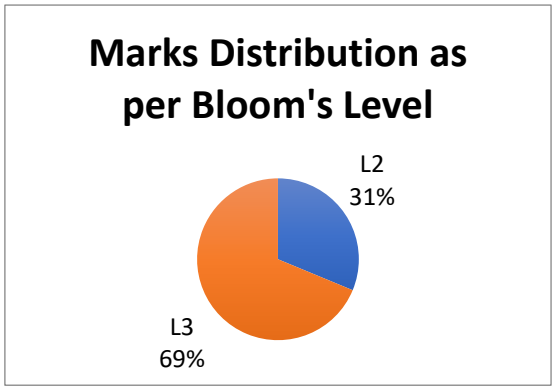
2 a	Consider the concurrently running processes A and B as given below. (Shared data: semaphore s1=0, s2=0;) **Process A:** do { Instruction a; Signal (s1); instruction b; Wait (s2); Instruction e; } while (true); **Process B:** do { Wait (s1); Instruction d; Instruction c; Signal (s2); } while (true) Illustrate the order in which these instructions (A, B, C and D) will get executed?	10	CO2	1.4.1	L3
b	Differentiate between one-to-one and many-to-many model used for multithreading implementation.	10	CO1	1.4.1	L2

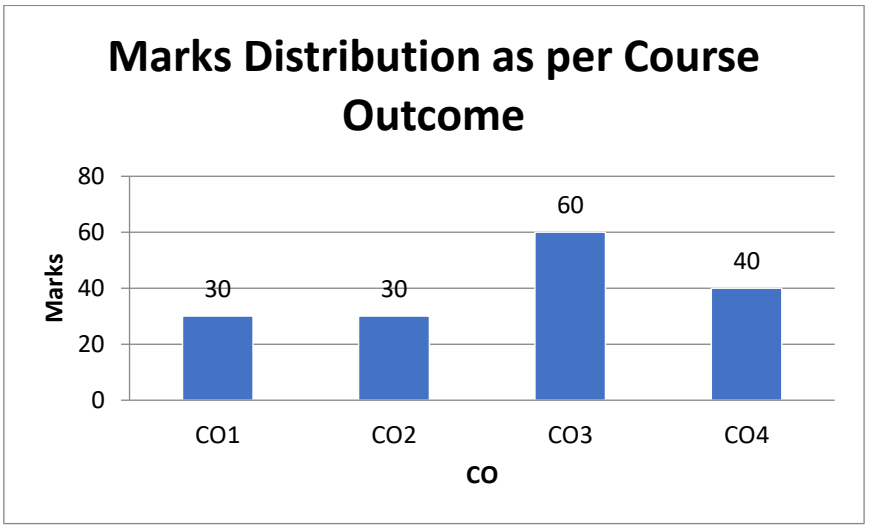
3 a	Assume two operations i) A(counter++) ii) B(counter--)
	Both are running following code:
	reg1= counter;
	reg1= Reg1+1;
	Counter = reg1;
	
	Check for occurrence of Race condition. Propose different approaches to solve race condition if it exists.
		10	CO1	1.4.1	L3

3.b	Every Saturday ram will get up at morning 6 o clock. His wife Swapna wants him to help her out with doing the house since guests were expected to arrive that evening. This would take Ram 45 minutes. His daughter Priya wants him to help her out in solving a math problem for 15 min, before she could face a test that afternoon in the tuition classes. Ram's son Arun, wants to be dropped at his friend's place as he is going to movie at 4pm. This would take an hour of travel. Ram's boss now calls him and asks him to call up a client and have a telecon on a particular business deal which would take him 20 minutes. In the evening 4 pm his colleague asks him a help to send one email to client so that in the night client will check and reply. It will take approximately 30 min. As age factor plays imp role Ram will be taking 5 min rest after each work.	10	CO2	1.4.1	L3

<table>
<tr><td></td><td>Implement above situation using FCFS, SJF and RR scheduling algorithms and identify the best one among them (with respect to Turn-around time). Priority will be ranging from professional to personal life.</td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="6">Unit-II</td></tr>
<tr><td>4a</td><td>Suppose that a disk drive has 5000 cylinders, numbered 0 to 4999. The drive is currently serving a request at cylinder 143, and the previous request was at cylinder 125. The queue of pending requests, in FIFO order, is: 86, 1470, 913, 1774, 948, 1509, 1022, 1750, 130 Starting from the current head position, what is the total distance (in cylinders) that the disk arm moves to satisfy all the pending requests for each of the following disk scheduling algorithms?
i)FCFS ii) SSTF iii) SCAN iv) LOOK</td><td>10</td><td>CO3</td><td>2.1.3</td><td>L3</td></tr>
<tr><td>b</td><td>Determine the total amount of internal fragmentation using i) First fit ii) Best fit iii) Worst fit techniques for the memory requests given below:
<table><tr><th colspan="2">Memory Requests</th><th></th><th colspan="3">Free regions</th></tr><tr><th>Jobs</th><th>Request</th><th></th><th>Free Blocks</th><th>Address</th><th>Size</th></tr><tr><td>Job 1</td><td>200 Bytes</td><td></td><td>Free region 1</td><td>2000-2900</td><td>900</td></tr><tr><td>Job 2</td><td>700 Bytes</td><td></td><td>Free region 2</td><td>5000-5500</td><td>500</td></tr><tr><td>Job 3</td><td>50 Bytes</td><td></td><td>Free region 3</td><td>6000-6800</td><td>800</td></tr><tr><td>Job 4</td><td>800 Bytes</td><td></td><td>Free region 4</td><td>7000-7100</td><td>100</td></tr><tr><td>Job 5</td><td>350 Bytes</td><td></td><td>Free region 5</td><td>8000-8600</td><td>600</td></tr><tr><td></td><td></td><td></td><td>Free region 6</td><td>9000-9300</td><td>300</td></tr></table></td><td>10</td><td>CO3</td><td>1.4.1</td><td>L2</td></tr>
<tr><td>5a</td><td>Consider the following snapshot of a system:
<table><tr><th></th><th colspan="4">Allocation</th><th colspan="4">Max</th><th colspan="4">Available</th></tr><tr><th></th><th>A</th><th>B</th><th>C</th><th>D</th><th>A</th><th>B</th><th>C</th><th>D</th><th>A</th><th>B</th><th>C</th><th>D</th></tr><tr><td>P0</td><td>0</td><td>0</td><td>1</td><td>2</td><td>0</td><td>0</td><td>1</td><td>2</td><td>1</td><td>5</td><td>2</td><td>0</td></tr><tr><td>P1</td><td>1</td><td>0</td><td>0</td><td>0</td><td>1</td><td>7</td><td>5</td><td>0</td><td colspan="4" rowspan="4"></td></tr><tr><td>P2</td><td>1</td><td>3</td><td>5</td><td>4</td><td>2</td><td>3</td><td>5</td><td>6</td></tr><tr><td>P3</td><td>0</td><td>6</td><td>3</td><td>2</td><td>0</td><td>6</td><td>5</td><td>2</td></tr><tr><td>P4</td><td>0</td><td>0</td><td>1</td><td>4</td><td>0</td><td>6</td><td>5</td><td>6</td></tr></table>Answer the following questions using the Banker’s algorithm:
a. Is the system in a safe state?
b. If a request from process P4 arrives for (0, 3, 2, 0), can the request be granted immediately?</td><td>10</td><td>CO3</td><td>2.1.3</td><td>L3</td></tr>
<tr><td>b</td><td>Determine the total number of page faults and page fault rate for the reference string given below using following page replacement algorithms. i) FIFO ii) Optimal iii) LRU (Assume Number of frames = 4)
Reference string:
3, 2, 5, 0, 1, 3, 5, 0, 7, 8, 2, 1, 2, 3, 4, 5, 8, 4, 7, 8</td><td>10</td><td>CO3</td><td>1.4.1</td><td>L3</td></tr>
</table>

<table>
<tr><td>6a</td><td>Suppose at time T0, we have the following resource allocation state:
<table><tr><th></th><th colspan="3">Allocation</th><th></th><th colspan="3">Request</th><th></th><th colspan="3">Available</th></tr><tr><th></th><th>A</th><th>B</th><th>C</th><th></th><th>A</th><th>B</th><th>C</th><th></th><th>A</th><th>B</th><th>C</th></tr><tr><td>P0</td><td>0</td><td>1</td><td>0</td><td></td><td>0</td><td>0</td><td>0</td><td></td><td>0</td><td>0</td><td>0</td></tr><tr><td>P1</td><td>2</td><td>0</td><td>0</td><td></td><td>2</td><td>0</td><td>2</td><td></td><td colspan="3" rowspan="4"></td></tr><tr><td>P2</td><td>3</td><td>0</td><td>3</td><td></td><td>0</td><td>0</td><td>0</td><td></td></tr><tr><td>P3</td><td>2</td><td>1</td><td>1</td><td></td><td>1</td><td>0</td><td>0</td><td></td></tr><tr><td>P4</td><td>0</td><td>0</td><td>2</td><td></td><td>0</td><td>0</td><td>2</td><td></td></tr></table>Using deadlock detection algorithm test whether the system is in a deadlock state.</td><td>10</td><td>CO3</td><td>1.4.1</td><td>L3</td></tr>
<tr><td>b</td><td>With neat diagram describe the Paging hardware with translation look-aside buffer (TLB). Also calculate the effective memory access time for the following specification.
Hit ratio = 60%.Memory access time = 200 nano seconds
TLB search time = 50 nano seconds</td><td>10</td><td>CO3</td><td>2.1.3</td><td>L3</td></tr>
<tr><td colspan="6">Unit III</td></tr>
<tr><td>7a</td><td>Discuss the strengths and weaknesses of implementing an access matrix using access lists that are associated with objects.</td><td>10</td><td>CO4</td><td>1.4.1</td><td>L2</td></tr>
<tr><td>b</td><td>Explain why a capability-based system such as Hydra provides greater flexibility than the ring-protection scheme in enforcing protection policies.</td><td>10</td><td>CO4</td><td>1.4.1</td><td>L3</td></tr>
<tr><td>8a</td><td>Explain overall architecture of Android Operating system. What are the benefits of using Android for Mobile application development</td><td>10</td><td>CO4</td><td>1.4.1</td><td>L2</td></tr>
<tr><td>b</td><td>Discuss design issues of network operating systems. Comment on use of dedicated routers/gateways against general purpose computers for networking.</td><td>10</td><td>CO4</td><td>1.4.1</td><td>L3</td></tr>
</table>





***CO5 is addresses through Course Activity.**

BL – Bloom's Taxonomy Levels (1- Remembering, 2- Understanding, 3 – Applying, 4 – Analysing, 5 – Evaluating, 6 - Creating)
CO – Course Outcomes
PO – Program Outcomes; PI Code – Performance Indicator Code

Competency addressed in the Course and corresponding Performance Indicators

Competency	Performance Indicators
1.4: Demonstrate competence in computer science engineering knowledge.	1.4.1. Apply knowledge of suitable data structures and / or programming paradigm to solve problems.
2.1 :Demonstrate an ability to identify and characterize an engineering problem	2.1.3. Identify the mathematical, engineering and other relevant knowledge that applies to a given problem.

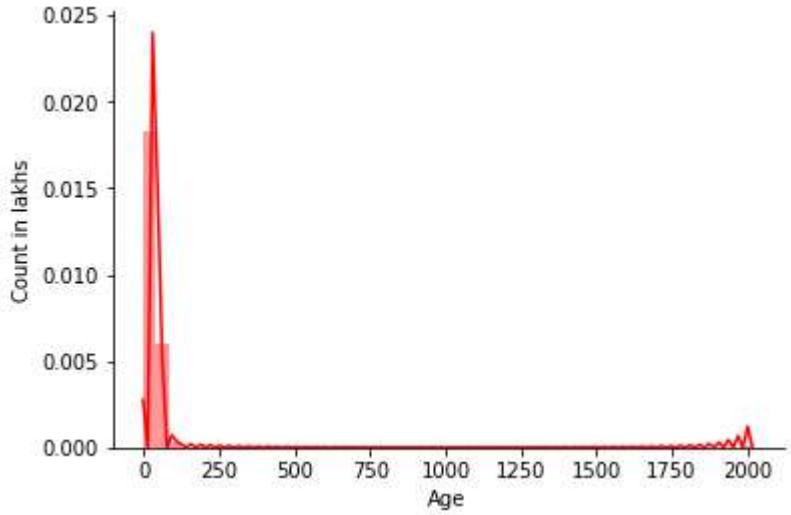
Model Question Paper
Total Duration (H:M): 3:00
Course : Data Mining and Analysis
Maximum Marks :100

<table>
<tr><th>Q.No.</th><th>Questions</th><th>Marks</th><th>CO</th><th>BL</th><th>PI</th></tr>
<tr><td>1a)</td><td>Answer the following questions with justification.
(i) Is noise ever interesting or desirable?
(ii) Can noise objects be outliers?
(iii) Are noise objects always outliers?
(iv) Are outliers always noise objects?
(v) Can noise make a typical value into an unusual one, or vice versa?</td><td>10</td><td>CO1</td><td>L3</td><td>1.1.3</td></tr>
<tr><td>1b)</td><td>Consider the following database of travel sequences for one working week of a traveler:
<table><tr><th>Day</th><th>Sequence</th></tr><tr><td>Mon</td><td>A, U, H, U, A</td></tr><tr><td>Tue</td><td>A, U, H, U, B</td></tr><tr><td>Wed</td><td>B, A</td></tr><tr><td>Thu</td><td>A, U, H, U, A</td></tr><tr><td>Fri</td><td>A, U, B</td></tr></table>(i) Use the Apriori algorithm to compute all frequent itemsets, and their support, with minimum support 3. Clearly indicate the steps of the algorithm, and the pruning that is performed.
(ii) Which of the frequent sequences are maximal?
(iii) Which of the frequent sequences are closed?</td><td>10</td><td>CO2</td><td>L3</td><td>2.1.4</td></tr>
<tr><td>2a)</td><td>(i) For the Traveler dataset given in Figure 3.a, answer the following questions:
[A] Which kind of plots are suitable for each of the 'gender', 'signup_method', 'first_device_type', 'first_browser' and 'timestamp_first_active' attributes.
[B] Which kind of plots are suitable for numerical data, give example.
[C] Classify each of the following attributes {'id', 'date_account_created', 'timestamp_first_active', 'gender', 'age', 'country_destination'} as qualitative (nominal or ordinal or binary) or quantitative (discrete or continuous). Some cases may have more than one interpretation, briefly indicate your reasoning if you think there may be some ambiguity.

(ii) List and briefly describe two other techniques for numerosity reduction.</td><td>6 + 4</td><td>CO1</td><td>L3</td><td>1.4.1</td></tr>
</table>

<table>
<tr>
<td>2b)</td>
<td>
(i) Draw a contingency table for each of the rules using the transactions shown in Table 2.b.
Rules: [A] {b} → {c}; [B] {a} → {d}; [C] {b} → {d};
[D] {e} → {c}; [E] {c} → {a};
<table>
<tr><th colspan="2">Table 2.b: Market basket transactions</th></tr>
<tr><th>Transaction ID</th><th>Item Brought</th></tr>
<tr><td>0001</td><td>{a, b, d, e}</td></tr>
<tr><td>0002</td><td>{b, c, d}</td></tr>
<tr><td>0003</td><td>{a, b, d, e}</td></tr>
<tr><td>0004</td><td>{a, c, d, e}</td></tr>
<tr><td>0005</td><td>{b, c, d, e}</td></tr>
<tr><td>0006</td><td>{b, d, e}</td></tr>
<tr><td>0007</td><td>{c, d}</td></tr>
<tr><td>0008</td><td>{a, b, c}</td></tr>
<tr><td>0009</td><td>{b, d, e}</td></tr>
<tr><td>0010</td><td>{b, d}</td></tr>
</table>
(ii) Use the contingency tables obtained in part (i) to compute and rank the rules in decreasing order according to the following measures:
[A] Support; [B] Confidence; [C] Interest (X → Y) = P(X, Y) / (P(X)*P(Y))
</td>
<td>5
+
5</td>
<td>CO2</td>
<td>L3</td>
<td>2.1.4</td>
</tr>
<tr>
<td>3a)</td>
<td>
<table>
<tr><th></th><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th></tr>
<tr><td>id</td><td>gxn3p5htnn</td><td>820tgsjxq7</td><td>4ft3gnwmtx</td><td>bjjt8pjhuk</td><td>87mebub9p4</td></tr>
<tr><td>date_account_created</td><td>2010-06-28</td><td>2011-05-25</td><td>2010-09-28</td><td>2011-12-05</td><td>2010-09-14</td></tr>
<tr><td>timestamp_first_active</td><td>20090319043255</td><td>20090523174809</td><td>20090609231247</td><td>20091031060129</td><td>20091208061105</td></tr>
<tr><td>date_first_booking</td><td>NaN</td><td>NaN</td><td>2010-08-02</td><td>2012-09-08</td><td>2010-02-18</td></tr>
<tr><td>gender</td><td>-unknown-</td><td>MALE</td><td>FEMALE</td><td>FEMALE</td><td>-unknown-</td></tr>
<tr><td>age</td><td>NaN</td><td>38</td><td>56</td><td>42</td><td>41</td></tr>
<tr><td>signup_method</td><td>facebook</td><td>facebook</td><td>basic</td><td>facebook</td><td>basic</td></tr>
<tr><td>signup_flow</td><td>0</td><td>0</td><td>3</td><td>0</td><td>0</td></tr>
<tr><td>language</td><td>en</td><td>en</td><td>en</td><td>en</td><td>en</td></tr>
<tr><td>affiliate_channel</td><td>direct</td><td>seo</td><td>direct</td><td>direct</td><td>direct</td></tr>
<tr><td>affiliate_provider</td><td>direct</td><td>google</td><td>direct</td><td>direct</td><td>direct</td></tr>
<tr><td>first_affiliate_tracked</td><td>untracked</td><td>untracked</td><td>untracked</td><td>untracked</td><td>untracked</td></tr>
<tr><td>signup_app</td><td>Web</td><td>Web</td><td>Web</td><td>Web</td><td>Web</td></tr>
<tr><td>first_device_type</td><td>Mac Desktop</td><td>Mac Desktop</td><td>Windows Desktop</td><td>Mac Desktop</td><td>Mac Desktop</td></tr>
<tr><td>first_browser</td><td>Chrome</td><td>Chrome</td><td>IE</td><td>Firefox</td><td>Chrome</td></tr>
<tr><td>country_destination</td><td>NDF</td><td>NDF</td><td>US</td><td>other</td><td>US</td></tr>
</table>
Fig. 3. a.: Sample Traveler dataset

(i) For the Traveler dataset given in Figure 3. a, write the appropriate one or two line python code for the following questions:
[A] Code: To find what is the percentage of data missing in each of the attributes?
[B] Inference: What kind of analysis do you infer from distribution plot shown below.
</td>
<td>5
+
5</td>
<td>CO1</td>
<td>L3</td>
<td>1.4.1</td>
</tr>
</table>



[C] **Code:** To find are there any outliers in the age attribute and specify how do you remove noise and outliers from the age attribute.

(ii) Write your **Inference** from these codes (`users` is a dataFrame depicted in Fig. 3.a).

[D]
```
users.gender.replace('-unknown-', np.nan,
inplace=True)
users.gender.value_counts(dropna =
False).plot(kind='bar',
color='red', rot=45)
```

[E]
```
users['date_first_active'] =  pd.to_datetime(
(users.timestamp_first_active),format='%Y%m%d')
```

[F]
```
gender0 = sum(users['gender'] == 'FEMALE')
gender1 = sum(users['gender'] == 'MALE')
gender0_0 = users.loc[users['gender'] ==
'FEMALE',
country_destination'].value_counts() / gender0 *
100
gender1_1 = users.loc[users['gender'] == 'MALE',
'country_destination'].value_counts() / gender1
* 100
```

<table>
<tr>
<td>3b)</td>
<td>
True or False/ Yes or No/ Short answers/Inference from analysis.

(i) [True/False] In comparison with the performance of a base classifier on a particular data set, bagging will generally not increase the error whereas as boosting may lead to an increase in the error.

(ii) [True/False] The use of weak classifiers prevents overfitting when we perform bagging.

(iii) [True/False] If one feature (compared to all others) is a very strong predictor of the class label of the output variable, then all of the trees in a random forest will have this feature as the root node.

(iv) [True/False] Is AdaBoost sensitive to outliers?

(v) [Yes/No] On a particular data set, we use the ensemble method approach to building a predictor and achieve state of the art performance. Is it possible for some of the individual models in this ensemble to have poor performance as measured on the training data?

(vi) [Yes/No] Consider an alternative way of learning a Random Forest where instead of randomly sampling the attributes at each node, we sample a subset of attributes for each tree and build the tree on these features. Would you prefer this method over the original or not, and why?

(vii) [Yes/No] Can the boosting technique be applied on regression problems? Can bagging be applied on regression problems?

(viii) While working on a recognition problem, we implemented four different classifiers. You are now trying to compare their performance. Their misclassification rates are as follows:
<table>
<tr><th></th><th>Classifier P</th><th>Classifier Q</th><th>Classifier R</th><th>Classifier S</th></tr>
<tr><td>Error rate on training data</td><td>25%</td><td>5%</td><td>10%</td><td>20%</td></tr>
<tr><td>Error rate on testing data</td><td>30%</td><td>20%</td><td>15%</td><td>25%</td></tr>
</table>
Write Classifier P, Q, R, or S.

[A] Which classifier has the best generalization performance, i.e., most likely would perform the best when applied to unseen data?

[B] Which classifier is underfitting the most?

[C] Which classifier is overfitting the most?
</td>
<td>10</td>
<td>CO3</td>
<td>L3</td>
<td>2.1.4</td>
</tr>
</table>

<table>
<tr>
<td>4a)</td>
<td>Given the contingency table, compute the following pairwise measures for clustering evaluation:
(i) Jaccard coefficient
(ii) Rand statistic
(iii) Fowlkes–Mallows measure

Contingency table
<table><tr><th>C\T</th><th>T1</th><th>T2</th><th>T3</th></tr><tr><td>C1</td><td>0</td><td>47</td><td>14</td></tr><tr><td>C2</td><td>50</td><td>0</td><td>0</td></tr><tr><td>C3</td><td>0</td><td>3</td><td>36</td></tr></table></td>
<td>10</td><td>CO3</td><td>L2</td><td>1.1.3</td>
</tr>
<tr>
<td>4b)</td>
<td>Solve the following HAC by considering the proximity matrix given, and consider inter-cluster similarity between clusters using:
(i) Nearest neighbor linkage
(ii) Farthest neighbor linkage
And draw the corresponding dendograms.

Proximity matrix
<table><tr><th></th><th>P1</th><th>P2</th><th>P3</th><th>P4</th><th>P5</th></tr><tr><td>P1</td><td>1.00</td><td>0.10</td><td>0.41</td><td>0.55</td><td>0.35</td></tr><tr><td>P2</td><td>0.10</td><td>1.00</td><td>0.64</td><td>0.47</td><td>0.98</td></tr><tr><td>P3</td><td>0.41</td><td>0.64</td><td>1.00</td><td>0.44</td><td>0.85</td></tr><tr><td>P4</td><td>0.55</td><td>0.47</td><td>0.44</td><td>1.00</td><td>0.76</td></tr><tr><td>P5</td><td>0.35</td><td>0.98</td><td>0.85</td><td>0.76</td><td>1.00</td></tr></table></td>
<td>10</td><td>CO2</td><td>L3</td><td>2.1.4</td>
</tr>
<tr>
<td>5a)</td>
<td>Consider the following data set:

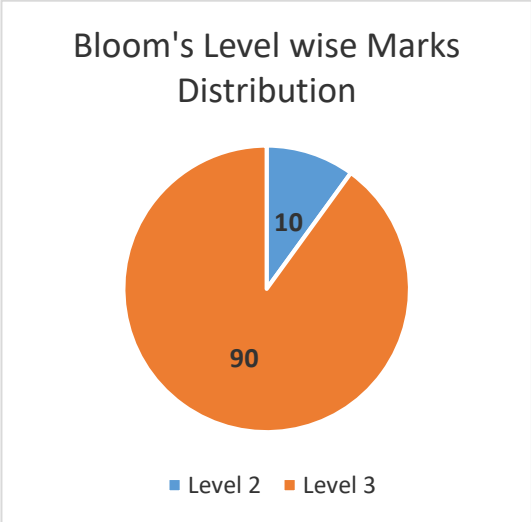
<table><tr><th>V</th><th>W</th><th>X</th><th>Y</th></tr><tr><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>0</td><td>1</td></tr><tr><td>1</td><td>0</td><td>0</td><td>1</td></tr><tr><td>1</td><td>1</td><td>0</td><td>0</td></tr><tr><td>1</td><td>1</td><td>1</td><td>0</td></tr></table>
Your task is to build a decision tree for classifying variable Y. (You can think of the data set as replicated many times, i.e. overfitting is not an issue here).

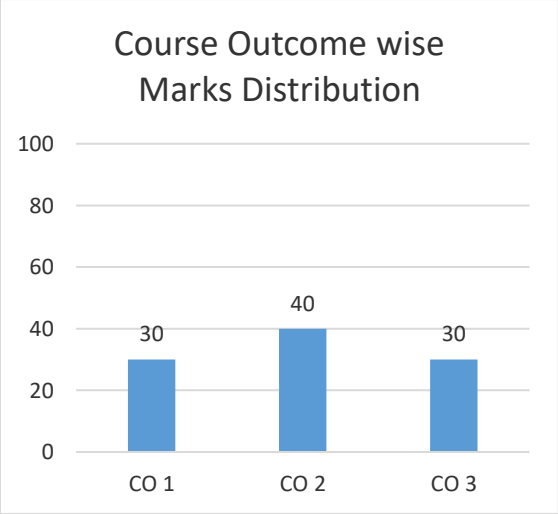
(i) Compute the information gains IG(Y | V), IG(Y | W) and IG(Y | X). Which attribute would Decision Tree classifier select first?

(ii) Write down the entire decision tree constructed by Decision Tree classifier, without pruning.

(iii) One idea for pruning would be to start at the root, and prune splits for which the information gain (or some other criterion) is less than some small ϵ.</td>
<td>10</td><td>CO2</td><td>L3</td><td>2.1.4</td>
</tr>
</table>

	This is called top-down pruning. What is the decision tree returned for **ϵ** = 0.0001? What is the training set error for this tree? **(iv)** Another option would be to start at the leaves, and prune subtrees for which the information gain (or some other criterion) of a split is less than some small **ϵ**. In this method, no ancestors of children with high information gain will get pruned. This is called bottom-up pruning. What is the tree returned for **ϵ** = 0.0001? What is the training set error for this tree?				
5b)	**i)** You are given a training set of five real-valued points and their 2-class classifications (+or -): (1.5, +), (3.2, +), (5.4, -), (6.2, -), (8.5, -). **[A]** What is the predicted class for a test example at point 4.0 using 3-NN? **[B]** What is the decision boundary associated with this Training set using 3-NN? (Hint: The boundary is defined by a single real value.) **[C]** True or False (Justify): For *any* 2-class, linearly-separable Training set (e.g., the one given above), a 3-NN classifier will always have 100% accuracy on the Training set. **(ii)** Say we have a Training set consisting of 30 positive examples and 10 negative examples where each example is a point in a two-dimensional, real-valued feature space. **[A]** What will the classification accuracy be on the Training set with 1-NN? **[B]** What will the classification accuracy be on the Training set with 40-NN?	10	CO3	L3	2.1.4





BL – Bloom's Taxonomy Levels (1- Remembering, 2- Understanding, 3 – Applying, 4 – Analysing, 5 – Evaluating, 6 - Creating)
CO – Course Outcomes
PO – Program Outcomes; PI Code – Performance Indicator Code