SEMESTER VIII

CDT 402	DEEP LEARNING FOR DATA SCIENCE	CATEGORY	L	T	P	CREDIT
		PCC	2	1	0	3

Preamble: Study of this course provides the learners an overview of the concepts and algorithms involved in deep learning. The course covers the basic concepts in neural networks, deep learning, optimization techniques, regularization techniques, convolutional neural networks, recurrent neural networks, autoencoders, generative models. The students will be able to implement deep learning algorithms to solve real-world problems.

Prerequisite: Sound knowledge in concepts of Machine learning.

Course Outcomes: After the completion of the course the student will be able to

CO 1	Illustrate the basic concepts of neural networks, deep learning and its practical issues **(Cognitive Knowledge Level : Apply)**
CO 2	Describe the standard regularization and optimization techniques for the effective training of deep neural networks. **(Cognitive Knowledge Level: Understand)**
CO 3	**Build convolutional Neural Network (CNN) models for different use cases. (Cognitive Knowledge Level: Apply)**
CO 4	Apply the concepts of Recurrent Neural Network (RNN), Long Short Term Memory(LSTM), Gated Recurrent Unit (GRU**). (Cognitive Knowledge Level: Apply)**
CO 5	Explain the concepts of auto encoder, generative models **(Cognitive Knowledge Level: Understand)**

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO6	PO 7	PO 8	PO 9	PO 10	PO1 1	PO12
CO1	✓	✓	✓	✓								✓
CO2	✓	✓	✓	✓								✓
CO3	✓	✓	✓	✓	✓							✓
CO4	✓	✓	✓	✓	✓							✓
CO5	✓	✓	✓	✓								✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and team work
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Lifelong learning

Assessment Pattern

Bloom's Category	**Continuous Assessment Tests**		**End Semester Examination Marks**
	Test1 (percentage)	**Test2 (percentage)**	
Remember	20	20	20
Understand	50	50	50
Apply	30	30	30
Analyse			
Evaluate			
Create			

Mark distribution

Total Marks	**CIE Marks**	**ESE Marks**	**ESE Duration**
150	50	100	3 hours

Continuous Internal Evaluation Pattern:

Attendance: 10 marks

Continuous Assessment Tests : 25 marks

Continuous Assessment Assignment: 15 marks

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks

First Internal Examination shall be preferably conducted after completing the first half of the syllabus and the Second Internal Examination shall be preferably conducted after completing the remaining part of the syllabus.

There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly covered module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly covered module), each with 7 marks. Out of the 7 questions in Part B, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 questions from each module of which a student should answer any one. Each question can have maximum 2 sub-divisions and carries 14 marks.

Syllabus

Module 1: Introduction to Neural Networks and Deep learning

Introduction, The Basic Architecture of Neural Networks - Single Computational Layer: The Perceptron, Multilayer Neural Networks. Activation functions – Sign, Sigmoid, Tanh, ReLU, leaky ReLU, Hard Tanh, Softmax. Loss function. Training a Neural Network with Backpropagation. Practical issues in neural network training. Overfitting, Underfitting, Hyper parameters and Validation sets, Estimators -Bias and Variance. Introduction to deep learning, Deep feed forward network.

Module 2: Training deep models

Introduction, setup and initialization issues, Vanishing and exploding gradient problems, Optimization techniques - Gradient Descent (GD), Stochastic GD, GD with momentum, GD with Nesterov momentum, AdaGrad, RMSProp, Adam. Regularization Techniques - L1 and L2

regularization, Early stopping, Dataset augmentation, Parameter tying and sharing, Ensemble methods, Dropout.

Module 3: Convolutional Neural Networks

Convolutional Neural Networks –Architecture, Convolution operation, Motivation, pooling. Variants of convolution functions, Structured outputs, Data types, Efficient convolution algorithms, Applications of Convolutional Networks, Pre-trained convolutional Architectures : AlexNet, ZFNet, ResNet

Module 4: Recurrent Neural Networks

Recurrent neural networks – Computational graphs. RNN design. Encoder – decoder sequence to sequence architectures. Language modeling example of RNN. Deep recurrent networks. Recursive neural networks. Challenges of training Recurrent Networks. Gated RNNs LSTM and GRU.

Module 5: Autoencoders and Generative models

Autoencoders, Variational AutoEncoder, Undercomplete Autoencoders, Regularized Autoencoders, Denoising Autoencoders, Applications of Autoencoders

Generative models - Boltzmann machines, Deep Belief Networks, Generative Adversarial Networks.

Reference Books

1. Goodfellow, I., Bengio,Y., and Courville, A., Deep Learning, MIT Press, 2016.
2. Neural Networks and Deep Learning, Aggarwal, Charu C., c Springer International Publishing AG, part of Springer Nature 2018
3. Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms (1st. ed.). Nikhil Buduma and Nicholas Locascio. 2017. O'Reilly Media, Inc.

Sample Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Suppose you have a 3-dimensional input x = (x1, x2, x3) = (2, 2, 1) fully connected with weights (0.5, 0.3, 0.2) to one neuron which is in the hidden layer with sigmoid activation function. Calculate the output of the hidden layer neuron.
2. Consider the case of the XOR function in which the two points {(0, 0), (1, 1)} belong to

one class, and the other two points {(1, 0), (0, 1)} belong to the other class. Design a multilayer perceptron for this binary classification problem.

3. Sketch the typical learning curves for the training and validation sets, for a setting where overfitting occurs at some point. Assume that the training set and the validation set are of the same size.

Course Outcome 2 (CO2):

1. Explain how L2 regularization improves the performance of deep feed forward neural networks.
2. Explain how L1 regularization method leads to weight sparsity.
3. Derive update rules for parameters in the multi-layer neural network through the gradient descent.

Course Outcome 3(CO3):

1. Give two benefits of using convolutional layers instead of fully connected ones for visual tasks.
2. Suppose that a CNN was trained to classify images into different categories. It performed well on a validation set that was taken from the same source as the training set but not on a testing set. What could be the problem with the training of such a CNN? How will you ascertain the problem? How can those problems be solved?
3. Weight sharing allows CNNs to deal with image data without using too many parameters. Does weight sharing increase the bias or the variance of a model?

Course Outcome 4 (CO4):

1. Illustrate the workings of the RNN with an example of a single sequence defined on a vocabulary of four words.
2. List the differences between LSTM and GRU
3. Show the steps involved in an LSTM to predict stock prices.Give one advantage of using an RNN rather than a convolutional network.

Course Outcome 5 (CO5):

1. Is an autoencoder for supervised learning or for unsupervised learning? Explain briefly.
2. List the difference between Boltzmann Machine and Deep Belief Network.

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: CDT 402

Course Name: Deep Learning for Data Science

Max.Marks:100 **Duration: 3 Hours**

PART A

Answer All Questions. Each Question Carries 3 Marks

1. Illustrate the limitation of a single layer perceptron with an example.
2. Specify the advantages of ReLU over sigmoid activation function.
3. Write weight updating rule in gradient descent when the error function is a) mean squared error b) cross entropy
4. List any three methods to prevent overfitting in neural networks.
5. Illustrate the strengths and weaknesses of convolutional neural networks.
6. What happens if the stride of the convolutional layer increases? What can be the maximum stride? Justify your answer.
7. List the differences between LSTM and GRU.
8. How does a recursive neural network work?
9. List the difference between Boltzmann Machine and Deep Belief Network.
10. How does the variational auto-encoder(VAE) architecture allow it to generate new data points, compared to auto-encoder, which cannot generate new data points? **(10x3=30)**

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) Explain back propagation algorithm for neural network training. **(9)**

    (b) How does bias and variance trade-off affect machine learning algorithms? **(5)**

OR

12. (a) With an example classification problem, explain the following terms: **(8)**
    a) Hyper parameters b) Training set c) Validation sets d) Bias

    (b) Compare overfitting and underfitting. How can it affect model generalization? **(6)**

13. (a) Differentiate gradient descent with and without momentum. Give equations for weight updation in GD with and without momentum. Illustrate plateaus, saddle points and slowly varying gradients. **(8)**

    (b) Describe the effect in bias and variance when a neural network is modified with more number of hidden units followed with dropout regularization. **(6)**

OR

14. (a) Explain how L2 regularization improves the performance of deep feed forward neural networks. **(7)**

    (b) Initializing the weights of a neural network with very small or large random numbers is not advisable. Justify. **(7)**

15. (a) Consider an activation volume of size 13×13×64 and a filter of size 3×3×64. Discuss whether it is possible to perform convolutions with strides 2, 3 and 5. Justify your answer in each case. **(6)**

    (b) Suppose that a CNN was trained to classify images into different categories. It performed well on a validation set that was taken from the same source as the training set but not on a testing set. What could be the problem with the training of such a CNN? How will you ascertain the problem? How can those problems be solved? **(8)**

OR

16. (a) Explain the following convolution functions a)tensors b) kernel flipping c) down sampling d) strides e) zero padding. **(10)**

    (b) What is the motivation behind convolution neural networks? **(4)**

17. (a) If we have a recurrent neural network (RNN), we can view it as a different type of network by "unrolling it through time". Briefly explain what that means. **(6)**

(b) Explain the architecture of GRU. **(8)**

OR

18. (a) The vanishing gradient problem is more pronounced in RNN than in traditional neural networks. Give reasons. Discuss a solution for the problem. **(6)**

(b) Show the steps involved in an LSTM to predict stock prices.Give one advantage of using an RNN rather than a convolutional network. **(8)**

19. (a) Generative Adversarial Networks(GANs) include a generator and a discriminator. Sketch a basic GAN using those elements, a source of real images, and a source of randomness. **(10)**

(b) The word “adversarial” in the acronym for GANs suggests a two-player game. What are the two players, and what are their respective goals? **(4)**

OR

20. (a) Explain auto encoder with an example. **(7)**

(b) Explain Generative Adversarial Networks using suitable diagram. **(7)**

Teaching Plan

No	Topic	No. of Lectures (**36 Hours**)
1	**Module 1: Introduction to neural network and Deep Learning**	**7**
1.1	Introduction, The Basic Architecture of Neural Networks - Single Computational Layer: The Perceptron.	1 hour
1.2	Multilayer Neural Networks.	1 hour
1.3	Activation functions - Sigmoid, Tanh, ReLU, leaky ReLU, Hard Tanh, Softmax. Loss function.	1 hour
1.4	Training a Neural Network with Backpropagation.	1 hour
1.5	Practical issues in neural network training	1 hour

1.6	Overfitting, Underfitting, Hyper parameters, Validation sets	1 hour
1.7	Estimators -Bias and Variance, Introduction to deep learning, Deep feed forward network	1 hour
2	**Module 2: Training deep models**	**8**
2.1	Introduction, setup and initialization issues	1 hour
2.2	Vanishing and exploding gradient problems	1 hour
2.3	Concepts of optimization, Gradient Descent (GD)	1 hour
2.4	Stochastic GD, GD with momentum, GD with Nesterov momentum	1 hour
2.5	AdaGrad, RMSProp, Adam	1 hour
2.6	Concepts of Regularization, L1 and L2 regularization	1 hour
2.7	Early stopping, Dataset augmentation	1 hour
2.8	Parameter tying and sharing, Ensemble methods, Dropout	1 hour
3	**Module 3: Convolutional Neural Network**	**8**
3.1	Convolutional Neural Networks, Architecture	1 hour
3.2	Convolution operation	1 hour
3.3	Motivation, pooling	1 hour
3.4	Variants of convolution functions	1 hour
3.5	Structured outputs, Data types	1 hour
3.6	Efficient convolution algorithms	1 hour
3.7	Applications of Convolutional Networks	1 hour
3.8	Pretrained Convolutional Architectures : AlexNet, ZFNet,ResNet	1 hour
4	**Module 4 : Recurrent Neural Network**	**7**
4.1	Recurrent neural networks – Computational graphs	1 hour
4.2	RNN design, Encoder – decoder sequence to sequence architectures	1 hour
4.3	Language modeling example of RNN	1 hour
4.4	Deep recurrent networks, Recursive neural networks	1 hour
4.5	Challenges of training Recurrent Networks	1 hour
4.6	LSTM	1 hour
4.7	GRU	1 hour
5	**Module 5 : Autoencoders and Generative models**	**6**
5.1	Autoencoders, Variational AutoEncoder	1 hour
5.2	Undercomplete Autoencoders, Regularized Autoencoders ,	1 hour
5.3	Denoising Autoencoders, Applications of Autoencoders	1 hour
5.4	Boltzmann machines	1 hour
5.5	Deep Belief Networks	1 hour
5.6	Generative Adversarial Networks.	1 hour

CDT404	COMPREHENSIVE COURSE VIVA	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PCC	1	0	0	1	2019

The objective of this Course viva is to ensure the basic knowledge of each student in the most fundamental core courses in the curriculum. The viva voce shall be conducted based on the core subjects studied from third to eighth semester. This course helps the learner to become competent in placement tests and other competitive examinations.

Guidelines

1. The course should be mapped with a faculty and classes shall be arranged for practicing questions based on the core courses listed in the curriculum.
2. The viva voce will be conducted by the same three member committee assigned for final project phase II evaluation. It comprises of Project coordinator, expert from Industry/research Institute and a senior faculty from a sister department.
3. The pass minimum for this course is 25.
4. The mark will be treated as internal and should be uploaded along with internal marks of other courses.
5. Comprehensive Viva should be conducted along with final project evaluation by the three member committee.

Mark Distribution

Total marks: 50, only CIE, minimum required to pass : **25**

CDD416	PROJECT PHASE II	CATEGORY	L	T	P	CREDIT
		PWS	0	0	12	4

Preamble: The course 'Project Work' is mainly intended to evoke the innovation and invention skills in a student. The course will provide an opportunity to synthesize and apply the knowledge and analytical skills learned, to be developed as a prototype or simulation. The project extends to 2 semesters and will be evaluated in the 7th and 8th semester separately, based on the achieved objectives. One third of the project credits shall be completed in 7th semester and two third in 8th semester. It is recommended that the projects may be finalized in the thrust areas of the respective engineering stream or as interdisciplinary projects. Importance should be given to address societal problems and developing indigenous technologies.

Course Objectives

- To apply engineering knowledge in practical problem solving.
- To foster innovation in design of products, processes or systems.
- To develop creative thinking in finding viable solutions to engineering problems.

Course Outcomes [COs]: After successful completion of the course, the students will be able to:

CO1	Model and solve real world problems by applying knowledge across domains (Cognitive knowledge level: **Apply**).
CO2	Develop products, processes or technologies for sustainable and socially relevant applications (Cognitive knowledge level: **Apply**).
CO3	Function effectively as an individual and as a leader in diverse teams and to comprehend and execute designated tasks (Cognitive knowledge level: **Apply**).
CO4	Plan and execute tasks utilizing available resources within timelines, following ethical and professional norms (Cognitive knowledge level: **Apply**).
CO5	Identify technology/research gaps and propose innovative/creative solutions (Cognitive knowledge level: **Analyze**).
CO6	Organize and communicate technical and scientific findings effectively in written and oral forms (Cognitive knowledge level: **Apply**).

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	2	2	2	1	2	2	2	1	1	1	1	2
CO2	2	2	2		1	3	3	1	1		1	1
CO3									3	2	2	1
CO4					2			3	2	2	3	2
CO5	2	3	3	1	2							1
CO6					2			2	2	3	1	1

Abstract POs defined by National Board of Accreditation			
PO #	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	PO7	Environment and Sustainability
PO2	Problem Analysis	PO8	Ethics
PO3	Design/Development of solutions	PO9	Individual and team work
PO4	Conduct investigations of complex problems	PO0	Communication
PO5	Modern tool usage	PO11	Project Management and Finance
PO6	The Engineer and Society	PO12	Lifelong learning

PROJECT PHASE II

Phase 2 Targets

- In depth study of the topic assigned in the light of the report prepared under Phase - I;
- Review and finalization of the approach to the problem relating to the assigned topic.
- Preparing a detailed action plan for conducting the investigation, including teamwork.
- Detailed Analysis/ Modeling / Simulation/ Design/ Problem Solving/Experiment as needed.
- Final development of product/ process, testing, results, conclusions and future directions.
- Preparing a paper for Conference Presentation/ Publication in Journals, if possible.
- Presenting projects in Project Expos conducted by the University at the cluster level and/ or state level as well as others conducted in India and abroad.
- Filing Intellectual Property Rights (IPR) if applicable.
- Preparing a report in the standard format for being evaluated by the Department Assessment Board.
- Final project presentation and viva voce by the assessment board including the external expert.

Evaluation Guidelines & Rubrics

Total: 150 marks (Minimum required to pass: 75 marks).

- Project progress evaluation by guide: 30 Marks.
- Two interim evaluations by the Evaluation Committee: 50 Marks (25 marks for each evaluation).
- Final evaluation by the Final Evaluation committee: 40 Marks
- Quality of the report evaluated by the evaluation committee: 30 Marks

(The evaluation committee comprises HoD or a senior faculty member, Project coordinator and project supervisor. The final evaluation committee comprises of Project coordinator, expert from Industry/research/academic Institute and a senior faculty from a sister department).

Evaluation by the Guide

The guide/supervisor must monitor the progress being carried out by the project groups on regular basis. In case it is found that progress is unsatisfactory it should be reported to the Department Evaluation Committee for necessary action. The presence of each student in the group and their involvement in all stages of execution of the project shall be ensured by the guide. Project evaluation by the guide: 30 Marks. This mark shall be awarded to the students in his/her group by considering the following aspects:

Project Scheduling & Distribution of Work among Team members: Detailed and extensive Scheduling with timelines provided for each phase of project. Work breakdown structure well defined. (5)

Literature survey: Outstanding investigation in all aspects. (4)

Student's Diary/ Daily Log: The main purpose of writing daily diary is to cultivate the habit of documenting and to encourage the students to search for details. It develops the students' thought process and reasoning abilities. The students should record in the daily/weekly activity diary the day to day account of the observations, impressions, information gathered and suggestions given, if any. It should contain the sketches & drawings related to the observations made by the students. The daily/weekly activity diary shall be signed after every day/week by the guide. (7)

Individual Contribution: The contribution of each student at various stages. (9)

Completion of the project: The students should demonstrate the project to their respective guide. The guide shall verify the results and see that the objectives are met. (5)

EVALUATION RUBRICS for PROJECT Phase II: Interim Evaluation - 1						
No.	**Parameters**	**Marks**	**Poor**	**Fair**	**Very Good**	**Outstanding**
2-a	Novelty of idea, and Implementation scope [CO5] [Group Evaluation]	**5**	The project is not addressing any useful requirement. The idea is evolved into a non-implementable one. The work presented so far is lacking any amount of original work by the team.	Some of the aspects of the proposed idea can be implemented. There is still lack of originality in the work done so far by the team. The project is a regularly done theme/topic without any freshness in terms of specifications, features, and/or improvements.	Good evidence of an implementable project. There is some evidence forthe originality of the work done by the team . There is fresh specifications/ features/improvements suggested by the team. The team is doing a design from fundamental principles, and there is some independent learning and engineering ingenuity.	The project has evolved into incorporating an outstandingly novel idea. Original work which is not yet reported anywhere else. Evidence for ingenious way of innovation which is also Implementable. Could be a patentable / publishable work.
			(0 – 1 Marks)	(2 – 3 Marks)	(4 Marks)	(5 Marks)
2-b	Effectiveness of task distribution among team members. [CO3] [Group Evaluation]	**5**	No task distribution of any kind. Members are still having no clue on what to do.	Task allocation done, but not effectively, some members do not have any idea of the tasks assigned. Some of the tasks were identified but not followed individually well.	Good evidence of task allocation being done, supported by project journal entries, identification of tasks through discussion etc. However, the task distribution seems to be skewed, and depends a few members heavily than others. Mostly the tasks are being followed by the individual members.	Excellent display of task identification and distribution backed by documentary evidence of team brainstorming, and project journal entries. All members are allocated tasks according to their capabilities, and as much as possible in an equal manner. The individual members are following the tasks in an excellent manner.
			(0 – 1 Marks)	(2 – 3 Marks)	(4 Marks)	(5 Marks)
2-c	Adherence to project schedule. [CO4] [Group Evaluation]	**5**	Little or no evidence of continued planning or scheduling of the project. The students did not stick to the plan what they were going to build nor plan on what materials / resources to use in the project. The students do not have any idea on the budget required even after the end of phase - I. No project journal kept or the journal.	There is some improvement in the primary plan prepared during phase I. There were some ideas on the materials /resources required, but not really thought out. The students have some idea on the finances required, but they have not formalized a budget plan. Schedules were not prepared. The project journal has no useful details on the project.	Good evidence of planning done and being followed up to a good extent after phase I. Materials were listed and thought out, but the plan wasn't followed completely. Schedules were prepared, but not detailed, and needs improvement. Project journal is presented but it is neither complete nor updated regularly.	Excellent evidence of enterprising and extensive project planning and follow-up since phase I. Continued use of project management/version control tool to trackthe project. Material procurement if applicable is progressing well. Tasks are updated and incorporated in the schedule. A well-kept project journal showed evidence for all the above, in addition to the interaction with the project guide.
			(0 - 1 Marks)	(2 - 3 Marks)	(4 Marks)	(5 Marks)

2-d	Interim Results. [CO6] [Group assessment]	**5**	There are no interim results to show.	The team showed some interim results, but they are not complete / consistent to the current stage, Some corrections are needed.	The interim results showed were good and mostly consistent/correct with respect to the current stage. There is room for improvement.	There were significant interim results presented which clearly shows the progress.
			(0 - 1 Marks)	(2 - 3 Marks)	(4 Marks)	(5 Marks)
2-e	Presentation [Individual assessment]	**5**	Very poor presentation and there is no interim results. The student has no idea about the project proposal.	Presentation is average, and the student has only a feeble idea about the team work.	Good presentation. Student has good idea about the team's project. The overall presentation quality is good.	Exceptionally good presentation. Student has excellent grasp of the project. The quality of presentation is outstanding.
			(0 - 1 Marks)	(2 - 3 Marks)	(4 Marks)	(5 Marks)
Phase-II Interim Evaluation - 1 Total Marks: 25						

EVALUATION RUBRICS for PROJECT Phase II: Interim Evaluation – 2

No	Parameters	Marks	Poor	Fair	Very Good	Outstanding
2-f	Application of engineering knowledge [CO1] [Individual Assessment]	**10**	The student does not show any evidence of applying engineering knowledge on the design and the methodology adopted. The student's contribution in application of engineering knowledge in the project is poor.	The student appears to apply some basic knowledge, but not able to show the design procedure and the methodologies adopted in a comprehensive manner.	The student is able to show some evidence of application of engineering knowledge in the design and development of the project to good extent.	Excellent knowledge in design procedure and its adaptation. The student is able to apply knowledge from engineering domains to the problem and develop solutions.
			(0 – 3 Marks)	(4 – 6 Marks)	(7 - 9 Marks)	(10 Marks)
2-g	Involvement of individual members [CO3] [Individual Assessment]	**5**	No evidence of any Individual participation in the project work.	There is evidence for some amount of individual contribution, but is limited to some of the superficial tasks.	The individual contribution is evident. The student has good amount of involvement in core activities of the project.	Evidence available for the student acting as the core technical lead and has excellent contribution to the project.
			(0 - 1 Marks)	(2 - 3 Marks)	(4 Marks)	(5 Marks)
2-h	Results and inferences upon execution [CO5] [Group Assessment]	**5**	None of the expected outcomes are achieved yet. The team is unable to derive any inferences on the failures/ issues observed. Any kind o f observations or studies are not made.	Only a few of the expected outcomes are achieved. A few inferences are made on the observed failures/issues. No further work suggested.	Many of the expected outcomes are achieved. Many observations and inferences are made, and attempts to identify the issues are done. Some suggestions are made for further work.	Most of the stated outcomes are met. Extensive studies are done and inferences drawn. Most of the failures are addressed and solutions suggested. Clear and valid suggestions made for further work.
			(0 - 1 Marks)	(2 - 3 Marks)	(4 Marks)	(5 Marks)
2-i	Documentation and presentation. .[CO6] [Individual assessment]	**5**	The individual student has no idea on the presentation of his/her part. The presentation is of poor quality.	Presentation's overall quality needs to be improved.	The individual's presentation performance is satisfactory.	The individual's presentation is done professionally and with great clarity. The individual's performance is excellent.
			(0 - 1 Marks)	(2 - 3 Marks)	(4 Marks)	(5 Marks)
Phase-II Interim Evaluation - 2 Total Marks: 25						

EVALUATION RUBRICS for PROJECT Phase II: Final Evaluation						
No	**Parameters**	**Marks**	**Poor**	**Fair**	**Very Good**	**Outstanding**
2-j	Engineering knowledge. [CO1] [Group Assessment]	10	The team does not show any evidence of applying engineering knowledge on the design and the methodology adopted.	The team is able to show some of the design procedure and the methodologies adopted, but not in a comprehensive manner.	The team is able to show evidence of application of engineering knowledge in the design and development of the project to good extent. There is scope for improvement.	Excellent knowledge in design procedure and its adaptation. The team is able to apply knowledge from engineering domains to the problem and develop an excellent solution.
			(0 – 3 Marks)	(4 – 6 Marks)	(7 - 9 Marks)	(10 Marks)
2-k	Relevance of the project with respect to societal and/or industrial needs. [Group Assessment][CO2]	5	The project as a whole do not have any societal / industrial relevance at all.	The project has some relevance with respect to social and/or industrial application. The team has however made not much effort to explore further and make it better.	The project is relevant to the society and/or industry. The team is mostly successful in translating the problem into an engineering specification and managed to solve much of it.	The project is exceptionally relevant to society and/or industry. The team has made outstanding contribution while solving the problem in a professional and/ or ethical manner.
			(0 - 1 Marks)	(2 - 3 Marks)	(4 Marks)	(5 Marks)
2-i	Innovation / novelty / Creativity [CO5] [Group Assessment]	5	The project is not addressing any useful requirement. The idea is evolved into a non-implementable one. The work presented so far is lacking any amount of original work by the team.	Some of the aspects of the proposed idea appears to be practical. There is still lack of originality in the work done. The project is a regularly done theme/topic without any freshness in terms of specifications, features, and/ or improvements.	Good evidence of an implementable project. There is some evidence for the originality of the work done by the team. There is fresh specifications/ features/improvements suggested by the team. The team is doing a design from fundamental principles, and there is some independent learning and engineering ingenuity. Could be translated into a product / process if more work is done.	The project has evolved into incorporating an outstandingly novel idea. Original work which is not yet reported anywhere else. Evidence for ingenious way of innovation which is also Implementable. Could be a patentable publishable work.
			(0 - 1 Marks)	(2 - 3 Marks)	(4 Marks)	(5 Marks)
2-m	Quality of results / conclusions / solutions. [CO1] [Group Assessment]	10	None of the expected outcomes are achieved. The team is unable to derive any inferences on the failures/issues observed. Any kind of observations or studies is not made.	Only a few of the expected outcomes are achieved. A few inferences are made on the observed failures/issues. No further work suggested.	Many of the expected outcomes are achieved. Many observations and inferences are made, and attempts to identify the issues are done. Some suggestions are made for further work.	Most of the stated outcomes are met. Extensive studies are done and inferences drawn. Most of the failures are addressed and solutions suggested. Clear and valid suggestions made for further work.
			(0 – 3 Marks)	(4 – 6 Marks)	(7 - 9 Marks)	(10 Marks)

2-n	Presentation - Part I Preparation of slides. [CO6] [Group Assessment].	5	The presentation slides are shallow and in a clumsy format. It does not follow proper organization.	Presentation slides follow professional style formats to some extent. However, its organization is not very good. Language needs to be improved. All references are not cited properly, or acknowledged. Presentation slides needs to be more professional.	Presentation slides follow a good style format and there are only a few issues. Organization of the slides is good. Most of references are cited properly. The flow is good and team presentation is neatly organized. Some of the results are not clearly shown. There is room for improvement.	The presentation slides are exceptionally good. Neatly organized. All references cited properly. Diagrams/Figures, Tables and equations are properly numbered, and l i s ted. Results/ inferences clearly highlighted and readable.
			(0 - 1 Marks)	(2 - 3 Marks)	(4 Marks)	(5 Marks)
	Presentation - Part II: Individual Communication [CO6] [Individual Assessment].	5	The student is not communicating properly. Poor response to questions.	The student is able to explain some of the content. The student requires a lot of prompts to get to the idea. There are language issues.	Good presentation/ communication by the student. The student is able to explain most of the content very well. There are however, a few areas where the student shows lack of preparation. Language is better.	Clear and concise communication exhibited by the student. The presentation is outstanding. Very confident and tackles all the questions without hesitation. Exceptional traits of communicator.
			(0 - 1 Marks)	(2 - 3 Marks)	(4 Marks)	(5 Marks)
Phase-II Final Evaluation, Marks: 40						

EVALUATION RUBRICS for PROJECT Phase II: Report Evaluation						
Sl. No.	**Parameters**	**Marks**	**Poor**	**Fair**	**Very Good**	**Outstanding**
2-o	Report [CO6]	30	The prepared report is shallow and not as per standard format. It does not follow proper organization. Contains mostly unacknowledged content. Lack of effort in preparation is evident. References are not cited. Unprofessional and inconsistent formatting.	Project report follows the standard format to some extent. However, its organization is not very good. Language needs to be improved. All references are not cited properly in the report. There is lack of formatting consistency.	Project report shows evidence of systematic documentation. Report is mostly following the standard style format and there are only a few issues. Organization of the report is good. Mostly consistently formatted. Most of references/sources are cited/ acknowledged properly.	The report is exceptionally good. Neatly organized. All references cited properly. Diagrams/Figures, Tables and equations are properly numbered, and listed and clearly shown. Language is excellent and follows professional styles. Consistent formatting and exceptional readability.
			(0 - 11 Marks)	(12 - 18 Marks)	(19 - 28 Marks)	(29 - 30 Marks)
Phase - II Project Report Marks: 30						

SEMESTER VIII

PROGRAM ELECTIVE III

AIT 424	Introduction to Business Analytics	CATEGORY	L	T	P	CREDIT
		PEC	2	1	0	3

Preamble: The course aims to introduce the fundamental concepts of business analytics to students. This involves basic concepts of business analytics, descriptive analytics, predictive analytics, forecasting techniques, prescriptive analytics and to apply the appropriate analytics for generating solutions.

Prerequisite: Basic knowledge in Probability and Statistical Modelling.

Course Outcomes: After the completion of the course the student will be able to gain an understanding of how managers use business analytics to formulate and solve business problems and to support managerial decision making.

CO 1	Explain the concept of Business Analytics process and the role of Business Analytics in decision making**. (Cognitive Knowledge level: Understand)**
CO 2	Use appropriate methods for solving problems in Descriptive analytics (**Cognitive knowledge level: Apply)**
CO 3	Use appropriate methods to solve problems using Predictive analytics techniques. (**Cognitive Knowledge level: Apply)**
CO 4	Use appropriate forecasting techniques to inference analyze business trends. (**Cognitive Knowledge level: Apply)**
CO 5	Formulate linear programming model for solving a problem **(Cognitive Knowledge level: Apply)**

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓				✓							✓
CO2	✓	✓	✓		✓							✓
CO3	✓	✓	✓		✓							✓
CO4	✓	✓	✓		✓							✓
CO5	✓	✓	✓		✓							✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and team work
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Lifelong learning

Assessment Pattern

Bloom's Category	**Continuous Assessment Tests**		**End Semester Examination Marks**
	Test1 (percentage)	**Test2 (percentage)**	
Remember	20	20	20
Understand	40	40	40
Apply	40	40	40
Analyse			
Evaluate			
Create			

Mark distribution

Total Marks	**CIE Marks**	**ESE Marks**	**ESE Duration**
150	50	100	3 hours

Continuous Internal Evaluation Pattern:

Attendance: 10 marks

Continuous Assessment Tests : 25 marks

Continuous Assessment Assignment: 15 marks

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks

First Internal Examination shall be preferably conducted after completing the first half of the syllabus and the Second Internal Examination shall be preferably conducted after completing the remaining part of the syllabus.

There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly covered module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly covered module), each with 7 marks. Out of the 7 questions in Part B, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 questions from each module of which a student should answer any one. Each question can have a maximum 2 sub-divisions and carries 14 marks.

Syllabus

Module -1 (Introduction To Business Analytics)

Business Analytics - Terminologies, Business Analytics Process, Importance, Relationship of BA process and Organization Decision-Making process, Managing BA Personnel, Data and Technology. Organization Structures aligning BA. Management Issues – Establishing an Information policy, Outsourcing BA, Data quality, Measuring BA contribution, Change Management in BA.

Module -2 (Descriptive Analytics)

Introduction to Descriptive analytics – Visualizing and Exploring Data – Descriptive Statistics - Sampling and Estimation - Probability Distribution for Descriptive Analytics - Marketing/Planning Case Study Example : Descriptive analytics step in the BA process.

Module -3 (Predictive Analytics)

Introduction to Predictive analytics – Predictive Modeling - Logic and Data Driven Models - Predictive Analysis Modeling and procedure. Data Mining: Simple Illustration of Data

Mining, Data Mining Methodologies. Prescriptive Analysis step in the BA Process - Analysis of Predictive analytics.

Module - 4 (Forecasting Techniques)

Introduction - Types of Variation in Time Series Data - Simple Regression Model - Multiple Regression Models - Simple Exponential Smoothing - Smoothing Averages - Fitting Models to Data - How to Select Models and Parameters for Models - Forecasting Practice Problems.

Module - 5 (Prescriptive Analytics)

Introduction to Prescriptive analytics - Prescriptive Modeling - Non Linear Optimization. Prescriptive step in the BA Analysis – Background Review and Prescriptive Analysis.

Linear Programming – Types of Linear Programming Problems/Models - Linear Programming Problems/Model Elements - Linear Programming Problems/Model Formulation Procedure.

Text Books

1. Marc J. Schniederjans, Dara G. Schniederjans and Christopher M. Starkey, "Business Analytics Principles, Concepts, and Applications - What, Why, and How", Pearson Ed, 2014.
2. James R. Evans, "Business Analytics - Methods, Models and Decisions", Pearson Ed, 2012

Reference Books

1. Christian Albright S and Wayne L. Winston, "Business Analytics - Data Analysis and Decision Making", Fifth edition, Cengage Learning, 2015.

Sample Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Compare business analytics and organization decision-making process.
2. Explain how business analytics can help an organization achieve a competitive advantage.

Course Outcome 2 (CO2):

1. Describe the sampling methods useful in BA. What is sampling estimation and describe how it can aid in the BA process.
2. The Homes Golf Ball Company has made a number of different golf products over the years. Research on thousands of balls revealed the mean flight distance of its Maximum Fly golf ball product to be 450 yards, with a standard error of the mean of 145 yards. The company is hoping to improve the product to fly an additional 290 yards. What is the probability of the improvement from 450 to 740 yards?

Course Outcome 3(CO3):

1. Discuss the logic-driven and data-driven models used in Business analytics.
2. With an investment of $100,000 in radio commercials and $300,000 in TV commercials, what is the prediction on dollar product sales. Use the formula

   $Y_p = -17150.4555 + 275.691X_1 + 48.341X_2$ where

   Y_p= the estimated number of dollars of product sales

   X_1 = the number of dollars to invest in radio commercials

   X_2 = the number of dollars to invest in TV commercials

Course Outcome 4 (CO4):

1. What is forecasting accuracy? Discuss the most commonly used forecast accuracy statistics.
2. Give the forecasting model formula for a weighted moving average. Using a two-value (k) moving average with equal weights of 0.5?

Time Period	Sales
1	49
2	56
3	67
4	78

Course Outcome 5 (CO5):

1. Explain how to formulate a linear programming model?
2. A trucking firm must transport exactly 900, 800, 700 and 1000 units of a product to four cities: A, B, C and D. The product is manufactured and supplied in two other cities X and Y, in the exact amounts to match the total demand. The production of units from the two cities is 1900 and 1500 units respectively to X and Y. The cost per unit to

transport the product between the manufacturing plants in cities X and Y and the demand market cities A, B, C and D are given as :

	DEMAND MARKET			
SUPPLY PLANT	A	B	C	D
X	0.65	0.70	0.80	0.90
Y	0.60	0.60	0.80	0.70

For example, in the table $0.655 is the cost to ship one unit from Supply Plant X to Demand Market A. The trucking firm needs to know how many units should be shipped from each supply city to each demand city in such a way that it minimizes total cost. What is the LP model formulation for this problem?

Model Question paper

QP CODE: PAGES:3

Reg No:_____

Name :_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SEVENTH SEMESTER B. TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: **AIT 424**

Course Name: Introduction to Business Analytics

Max.Marks:100 **Duration: 3 Hours**

PART A

Answer all Questions. Each question carries 3 Marks

1. Explain the relationship of business intelligence to the subject of business analytics.
2. Justify the statement: “Establishing an information policy affect BA”.
3. Differentiate skewedness and kurtosis.
4. What is the 99 percent confidence interval for a problem with a mean value of 120 and a standard error of the mean 20?
5. Illustrate the importance of establishing clusters in BA.

6. How are neural networks helpful in determining both associations and classification tasks required in BA analyses?
7. Differentiate between additive time series model and multiplicative time series model.
8. What is meant by absolute deviation?
9. List the commonly used prescriptive analytics in the business analytics process.
10. How are prescriptive and descriptive analytics related?

(10x3=30)

PART B

Answer any one Question from each module. Each question carries 14 Marks

11. (a) The complete business analytic process involves the three major component steps applied sequentially to a source of data. Justify. (8 marks)

    (b) Compare business analytics and organization decision-making process. (6 marks)

OR

12. (a) Explain how business analytics can help an organization to achieve a competitive advantage. (7 marks)

    (b) Discuss the general management issues related to a BA program. (7 marks)

13. (a) Describe various types of statistical charts and how to apply them. (8 marks)

    (b) Discuss the use of confidence intervals and probability distributions. (6 marks)

OR

14. (a) Describe the sampling methods useful in BA. What is sampling estimation and describe how it can aid in the BA process. (8 marks)

    (b) The Homes Golf Ball Company has made a number of different golf products over the years. Research on thousands of balls revealed the mean flight distance of its Maximum Fly golf ball product to be 450 yards, with a standard error of the mean of 145 yards. The company is hoping to improve the product to fly an additional 290 yards. What is the probability of the improvement from 450 to 740 yards? (6 marks)

15. (a) Discuss the logic-driven and data-driven models used in Business analytics. (7 marks)

(b) With an investment of $100,000 in radio commercials and $300,000 in TV commercials, what is the prediction on dollar product sales. Use the formula (7 marks)

Y_p= -17150.4555 + 275.691X_1 + 48.341X_2 where

Y_p= the estimated number of dollars of product sales

X_1 = the number of dollars to invest in radio commercials

X_2 = the number of dollars to invest in TV commercials

OR

16. (a) Explain how data mining is an ideal predictive analytics tool used in the BA process. (7 marks)

(b) Assume for this problem the following table would have held true for the resulting marketing/planning case study problem. Which combination of variables is estimated here to be the best predictor set? Explain why. (7 marks)

Variable Combination	R –Square	R –Square (Adjusted)	F-Ratio
POS/radio	0.057	0.009	2.977
POS/TV	0.120	0.100	3.662
POS/radio/TV	0.179	0.101	4.315
Radio/TV	0.879	0.853	122.555

17. (a) What is forecasting accuracy? Discuss the most commonly used forecast accuracy statistics. (8 marks)

(b) Give the forecasting model formula for a weighted moving average. Using a two-value (k) moving average with equal weights of 0.5? (6 marks)

Time Period	Sales
1	49
2	56
3	67
4	78

OR

18. (a) Use the following data to construct a linear regression model for the auto insurance premium as a function of driving experience. (6 marks)

Driving Experience (in years)	5	2	12	9	15	6	25	16
Monthly auto insurance premium($)	64	87	50	71	44	56	42	60

(b) Explain multiple regression models with an example. Discuss the limitations on the use of multiple regression models in forecasting time series data. (8 marks)

19. (a) Explain how to formulate a linear programming model? (7 marks)

(b) A trucking firm must transport exactly 900, 800, 700 and 1000 units of a product to four cities: A, B, C and D. The product is manufactured and supplied in two other cities X and Y, in the exact amounts to match the total demand. The production of units from the two cities is 1900 and 1500 units respectively to X and Y. The cost per unit to transport the product between the manufacturing plants in cities X and Y and the demand market cities A, B, C and D are given as : (7 marks)

	DEMAND MARKET			
SUPPLY PLANT	A	B	C	D
X	0.65	0.70	0.80	0.90
Y	0.60	0.60	0.80	0.70

For example, in the table \$0.655 is the cost to ship one unit from Supply Plant X to Demand Market A. The trucking firm needs to know how many units should be shipped from each supply city to each demand city in such a way that it minimizes total cost. What is the LP model formulation for this problem?

OR

20. (a) Explain the linear programming complications that prevent the simplex method from generating a desired optimal solution? (8 marks)

(b) Describe the five necessary assumptions that need to be met for Linear Programming to be used in a modeling situation. (6 marks)

Teaching Plan

	Topics	**No. of Lecture Hours (33)**
	Module - 1 (Introduction To Business Analytics)	**(6 hours)**
1.1	Business Analytics - Terminologies, Business Analytics Process	1 hour
1.2	Relationship of BA process and Organization Decision-Making process	1 hour
1.3	Managing BA Personnel, Data and Technology	1 hour
1.4	Organization Structures aligning BA.	1 hour
1.5	Management Issues – Establishing an Information policy, Outsourcing BA	1 hour

1.6	Data quality, Measuring BA contribution, Change Management in BA	1 hour
Module - 2 (Descriptive Analytics)		**(6 hours)**
2.1	Introduction to Descriptive analytics	1 hour
2.2	Visualizing and Exploring Data, Descriptive Statistics	1 hour
2.3	Sampling and Estimation	1 hour
2.4	Probability Distribution for Descriptive Analytics	1 hour
2.5	Marketing/Planning Case Study Example	1 hour
2.6	Descriptive analytics step in the BA process	1 hour
Module - 3 (Predictive Analytics)		**(7 hours)**
3.1	Introduction to Predictive analytics, Predictive Modeling	1 hour
3.2	Logic and Data Driven Models	1 hour
3.3	Predictive Analysis Modeling and procedure	1 hour
3.4	Data Mining: Simple Illustration of Data Mining	1 hour
3.5	Data Mining Methodologies	1 hour
3.6	Prescriptive Analysis step in the BA Process	1 hour
3.7	Analysis of Predictive analytics.	1 hour
Module - 4 (Forecasting Techniques)		**(7 hours)**
4.1	Introduction - Types of Variation in Time Series Data	1 hour
4.2	Simple Regression Model	1 hour
4.3	Multiple Regression Models	1 hour
4.4	Simple Exponential Smoothing	1 hour
4.5	Smoothing Averages, Fitting Models to Data	1 hour
4.6	How to Select Models and Parameters for Models	1 hour
4.7	Forecasting Practice Problems	1 hour
Module - 5 (Prescriptive Analytics)		**(7 hours)**
5.1	Introduction to Prescriptive analytics - Prescriptive Modeling	1 hour
5.2	Non Linear Optimization	1 hour
5.3	Prescriptive step in the BA Analysis	1 hour
5.4	Background Review and Prescriptive Analysis	1 hour
5.5	Linear Programming – Types of Linear Programming Problems/Models	1 hour

5.6	Linear Programming Problems/Model Elements	1 hour
5.7	Linear Programming Problems/Model Formulation Procedure.	1 hour

CDT 464	BIG DATA SECURITY	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PEC	2	1	0	3	2019

Preamble:
The course explores the foundations of big data, including its foundations in computing technology and statistics. The course also gives an understanding of the nature of underlying technical challenges and statistical assumptions used to understand relationships in a variety of applied fields, with a focus on the fields of fraud detection and communication monitoring.

Course Outcomes: After the completion of the course, the student will be able to

CO#	Course Outcomes
CO1	Explain the basics of Big Data and their challenges. **(Cognitive knowledge level: Understand)**
CO2	Explain the difference between predictive analytics and descriptive analytics **(Cognitive knowledge level: Understand)**
CO3	Trace out the role played by authentication in security**(Cognitive knowledge level: Apply)**
CO4	Describe the security concerns of big-data. **(Cognitive knowledge level: Understand)**
CO5	Escalate the applications of security analytics. **(Cognitive knowledge level: Apply)**

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	P10	P11	P12
CO1	✓	✓	✓									✓
CO2	✓	✓	✓									✓
CO3	✓	✓	✓		✓							✓

CO4	✓	✓	✓		✓							✓
CO5	✓	✓	✓		✓							✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and teamwork
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Lifelong learning

Assessment Pattern

Bloom's Category	**Test 1 (Marks in percentage)**	**Test 2 (Marks in percentage)**	**End Semester Examination (Marks in percentage)**
Remember	30	30	30
Understand	30	30	30
Apply	40	40	40
Analyze			
Evaluate			
Create			

Mark Distribution

Total Marks	**CIE Marks**	**ESE Marks**	**ESE Duration**
150	50	100	3

Continuous Internal Evaluation Pattern:

Attendance : **10 marks**
Continuous Assessment Test : **25 marks**
Continuous Assessment Assignment : **15 marks**

Internal Examination Pattern: Each of the two internal examinations has to be conducted out of 50 marks. The first series test shall be preferably conducted after completing the first half of the syllabus. The second series test shall be preferably conducted after completing the remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly completed module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly completed module), each with 7 marks. Out of the 7 questions, a student should answer any 5.

End Semester Examination Pattern: There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 questions from each module of which a student should answer anyone. Each question can have a maximum 2 subdivisions and carries 14 marks.

Syllabus

Module 1 (Introduction to Big Data)

Introduction to Big Data, Evolution of Big data, Characteristics. Big Data Analytics, Big Data framework - fundamental concepts of Big Data management and analytics - Current challenges and trends in Big Data Acquisition.

Module 2 (Data Analytics)

Predictive Analytics: Regression, Decision Tree, Neural Networks - Descriptive Analytics: Association Rules, Sequence Rules, Survival Analysis: Survival Analysis Measurements, Kaplan Meir Analysis, Parametric Survival Analysis - Social Network Analytics: Social Network Learning Relational Neighbor Classification

Module 3 (Introduction to Security Analytics)

Introduction to Security Analytics – Techniques in Analytics – Analysis in everyday life – Challenges in Intrusion and Incident Identification – Simulation and Security Process, Analytical Software’s and tools, Malware Analysis – static and dynamic analysis - Security Intelligence – Security Breaches

Module 4(Applications of Security Analytics)

Access Analytics – Analysis of Log file -Security analysis with text mining –Machine Learning and data mining applications for security: Intrusion detection and network anomaly detection. Big data analytics for security: Analyzing DDOS – Distributed Denial of Service attack: counter based method, and access pattern based method – Machine learning for Ransomware detection and prevention.

Module 5 (Big Data Privacy and Applications)

Data Masking – Privately Identified Information (PII) -Privacy preservation in Big Data- Popular Big Data Techniques and tools- Map Reduce paradigm and the Hadoop system – Applications- Social Media Analytics- Recommender Systems- Fraud Detection.

Text Books

1. Bart Baesens, "Analytics in a Big Data World: The Essential Guide to Data Science and its Applications", John Wiley & Sons, 2014
2. Min Chen, Shiwen Mao, Yin Zhang, Victor CM Leung, “Big Data: Related Technologies, Challenges and Future Prospects” , Springer, 2014.
3. Michael Minelli, Michele Chambers, AmbigaDhiraj ,“Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends”, John Wiley & Sons, 2013.

References

1. Raj, Pethuru, “Handbook of Research on Cloud Infrastructures for Big Data Analytics”, IGI Global.
2. Douglas R. Stinson ,“Cryptography Theory and Practice ”, Chapman & Hall/CRC, 3rd Edition, 2006.
3. Mark Talabis, Robert McPherson, I Miyamoto and Jason Martin, “Information Security Analytics: Finding Security Insights, Patterns, and Anomalies in Big Data”, Syngress Media, U.S., 2014.

Sample Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Explain the Evolution of Big Data and their characteristics

2. Describe any five characteristics of Big Data.

Course Outcome 2 (CO2):

1. Describe the prediction error and regression techniques

2. Explain the three categories of Prediction methodologies.

Course Outcome 3 (CO3):

1. Identify the various challenges in Intrusion and Incident Identification.

Course Outcome 4 (CO4):

1. How machine learning helps in Ransom ware detection and prevention.

Course Outcome 5 (CO5):

1. What is Privacy preservation? Discuss its importance in Big Data.

Model Question Paper

QP CODE:
Reg No:_____
Name :_____ **PAGES : 3**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

EIGTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: CDT 464

Course Name: BIG DATA SECURITY

Max.Marks:100 **Duration: 3 Hours**

PART A

Answer all Questions. Each question carries 3 Marks

1. Explain the Evolution of Big Data.

2. Describe any five characteristics of Big Data
3. How decision trees aid in big data analytics.
4. What is regression analysis? How is it done?
5. Why data Analytics is needed?
6. How is security analytics done?
7. Explain DDOS attack.
8. How does analysis of log files assist in security analytics?
9. Define Data masking.
10. How map reduce is performed in hadoop?

PART B

Answer any one Question from each module.

Each question carries 14 Marks

11. a) Illustrate the various phases involved in Big Data Analytics with a neat diagram. (7)
    b) Explain the trends in big data acquisition. (7)

OR

12. a) Describe the challenges in big data acquisition (7)

    b) Explain Big data framework. (7)
13. a) Illustrate Kaplan Meir Analysis with an example (8)
    b) Describe Social network Analytics (6)

OR

14. a) Compare predictive and descriptive analysis. (7)
    b) Describe Parametric Survival Analysis. (7)

15. a) Compare and contrast static and dynamic malware analysis. (7)
    b) Describe the various security breaches possibilities in big data scenarios. (7)

OR

16. a) Summarize various challenges in Intrusion and Incident Identification. (8)
    b) Briefly explain various methods used in security analytics. (6)
17. a) Differentiate between counter based method and access pattern based method. (8)

b) Describe how network anomaly detection is done (6)

OR

18. a) How machine learning helps in Ransomware detection and prevention. (8)

b) Explain the scope of Security analysis with text mining. (6)

19. a)Describe Popular Big Data Techniques and tools. (8)

b) Explain the significance of Privately Identified Information. (6)

OR

20.a) How is Privacy preservation in Big Data achieved (8)

b) Describe the role of Recommender Systems (6)

TEACHING PLAN

Sl.No.	Contents	No of Lecture Hrs (35)
Module 1 (Introduction to Big Data) (6 hrs)		
1.1	Introduction to Big Data	1 hour
1.2	Evolution of Big data, Characteristics	1 hour
1.3	Big Data Analytics, Big Data framework	1 hour
1.4	Fundamental concepts of Big Data management and analytics	1 hour
1.5	Current challenges in Big Data Acquisition	1 hour
1.6	Trends in Big Data Acquisition	1 hour
Module 2 (Data Analytics) (8 hrs)		
2.1	Predictive Analytics: Regression, Decision Tree	1 hour

2.2	Neural Networks	1 hour
2.3	Descriptive Analytics: Association Rules, Sequence Rules.	1 hour
2.4	Survival Analysis: Survival Analysis Measurements	1 hour
2.5	Kaplan Meir Analysis	1 hour
2.6	Parametric Survival Analysis	1 hour
2.7	Social Network Analytics	1 hour
2.8	Social Network Learning Relational Neighbor Classification	1 hour
Module 3 (Introduction to Security Analytics) (8 hrs)		
3.1	Introduction to Security Analytics	1 hour
3.2	Techniques in Analytics – Analysis in everyday life	1 hour
3.3	Challenges in Intrusion and Incident Identification	1 hour
3.4	Simulation and Security Process	1 hour
3.5	Analytical Softwares and tools	1 hour
3.6	Malware Analysis	1 hour
3.7	Static and dynamic analysis	1 hour
3.8	Security Intelligence – Security Breaches	1 hour
Module 4(Applications of Security Analytics) (7 hrs)		
4.1	Access Analytics – Analysis of Log file	1 hour
4.2	Security analysis with text mining.	1 hour

4.3	Machine Learning and data mining applications for security:	1 hour
4.4	Intrusion detection and network anomaly detection.	1 hour
4.5	Big data analytics for security: Analyzing DDOS –	1 hour
4.6	Distributed Denial of Service attack: counter based method, and access pattern based method	1 hour
4.7	Machine learning for Ransom ware detection and prevention.	1 hour
Module 5 (Big Data Privacy and Applications) (6 hrs)		
5.1	Data Masking – Privately Identified Information (PII).	1 hour
5.2	Privacy preservation in Big Data.	1 hour
5.3	Popular Big Data Techniques and tools- Map Reduce paradigm	1 hour
5.4	Hadoop system.	1 hour
5.5	Applications- Social Media Analytics	1 hour
5.6	Recommender Systems- Fraud Detection.	1 hour

CST424	PROGRAMMING PARADIGMS	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PEC	2	1	0	3	2019

Preamble: The course provides the learners a clear understanding of the main constructs of contemporary programming languages and the various systems of ideas that have been used to guide the design of programming languages. This course covers the concepts of Names, Bindings & Scope, Statement-Level Control Structures, Sub Programs, Support for Object Oriented Programming, Exception Handling, Concurrency Control, Functional Programming and Logic Programming. This course helps the learners to equip with the knowledge necessary for the critical evaluation of existing and upcoming programming languages. It also enables the learner to choose the most appropriate language for a given programming task, apply that language's approach to structure or organize the code, classify programming languages based on their features and to design new generation languages.

Prerequisite: Sound knowledge in Programming in C and Object-Oriented Programming.

Mapping of course outcomes with program outcomes

CO1	Explain the criteria for evaluating programming languages and compare Imperative, Functional and Logic programming languages (**Cognitive Knowledge Level: Understand**)
CO2	Illustrate the characteristics of data types and variables (**Cognitive Knowledge Level: Apply**)
CO3	Comprehend how control flow structures and subprograms help in developing the structure of a program to solve a computational problem (**Cognitive Knowledge Level: Apply)**
CO4	Explain the characteristics of Object-Oriented Programming Languages (**Cognitive Knowledge Level: Understand)**
CO5	Compare concurrency constructs in different programming languages (**Cognitive Knowledge Level: Understand)**

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓									✓
CO2	✓	✓										✓
CO3	✓	✓	✓	✓								✓
CO4	✓	✓										✓
CO5	✓	✓	✓									✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and team work
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Life long learning

Assessment Pattern

Bloom's Category	**Continuous Assessment Tests**		**End Semester Examination Marks (%)**
	Test 1 (%)	**Test 2 (%)**	
Remember	**30**	**30**	**30**
Understand	**40**	**40**	**40**

Apply	**30**	**30**	**30**
Analyze			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	**50**	**100**	**3**

Continuous Internal Evaluation Pattern:

Attendance **10 marks**

Continuous Assessment Tests (Average of Internal Tests 1 & 2) **25 marks**

Continuous Assessment Assignment **15 marks**

Internal Examination Pattern

Each of the two internal examinations has to be conducted out of 50 marks. First series test shall be preferably conducted after completing the first half of the syllabus and the second series test shall be preferably conducted after completing the remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the two completed modules and 1 question from the partly completed module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed two modules and 1 question from the partly completed module), each with 7 marks. Out of the 7 questions, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 full questions from each module of which student should answer any one. Each question can have maximum 2 sub-divisions and carries 14 marks.

Course Level Assessment Questions

Course Outcome1 (CO1):

1. Compare any three programming languages based on the language evaluation criteria. Prepare a list of characteristics that affect the language evaluation criteria.
2. Identify the advantages and disadvantages of imperative, functional and logic programming languages.

Course Outcome 2 (CO2):

1. Two most important design issues that are specific to character string types are
   (1) whether a string is simply a special kind of character array or a primitive type.
   (2) whether strings have static or dynamic length.
   Identify the implementations options for the above two cases.
2. Consider the following records of a particular language. Let the size of each char variable be 1 byte, int be 4 bytes and and Boolean be 1 bit.

```
Struct Student
{
        int id;
        char name[2];
        int age;
        boolean scholarship;
}
```

   Draw and comment on the possible memory layouts for the record for a 32-bit aligned machine

Course Outcome 3(CO3):

1. Explain three situations where a combined counting and logical looping statement is needed.
2. Describe the ways that aliases can occur with pass-by-reference parameters.
3. Identify the two fundamental design considerations for parameter-passing methods.
4. What will be the output of the given program segment if it uses the following parameter passing mechanisms:
   a) call by reference
   b) call by value

```
x : integer – – global
procedure foo(y : integer)
y := 3
print x
. . .
```

```
x := 2
foo(x)
print x
```

Course Outcome 4 (CO4):

1. Describe the role of a virtual method table in implementing dynamic method binding.
2. Identify the merits and demerits of inheritance.

Course Outcome 5 (CO5):

1. Evaluate the use of semaphores and monitors for providing competition synchronization and cooperation synchronization.

Syllabus

Module – 1

Introduction – Role of Programming Languages, Programming Domains, Language Evaluation Criteria, Influence on Language Design, Language Design Trade-offs, Implementation Methods. Names, Bindings & Scope – Names, Variables, Concept of Binding, Scope and Lifetime, Referencing Environments.

Module - 2

Data Types – Primitive Data Types, Character String Types, User-Defined Ordinal Types, Array Types, Record Types, List Types, Pointer & Reference Types, Type Checking, Strong Typing, Type Equivalence. Expressions – Arithmetic Expressions, Overloaded Operators, Type Conversions, Relational and Boolean Expressions, Short-Circuit Evaluation. Assignment - Assignment Statements, Mixed-mode Assignment.

Module - 3

Statement-Level Control Structures – Selection Statements, Iterative Statements, Unconditional Branching, Guarded Commands. Subprograms – Design Issues of Subprograms, Local Referencing Environments, Parameter Passing Methods, Subprograms as Parameters, Overloaded Subprograms, Closures, Co-routines

Module - 4

Support for Object Oriented Programming – Inheritance, Dynamic Binding, Design Issues for Object Oriented Languages, Support for Object Oriented Programming in C++, Implementation of Object-oriented Constructs. Exception Handling – Basic Concepts, Design Issues.

Module - 5

Concurrency – Subprogram Level Concurrency, Semaphores, Monitors, Message Passing. Functional Programming Languages – Introduction to LISP and Scheme, Comparison of

Functional and Imperative Languages. Logic Programming Languages – Basic Elements of Prolog, Applications of Logic Programming.

Text Books

1. Robert W Sebesta, Concepts of Programming Languages, 10th Edition, Pearson.
2. Scott M L, Programming Language Pragmatics, 3rd Edition, Morgan Kauffman Publishers.

ReferenceBooks

1. Kenneth C. Louden, Programming Languages: Principles and Practice, 2nd Edition, Cengage Learning.
2. Tucker A. B. and R. E. Noonan, Programming Languages: Principles and Paradigms, 2nd Edition. –TMH.
3. Ravi Sethi, Programming Languages: Concepts & Constructs, 2nd Edition., Pearson Education.
4. David A. Watt, Programming Language Design Concepts, Wiley Dreamtech.

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: CST424

Course Name: Programming Paradigms

Max. Marks : 100 **Duration: 3 Hours**

PART A

Answer All Questions. Each Question Carries 3 Marks

1. Differentiate between readability and writability.
2. Define binding and binding time.
3. What are the advantages of user-defined enumeration types?
4. Define narrowing and widening conversions.
5. Why for statement in C language is more flexible than that of older languages?

6. What are the advantages and disadvantages of dynamic local variables in subprograms?

7. Illustrate the concept of dynamic method binding with an example.

8. Is it mandatory to use constructors in object-oriented languages? Justify your answer.

9. What are the applications of logic programming languages?

10. Explain the working of let and let-rec constructs in Scheme.

(10x3=30)

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) Explain different criteria used for evaluating languages. **(7)**

(b) Consider the following pseudocode: **(7)**

```
x : integer := 3
y : integer := 4
procedure add
x := x + y
procedure second(P : procedure)
x : integer := 5
P()
procedure first
y : integer := 6
second(add)
first()

write integer(x)
```

(a) What does this program print if the language uses static scoping? Give reasons.

(b) What does it print if the language uses dynamic scoping? Give reasons.

OR

12. (a) With respect to storage binding, explain the meanings, purposes, advantages and disadvantages of four categories of scalar variables. **(7)**

(b) What is meant by referencing environment of a statement? Show the **(7)**

referencing environment at the indicated program points (1), (2), (3) & (4) for the following program segment. Assume that the programming language is statically scoped.

```
program example;
        var a, b : integer;
        procedure sub1;
                var x, y: integer;
                        begin { sub1 }
                        ……..                    (1)
                        end { sub1 }
        procedure sub2;
                var x : integer;
                ……..
                procedure sub3;
                        var x: integer;
                                begin { sub3 }
                                ……..            (2)
                                end { sub3 }
                        begin { sub2 }
                                ……..            (3)
                        end { sub2}
        begin {example}
                                ……..            (4)
        end {example }
```

13. (a) Explain any two issues associated with the pointer data types and also indicate how dangling pointer problem can be solved. **(7)**

(b) Describe the lazy and eager approaches for reclaiming garbage. **(7)**

OR

14. (a) What is meant by side effect and illustrate the advantages of referential transparency? **(8)**

(b) Explain the terms: compound assignment operator, coercion and short circuit evaluation. **(6)**

15. (a) Illustrate the different categories of iteration control statements. **(8)**

(b) Explain the techniques used for identifying the correct referencing environment for a subprogram that was sent as a parameter. **(6)**

OR

16. (a) Describe the implementation models of Parameter passing. **(10)**

(b) Differentiate coroutines from conventional subprograms. **(4)**

17. (a) What is meant by an exception handler? Explain how exceptions are handled in object-oriented languages. **(7)**

(b) Describe the design issues in object-oriented languages. **(7)**

OR

18. (a) Illustrate how a virtual method table can be used for implementing dynamic method binding. **(7)**

(b) Explain the different categories, merits and demerits of inheritance. **(7)**

19. (a) Compare functional and imperative programming languages. **(7)**

(b) Explain the role of monitors in concurrency. **(7)**

OR

20. (a) Explain the searching strategies used in Prolog. Why backward chaining is preferred over forward chaining in Prolog? **(10)**

(b) **(let ((a 6)**
(b 8)
(square (lambda (x) (* x x)))
(plus +))
(sqrt (plus (square a) (square b)))) **(4)**
Write the output of the above code? Explain how let and lambda construct works?

Teaching Plan

No	Contents	No. of Lecture Hours (36 hrs.)
Module-1 (7 hours)		
1.1	Introduction: Reasons for studying Concepts of programming languages, Programming Domains	1 hour
1.2	Language Evaluation Criteria	1 hour
1.3	Influence on Language Design, Language Design Trade-offs	1 hour
1.4	Implementation Methods	1 hour
1.5	Names, Variables	1 hour
1.6	Concept of Binding	1 hour
1.7	Scope and Lifetime, Referencing Environments	1 hour
Module-2 (7 hours)		
2.1	Primitive Data Types, Character String Types	1 hour
2.2	User-Defined Ordinal Types, Array Types	1 hour
2.3	Record Types, List Types, Pointer and Reference Types	1 hour
2.4	Implementation of pointer and reference types, Type Checking, Strong Typing, Type Equivalence	1 hour
2.5	Expressions and Assignment Statements, Arithmetic Expressions	1 hour
2.6	Overloaded Operators, Type Conversions	1 hour
2.7	Relational and Boolean Expressions, Short-Circuit Evaluation, Assignment Statements, Mixed-mode Assignment	1 hour
Module-3 (8 hours)		
3.1	Selection Statements, Iterative Statements	1 hour
3.2	Unconditional Branching	1 hour

3.3	Guarded Commands	1 hour
3.4	Subprograms: Design Issues of Subprograms	1 hour
3.5	Local Referencing Environments	1 hour
3.6	Parameter Passing Methods	1 hour
3.7	Subprograms as Parameters, Overloaded Subprograms	1 hour
3.8	Closures, Co-routines	1 hour
Module-4 (7 hours)		
4.1	Inheritance	1 hour
4.2	Dynamic Binding	1 hour
4.3	Design Issues for Object Oriented Languages	1 hour
4.4	Support for Object Oriented Programming in C++	1 hour
4.5	Implementation of Object-Oriented Constructs	1 hour
4.6	Exception Handling – Basic Concepts	1 hour
4.7	Exception Handling - Design Issues	1 hour
Module-5 (7 hours)		
5.1	Subprogram Level Concurrency	1 hour
5.2	Semaphores, Monitors	1 hour
5.3	Message Passing	1 hour
5.4	Introduction to LISP and Scheme	1 hour
5.5	Comparison of Functional and Imperative Languages	1 hour
5.6	Basic Elements of Prolog	1 hour
5.7	Applications of Logic Programming	1 hour

CST434	NETWORK SECURITY PROTOCOLS	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PEC	2	1	0	3	2019

Preamble: This course helps the learners to explore various network and system security protocols. This course covers authentication protocols, firewalls and security protocols from different layers such as data link, network, transport and application. The concepts covered in this course enable the learners in effective use of security protocols for securing network applications.

Prerequisite: A fundamental knowledge in the concepts of Security in Computing.

Course Outcomes: After the completion of the course, the student will be able to

CO1	Explain authentication protocols, X.509 authentication service and Public Key Infrastructure (PKI).**(Cognitive Knowledge Level: Understand)**
CO2	Identify the security mechanisms in E mail security services. **(Cognitive Knowledge Level: Understand)**
CO3	Summarize the network and transport layer security services provided in a secure communication scenario. **(Cognitive Knowledge Level: Apply)**
CO4	Describe real time communication security and application layer security protocols. **(Cognitive Knowledge Level: Apply)**
CO5	Explain the concepts of firewalls and wireless network security. **(Cognitive Knowledge Level: Understand)**

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓		✓									✓
CO2	✓	✓	✓									✓
CO3	✓	✓	✓									✓
CO4	✓	✓	✓			✓						✓
CO5	✓	✓	✓									✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and team work
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Life long learning

Assessment Pattern

Bloom's Category	Test 1 (%)	Test 2 (%)	End Semester Examination (%)
Remember	20	20	20
Understand	50	50	50
Apply	30	30	30
Analyse			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	50	100	3

Continuous Internal Evaluation Pattern:

Attendance : **10 marks**

Continuous Assessment Test : **25 marks**

Continuous Assessment Assignment : **15 marks**

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks. First series test shall be preferably conducted after completing the first half of the syllabus and the second series test shall be preferably conducted after completing remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly completed module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly completed module), each with 7 marks. Out of the 7 questions, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 questions from each module of which a student should answer any one. Each question can have maximum 2 sub-divisions and carries 14 marks.

Syllabus

Module-1 (Authentication Protocols)

Authentication Protocols – Mutual authentication, One way authentication. Kerberos – Kerberos Version 4, Kerberos Version 5. X.509 Authentication service. Public Key Infrastructure (PKI) – Trust models, Revocation.

Module-2 (E-mail Security)

Pretty Good Privacy (PGP) – Operational Description, Cryptographic keys and key rings, Message format, PGP message generation, PGP message reception, Public key management. S/MIME – Functionality, Messages, Certificate processing, Enhanced security services.

Module-3 (Network Layer Security and Web Security)

Internet Protocol Security (IPSec) – Overview, IP security architecture, Authentication Header (AH), Encapsulating Security Payload (ESP), Combining Security Associations, Key management. Internet Key Exchange (IKE) - Phases. Web Security – Web security considerations. Secure Socket Layer and Transport Layer Security (SSL/TLS) – SSL Architecture, SSL protocols, Cryptographic computations, Transport layer security.

Module-4 (Real-time Security and Application Layer Security)

Real-time communication security – Perfect Forward Secrecy (PFS), Denial-of-Service protection, Endpoint identifier hiding, Live partner reassurance. Hypertext Transfer Protocol Secure (HTTPS) – Connection initiation, Closure. Secure Shell (SSH) – Transport layer protocol, User authentication protocol, Connection protocol. Secure Electronic Transaction (SET) – Overview, Features, Participants, Dual signature, Payment processing.

Module-5 (System Security and Wireless Security)

Firewalls – Firewall characteristics, Types of Firewalls, Firewall configurations, Encrypted Tunnels, Trusted systems – Data access control, The concept of Trusted Systems, Trojan horse defense. IEEE 802.11i wireless LAN security - Services, Phases of operation, Wired Equivalent Privacy (WEP), Wi-Fi Protected Access (WPA), WPA2.

Text Books

1. William Stallings, Cryptography and Network Security Principles and Practice, 4/e, Pearson Ed.
2. C. Kaufman, R. Perlman and M. Speciner, "Network Security: Private Communication in a Public World", 2/e, PHI.

References

1. Behrouz A. Forouzan, DebdeepMukhopadhyay, "Cryptography and Network Security", 3/e, Tata McGraw Hill.
2. Tyler Wrightson, "Wireless Network Security A Beginner's Guide", 2012, Tata McGraw Hill.
3. William Stallings, "Network Security Essentials: Applications and Standards", 4/e, Prentice Hall.
4. Schiller J., Mobile Communications, 2/e, Pearson Education.
5. Roberta Bragg et. al., "Network Security: The Complete Reference", Tata McGraw Hill

Sample Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Identify the threats associated with user authentication over a network or Internet.
2. In the context of Kerberos, mention the significance of a realm.

Course Outcome 2 (CO2):

1. Mention the use of R64 conversion for an e-mail application.
2. Give the general structure of Private and Public Key rings in PGP.

Course Outcome 3 (CO3):

1. In AH protocol, identify the fields in an IP header which are included in MAC calculation. For each of the fields in the IP header, indicate whether the field is immutable, mutable but predictable, or mutable. Justify your decision for each field.

2. Is it possible for the receiver to reorder SSL record blocks that arrive out of order? If so, explain how it can be done. If not, why?

Course Outcome 4 (CO4):

1. Devise a protocol based on a pre-shared secret key that hides identities and gives Perfect Forward Secrecy (PFS) for identity hiding. Make two variants, one in which an active attacker can learn only the initiator's identity, and one in which an active attacker can learn only the target's identity.
2. Explain the tasks performed by the payment gateway during Payment Authorization in SET.

Course Outcome 5 (CO5):

1. List the weaknesses of a packet-filtering router.
2. Give the relevance of pair wise keys and group keys in IEEE 802.11i.
3. State the design goals of firewalls.

Model Question Paper

QP CODE: **PAGES: ___**

Reg No:_____
Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: CST434
Course Name: NETWORK SECURITY PROTOCOLS

Max Marks: 100 **Duration: 3 Hours**

PART A
(Answer All Questions. Each question carries 3 marks)

1. List any three requirements of Kerberos.
2. Specify the significance of key pair recovery. When is the key pair updated?
3. Why does PGP generate signature before applying compression?
4. List the four principal services provided by S/MIME.
5. Explain the significance of Alert protocol in SSL and list out any three Alert messages with their uses.
6. Specify the purpose of MAC during the change cipher spec TLS exchange.

7. What is the advantage, if any, of not including the MAC in the scope of packet encryption in SSH packets?

8. Givethe significance of dual signature in SET.

9. List the IEEE 802.11i services.

10. How is the concept of association related to that of mobility in wireless networks? **(10x3=30)**

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) Describe the requirements for a public-key certificate scheme. **(8)**

    (b) Explain the significance of chain of certificates. **(6)**

OR

12. (a) Specify the purpose of the X.509 standard. How is an X.509 certificate revoked? **(8)**

    (b) Describe the management functions of a PKI. What is a cross certificate? **(6)**

13. (a) List the services provided by PGP and explain how authentication and confidentiality are provided. **(8)**

    (b) Explain the functionalities provided by S/MIME. **(6)**

OR

14. (a) Give the format of a PGP message and specify the significance of each field in the message. **(8)**

    (b) Explain the enhanced security services provided in S/MIME. **(6)**

15. (a) Explain the parameters that identify an SSL session state. **(8)**

    (b) Differentiate between transport mode and tunnel mode in IPSec. **(6)**

OR

16. (a) The IPsec architecture document states that when two transport mode SAs arebundled to allow both AH and ESP protocols on the same end-to-end flow, only one ordering of security protocols seems appropriate: performing the ESP protocol before performing the AH protocol. Why is this approach **(8)**

recommended rather than authentication before encryption?

(b) List and explain the purpose each Alert Codes supported by SSL. **(6)**

17. (a) Illustrate the significance ofperfect forward secrecy. **(6)**

(b) Explain the key features provided by SET. **(8)**

OR

18. (a) List and explain the SSH protocols. **(8)**

(b) "The HTTPS capability is built into all modern web browsers". Justify. **(6)**

19. (a) Explain the phases of operations in IEEE 802.11i. **(8)**

(b) Give the significances of Encrypted Tunnels **(6)**

OR

20. (a) Compare the features of three types of firewalls. **(8)**

(b) Compare the Wireless LAN protocols WEP, WPA and WPA2 **(6)**

TEACHING PLAN

No	Contents	No.of Lecture Hours (35 Hrs)
Module-1 (Authentication Protocols)(7hrs)		
1.1	Authentication Protocols – Mutual authentication, One way authentication	1
1.2	Kerberos –Version 4	1
1.4	Differences between Kerberos Version 4 and Version 5, Kerberos Version 5	1
1.5	X.509 Authentication service – Certificates, Authentication Procedures, X.509 Version 3	1
1.6	Public Key Infrastructure (PKI) – Trust models	1
1.7	Public Key Infrastructure (PKI) – Revocation	1

Module-2 (E-mail Security) (6 hrs)		
2.1	Pretty Good Privacy (PGP) – Operational Description	1
2.2	Cryptographic keys and key rings, Message format	1
2.3	PGP message generation, PGP message reception	1
2.4	PGP -Public key management	1
2.5	S/MIME – Overview of MIME, Functionality, Messages	1
2.6	S/MIME - Certificate processing, Enhanced security services	1
Module-3 (Network Layer Security and Web Security)(8 hrs)		
3.1	Internet Protocol Security (IPSec) – Overview, IP security architecture	1
3.2	Authentication Header (AH)	1
3.3	Encapsulating Security Payload (ESP)	1
3.4	Combining Security Associations, Key management	1
3.5	Internet Key Exchange (IKE) – Phases	1
3.6	Web Security – Web security considerations. Secure Socket Layer and Transport Layer Security (SSL/TLS) – SSL Architecture	1
3.7	SSL Protocols - Record Protocol, Change Cipher Spec Protocol, Alert Protocol	1
3.8	SSL Handshake Protocol, Cryptographic computations, Transport Layer Security	1
Module-4 (Real-time Security and Application Layer Security) (8hrs)		
4.1	Real-time communication security – Perfect Forward Secrecy (PFS)	1
4.2	Denial-of-Service protection, Endpoint identifier hiding, Live partner reassurance	1
4.3	Hypertext Transfer Protocol Secure (HTTPS) – Connection initiation, Closure	1
4.4	Secure Shell (SSH) – Transport layer protocol	1
4.5	User authentication protocol	1

4.6	Connection protocol	1
4.7	Secure Electronic Transaction (SET) – Overview, Features, Participants	1
4.8	Dual signature, Payment processing	1
Module-5 (System Security and Wireless Security) (6 hrs)		
5.1	Firewalls – Firewall characteristics, Types of Firewalls	1
5.2	Firewalls – Firewall configurations, Encrypted Tunnels	1
5.3	Trusted systems – Data Access Control, The Concept of Trusted Systems, Trojan Horse Defense	1
5.4	IEEE 802.11i wireless LAN security - Services, Phases of operation	1
5.5	Wired Equivalent Privacy (WEP)	1
5.6	Wi-Fi Protected Access (WPA), WPA2	1

CST444	SOFT COMPUTING	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PEC	2	1	0	3	2019

Preamble: This course enables the learners to understand the concepts of Soft Computing techniques and its applications. It covers Artificial Neural Networks, operations and models of fuzzy logic, genetic algorithms and multi objective optimization techniques. This course helps the students to develop algorithms and solutions for different real world applications.

Prerequisite: NIL.

Mapping of course outcomes with program outcomes

CO1	Describe soft computing techniques and the basic models of Artificial Neural Network **(Cognitive Knowledge Level: Understand)**
CO2	Solve practical problems using neural networks **(Cognitive Knowledge Level: Apply)**
CO3	Illustrate the operations, model and applications of fuzzy logic **(Cognitive Knowledge Level: Apply)**
CO4	Illustrate the concepts of Genetic Algorithm **(Cognitive Knowledge Level: Apply)**
CO5	Describe the concepts of multi-objective optimization models and the need for using hybrid soft computing approaches**(Cognitive Knowledge Level: Understand)**

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓									✓
CO2	✓	✓	✓	✓								✓
CO3	✓	✓	✓	✓								✓
CO4	✓	✓	✓	✓								✓
CO5	✓	✓	✓									✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and team work
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Life long learning

Assessment Pattern

Bloom's Category	**Continuous Assessment Tests**		**End Semester Examination Marks (%)**
	Test 1 (%)	**Test 2 (%)**	
Remember	**30**	**30**	**30**
Understand	**30**	**30**	**30**
Apply	**40**	**40**	**40**
Analyze			
Evaluate			
Create			

Mark Distribution

Total Marks	**CIE Marks**	**ESE Marks**	**ESE Duration**
150	**50**	**100**	**3**

Continuous Internal Evaluation Pattern:

Attendance	**10 marks**
Continuous Assessment Tests(Average of Internal Tests 1 & 2)	**25 marks**
Continuous Assessment Assignment	**15 marks**

Internal Examination Pattern

Each of the two internal examinations has to be conducted out of 50 marks. First series test shall be preferably conducted after completing the first half of the syllabus and the second series test shall be preferably conducted after completing remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly completed module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly completed module), each with 7 marks. Out of the 7 questions, a student should answer any 5.

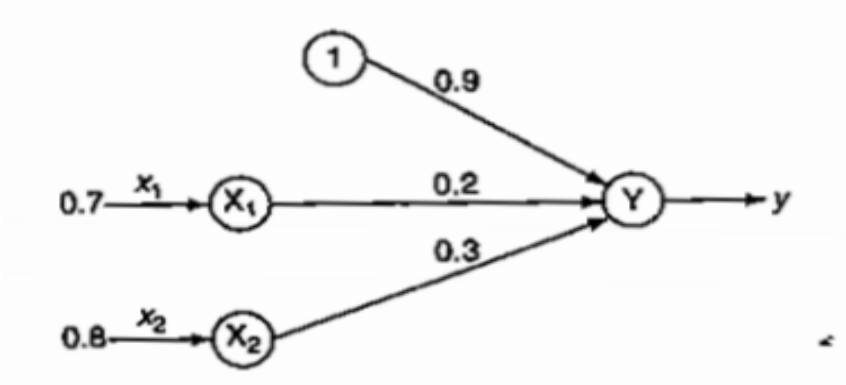
End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 full questions from each module of which student should answer any one. Each question can have maximum 2 sub-divisions and carries 14 marks.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Describe the necessity of Activation function? Examine the various aspects of sigmoidal activation function. List the drawbacks. Calculate the net output of the following neural network using the bipolar and binary sigmoidal activation function.



2. Explain the architecture of McCulloch-Pitts Neuron network model. Implement NAND(NOT-AND) gate function using M-P Neuron Model(with binary input).

Course Outcome 2(CO2):

1. Find the weights required to perform classification of patterns shown below using perceptron network. The patterns (1,1,-1) and (1, -1,-1) are belonging to the target class -1. The patterns (-1,1,1) and (-1,-1,1) are belonging to the target class +1. Assume suitable learning rate and initial weights.
2. Explain the architecture and training algorithm of Adaline network . Use Adaline nerwork to train NOR logic function with bipolar inputs and targets. Perform 2 epochs of training.

Course Outcome 3(CO3):

1. There is an imprecise relationship between the ambient temperature for clay masonry bricks and their compressive strengths. Let X be a fuzzy set of fracture strengths and Y be a fuzzy set of temperatures with the following membership functions:

$$X=\left\{\frac{1.0}{1500}+\frac{0.8}{2175}+\frac{0.6}{7000}+\frac{0.5}{12750}+\frac{0.3}{16500}+\frac{0.1}{20000}\right\}$$

$$Y=\left\{\frac{0.2}{20}+\frac{0.4}{25}+\frac{0.5}{32}+\frac{1.0}{50}+\frac{0.6}{90}+\frac{0.3}{105}\right\}$$

(a) Find the Cartesian Product of X and Y and represent it as relation R.
Suppose there is a second fuzzy set of masonry lengths given as

$$Z=\left\{\frac{0.4}{1500}+\frac{0.5}{2175}+\frac{0.6}{7000}+\frac{0.8}{12750}+\frac{0.9}{16500}+\frac{1.0}{20000}\right\}$$

(b) Find S=ZoR using max-min composition (c) Find T=ZoR using max-product composition

2. Given two universes X={x1,x2,x3,x4,x5} and Y={y1,y2,y3,y4,y5},the fuzzy sets A defined on X and fuzzy set B defined on Y are given below:

$A = \left\{\frac{0.4}{x1} + \frac{0.7}{x2} + \frac{1}{x3} + \frac{0.8}{x4} + \frac{0.6}{x5}\right\}$ $B = \left\{\frac{0.2}{y1} + \frac{0.6}{y2} + \frac{1}{y3} + \frac{0.9}{y4} + \frac{0.7}{y5}\right\}$

(i)Find the relation R = A × B

Consider another fuzzy set C defined on the universe V={v1,v2,v3}, $C = \left\{\frac{0.4}{v1} + \frac{1}{v2} + \frac{0.8}{v3}\right\}$

(ii) Find P = B × C. Using max-min composition, Find RoP.

Course Outcome 4(CO4):

1. Illustrate the various types of cross over with suitable examples.
2. Using Genetic algorithm with Roulette wheel selection method maximize the function f(x)=x2 over {0, 1, 2, ..., 31} with initial x values of (13, 24, 8, 19). Show one crossover and mutation.

Course Outcome 5(CO5):

1. Explain strong dominance and weak pareto-optimality.
2. What are the different classifications of neuro-fuzzy hybrid systems?

Syllabus

Module – 1 (Introduction to Soft Computing & Artificial Neural Network)

Introduction to Soft Computing. Difference between Hard Computing & Soft Computing. Applications of Soft Computing. Artificial Neurons Vs Biological Neurons. Basic models of artificial neural networks – Connections, Learning, Activation Functions. McCulloch and Pitts Neuron. Hebb network.

Module – 2 (Supervised Learning Network)

Perceptron Networks– Learning rule, Training and testing algorithm. Adaptive Linear Neuron– Architecture, Training and testing algorithm. Back propagation Network – Architecture, Training and testing algorithm.

Module - 3 (Fuzzy Logic & Defuzzification)

Fuzzy sets – properties, operations on fuzzy set. Fuzzy membership functions, Methods of membership value assignments – intuition, inference, Rank Ordering. Fuzzy relations– operations on fuzzy relation. Fuzzy Propositions. Fuzzy implications. Defuzzification– Lamda cuts, Defuzzification methods.

Module - 4 (Fuzzy Inference System & Genetic Algorithm)

Fuzzy Inference Systems - Mamdani and Sugeno types. Fuzzy Logic Controller. Concepts of genetic algorithm. Operators in genetic algorithm - coding, selection, cross over, mutation. Stopping condition for genetic algorithm.

Module - 5 (Multi Objective Optimization & Hybrid Systems)

Multi objective optimization problem. Principles of Multi- objective optimization, Dominance and pareto-optimality. Optimality conditions. Neuro-fuzzy hybrid systems. Genetic – neuro hybrid systems.

Text Books

1. S.N.Sivanandam and S.N. Deepa, Principles of Soft Computing , 2nd Edition, John Wiley & Sons.
2. Kalyanmoy Deb, Multi-objective Optimization using Evolutionary Algorithms, 1st Edition, John Wiley & Sons.

ReferenceBooks

1. Timothy J Ross, Fuzzy Logic with Engineering Applications, John Wiley & Sons, 2016.
2. T.S.Rajasekaran, G.A.Vijaylakshmi Pai “Neural Networks, Fuzzy Logic & Genetic Algorithms Synthesis and Applications”, Prentice-Hall India.
3. Simon Haykin, “Neural Networks- A Comprehensive Foundation”, 2/e, Pearson Education.
4. Zimmermann H. J, “Fuzzy Set Theory & Its Applications”, Allied Publishers Ltd.

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: CST 444

Course Name: Soft Computing

Max. Marks : 100 **Duration: 3 Hours**

PART A

Answer All Questions. Each Question Carries 3 Marks

1. Explain the architecture of a simple Artificial Neural network? Compare it with a biological neuron.

2. A 4-input neuron has weights 1, 2, 3 and 4. The transfer function is linear with the constant of proportionality being equal to 2. The inputs are 4, 10, 5 and 20 respectively. Predict the output?

3. Explain the Widrow-Hoff learning rule for supervised learning in neural networks with help of an example. Why is it sometimes called the LMS learning rule?

4. Implement one epoch of Adaline algorithm for AND logic function with binary inputs and bipolar outputs. Initial weights are w1=0.2, w2=0.1 and learning rate parameter η=0.2.

5. Consider two fuzzy sets $A = \left\{\frac{0.2}{0} + \frac{0.3}{1} + \frac{1}{2} + \frac{0.1}{3} + \frac{0.5}{4}\right\}$ $B = \left\{\frac{0.1}{0} + \frac{0.25}{1} + \frac{0.9}{2} + \frac{0.7}{3} + \frac{0.3}{4}\right\}$ Find the following: (a) Algebraic sum (b) Algebraic product(c) Bounded sum.

6. Using your own intuition and definition of universe of discourse, plot membership

functions for liquid level (Empty, very less, less, full, very full) in a tank.

7. Explain Stochastic Universal Sampling with an example.

8. Explain any two mutation methods.

9. Differentiate between linear and nonlinear Multi Objective Optimization Problem.

10. What are the characteristics of neuro fuzzy hybrid systems?

(10x3=30)

Part B

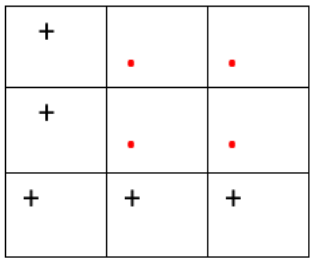
(Answer any one question from each module. Each question carries 14 Marks)

11. (a) Implement XOR function using M-P Neuron Model (with binary input).Why M-P neuron is widely used in processing binary data? **(8)**

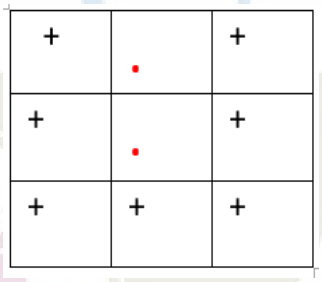
(b) Using Hebb Network calculate the weight required to perform the following classification of given input pattern. **(6)**

L □belongs to the members of the class(+)□target value +1

U□does not belongs to members of class(.)□target value -1



L



U

OR

12. (a) Compare the three learning approaches in Artificial Neural Network. How is the critic information used in learning process. **(8)**

(b) Define Hebb Law. Design a Hebb Network to implement logical AND function. Use bipolar input and targets. **(7)**

13. (a) Discuss the training algorithm and explain the weight updates in back propagation networks. **(10)**

(b) Implement one epoch of Perceptron training algorithm for OR logic function with binary input and bipolar output. **(4)**

OR

14. (a) Explain how synaptic weights are adapted iteration by iteration using error correction rule in Perceptron convergence algorithm for an OR gate with bipolar inputs and outputs. Initial weights are all zero and learning rate parameter η=0.1. **(10)**

(b) Explain Perceptron convergence theorem and discuss Perceptron algorithm based on XOR logic function. **(4)**

15. (a) Three fuzzy sets are defined as follows: **(10)**

$A = \left\{\frac{0.1}{30} + \frac{0.2}{60} + \frac{0.3}{90} + \frac{0.4}{120}\right\}, B = \left\{\frac{1}{1} + \frac{0.2}{2} + \frac{0.5}{3} + \frac{0.7}{4} + \frac{0.3}{5} + \frac{0}{6}\right\},$

$$C = \left\{\frac{0.33}{100} + \frac{0.65}{200} + \frac{0.92}{300} + \frac{0.21}{400}\right\}$$

Find: (i)$R = A \times B$ (ii)$S = B \times C$ (iii)$T = RoS$, using Max-Min composition (iv)$T = RoS$, using Max-Product composition.

(b) For the fuzzy sets given $A = \left\{\frac{0.5}{x_1} + \frac{0.2}{x_2} + \frac{0.9}{x_3}\right\}$ and$B = \left\{\frac{1}{y_1} + \frac{0.5}{y_2} + \frac{1}{y_3}\right\}$. Find relation R by performing Cartesian product over the given fuzzy sets. **(4)**

OR

16. (a) Using inference approach, find the membership values for each of the triangular shapes (I, R, IR, T) for a triangle with angles 120°, 50°, 10°. **(8)**

(b) Using Zadeh's notation, determine the ʎ - cut sets for the given fuzzy sets: **(6)**

$$S_1 = \left\{\frac{0}{0} + \frac{0.5}{20} + \frac{0.65}{40} + \frac{0.85}{60} + \frac{1.0}{80} + \frac{1.0}{100}\right\}$$

$$S_2 = \left\{\frac{0}{0} + \frac{0.45}{20} + \frac{0.6}{40} + \frac{0.8}{60} + \frac{0.95}{80} + \frac{1.0}{100}\right\}$$

Express the following for $\lambda = 0.5$: a) $S_1 \cup S_2$ b) $\acute{S_2}$ c) $S_1 \acute{\cap} S_2$

17. (a) Differentiate between value encoding and permutation encoding. **(8)**

(b) Explain the stopping conditions for genetic algorithm. **(6)**

OR

18. (a) Apply Mamdani fuzzy model to design a controller to determine the wash time of a domestic washing machine. Assume input is dirt and grease of the cloth. Use three descriptors for input variable and five descriptors for output variables .Derive the set of rules for controller action and defuzzification. Design should be supported by figure wherever possible. **(10)**

(b) Explain Single-Point Crossover and Two-Point Crossover with example. **(4)**

19. (a) Explain convex and non convex MOOP? How to find a non dominated set. **(10)**

(b) What are the properties of dominance relation? **(4)**

OR

20. (a) Explain Genetic Neuro-Hybrid System with block diagram. Also write the advantages of Genetic- Neuro Hybrid systems. **(8)**

(b) Discuss the classification of Neuro-Fuzzy Hybrid System. **(6)**

Teaching Plan

No	Contents	No. of Lecture Hours (35 hrs)
Module-1 (Introduction to Soft Computing & Artificial Neural Network) (6 hours)		
1.1	Introduction to Soft Computing	1 hour
1.2	Difference between Hard Computing & Soft Computing & Applications of Soft Computing	1 hour
1.3	Artificial Neurons Vs Biological Neurons, Basic models of artificial neural networks	1 hour
1.4	Activation Functions	1 hour
1.5	McCulloch and Pitts Neuron	1 hour
1.6	Hebb network	1 hour
Module-2 (Supervised Learning Network) (7 hours)		
2.1	Perceptron networks – Learning rule, Training and testing algorithm	1 hour
2.2	Perceptron networks – Problems	1 hour
2.3	Adaptive Linear Neuron (Lecture I)	1 hour
2.4	Adaptive Linear Neuron (Lecture II)	1 hour
2.5	Adaptive Linear Neuron-Problems (Lecture III)	1 hour
2.6	Back propagation Network (Lecture I)	1 hour
2.7	Back propagation Network (Lecture II)	1 hour
Module-3 (Fuzzy Logic & Defuzzification) (8 hours)		
3.1	Introduction to Fuzzy Set, Properties & operations on fuzzy sets	1 hour
3.2	Fuzzy membership functions, Fuzzification	1 hour
3.3	Methods of membership value assignments	1 hour
3.4	Fuzzy relations, Operations on Fuzzy Relation	1 hour

3.5	Fuzzy Propositions & Fuzzy Implications	1 hour
3.6	Lamda cuts for fuzzy sets	1 hour
3.7	Defuzzification methods(Lecture I)	1 hour
3.8	Defuzzification methods(Lecture II)	1 hour
Module-4 (Fuzzy Inference System & Genetic Algorithm) (6 hours)		
4.1	Fuzzy Inference Systems - Mamdani type	1 hour
4.2	Fuzzy Inference Systems - Sugeno type	1 hour
4.3	Fuzzy Logic Controller	1 hour
4.4	Introduction to genetic algorithm, operators in genetic algorithm - coding	1 hour
4.5	Selection, Cross over	1 hour
4.6	Mutation, stopping condition for genetic algorithm	1 hour
Module-5 (Multi-Objective Optimization & Hybrid System) (8 hours)		
5.1	MOOP-Linear &Non linear, Convex & Non Convex	1 hour
5.2	Principles of MOO-Illustrating Pareto Optimal Solutions, Objectives in MOO	1 hour
5.3	Dominance & Pareto-Optimality-Concept of Domination	1 hour
5.4	Properties of Dominance Relation, Pareto Optimality	1 hour
5.5	Procedure for finding a non dominated set	1 hour
5.6	Optimality Conditions	1 hour
5.7	Neuro Fuzzy hybrid system-Classification& characteristics	1 hour
5.8	Genetic –neuro hybrid systems	1 hour

CST454	FUZZY SET THEORY AND APPLICATIONS	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PEC	2	1	0	3	2019

Preamble: This course equips the students to understand the concepts of fuzziness and its use in building better solutions to problems. The course covers basic concepts of fuzzy sets, fuzzy relations, fuzzy logic and building of fuzzy approximation-based solutions. It helps students to design and develop fuzzy based solutions to real world applications.

Prerequisite: Basic knowledge in set theory.

Course Outcomes: After the completion of the course, the student will be able to

CO1	Explain fuzzy logic based problem solving **(Cognitive Knowledge Level: Understand)**
CO2	Summarize the concepts of crisp sets, crisp relations, crisp logic with fuzzy sets, fuzzy relations and fuzzy logic**(Cognitive Knowledge Level: Apply)**
CO3	Develop fuzzy systems by selecting appropriate membership functions, fuzzification and defuzzification methods **(Cognitive Knowledge Level: Apply)**
CO4	Develop solutions using graphical and rule-based methods**(Cognitive Knowledge Level: Apply)**
CO5	Make use of fuzzy logic inference to solve real world problems**(Cognitive Knowledge Level: Apply)**

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓											✓
CO2	✓	✓	✓									✓
CO3	✓	✓	✓	✓	✓							✓
CO4	✓	✓	✓	✓	✓							✓

CO5	✓	✓		✓	✓							✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and team work
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Life long learning

Assessment Pattern

Bloom's Category	**Continuous Assessment Tests**		**End Semester Examination Marks (%)**
	Test 1 (%)	**Test 2 (%)**	
Remember	**20**	**20**	**20**
Understand	**50**	**50**	**50**
Apply	**30**	**30**	**30**
Analyze			
Evaluate			
Create			

Mark Distribution

Total Marks	**CIE Marks**	**ESE Marks**	**ESE Duration**
150	**50**	**100**	**3**

Continuous Internal Evaluation Pattern:

Attendance	**10 marks**
Continuous Assessment Tests(Average of Internal Tests1&2)	**25 marks**
Continuous Assessment Assignment	**15 marks**

Internal Examination Pattern

Each of the two internal examinations has to be conducted out of 50 marks. First series test shall be preferably conducted after completing the first half of the syllabus and the second series test shall be preferably conducted after completing remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly completed module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly completed module), each with 7 marks. Out of the 7 questions, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 full questions from each module of which student should answer any one. Each question can have maximum 2 sub-divisions and carries 14 marks.

Syllabus

Module – 1 (Basic Fuzzy Set Theory)

The case for imprecision, Utility and Limitations of Fuzzy Systems, Fuzzy Sets and Membership, Classical Sets – Properties, Operations, Fuzzy Sets – Properties and Operations, Classical Relations – Cartesian Product, Operations and Properties of Crisp Relations, Composition, Fuzzy Relations – Cardinality, Operations, Properties, Fuzzy Cartesian Product and Composition.

Module – 2 (Fuzzy Membership Functions)

Tolerance and Equivalence Relations – Crisp and Fuzzy, Similarity Methods – Cosine, Min-max, Fuzzy Membership Functions – Features, Fuzzification, Defuzzification to Crisp Sets, λ-Cutsfor Fuzzy Relations, Linguistic Hedges.

Module - 3 (Fuzzification and Defuzzification Methods)

Development of Membership Functions –Intuition, Inference, Rank ordering, Inductive reasoning. Defuzzification to Scalars - Max membership principle, Centroid method, Weighted average method, Mean max membership, Center of sums, Center of largest area, First (or last) of maxima.

Module - 4 (Fuzzy Inference)

Classical Logic, Fuzzy Logic, Approximate Reasoning, Fuzzy (Rule-Based) Systems - Multiple conjunctive antecedents, Multiple disjunctive antecedents, Aggregation of fuzzy rules, Graphical Techniques of Inference.

Module - 5 (Fuzzy Applications)

Applications of Fuzzy Systems - Fuzzy Classification, Fuzzy Pattern Recognition, Fuzzy Control Systems, Fuzzy Systems and Neural Networks, Fuzzy Clustering, Fuzzy Databases and Information retrieval systems.

Text Books

1. Fuzzy Logic with Engineering Applications – Timothy J. Ross, Third Edition, John Wiley and Sons, 2010
2. Fuzzy Sets and Fuzzy Logic: Theory and Applications - George J. Klir and Bo Yuan , Prentice Hall, 1995.

Reference Books

1. Kenneth H. Rosen, Discrete Mathematics and Its Applications with Combinatorics and GraphTheory, Seventh Edition, MGH,2011
2. Trembly J.P and Manohar R, "Discrete Mathematical Structures with Applications to Computer Science", TataMc Graw Hill Pub. Co. Ltd., New Delhi,2003.
3. Bernard Kolman, Robert C. Busby, Sharan Cutler Ross, "Discrete Mathematical Structures", Pearson Education Pvt Ltd., New Delhi,2003
4. Kenneth H .Rosen, "Discrete Mathematics and its Applications", 5/e, TataMc Graw Hill Pub. Co. Ltd, New Delhi2003

Course Level Assessment Questions

Course Outcome1 (CO1):

1. What are the limitations of crisp systems?
2. Explain the difference between randomness and fuzziness.
3. Find some examples of prospective fuzzy variables in daily life.

Course Outcome 2(CO2):

1. The strength of two types of concrete needs to be compared. Four concrete masonry units (CMUs) from each type of concrete are stressed until they fail. The lowest stress at failure of a CMU is denoted 1, and the highest stress at failure is denoted 4,so the CMUs are rank ordered by failure stress, that is, X = {1, 2, 3, 4}. Since "failure" of CMUs is fuzzy, the membership value for a specific CMU represents the judgment that the CMU really failed. The following fuzzy sets represent the failure estimates for the two different concrete types:

$$A = \left\{\frac{0 \cdot 15}{1} + \frac{0.25}{2} + \frac{0 \cdot 6}{3} + \frac{0.9}{4}\right\}$$

$$B = \left\{\frac{0.2}{1} + \frac{0.3}{2} = + \frac{0.5}{3} + \frac{0.8}{4}\right\}$$

Calculate the union, intersection and difference for the two concrete types.

2. An engineer is testing the properties, strength and weight of steel. Suppose he has two fuzzy sets A , defined on a universe of three discrete strengths, $\{s1, s2, s3\}$, and B, defined on a universe of three discrete weights, $\{w1,w2,w3\}$. Suppose A and B represent a "high-strength steel" and a "near-optimum weight," respectively, as shown below

$$A = \left\{\frac{1}{s_1} + \frac{0.5}{s_2} + \frac{0.2}{s_3}\right\}$$

$$B = \left\{\frac{1}{w_1} + \frac{0.5}{w_2} + \frac{0.2}{w_3}\right\}$$

a) Find the fuzzy relation for the Cartesian product, R, of A and B
b) Introducing another fuzzy set, C, which represents a set of "moderately good" steel strengths

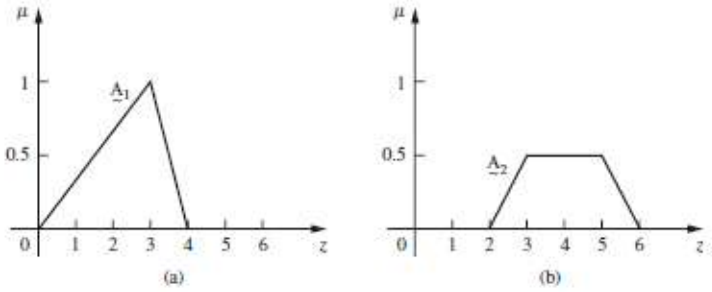
$$C = \left\{\frac{0.1}{s_1} + \frac{0.6}{s_2} + \frac{1}{s_3}\right\}$$

Find C◦R using max–min composition

Course Outcome 3(CO3):

1. Using your own intuition and your own definitions of the universe of discourse,plot fuzzy membership functions for "age of people" who are:
   (i) very young
   (ii) young
   (iii) middle-aged
   (iv) old
2. a) Define membership functions for approximately isosceles triangle, approximately equilateral and approximately right-angled triangles.

   b) Find the membership value for the triangle represented by the angles 80°, 75°, 25°, in the above triangles.
3. In metallurgy, materials are made with mixtures of various metals and other elements to achieve certain desirable properties. In a particular preparation of steel, three elements, namely, iron, manganese, and carbon, are mixed in two different proportions. The samples obtained from these two different proportions are placed on a normalized scale and are represented as fuzzy sets A1 and A2. Do a logical union of the membership functions A1 and A2 and find the defuzzified value of the resulting membership function.



Course Outcome 4(CO4): .

1. Consider the following two discrete fuzzy sets, which are defined on universeX = {−5, 5}:

$$A = \text{"zero"} = \left\{\frac{0}{-2} + \frac{0.5}{-1} + \frac{1}{0} + \frac{0.5}{1} + \frac{0}{2}\right\}$$

$$B = \text{"positive medium"} = \left\{\frac{0}{0} + \frac{0.6}{1} + \frac{1}{2} + \frac{0.6}{3} + \frac{0}{4}\right\}$$

   Construct the relation for IF *x* is "zero" THEN *y* is "positive medium"

2. A metro train system uses fuzzy logic in ensuring smooth ride on the train. The metro train system has fixed stops and the distance between the stops are known. The system uses fuzzy logic in deciding the pressure applied on the brakes. The amount of pressure applied depends on the distance to the next stop and the speed of the train. Design appropriate membership functions for the input and illustrate the use of Mamdani Inference in arriving at the brake pressure.

Course Outcome 5(CO5):

1. A fuzzy systems needs to be designed to provide a rating for a web store as "excellent", "good" or "poor". The web store can be rated based on the products available, the customer service and the discount provided. Design appropriate membership functions and fuzzy rules for generating the fuzzy based rating system.
2. Design a fuzzy control system for an air-conditioning application. Make appropriate decisions regarding inputs and outputs.

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: CST454

Course Name: Fuzzy Set Theory and Applications

Max.Marks:100 **Duration: 3 Hours**

PART A

Answer All Questions. Each Question Carries 3 Marks

1. Illustrate where a fuzzy logic based application is suitable.

2. Consider a LAN using Ethernet protocol with maximum bandwidth of 10 Mbps. Traffic rates can be represented using two fuzzy variables, Quiet and Congested. If the universal set X = {0,1,2,3,4,5,6,7,8,9,10} represents bandwidth usage in Mbps, then draw possible membership functions for the fuzzy variables.

3. Define fuzzy tolerance and equivalence relations.

4. Given two data points, illustrate how a similarity measure between them can be computed.

5. Define a convex normalized fuzzy set.

6. How does augmented query help in information retrieval.

7. Given the propositions
   (i) $C \vee D$
   (ii) $\sim H \Rightarrow (A \wedge \sim B)$
   (iii) $(C \vee D) \Rightarrow \sim H$

(iv) $(A \wedge \sim B) \Rightarrow (R \vee S)$

Infer $(R \vee S)$ from the above propositions and state the tautologies used.

8. Write a predicate logic statement for "Ram likes all kinds of food".

9. Given the relation R below, find λ-cut for the relation using suitable λ value.

$$\begin{bmatrix} 1 & 0.8 & 0 & 0.1 & 0.2 \\ 0.8 & 1 & 0.4 & 0 & 0.9 \\ 0 & 0.4 & 1 & 0 & 0 \\ 0.1 & 0 & 0 & 1 & 0.5 \\ 0.2 & 0.9 & 0 & 0.5 & 1 \end{bmatrix}$$

10. Define maximum approaching degree.

(10x3=30)

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) An engineer is testing the properties, strength and weight of steel. Suppose he has two fuzzy sets A, defined on the universe of three discrete strengths { s1, s2, s3 } and B, defined on the universe of discrete weights { w1, w2, w3}. Suppose A represents a "high-strength steel" and B a "near-optimum weight". **(4)**

$$A = \left\{\frac{1}{s1} + \frac{0.5}{s2} + \frac{0.2}{s3}\right\}, B = \left\{\frac{1}{w1} + \frac{0.5}{w2} + \frac{0.3}{w3}\right\}$$

Find fuzzy Cartesian product, R, of A and B.

(b) Let a fuzzy set C= $\left\{\frac{0.1}{s1} + \frac{0.6}{s2} + \frac{1}{s3}\right\}$ be introduced, which represents a set of "moderately good" steel strength. Find the max-min composition of C and R. **(5)**

(c) Define 5 operations associated with crisp relations. **(5)**

OR

12. (a) How is excluded middle axiom different for crisp and fuzzy sets? **(4)**

(b) Differentiate between crisp and fuzzy sets with respect to their membership functions. **(4)**

(c) Illustrate any 4 operations associated with a fuzzy relation. **(6)**

13. (a) A structural designer is considering four different kinds of structural beams { S1, S2, S3, S4} for a new building. Laboratory experiments on the deflection resistance for these four kinds of beams have been performed, and the engineer wants to determine their suitability in the new structure. The following data have been observed based on the overall deflection capacity of each beam type: **(10)**

		S1	**S2**	**S3**	**S4**
No deflection	X_1	0.3	0.6	0.5	0.8
Some deflection	X_2	0.6	0.3	0.5	0.2
Excessive deflection	X_3	0.1	0.1	0	0

Use cosine amplitude method to determine the similarity of the four beam types.

(b) Given a fuzzy set "tall" = $\left\{\frac{0.1}{s1} + \frac{0.6}{s2} + \frac{1}{s3}\right\}$, illustrate how the fuzzy set "very tall" be defined? **(4)**

OR

14. (a) Define tolerance and equivalence relations. Check whether the relation R given below is tolerance or equivalence relation. **(4)**

$$\begin{bmatrix} 1 & 0.8 & 0 & 0.1 & 0.2 \\ 0.8 & 1 & 0.4 & 0 & 0.9 \\ 0 & 0.4 & 1 & 0 & 0 \\ 0.1 & 0 & 0 & 1 & 0.5 \\ 0.2 & 0.9 & 0 & 0.5 & 1 \end{bmatrix}$$

(b) Given the following data regarding three cities and the quality of their bridges, find the similarity between the cities using max-min method. **(10)**

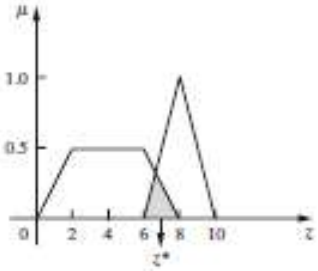
		C1	**C2**	**C3**
Poor	Q_1	0.00	0.10	0.10
Fair	Q_2	0.04	0.04	0.08
Good	Q_3	0.02	0.04	0.06

15. (a) Explain the process of developing membership functions using the inference method. **(6)**

(b) The following raw data were determined in a pair wise comparison of new premium car preferences in a poll of 100 people. When it was compared with a Porsche (P), 79 of those polled preferred a BMW (B), 85 preferred a Mercedes (M), 59 preferred a Lexus (L), and 67 preferred an Infinity (I). When a BMW was compared, the preferences were 21 – P, 23 – M, 37 – L, and 45 – I. When a Mercedes was compared, the preferences were 15 – P, 77 – B, 35 – L, and 48 – I. When a Lexus was compared, the preferences were 41 – P, 63 – B, 65 – M, and 51 – I. Finally, when an Infinity was compared, the preferences were 33 – P, 55 – B, 52 – M, and 49 – L. Using rank ordering, plot the membership function for "most preferred car." **(8)**

OR

16. (a) **1.** Defuzzify the following region using centroid method. **(9)**



(b) 2. Defuzzify the region given in 16(a) using weighted average method. **(5)**

17. (a) For a distillation process, the objective is to separate components of a mixture in the input stream. The relationship between the input variable, temperature, and the output variable, distillate fractions, is not precise but the human operator of this process has developed an intuitive understanding of this relationship. The universe for each of these variables is **(8)**

X = universe of temperatures (degree fahrenheit) = {160, 165, 170, 175, 180, 185, 190, 195}.
Y = universe of distillate fractions (percentage) = {77, 80, 83, 86, 89, 92, 95, 98}.
Given two fuzzy sets
A = "temperature of input steam is hot"= $\left\{\frac{0}{175} + \frac{0.7}{180} + \frac{1}{185} + \frac{0.4}{190}\right\}$

B = "separation of mixture is good" $=\left\{\frac{0}{89} + \frac{0.5}{92} + \frac{0.8}{95} + \frac{1}{98}\right\}$.

Find the fuzzy relation corresponding to " IF x is $\tilde{A}$, THEN y is $\tilde{B}$

(b) Show how inference is done using Generalized Modus Ponens **(6)**

OR

18. (a) Illustrate how graphical inference is done using Mamdani method. **(6)**

(b) A restaurant uses a fuzzy inference system to calculate the tips given to its employees. The tips are based on the timeliness of service and quality of service of the waiters. Design appropriate membership functions for the input and illustrate the use of Sugeno Inference in arriving at the tip amount. **(8)**

19. (a) Explain fuzzy pattern recognition using multiple features. **(7)**

(b) Describe how fuzziness in information retrieval can enhance the quality of search results. **(7)**

OR

20. (a) Design a fuzzy control system for an air-conditioning system. **(7)**

(b) Illustrate how the join operation is performed in fuzzy databases. **(7)**

Teaching Plan

No	Contents	No. of Lecture Hours (36 hrs)
Module-1(Basic Fuzzy Set Theory) (6 hours)		
1.1	Introduction to Fuzzy Concepts – Case for imprecision- utility and limitations of Fuzzy Systems	1 hour
1.2	Classical Sets – Properties, Operations	1 hour
1.3	Fuzzy Sets – Properties, Operations	1 hour
1.4	Classical Relations – Properties, Operations – Cartesian Product,	1 hour

	Composition	
1.5	Fuzzy Relations – Properties, Operations, Cardinality	1 hour
1.6	Fuzzy Cartesian Product, Fuzzy Composition	1 hour
Module-2 (Fuzzy Membership Functions) (6 hours)		
2.1	Tolerance and Equivalence Relations - Crisp	1 hour
2.2	Tolerance and Equivalence Relations - Fuzzy	1 hour
2.3	Similarity Methods – Cosine, Minmax	1 hour
2.4	Fuzzy Membership Functions- Features	1 hour
2.5	Fuzzification, Defuzzification to crisp sets – λ-cuts	1 hour
2.6	Linguistic Hedges	1 hour
Module-3 (Fuzzification and Defuzzification Methods) (7 hours)		
3.1	Development of Membership Functions – Intuition, Inference	1 hour
3.2	Development of Membership Functions – Rank Ordering	1 hour
3.3	Development of Membership Functions – Inductive reasoning	1 hour
3.4	Defuzzification – Max membership principle, weighted average method, mean max membership	1 hour
3.5	Defuzzification – Centroid method	1 hour
3.6	Defuzzification – Center of Sums, Center of Largest area, First/Last of maxima	1 hour
3.7	Defuzzification - exercises	1 hour
Module-4 (Fuzzy Inference) (9 hours)		
4.1	Classical Logic – Propositional Logic	1 hour
4.2	Classical Logic – Predicate Logic	1 hour
4.3	Fuzzy Logic	1 hour
4.4	Fuzzy Approximation based reasoning	1 hour
4.5	Fuzzy Rule based systems	1 hour
4.6	Multiple conjunctive and disjunctive antecedents, aggregation	1 hour
4.7	Graphical Techniques for Inference	1 hour
4.8	Illustration of Graphical Techniques for Inference	1 hour

4.9	Fuzzy Inference - Exercises	1 hour
Module-5 (Fuzzy Applications) (8 hours)		
5.1	Fuzzy Control Systems	1 hour
5.2	Illustration of Fuzzy Control Systems	1 hour
5.3	Fuzzy Classification	1 hour
5.4	Fuzzy Pattern Recognition	1 hour
5.5	Fuzzy Systems and Neural Networks	1 hour
5.6	Fuzzy Clustering	1 hour
5.7	Fuzzy Databases	1 hour
5.8	Fuzzy Information Retrieval Systems	1 hour

CST474	COMPUTER VISION	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PEC	2	1	0	3	2019

Preamble: Computer vision is a field of artificial intelligence (AI) that enables computers and systems to derive meaningful information from digital images, videos and other visual inputs. The curriculum covers the basics of image formation, key computer vision concepts, methods, techniques, pattern recognition, various problems in designing computer vision and object recognition systems. This course enables the learners to understand the fundamentals of computer vision and develop applications in computer vision.

Prerequisite: Nil

Course Outcomes: After the completion of the course, the student will be able to

CO1	Summarize basic concepts, terminology, theories, models and methods in the field of computer vision. **(Cognitive Knowledge Level: Understand)**
CO2	Explain basic methods of computer vision related to multi-scale representation, edge detection, detection of other primitives, stereo, motion and object recognition. **(Cognitive Knowledge Level: Understand)**
CO3	Describe principles of Segmentation, Motion Segmentation and Classification **(Cognitive Knowledge Level: Understand)**
CO4	Select appropriate object Tracking and detection methods for computer vision applications **(Cognitive Knowledge Level: Understand).**
CO5	Implement a computer vision system for a specific problem **(Cognitive Knowledge Level: Apply)**

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓		✓									✓
CO2	✓		✓									✓
CO3	✓		✓									✓

CO4	✓		✓									✓
CO5	✓	✓	✓	✓	✓							✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and team work
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Lifelong learning

Assessment Pattern

Bloom's Category	**Continuous Assessment Tests**		**End Semester Examination Marks (%)**
	Test 1 (%)	**Test 2 (%)**	
Remember	**30**	**30**	**30**
Understand	**50**	**50**	**50**
Apply	**20**	**20**	**20**
Analyze			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	50	100	3

Continuous Internal Evaluation Pattern:

Attendance **10 marks**

Continuous Assessment Tests(Average of Internal Tests1&2) **25 marks**

Continuous Assessment Assignment **15 marks**

Internal Examination Pattern

Each of the two internal examinations has to be conducted out of 50 marks. First series test shall be preferably conducted after completing the first half of the syllabus and the second series test shall be preferably conducted after completing remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly completed module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly completed module), each with 7 marks. Out of the 7 questions, a student should answer any5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 full questions from each module of which student should answer any one. Each question can have maximum 2 sub-divisions and carries 14 marks.

Syllabus

Module – 1 (Image Formation and Filtering)

Geometric Camera Models - Pinhole perspective, Intrinsic and Extrinsic Parameters, Geometric Camera Calibration. Linear Filters- Linear Filters and Convolution, Shift Invariant Linear Systems. Filters as Templates - Normalized Correlation and Finding Patterns.

Module - 2(Local Image Features and Stereo Vision)

Image Gradients - Computing the Image Gradient, Gradient Based Edge and Corner Detection. Stereopsis- Binocular Camera Geometry, Epipolar Constraint, Binocular Reconstruction, Local Methods for Binocular Fusion, Global Methods for Binocular Fusion.

Module - 3 (Segmentation)

Segmentation - Background subtraction, Interactive segmentation, Forming image regions. Segmentation by clustering - Watershed Algorithm. Motion Segmentation by Parameter Estimation- Optical Flow and Motion, Flow Models, Motion Segmentation with Layers.

Module- 4 (Classification and Tracking)

Classification - Classification Basics, Two-class and Multiclass classifiers, Error, Overfitting and Regularization, Cross Validation, Classifying Images of Single Objects.
Tracking - Tracking Basics, Simple Tracking Strategies, Tracking by detection, Tracking Linear Dynamical models with Kalman filters.

Module - 5 (Finding Objects and other Applications)

Object detection - The Sliding Window Method. Object Recognition -Goals of Object Recognition System. Applications - Robot Navigation by stereo vision, Face detection, Face recognition, Activity Recognition, Tracking people.

Text Books

1. Forsyth, David, and Jean Ponce. Computer vision: A modern approach. Prentice hall, 2011.

Reference Books

1. Szeliski, Richard, Computer vision: algorithms and applications. Springer Science & Business Media, 2010.
2. Medioni, Gerard, Emerging topics in computer vision. and Sing Bing Kang. Prentice Hall PTR, 2004.
3. Trucco, Emanuele, and Alessandro Verri, Introductory techniques for 3-D computer vision. Vol. 201. Englewood Cliffs: Prentice Hall, 1998.
4. Faugeras, Olivier, and Olivier Autor Faugeras, Three-dimensional computer vision: a geometric viewpoint. MIT press, 1993.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Explain the relationship between coordinates involved in a pinhole camera imaging setup.
2. Explain the basic principle behind geometric camera calibration**.**
3. Describe how linear filters can be used for smoothing digital images.
4. How does normalised correlation help in matching patterns in images?

Course Outcome 2 (CO2):

1. Describe edge detection methods for computer vision.
2. List any five applications of object recognition.
3. Explain how the epipolar constraint simplifies the correspondence search between two stereo images.
4. List and explain the different methods used for binocular fusion.
5. Explain the different corner detection methods.

Course Outcome 3 (CO3):

1. Explain the principle of background subtraction.
2. Describe the watershed algorithm for image segmentation.
3. What is meant by optical flow? How can it be utilized for segmenting images?
4. Describe motion segmentation with layers.
5. What is overfitting in the context of classification?
6. Explain the principle behind classification of single images.

Course Outcome 4 (CO4):

1. Explain 'Mean Shift Algorithm' to track an object using matching.

2. Describe an algorithm to track a moving object (dynamic object).

3. Explain the sliding window method for object detection.

4. Assume that we have the dynamics

$$x_i \sim N(d_i x_{i-1}, \sigma_{d_i}^2)$$

$$y_i \sim N(m_i x_i, \sigma_{m_i}^2)$$

a. $P(x_i|x_{i-1})$ is a normal density with mean $d_i x_{i-1}$ and variance $\sigma_{d_i}^2$. Whatis$(x_{i-1}|x_i)$?

b. Show how to obtain a representation of $P(x_i|y_{i+1}, \ldots, y_N)$ usinga KalmanFilter.

Course Outcome 5(CO5):

1. Explain how to implement a computer vision system.
2. Illustrate a computer vision system with the help of a neat diagram.
3. Discuss the components of a computer vision system for object recognition.
4. Explain how activity recognition can be done using computer vision.
5. Illustrate a face recognition system with the help of a diagram.

Assignment Questions

6. Implement a voxel-based approach to visual hull construction.
7. Implement a computer vision system for object recognition.

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 3**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: CST474

Course Name: COMPUTER VISION

Max.Marks:100 **Duration: 3 Hours**

PART A

Answer All Questions. Each Question Carries 3 Marks

1. State three properties of shift invariant linear systems.
2. Explain the term normalized correlation.
3. What is image rectification? Mention its significance?
4. Illustrate epipolar geometry and showepipolar lines and epipoles.
5. Explain the term flow model.
6. How does background subtraction help in segmenting an image?
7. What is a Kalman filter? Give its applications.
8. State any three simple tracking strategies.
9. State the goals of an object recognition system.
10. Explain the task of face recognition. **(10x3=30)**

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) Demonstrate the relationship between a point in the world coordinate frame and its corresponding image point using camera parameters. **(9)**

    (b) Show that convolving a function with a δ function simply reproduces the original function. **(5)**

OR

12. (a) What is linear filtering? Explain two applications of linear filtering to image processing. **(7)**

    (b) Explain an application of normalised correlation to find patterns. **(7)**

13. (a) Show that smoothing an image and then computing the gradient is same as convolving an image with the derivative of a smoothing function. **(5)**

    (b) State the epipolar constraint and derive its representations using the Essential matrix and the Fundamental matrix. **(9)**

OR

14. (a) Explain the algorithm for computing edges using gradients. **(9)**

    (b) Define binocular fusion. Explain two local methods for binocular fusion. **(5)**

15. (a) Discuss the different interactive segmentation approaches. **(7)**

    (b) What is meant by optical flow? How can it be utilized for segmenting images? **(7)**

OR

16. (a) Explain the Watershed algorithm. **(7)**

    (b) How can we perform motion segmentation by parameter estimation? **(7)**

17. (a) Explain tracking algorithm using Kalman filtering. **(7)**

    (b) Illustrate the tracking by detection algorithm. **(7)**

OR

18. (a) Explain the various kinds of errors in classification and the relationship between them. **(7)**

    (b) What is overfitting and how does regularization help to minimise it? **(7)**

19. (a) Explain human activity recognition with appearance features. **(7)**

(b) Describe the Sliding window method for detecting objects in images. **(7)**

OR

20. (a) Explain the principle of detecting faces in an image. **(7)**

(b) What are the various strategies for object recognition? **(7)**

Teaching Plan

No	Contents	No. of Lecture Hours (36hrs)
Module 1 Image Formation and Filtering (7)		
1.1	Geometric Camera model - Pinhole perspective	1
1.2	Geometric Camera model - Intrinsic Parameters	1
1.3	Geometric Camera model - Extrinsic Parameters	1
1.4	Geometric Camera Calibration – Linear Approach	1
1.5	Linear Filters and Convolution	1
1.6	Shift Invariant Linear Systems - Discrete convolution	1
1.7	Normalized Correlation and Finding patterns	1
Module 2 Local Image Features and Stereo Vision (8)		
2.1	Local Image Features - Computing the Image Gradient	1
2.2	Gradient Based Edge Detection	1
2.3	Gradient Based Corner Detection	1
2.4	Stereopsis - Binocular Camera Geometry and Epipolar Constraint	1
2.5	Essential Matrix and Fundamental Matrix	1
2.6	Binocular Reconstruction	1
2.7	Local Methods for Binocular Fusion	1
2.8	Global Methods for Binocular Fusion	1
Module 3 Segmentation (6)		

3.1	Segmentation basics	1
3.2	Applications - Background Subtraction, Interactive Segmentation	1
3.3	Forming Image Regions	1
3.4	Segmentation by clustering - The Watershed Algorithm	1
3.5	Motion Segmentation by Parameter Estimation - Optical Flow and Motion	1
3.6	Flow Models and Motion Segmentation with Layers	1
Module 4 Classification and Tracking (8)		
4.1	Classification Basics, Two-class and Multiclass classifier	1
4.2	Error, Overfitting and Regularization	1
4.3	Cross Validation, Classifying Images of Single Objects	1
4.4	Tracking Basics, Simple Tracking Strategies	1
4.5	Tracking by detection	1
4.6	Linear Dynamical models	1
4.7	The Kalman Filter background	1
4.8	Kalman filter algorithm	1
Module 5 Finding Objects and other Applications (7)		
5.1	Detecting Objects in Images- The Sliding Window Method	1
5.2	Object Recognition - Goals of Object Recognition System	1
5.3	Application of binocular stereo vision - Robot Navigation	1
5.4	Face detection	1
5.5	Face recognition	1
5.6	Activity recognition	1
5.7	Tracking people	1

SEMESTER VIII

PROGRAM ELECTIVE IV

AMT 416	HUMAN COMPUTER INTERACTION	CATEGORY	L	T	P	CREDIT
		PEC	2	1	0	3

Preamble:.This course provides an overview of Human-Computer Interaction (HCI), with an understanding of user interface design in general. The course covers topics which include user-centered design, human cognitive and physical abilities, prototyping and evaluation techniques, graphical design fundamentals and emerging areas of HCI research including mobile interaction, augmented-reality and ubiquitous computing. This course helps the learners to design and evaluate interactive systems by following the fundamental principles of human-computer interaction.

Prerequisite: Skill in any programming language. Exposure to web designing is preferred.

Course Outcomes: After the completion of the course the student will be able to

CO 1	Describe the usability based on a variety of classic universal user-centric models**. (Cognitive Knowledge level: Understand)**
CO 2	Comprehend the different interaction styles and the methodologies for designing interactive systems. **(Cognitive Knowledge level: Understand)**
CO 3	Investigate the core and complex user experience design issues. **(Cognitive Knowledge level: Understand)**
CO 4	Examine the evaluation methodologies of interactive system design. **(Cognitive Knowledge level: Apply)**
CO 5	Explore the different contexts and suggest suitable designs for applications related to web, mobile and wearable computing. **(Cognitive Knowledge level: Apply)**

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓	✓								✓
CO2	✓	✓	✓	✓								✓
CO3	✓	✓	✓	✓		✓						✓
CO4	✓	✓	✓	✓	✓							✓
CO5	✓	✓	✓	✓	✓	✓						✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and team work
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Life long learning

Assessment Pattern

Bloom's Category	**Continuous Assessment Tests**		**End Semester Examination Marks**
	Test1 (percentage)	**Test2 (percentage)**	
Remember	20	20	20
Understand	60	60	60
Apply	20	20	20
Analyse			
Evaluate			
Create			

Mark distribution

Total Marks	**CIE Marks**	**ESE Marks**	**ESE Duration**
150	50	100	3 hours

Continuous Internal Evaluation Pattern:

Attendance : 10 marks

Continuous Assessment Tests : 25 marks

Continuous Assessment Assignment: 15 marks

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks

First Internal Examination shall be preferably conducted after completing the first half of the syllabus and the Second Internal Examination shall be preferably conducted after completing remaining part of the syllabus.

There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly covered module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly covered module), each with 7 marks. Out of the 7 questions in Part B, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 questions from each module of which a student should answer any one. Each question can have maximum 2 sub-divisions and carries 14 marks.

Syllabus

Module -1(Introduction to HCI and Usability)

Introduction- - Components of Interaction – Ergonomics Designing Interactive systems – Understanding Users cognition and cognitive frameworks, User Centered approaches, Usability goals and measures, Universal Usability-Diverse Cognitive and Perceptual abilities, Personality differences, Cultural and International diversity, Users with disabilities- Older Adult users and Children. Guidelines, Principles and Theories.

Module -2 (Design Process and Interaction Styles)

HCI patterns, Design frameworks, Design methods, Prototyping. Understanding interaction styles - Direct Manipulation and Immersive environments, Fluid navigation - Navigation by Selection, Small Displays, Content Organization, Expressive Human and Command Languages-Speech Recognition, Traditional Command Languages, Communication and Collaboration-Models of Collaboration, Design considerations.

Module -3 (User Experience Design)

Frameworks for User Centric Computing, Computational models of users, Advancing the User Experience- Display Design, View (Window) Management, Animation, Webpage Design, Color. Timely user Experience-Models of System Response Time (SRT) Impacts, Frustrating Experiences, Information Search- Five Stage Search Framework, Data Visualization-Tasks in Data Visualization, Challenges

Module -4 (Cognitive Systems and Evaluation of HCI)

Cognitive Models- Goal and task hierarchies, GOMS Model. Introducing Evaluation-Types of Evaluation, Other Issues to Consider When Doing Evaluation. Conducting Experiments. Usability testing – Heuristic evaluation and walkthroughs, Analytics and predictive models.

Module -5 (Contexts for Designing UX)

Designing apps and websites – Website and app development, The information architecture of apps and websites. Social media -Social Networking, Sharing with others. Collaborative environments- Issues for cooperative working, Technologies to support cooperative working, AI and Interface Agents, Ubiquitous computing -Blended Spaces. Mobile Computing – Designing for Mobiles. Wearable Computing- Smart Materials, Material Design.

Text Book

1. Ben Shneiderman, Catherine Plaisant, Maxine Cohen, Steven Jacobs, NiklasElmqvist"Designing the User Interface: Strategies for Effective Human-Computer Interaction", Sixth Edition, Pearson Education, 2017.
2. Preece, J., Sharp, H., Rogers, Y., "Interaction Design: Beyond Human-Computer Interaction", Fifth Edition, Wiley, 2019.

3. David Benyon, "Designing User Experience: A guide to HCI, UX and interaction design", 4th Edition, Pearson, 2018.

Reference Books

1. Alan Dix, Janet Finlay, Gregory Abowd, Russell Beale, "Human Computer Interaction", Third Edition, Prentice Hall, 2004.
2. The essential guide to user interface design, Wilbert O Galitz, Wiley Dream Tech
3. Jonathan Lazar Jinjuan Heidi Feng, Harry Hochheiser, "Research Methods in Human-Computer Interaction", Wiley, 2010.
4. Samit Bhattacharya, "Human-Computer Interaction: User-Centric Computing for Design", McGraw-Hill India, 1st Edition, 2019.

Sample Course Level Assessment Questions

Course Outcome 1 (CO1):

1. What are the general principles of user interface design?
2. How can designers encourage novice users to use a system?
3. Define user interface. List and explain the benefits of good design.

Course Outcome 2 (CO2):

1. Design a touch screen music jukebox, which allows the user to select from a menu of the five most popular songs of the week. Draw a sketch of this interface for each of the following menu types—binary menu, multiple-item menu, check boxes, pull-down menus. Argue which design serves the user best.
2. List several situations when command languages can be attractive for users.

Course Outcome 3(CO3):

1. Explain how data visualization caters to the perceptual abilities of humans.
2. Demonstrate the five stage framework in designing the advanced search interface.

Course Outcome 4 (CO4):

1. Discuss the GOMS Model
2. Explain how Fitt's Law predictive model has been influential in HCI and Interaction design.

Course Outcome 5 (CO5):

3. Distinguish between GUI and Web user interface.
4. List the issues faced for cooperative working.

Model Question paper

			Total Pages: 2
Reg No.:_____		**Name:_____**	

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY
FIFTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: AMT 416

Course Name: HUMAN COMPUTER INTERACTION

Max. Marks: 100		**Duration: 3 Hours**

PART A

		Answer all questions, each carries 3 marks.	Marks
1		Define Principles, Standards, Guidelines and Rules.	(3)
2		Explain the term Universal Usability.	(3)
3		Prototyping will solve all problems associated with user interface design. Justify this statement.	(3)
4		List the three principles of direct manipulation.	(3)
5		Describe frustrating experiences.	(3)
6		List any three reasons for using animation in display design.	(3)
7		Explain how Fitt's Law predictive model has been influential in HCI and Interaction design.	(3)
8		Coordination is a task concept that describes how information objects change based on user actions. Cite any two coordination that should be supported by interface designers.	(3)
9		Discuss any three principles of designing rich web interface.	(3)
10		Summarize three guidelines for developing applications for pocket PCs.	(3)

PART B

Answer any one full question from each module, each carries 14 marks.

		Module I	
11	a)	Explain the relationship between the user experience and usability.	(7)
	b)	Describe user-centered design. What are its benefits?	(7)
		OR	
12	a)	Explain the difference between good and poor interaction design.	(4)
	b)	What is cognitive and perceptual ability? Discuss with an example cognitive perception.	(10)
		Module II	

13	a)	Outline the various interface styles used in interactive systems.	(7)
	b)	Discuss the obstacles to speech recognition and production.	(7)
		OR	
14	a)	Data entry is challenging for small devices. Explain the ways in which this issue can be addressed?	(7)
	b)	Explain the different phases involved in an interactive design process.	(7)
		Module III	
15	a)	How do rule and heuristics help interface designers in taking account of cognitive psychology? Illustrate your answer with the design of Microsoft Office Word.	(8)
	b)	Discuss three human values that are necessary to be understood by interface designers in order to ensure a timely user experience. State any three system response time (SRT) guidelines.	(6)
		OR	
16	a)	Explain how data visualization caters to the perceptual abilities of humans.	(9)
	b)	Colour displays are attractive to users and can often improve task performance, but the danger of misuse is high. List five guidelines for using colour and give an example of each.	(5)
		Module IV	
17	a)	What is meant by design evaluation? Describe the approaches to expert analysis.	(8)
	b)	What is a cognitive model? Classify cognitive models and discuss the same.	(6)
		OR	
18	a)	How are download delays masked by well-designed websites?	(7)
	b)	Discuss the GOMS Cognitive task analysis model.	(7)
		Module V	
19	a)	List and explain the key attributes of wearable computing.	(8)
	b)	Describe how the UCAMP framework helps designers of wearable systems to focus on the key design issues.	(6)
		OR	
20	a)	Illustrate any two applications of agent-based interaction.	(8)
	b)	Describe the main types of technologies that support cooperative working.	(6)
		****	

Teaching Plan

Topics		No. of Lecture Hours (36 Hours)
Module -1 (Introduction to HCI and Usability)		**(8 hours)**
1.1	Introduction- - Components of Interaction – Ergonomics	1 hour
1.2	Designing Interactive systems – Understanding Users cognition and cognitive frameworks	1 hour
1.3	User Centered approaches, Usability goals and measures	1 hour
1.4	Universal Usability	1 hour
1.5	Diverse Cognitive and Perceptual abilities	1 hour
1.6	Personality differences, Cultural and International diversity,	1 hour
1.7	Users with disabilities- Older Adult users and Children.	1 hour
1.8	Guidelines, Principles and Theories.	1 hour
Module -2 Design Process and Interaction Styles		**(6 hours)**
2.1	HCI patterns, Design frameworks. Design considerations.	1 hour
2.2	Understanding interaction styles- Direct Manipulation and Immersive environments,	1 hour
2.3	Fluid navigation -Navigation by Selection, Small Displays, Content Organization	1 hour
2.4	Expressive Human and Command Languages-Speech Recognition, Traditional Command Languages	1 hour
2.5	Communication and Collaboration-Models of Collaboration	1 hour
2.6	Design methods, Prototyping	1 hour
Module -3 User Experience Design		**(7 hours)**
3.1	Frameworks for User Centric Computing	1 hour
3.2	Computational models of users,	1 hour
3.3	Advancing the User Experience- Display Design, View (Window) Management,	1 hour
3.4	Animation, Webpage Design, Color	1 hour

3.5	Timely user Experience-Models of System Response Time (SRT) Impacts, Frustrating Experiences.	1 hour
3.6	Information Search- Five Stage Search Framework,	1 hour
3.7	Data Visualization-Tasks in Data Visualization, Challenges	1 hour
Module -4 Cognitive Systems and Evaluation of HCI		**(7 hours)**
4.1	Cognitive Models- Goal and task hierarchies.	1 hour
4.2	GOMS Model.	1 hour
4.3	Introducing Evaluation- Types of Evaluation	1 hour
4.4	Other Issues to Consider When Doing Evaluation.	1 hour
4.5	Conducting Experiments	1 hour
4.6	Usability testing – Heuristic evaluation and walkthroughs	1 hour
4.7	Analytics and predictive models	1 hour
Module -5 Contexts for Designing UX		**(8 hours)**
5.1	Designing apps and websites – Website and app development	1 hour
5.2	The information architecture of apps and websites.	1 hour
5.3	Social media -Social Networking, Sharing with others.	1 hour
5.4	Collaborative environments- Issues for cooperative working, Technologies to support cooperative working	1 hour
5.5	AI and Interface Agents	1 hour
5.6	Ubiquitous computing -Blended Spaces.	1 hour
5.7	Mobile Computing – Designing for Mobiles.	1 hour
5.8	Wearable Computing- Smart Materials, Material Design.	1 hour

AIT426	Mining of Massive Data Sets	Category	L	T	P	Credit	Year of Introduction
		PEC	2	1	0	3	2019

Preamble:

This course introduces concepts in mining of massive data sets. It covers different mining algorithms, distributed file systems and map-reduce as a tool for creating parallel algorithms that succeed on very large amounts of data.

Prerequisite: Sound knowledge in Data Analytics

Course Outcomes: After the completion of the course the student will be able to

CO1	Describe the process of extracting useful features in developing models(**Cognitive Knowledge level: Understand)**	
CO2	Make use of the concepts of MapReduce methodology for exploiting parallelism in computing clouds. (**Cognitive Knowledge level: Apply**)	
CO3	Explain applications of hashing that make management of stream data (Cognitive Knowledge level: **Understand)**	
CO4	Examines the problem of clustering to analyse a large amount of data and partition it into subsets.(Cognitive Knowledge level: **Apply**	
CO5	Describe on-line advertising, social networks and algorithms for their analysis. (Cognitive Knowledge level: **Understand**)	

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓										✓
CO2	✓	✓	✓	✓								✓
CO3	✓	✓										✓
CO4	✓	✓	✓	✓								✓
CO5	✓	✓	✓									✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and team work
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Life long learning

Assessment Pattern

Bloom's Category	Continuous Assessment Tests		End Semester Examination Marks (%)
	Test 1 (%)	Test 2 (%)	
Remember	**20**	**20**	**20**
Understand	**60**	**60**	**60**
Apply	**20**	**20**	**20**
Analyze			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	**50**	**100**	**3**

Continuous Internal Evaluation Pattern:

Attendance	**10 marks**
Continuous Assessment Tests(Average of Internal Tests1&2)	**25 marks**
Continuous Assessment Assignment	**15 marks**

Internal Examination Pattern

Each of the two internal examinations has to be conducted out of 50 marks. First series test shall be preferably conducted after completing the first half of the syllabus and the second series test shall be preferably conducted after completing remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly completed module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly completed module), each with 7 marks. Out of the 7 questions, a student should answer any5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 full questions from each module of which student should answer any one. Each question can have maximum 2 sub-divisions and carries 14 marks.

Syllabus

Module -1(Data Mining)

Data Mining- Statistical Modeling, Machine Learning, Computational Approaches to Modeling, Summarization, Feature Extraction. Statistical Limits on Data Mining- Total Information Awareness, Bonferroni's Principle, Importance of Words in Documents, Hash Functions, Secondary Storage, The Base of Natural Logarithms, Power Laws.

Module -2(MapReduce and the New Software Stack)

Distributed File Systems, MapReduce-The Map Tasks, Grouping by Key, The Reduce Tasks, Combiners, Details of MapReduce Execution, Details of MapReduce Execution, Algorithms Using MapReduce- Matrix-Vector Multiplication by MapReduce, If the Vector v Cannot Fit in Main Memory, Relational-Algebra Operations, Computing Selections by MapReduce, Computing Projections by MapReduce, Union, Intersection, and Difference by MapReduce, Computing

Natural Join by MapReduce, Grouping and Aggregation by MapReduce, Matrix Multiplication, Matrix Multiplication with One MapReduce Step, Extensions to MapReduce.

Module -3 (Mining Data Streams)

The Stream Data Model, Sampling Data in a Stream, Filtering Streams, Counting Distinct Elements in a Stream, Estimating Moments- Definition of Moments, The Alon-Matias-Szegedy Algorithm for Second Moments, Counting Ones in a Window.

Module -4(Clustering)

Introduction to Clustering Techniques-Points, Spaces, and Distances, Clustering Strategies, The Curse of Dimensionality, Hierarchical Clustering, K-means Algorithms, The CURE Algorithm, Clustering in Non-Euclidean Spaces, Clustering for Streams and Parallelism- The Stream-Computing Model, A Stream-Clustering Algorithm, Initializing Buckets, Merging Buckets, Answering Queries, Clustering in a Parallel Environment

Module -5 (Advertising on the Web)

Issues in On-Line Advertising, On-Line Algorithms, The Matching Problem, The Adwords Problem, Adwords Implementation. Mining Social-Network Graphs:Social Networks as Graphs, Clustering of Social-Network Graphs.

Text Book

1. Anand Rajaraman, Jure Leskovec, and Jeffrey D. Ullman “Mining of Massive Datasets”

Course Level Assessment Questions

Course Outcome1 (CO1):

1. Explain the Total Information Awareness
2. Discuss the Base of Natural Logarithms

Course Outcome 2(CO2):

1. Explain the Cluster Computing
2. What are the Applications of MapReduce

Course Outcome 3(CO3):

1. Demonstrate the Bloom Filters.
2. Discuss Moments of Streams.

Course Outcome 4(CO4): .

1. How CURE Algorithm works.
2. Differentiate Centroids and Clustroids

Course Outcome 5(CO5):

1. Discuss how Greedy Algorithms work.
2. Explain Competitive Ratio

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: AIT426

Course Name: Mining of Massive Data Sets

Max.Marks:100 **Duration: 3 Hours**

PART A

Answer All Questions. Each Question Carries 3 Marks

1. State Bonferroni's Principle.
2. Explain Power laws.
3. Illustrate the concepts of Map Reduce.
4. Explain Distributed File Systems.
5. List the Issues in Stream Processing.
6. How counting distinct elements is done in a stream.

7. How CURE algorithm begin in clustering.

8. To Merge two consecutive buckets, what things are done in clustering.

9. What are the Issues in On-Line Advertising.

10. List any three essential characteristics of a social network.

(10x3=30)

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) Discuss Some of the important kinds of feature extraction from large-scale data. **(7)**

    (b) Explain Term Frequency times Inverse Document Frequency (TF.IDF) measure of word importance in a documents. **(7)**

OR

12. (a) Explain Hash Function and their use with an example in data mining. **(10)**

    (b) Demonstrate Statistical Modelling in data mining with an example. **(4)**

13. (a) What are the relational algebra operations used in data base queries of map reduce. **(8)**

    (b) Illustrate Matrix-Vector Multiplication by MapReduce. **(6)**

OR

14. (a) How Grouping and Aggregation can be done by MapReduce. **(7)**

    (b) Design MapReduce algorithms to take a very large file of integers and produce as output: **(7)**

    (a) The largest integer.

    (b) The average of all the integers.

    (c) The same set of integers, but with each integer appearing only once.

(d) The count of the number of distinct integers in the input.

15. (a) Why the Alon-Matias-Szegedy Algorithm Works. **(7)**

(b) How stream data arises naturally in the stream data model. **(7)**

OR

16. (a) Explain The Flajolet-Martin Algorithm. **(7)**

(b) With a motivating example explain sampling data in a stream **(7)**

17. (a) Illustrate K-means Algorithm. **(8)**

(b) Perform a hierarchical clustering of the one-dimensional set of points 1, 4, 9, 16, 25, 36, 49, 64, 81, assuming clusters are represented by their centroid (average), and at each step the clusters with the closest centroids are merged. **(6)**

OR

18. (a) Explain Clustering in non- Euclidean spaces. **(7)**

(b) Summarize the alternative rules for controlling hierarchical clustering in a Euclidean space. **(7)**

19. (a) Illustrate the problem of matching ads to search queries with an example. **(7)**

(b) Explain the Adwords Problem. **(7)**

OR

20. (a) Explain the clustering of social networks graph. **(7)**

(b) What are the varieties of social networks when we consider social network as a graph. **(7)**

Teaching Plan

No	Contents	No. of Lecture Hours (36 hrs)
Module -1 (Data Mining)(6 hours)		
1.1	Data Mining- Statistical Modeling, Machine Learning	1 hour
1.2	Computational Approaches to Modeling, Summarization, Feature Extraction	1 hour
1.3	Statistical Limits on Data Mining- Total Information Awareness, Bonferroni's Principle	1 hour
1.4	Importance of Words in Documents, Hash Functions	1 hour
1.5	Indexes, Secondary Storage	1 hour
1.6	The Base of Natural Logarithms, Power Laws	1 hour
Module -2 (MapReduce and the New Software Stack) (8 hours)		
2.1	Distributed File Systems, The Communication Cost Model	1 hour
2.2	MapReduce-The Map Tasks, Grouping by Key, The Reduce Tasks, Combiners	1 hour
2.3	Details of MapReduce, Coping with Node Failures	1 hour
2.4	Algorithms Using MapReduce- Matrix-Vector Multiplication by MapReduce, If the Vector v Cannot Fit in Main Memory	1 hour
2.5	Relational-Algebra Operations, Computing Selections by MapReduce, Computing Projections by MapReduce	1 hour
2.6	Union, Intersection, and Difference by MapReduce, Computing Natural Join by MapReduce	1 hour
2.7	Grouping and Aggregation by MapReduce, Matrix Multiplication, Matrix Multiplication with One MapReduce Step	1 hour
2.8	Extensions to MapReduce	1 hour
Module -3 (Mining Data Streams) (7 hours)		
3.1	The Stream Data Model	1 hour
3.2	Sampling Data in a Stream	1 hour
3.3	Filtering Streams	1 hour

3.4	Counting Distinct Elements in a Stream	1 hour
3.5	Estimating Moments- Definition of Moments, The Alon-Matias-Szegedy Algorithm for Second Moments	1 hour
3.6	Counting Ones in a Window	1 hour
3.7	Storage Requirements for the DGIM Algorthim	1 hour
Module -4(Clustering) (8 hours)		
4.1	Introduction to Clustering Techniques-Points, Spaces, and Distances, Clustering Strategies, The Curse of Dimensionality	1 hour
4.2	Hierarchical Clustering	1 hour
4.3	K-means Algorithms	1 hour
4.4	The CURE Algorithm	1 hour
4.5	Clustering in Non-Euclidean Spaces	1 hour
4.6	Clustering for Streams and Parallelism- The Stream-Computing Model, A Stream-Clustering Algorithm	1 hour
4.7	Initializing Buckets, Merging Buckets	1 hour
4.8	Answering Queries, Clustering in a Parallel Environment	1 hour
Module -5 (Advertising on the Web) (7 hours)		
5.1	Issues in On-Line Advertising	1 hour
5.2	On-Line Algorithms	1 hour
5.3	The Matching Problem	1 hour
5.4	The Adwords Problem	1 hour
5.5	Adwords Implementation	1 hour
5.6	Mining Social-Network Graphs: Social Networks as Graphs	1 hour
5.7	Clustering of Social-Network Graphs	1 hour

AIT 456	INTODUCTION TO REINFORCEMENT LEARNING	CATEGORY	L	T	P	CREDIT
		PEC	2	1	0	3

Preamble: This course covers fundamental principles and techniques in reinforcement learning. Reinforcement learning is concerned with building programs that learn how to predict and act in a stochastic environment, based on past experience. Applications of reinforcement learning range from classical control problems, such as dynamical system control, to game playing, inventory control, and many other fields. Topics include Markov decision process, dynamic programming, Monte Carlo, temporal difference, function approximation reinforcement learning algorithms, and applications of reinforcement learning. This course enables the leaners to apply reinforcement learning on real world applications and research problems.

Prerequisite: Computational Fundamental for Machine Learning

Course Outcomes: After the completion of the course the student will be able to

CO 1	Solve computational problems using probability and random variables **(Cognitive Knowledge Level : Apply)**
CO 2	Apply policy iteration and value iteration reinforcement learning algorithms **(Cognitive Knowledge Level: Apply)**
CO 3	Employ Monte Carlo reinforcement learning algorithms. **(Cognitive Knowledge Level: Apply)**
CO 4	Apply temporal-difference reinforcement learning algorithms **(Cognitive Knowledge Level: Apply)**
CO 5	Apply on-policy and off-policy reinforcement learning algorithms with function approximation **(Cognitive Knowledge Level: Understand)**

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓									✓
CO2	✓	✓	✓	✓								✓
CO3	✓	✓	✓	✓								✓

CO4	✓	✓	✓	✓								✓
CO5	✓	✓	✓	✓	✓							✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and team work
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Life long learning

Assessment Pattern

Bloom's Category	**Continuous Assessment Tests**		**End Semester Examination Marks**
	Test1 (percentage)	**Test2 (percentage)**	
Remember	30	30	30
Understand	30	30	30
Apply	40	40	40
Analyze			
Evaluate			
Create			

Mark distribution

Total Marks	**CIEMarks**	**ESEMarks**	**ESE Duration**
150	50	100	3 hours

Continuous Internal Evaluation Pattern:

Attendance: **10 marks**

Continuous Assessment Tests :**25 marks**

Continuous Assessment Assignment: **15 marks**

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks
First Internal Examination shall be preferably conducted after completing the first half of the syllabus and the Second Internal Examination shall be preferably conducted after completing remaining part of the syllabus.

There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly covered module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly covered module), each with 7 marks. Out of the 7 questions in Part B, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 questions from each module of which a student should answer any one. Each question can have maximum 2 sub-divisions and carries 14 marks.

Syllabus

Module 1: Review of Probability Concepts

Probability concepts review - Axioms of probability, concepts of random variables, probability mass function, probability density function, cumulative density functions, Expectation of random variables. Concepts of joint and multiple random variables, conditional and marginal distributions. Correlation

and independence.

Module 2: Markov Decision Process

Introduction to Reinforcement Learning (RL) terminology - Examples of RL, Elements of RL, Limitations and Scope of RL. Finite Markov Decision Processes - The Agent–Environment Interface, Goals and Rewards, Returns and Episodes, Policies and Value Functions.

Module 3: Prediction and Control

Dynamic Programming - Policy Evaluation (Prediction), Policy Improvement, Policy Iteration, Value Iteration. Monte Carlo Prediction, MonteCarlo Estimation of Action Values, Monte Carlo Control, Monte Carlo Control without Exploring Starts, Off-policy Prediction via Importance Sampling.

Module 4: Temporal-Difference (TD) Methods for Model Free Prediction And Control

TD Methods - TD Prediction, Advantages of TD Prediction Methods, Optimality of TD(0), Sarsa: On-policy TD Control, Q-learning: Off-policy TD Control, Expected Sarsa. n-step TD Prediction, n-step Sarsa, n-step Off-policy Learning. Off -policy Learning without Importance Sampling – The n step Tree Backup Algorithm

Module 5: Function Approximation Method

On-policy Prediction with Approximation - Value-function Approximation, The Prediction Objective, Stochastic-gradient Methods, Linear Methods.

Eligibility Traces - The λ-return, TD(λ), n-step Truncated λ-return Methods, Sarsa(λ).

Reference Books

1. Richard S. Sutton and Andrew G. Barto, Reinforcement Learning: An Introduction, , 2nd Edition
2. 2 Alberto Leon-Garcia, Probability, Statistics, and Random Processes for Electrical Engineering, 3rd Edition,

Sample Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Let ***J*** and ***T*** be independent events, where ***P(J)=0.4***and ***P(T)=0.7***. Find ***P(J*** $\cap$ T), P(J $\cup$ T) and ***P(J*** $\cap$ ***T')***
2. 2 Let ***A*** and ***B*** be events such that ***P(A)=0.45*** ,***P(B)=0.35*** and ***P(A*** $\cup$ B)=0.5. Find ***P(A*** | B).

3. A random variable **R**has the probability distribution as shown in the following table:

r	1	2	3	4	5
P(R=r)	0.2	a	b	0.25	0.15

Given that ***E(R)=2.85***, find ***a*** and ***b***and ***P(R>2)***.

A biased coin (with probability of obtaining a head equal to ***p> 0***) is tossed repeatedly and independently until the first head is observed. Compute the probability that the first head appears at an even numbered toss.

Course Outcome 2 (CO2):

1. What are the main differences between supervised learning and reinforcement learning?
2. Give examples of Markovian and non-Markovian environments?
3. Define the optimal state-value function $\mathbf{V_*(s)}$ for an MDP.

Course Outcome 3(CO3):

1. Explain policy iteration and value iteration? What are their similarities and differences?
2. Why Monte Carlo methods for learning value functions require episodic tasks? How is it that n-step TD methods avoid this limitation and can work with continuing tasks?
3. List any three uses of the depth parameter in the Monte-Carlo tree search procedure.

Course Outcome 4 (CO4):

1. Draw the backup diagram for 2-step Sarsa. Write the corresponding learning rule for 2- step Sarsa.
2. Why is Sarsa an on-policy algorithm while Q-learning is an off-policy algorithm?
3. How would you differentiate between learning algorithms using on-policy from those that use off-policy?
4. When using Temporal Difference learning, why is it better to learn action values (Q-values) rather than state values (V-values)?

Course Outcome 5 (CO5):

1. How do you deal with a large possible action space in reinforcement learning?
2. List any two benefits of policy gradient methods over value function-based methods.
3. What is the relation between Q-learning and policy gradients methods?

Model Question paper

QP CODE: PAGES:5

Reg No:_____

Name :_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: **AIT - 456**

Course Name: Introduction to Reinforcement Learning

Max.Marks:100 **Duration: 3 Hours**

PART A

Answer all questions, each carries 3 marks.

1 The first three digits of a telephone number are 452. If all the sequences of the remaining four digits are equally likely, what is the probability that a randomly selected telephone number contains seven distinct digits?

2 If X is a discrete uniform random variable, i.e., $P(X = k) = 1/n$ for $k = 1, 2, \ldots, n$, find $E(X)$and $Var(X)$.

3 Explain the Limitations and Scope of RL ?

4 Write down the Bellman expectation equation for state-value functions.

5 What is Monte Carlo Prediction ?

6 List any three advantages of Monte Carlo methods over dynamic programming techniques?

7 Draw the backup diagram for 2-step Q-learning. Write the corresponding learning rule

for 2-step Q-learning.

8 Why Monte Carlo methods for learning value functions require episodic tasks. How does **n**-step TD methods avoid this limitation and can work with continuing tasks?

9 What is Stochastic-gradient Methods

10 Value function based methods are oriented towards finding deterministic policies whereas policy search methods are geared towards finding stochastic policies. True or false? Justify.

(10x3=30)

Part B

Answer any one Question from each module. Each question carries 14 Marks

11 a) Three players play 10 independent rounds of a game, and each player has probability 1/3 of winning each round. Find the joint distribution of the numbers of games won by each of the three players. **(7)**

b) exponential random variables with parameter λ. Show that $X + Y$ and $X/Y$are independent. **(7)**

12 a) An experiment consists of throwing a fair coin four times. Find the probability mass function and the cumulative distribution function of the following random variables: **(7)**

i) the number of heads before the first tail

ii) the number of heads following the first tail

iii) the number of heads minus the number of tails

iv) the number of tails times the number of heads.

b) Let X be a continuous random variable with probability density function on $0 <= x <= 1$ defined by $f(x) = 3x2$. Find the pdf of $Y = X2$. **(7)**

13 a) What is the difference between a state value function **V(s)** and a state-action value function **Q(s,a)**? **(4)**

b) Consider designing a recycling robot whose job is to collect empty bottles around the building. The robot has a sensor to detect when a bottle is in front of it, and a gripper to pick up the bottle. It also senses the level of its battery. The robot can navigate, as well as pick up a bottle and throw a bottle it is holding in the trash. There is a battery charger in the building, and the robot should not run out of battery. **(10)**

i. Describe this problem as an MDP. What are the states and actions?

ii. Suppose that you want the robot to collect as many bottles as possible, while

not running out of battery. Describe what rewards would enable it to achieve this task.

OR

14 a) Define the state-value function $V\pi(s)$ for a discounted MDP. **(5)**

b) Consider a 4x4 grid world where the agent starts in the top left, the bottom righ state is terminal, rewards are always **-1**, $\gamma = 1$, and state transitions ar deterministic. Consider the policy that always chooses the action to move down except when it is on the bottom row, at which point it chooses the action to mov right. Starting with **v0(s) = 0** for all **s**, compute **v1, v2, ... , v7**. **(10)**

15 a) During a single iteration of the Value Iteration algorithm, we typically iterate over the states in **S** in some order to update **Vt(s)** to **Vt+1(s)** for all states **s**. Is it possible to do this iterative process in parallel? Explain why or why not. **(7)**

b) Explain n-step TD Prediction, n-step Sarsa and n-step Off-policy Learning **(7)**

OR

16 a) Suppose you are given a finite set of transition data. Assuming that the Markov model that can be formed with the given data is the actual MDP from which the data is generated, will the value functions calculated by the MC and TD methods necessarily agree? Justify. **(4)**

b) With respect to the expected Sarsa algorithm, is exploration required as it is in the normal Sarsa and Q-learning algorithms? Justify. **(5)**

c) For a specific MDP, suppose we have a policy that we want to evaluate through the use of actual experience in the environment alone and using Monte Carlo methods. We decide to use the first-visit approach along with the technique of always picking the start state at random from the available set of states. Will this approach ensure complete evaluation of the action value function corresponding to the policy? **(5)**

17 a) Consider the following **Q[S,A]** table **(9)**

	State 1	State 2
Action 1	1.5	2.5
Action 2	4	3

Assume the discount factor, **γ= 0.5**, and the step size, **α = 0.1**. After the experience**(s, a, r, s') = (1, 1, 5, 2),** which value of the table gets updated and what is its new value?

b) What is the difference between Q-learning and Sarsa? (5)

OR

18 a) Consider the following **Q[S,A]** table Assume the discount factor, **γ= 0.5**, and the step size, **α = 0.1**. After the experience **(s, a, r, s', a')=(1, 1, 5, 2, 1),** which value of the table gets updated and what is its new value? (9)

	State 1	State 2
Action 1	1.5	2.5
Action 2	4	3

b) For Q-learning to converge we need to correctly manage the exploration vs.exploitation tradeoff. What property needs to be hold for the exploration strategy? (5)

19 a) Given the following sequence of states observed from the beginning of an episode, **s2, s1, s3, s2, s1, s2, s1, s6**, what is the eligibility value, **e7(s1)**, of state **s1**at time step 7 given trace decay parameter **λ**, discount rate **γ**, and initial value, **e0(s1) = 0**, when accumulating traces are used? What is the eligibility value if replacing traces are used? (8)

b) Suppose that we are using a policy gradient method to solve a reinforcement learning problem and the policy returned by the method is not optimal. Give three plausible reasons for such an outcome? (6)

20 a) Suppose that we have a Q-value function represented as a sigmoid function of a set of features: (8)

$$Q(\phi, a) = \frac{1}{1 + e^{\theta^T \phi}}$$

Write down the update rule that Sarsa would give for this function.

b) Suppose that in a particular problem, the agent keeps going back to the same state in a loop. What is the maximum value that can be taken by the eligibility trace of such a state if we consider accumulating traces with **λ = 0.25** and **γ = 0.8**? (6)

(14X5=70)

Teaching Plan

No	Topic	No. of Lectures (**35 Hours**)
1	**Module 1: Review of Probability Concepts**	**8**
1.1	Axioms of probability, concepts of random variables,	1 hour
1.2	probability mass function	1 hour
1.3	probability density function	1 hour
1.4	cumulative density functions	1 hour
1.5	Expectation of random variables	1 hour
1.6	Concepts of joint and multiple random variables	1 hour
1.7	conditional and marginal distributions	1 hour
1.8	Correlation and independence.	1 hour
2	**Module 2: Markov Decision Process**	**6**
2.1	Introduction to Reinforcement Learning (RL) terminology - Examples of RL, Elements of RL, Limitations and Scope of RL.	1 hour
2.2	Finite Markov Decision Processes	1 hour
2.3	The Agent–Environment Interface	1 hour
2.4	Goals and Rewards	1 hour
2.5	Returns and Episodes	1 hour
2.6	Policies and Value Functions	1 hour
3	**Module 3: Prediction and Control**	**7**
3.1	Dynamic Programming - Policy Evaluation (Prediction),	1 hour
3.2	Policy Improvement	1 hour
3.3	Policy Iteration, Value Iteration	1 hour
3.4	Monte Carlo Prediction	1 hour
3.5	MonteCarlo Estimation of Action Values,	1 hour
3.6	Monte Carlo Control without Exploring Starts	1 hour
3.7	Off-policy Prediction via Importance Sampling	1 hour
4	**Module 4: Temporal-Difference (TD) Methods**	**8**
4.1	TD Prediction, Advantages of TD Prediction Methods	1 hour
4.2	Optimality of TD(0)	1 hour
4.3	Sarsa: On-policy TD Control	1 hour
4.4	Q-learning: Off-policy TD Control	1 hour

4.5	Expected Sarsa	1 hour
4.6	n-step TD Prediction, n-step Sarsa	1 hour
4.7	n-step Off-policy Learning	1 hour
4.8	Off -policy Learning without Importance Sampling – The n step Tree ackupAlgorithm	1 hour
5	**lodule 5: Function Approximation Method**	**6**
5.1	Value-function Approximation,	1 hour
5.2	The Prediction Objective	1 hour
5.3	Stochastic-gradient Methods	1 hour
5.4	Linear Methods.	1 hour
5.5	Eligibility Traces - The λ-return, TD(λ)	1 hour
5.6	step Truncated λ-return Methods, Sarsa(λ).	1 hour

AIT 476	BIO-INSPIRED OPTIMIZATION TECHNIQUES	Category	L	T	P	Credit
		PEC	3	0	0	3

Preamble:

The aim of this course is to provide the students with the knowledge and skills required to design and implement Bio-inspired optimization techniques to problems for which a direct solution is impractical or unknown. This course covers concepts of evolutionary algorithms like genetic algorithms and various swarm optimization techniques like ACO, PSO. The learners will be able to provide Bio-inspired optimization solutions to real world problems.

Course Outcomes: After the completion of the course the student will be able to

CO1	Understand the fundamentals in bio-inspired optimization techniques which influence computing **(Cognitive Knowledge Level: Understand)**
CO2	Make use of the concepts of Evolutionary Algorithms, genetic algorithms in various domains. **(Cognitive Knowledge Level: Apply)**
CO3	Comprehend the concepts of Swarm Intelligence and collective systems such as ACO, PSO **(Cognitive Knowledge Level: Understand)**
CO4	Illustrate the concepts of biologically inspired algorithmic design**(Cognitive Knowledge Level: Understand)**
CO5	Select the most appropriate types of algorithms for different data analysis problems **(Cognitive Knowledge Level: Understand)**

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓											✓
CO2	✓	✓			✓							✓
CO3	✓	✓			✓							✓
CO4	✓	✓			✓							✓

CO5	✓	✓			✓							✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and team work
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Life long learning

Assessment Pattern

Bloom's Category	**Continuous Assessment Tests**		**End Semester Examination Marks (%)**
	Test 1 (%)	**Test 2 (%)**	
Remember	**20**	**20**	**20**
Understand	**70**	**70**	**70**
Apply	**10**	**10**	**10**
Analyze			
Evaluate			
Create			

Mark Distribution

Total Marks	**CIE Marks**	**ESE Marks**	**ESE Duration**
150	**50**	**100**	**3**

Continuous Internal Evaluation Pattern:

Attendance	**10 marks**
Continuous Assessment Tests(Average of Internal Tests1&2)	**25 marks**
Continuous Assessment Assignment	**15 marks**

Internal Examination Pattern

Each of the two internal examinations has to be conducted out of 50 marks. First series test shall be preferably conducted after completing the first half of the syllabus and the second series test shall be preferably conducted after completing remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly completed module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly completed module), each with 7 marks. Out of the 7 questions, a student should answer any5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 full questions from each module of which student should answer any one. Each question can have maximum 2 sub-divisions and carries 14 marks.

Syllabus

Module – 1 (Optimization Techniques) (7 hours)

Optimization Techniques: Understanding optimization process- Objective function, minima & maxima, convergence. Optimization methods- conventional methods, Gradient descent algorithm-drawbacks. Introduction to Optimization Problems – classification- Single and Muti- objective Optimization – Classical Techniques – Overview of various Optimization methods . Bio-inspired Computing (BIC): Motivation – Overview of BIC – usage of BIC – merits and demerits of BIC.

Module– 2(Evolutionary Computing) (7 hours)

Evolutionary Computing: Motivation, Genetic Algorithm and Genetic Programming: Basic concept – encoding – representation – fitness function – Population, Operators – Selection, Mutation, Crossover, Reproduction – Types of Evolutionary Algorithms, Differences between GA and Traditional optimization methods – Applications.

Module- 3 (Ant Colony Systems) (8 hours)

Swarm intelligent systems – Background. Ant colony systems – Biological systems, Development of the ant colony system- - Working of ACO Algorithm - Pheromone updating- Types of ant systems- ACO algorithms for TSP.

Module- 4 (Particle Swarm Optimization) (7 hours)

Foraging for food – Clustering of objects – Collective Prey retrieval – Scope of Swarm Robotics – Social Adaptation of Knowledge: Particle Swarm – Particle Swarm Optimization (PSO) – Particle Swarms for Dynamic Optimization Problems – Bee-inspired optimization, Artificial Bee Colony (ABC) Optimization , applications.

Module- 5 (Case Studies) (6 hours)

Other Swarm Intelligence algorithms: Fish Swarm – Bacteria foraging – Intelligent Water Drop Algorithms – Applications of biologically inspired algorithms in engineering. Case Studies: ACO for NP-hard problems – Routing problems – Assignment problems – Scheduling problems.

ReferenceBooks

1. A. E. Elben and J. E. Smith, “Introduction to Evolutionary Computing”, Springer,2010.
2. S. N. Sivanandam and S.N. Deepa, Principles of Soft Computing , 2nd Edition, John Wiley & Sons.
3. Marco Dorigo and Thomas Stutzle, “Ant Colony optimization”, Prentice Hall of India, New Delhi,2005.
4. FloreanoD. and Mattiussi C., "Bio-Inspired Artificial Intelligence: Theories,Methods, and Technologies", MIT Press, Cambridge, MA, 2008.
5. Leandro Nunes de Castro, " Fundamentals of Natural Computing, BasicConcepts, Algorithms and Applications", Chapman & Hall/ CRC, Taylor andFrancis Group, 2007.
6. SatyobrotoTalukder, Blekinge Institute of Technology, Mathematical Modelling and Applications of Particle Swarm Optimization, February 2011.
7. Christian Blum and Daniel Merkle, “Swarm Intelligence Introduction and Application”, Springer 2008.

Course Level Assessment Questions

Course Outcome1 (CO1):

1. Explain the need for bio-inspired computation algorithms.
2. Differentiate between Bio-inspired optimization and other optimization techniques.

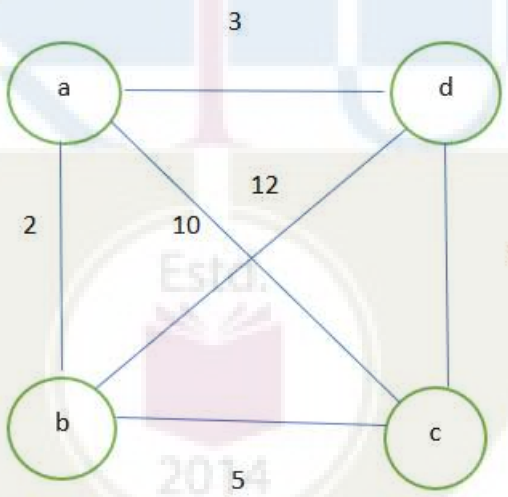
Course Outcome 2(CO2):

1. Describe how the Roulette wheel is used for selection. Draw the Roulette wheel for six chromosomes corresponding to the table given below.

Chromosome #	*Fitness*
1	10
2	5
3	25
4	15
5	30
6	20

Course Outcome 3(CO3):

1. Consider an Ant Colony System based on the Ant Quantity model for solving the following Travelling Salesman Problem. Compute the pheromone content at each of the edges after 4 steps (1 iteration). Assume pheromone decay factor ρ= 0.1, Q = 120. Assume an initial pheromone of 50 units at each of the edges and that three ants k1, k2 and k3 follow the paths given below in the first iteration. k1= a b c d a; k2=a c b d a; k3=a d c b a



Course Outcome 4(CO4): .

1. Consider a particle swarm optimization system composed of three particles and maximum velocity 10. Assume that both the random numbers r1 and r2 used for computing the movement of the particle towards the individual best position and social best position are 0.5. Also assume that the space of solutions is the two dimensional real valued space and the current state of swarm is as follows: Position of particles: x1 = (4,4); x2 = (8,3); x3 = (6,7) Individual best positions : x14,4 = (*); x27,3 = (*); x35,6 = (*) Velocities: v1 = (2,2);

v2 = (3,3); v3 = (4,4) .What would be the next position of each particle after one iteration of the PSO algorithm if the inertia parameter ω that is used along with current velocity update formula is 0.8 ?

Course Outcome 5(CO5):

1. Discuss applications of bio-optimization techniques (ACO) for solving NP-hard problems.

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: AIT 476

Course Name: Bio-Inspired Optimization Techniques

Max.Marks:100 **Duration: 3 Hours**

PART A

Answer All Questions. Each Question Carries 3 Marks

1. Differentiate Optimization and Constraint Satisfaction problems.
2. Define bio-Inspired Optimization.
3. Specify the importance of objective (fitness) function in genetic algorithm.
4. Compare Single-Point Crossover and Two-Point Crossover.
5. Describe how pheromone is updated.

6. Define Swarm Intelligence and list the algorithms under SI.

7. What is the significance of pbest and gbest particles in solving problems with particle swarm optimization?

8. List the scope of swarm robotics.

9. What is Fish Swarm optimization algorithm.

10. Define an assignment problem? List the different types of Assignment problems.

(10x3=30)

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) Discuss about Optimization, modelling, and simulation problems. **(7)**

(b) Differentiate between Bio-inspired optimization and other optimization techniques **(7)**

OR

12. (a) What is Bio-Inspired Computing? Explain the working of BIC algorithms. **(7)**

(b) Discuss the merits and demerits of BIC. **(7)**

13. (a) Explain any procedure to map a solution to the corresponding chromosome and vice versa in genetic algorithms. Also illustrate it with an example: **(7)**

(b) Describe two methods used to select individuals from a population for the mating pool in Genetic Algorithms. **(7)**

OR

14. (a) Explain any two mutation methods. **(4)**

(b) Differentiate between value encoding and permutation encoding. **(10)**

15. (a) Describe Ant Colony System. What are the different types of Ant systems? **(7)**

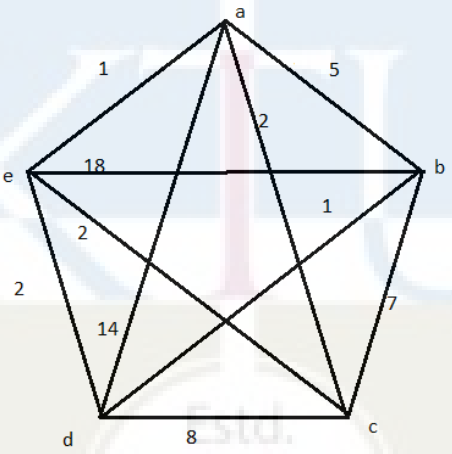
(b) Using the equation $T_{ij}(t+1)=(1-\rho)T_{ij}(t) + \Delta T_{ij}(t,t+1)$, compute the T_{ij} of the edge when 10 ants uses the edges, using the following models: **(7)**

i. Ant Density Model (Constant Q=10)

ii. Ant Quantity Model(Constant Q=100) , where Q is the constant related to the pheromone updation

OR

16. (a) Consider the TSP with the following edge costs. Given the evaporation factor ρ =0.02 and initial pheromone at all edges Tij=100. **(4)**



Compute the cost of the best tour?

(b) Describe ACO algorithm for TSP problems. **(10)**

17. (a) Illustrate Artificial Bee Colony optimization **(10)**

(b) List the advantages of Particle Swarm Optimization (PSO). **(4)**

OR

18. (a) Discuss Particle Swarm Optimization (PSO). **(6)**

(b) Explain the working of Particle Swarm Optimization (PSO) Algorithm. **(8)**

19. (a) Describe the working of Bacteria Foraging Algorithms. **(7)**

(b) Explain Intelligent Water Drop Algorithms . **(7)**

OR

20. (a) Discuss the different types of routing problems. **(6)**

(b) Discuss any four Applications of biologically inspired algorithms in engineering. **(8)**

Teaching Plan

No	Contents	No. of Lecture Hours (35hrs)
Module-1(Optimization Techniques) (7 hours)		
1.1	Understanding optimization process- Objective function, minima & maxima, convergence. Optimization methods- Gradient descent algorithm- drawbacks.	2hour
1.2	Introduction to Optimization Problems – classification- Single and Muti-objective Optimization	1 hour
1.3	Classical Techniques	1 hour
1.4	Overview of various Optimization methods	1 hour
1.5	Bio- inspired Computing (BIC): Motivation – Overview of BIC	1 hour
1.6	Usage of BIC – merits and demerits of BIC.	1 hour
Module-2 (Evolutionary Computing) (7hours)		
2.1	Evolutionary Computing: Motivation, Genetic Algorithm and Genetic Programming: Basic concepts	1 hour
2.2	Encoding – Representation	1 hour

2.3	Fitness function, Population, Reproduction	1 hour
2.4	Operators - Selection, Mutation	1 hour
2.5	Crossover, Reproduction	1 hour
2.6	Types of Evolutionary Algorithms	1 hour
2.7	Differences between GA and Traditional optimization methods – Applications.	1 hour
Module-3 (Ant colony systems) (8 hours)		
3.1	Swarm intelligent systems	1 hour
3.2	Background…	1 hour
3.3	Ant colony systems – Biological systems	1 hour
3.4	Development of the ant colony system	1 hour
3.5	Working of ACO Algorithm	1 hour
3.6	Pheromone updating	1 hour
3.7	Types of ant systems	1 hour
3.8	ACO algorithms for TSP	1 hour
Module-4 (Particle Swarm Optimization)) (7 hours)		
4.1	Foraging for food	1 hour
4.2	Clustering of objects	1 hour
4.3	Collective Prey retrieval	1 hour
4.4	Scope of Swarm Robotics	1 hour
4.5	Particle Swarm — Particle Swarms for Dynamic Optimization Problems	1 hour
4.6	Particle Swarm Optimization (PSO)	1 hour
4.7	Bee-inspired optimization, Artificial Bee Colony (ABC) Optimization , Applications	1 hour
Module-5 (CASE STUDIES) (6 hours)		
5.1	Other Swarm Intelligence algorithms: Fish Swarm	1 hour
5.2	Bacteria foraging	1 hour
5.3	Intelligent Water Drop Algorithms	1 hour
5.4	Applications of biologically inspired algorithms in engineering	1 hour

5.5	Case Studies: ACO for NP-hard problems – Routing problems – Assignment problems	1 hour
5.6	Scheduling problems	1 hour

CST436	PARALLEL COMPUTING	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PEC	2	1	0	3	2019

Preamble: This course helps the learners to understand basic and advanced concepts of parallel computing. It covers Principles of Parallel Algorithm Design, Communication operations, Programming Using the Message Passing Paradigm, Programming Shared Address Space Platforms Thread Basics, and GPU Programming. This course enables a learner to design solutions to complex real world problems using parallel computing paradigms including thread parallelism, shared memory program, message passing interfaces, and vector processing.

Prerequisite: Knowledge in Computer Organization and Architecture.

Course Outcomes: After the completion of the course the students will be able to

CO1	Summarize the key parallel computational models **(Cognitive Knowledge Level : Understand)**
CO2	Appreciate and apply parallel and distributed algorithms in problem Solving **(Cognitive Knowledge Level :Apply)**
CO3	Appreciate the communication models for parallel algorithm development (**Cognitive Knowledge Level : Understand**)
CO4	Develop parallel algorithms using message passing paradigm **(Cognitive Knowledge Level : Apply)**
CO5	Formulate parallel algorithms for shared memory architectures. **(Cognitive Knowledge Level : Apply)**
CO6	Demonstrate the fundamental skills of heterogeneous computing with GPUs**(Cognitive Knowledge Level : Apply)**

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓										✓
CO2	✓	✓	✓									✓
CO3	✓	✓										✓
CO4	✓	✓	✓	✓	✓							✓

CO5	✓	✓	✓	✓	✓							✓
CO6	✓	✓	✓	✓	✓							✓

Abstract POs Defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and teamwork
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Lifelong learning

Assessment Pattern

Blooms Category	Continuous Assessment Tests		End Semester Examination Marks
	Test 1 (Percentage)	Test 2 (Percentage)	
Remember	30	20	20
Understand	50	40	40
Apply	20	40	40

Analyze			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	**50**	**100**	**3 Hours**

Continuous Internal Evaluation Pattern:

Attendance	**10 marks**
Continuous Assessment Tests	**25 marks**
Continuous Assessment Assignment	**15 marks**

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks. First Internal Examination shall be preferably conducted after completing the first half of the syllabus, and the Second Internal Examination shall be preferably conducted after completing the remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly covered module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly covered module), each with 7 marks. Out of the 7 questions in Part B, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 questions from each module of which a student should answer anyone. Each question can have a maximum 2 subdivisions and carries 14 marks.

Syllabus

Module- 1 (Principles of Parallel Algorithm Design)

Basic Introduction to Parallel Processing platforms. Preliminaries, Decomposition Techniques, Characteristics of Tasks and Interactions, Mapping Techniques for Load Balancing, Methods for Containing Interaction Overheads, Parallel Algorithm Models.

Module- 2 (Communication Operations)

Basic Communication Operations - One-to-All Broadcast and All-to-One Reduction, All-to-All Broadcast and Reduction, All-Reduce and Prefix-Sum Operations, Scatter and Gather, All-to-All Personalized Communication, Circular Shift, Improving the Speed of Some Communication Operation

Module-3 (Programming Using the Message Passing Paradigm)

Principles of Message-Passing Programming, The Building Blocks: Send and Receive Operations, MPI: The Message Passing Interface, Overlapping Communication with Computation, Collective Communication and Computation Operations, Groups and Communicators.

Module 4 (Programming Shared Address Space Platforms Thread Basics)

Thread Basics, Why Threads? The POSIX Thread Application Programme Interface, Synchronization Primitives in POSIX, Controlling Thread and Synchronization Attributes, Thread Cancellation, Composite Synchronization Constructs, OpenMP: a Standard for Directive Based Parallel Programming, Specifying Concurrent Tasks in OpenMP, Synchronization Constructs in OpenMP, Data Handling in OpenMP, OpenMP Library Functions, OpenMP Applications: Parallel algorithm development for Matrix multiplication

Module 5 (GPU Programming)

Heterogeneous Parallel Computing, Architecture of a Modern GPU, Speeding up Real Applications, Data parallel computing, CUDA C Program Structure, A Vector Addition Kernel, Device Global Memory and Data Transfer, Kernel Functions and Threading, Kernel Launch, CUDA Thread Organization, Mapping Threads to Multidimensional Data, Synchronization and Transparent Scalability, Resource Assignment, Querying Device Properties, Thread Scheduling and Latency Tolerance, Importance of Memory Access Efficiency, Cuda Memory Types, Tiling for Reduced Memory Traffic, Tiled Matrix Multiplication Kernel, Boundary Checks

Text Books

1. Ananth Grama, Anshul Gupta, George Karypis, Vipin Kumar, Introduction to Parallel Computing, 2nd Ed, Addison-Wesley, 2003
2. David B. Kirk, Wen-mei W. Hwu, Programming Massively Parallel Processors: A Hands-on Approach, 3rd Ed., Morgan Kaufman, 2016.

References

1. Steven Brawer, Introduction to Parallel Computing, Academic Press, (1989)
2. Barbara Chapman, Gabriele Jost, Ruud van der Pas, Using OpenMP: Portable Shared Memory Paralwlel Programming , MIT Press, 2008.
3. William Gropp, Ewing Lusk, Anthony Skjellum Using MPI: Portable Parallel Programming with the Message-Passing Interface, 3rd Ed, MIT Press, 2014.
4. Thomas Rauber, Gudula Rünger, Parallel Programming for Multicore and Cluster Systems, Springer, 2010

Sample Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Differentiate between static and dynamic task mapping
2. Explain partitioning of data with an example

Course Outcome 2 (CO2):

1. Explain the handshaking sequence of Blocking Non-Buffered Send/Receive operation with a neat diagram.
2. In the algorithm, assume a decomposition such that each execution of Line 7 is a task. Draw a task-dependency graph and a task-interaction graph.

```
1.procedure FFT_like_pattern(A, n)
2.begin
3.m := log2 n;
4.for j := 0 to m - 1 do
5.k := 2j;
6.for i := 0 to n - 1 do
7.A[i] := A[i] + A[i XOR 2j];
8.end  // for
9.end    // FFT_like_pattern
```

Course Outcome 3 (CO3):

1. Write a procedure for performing all-to-all reduction on a mesh
2. Give a hypercube algorithm to compute prefix sums of n numbers if p is the number of nodes and n/p is an integer greater than 1. Assuming that it takes time t_{add} to add two numbers and time ts to send a message of unit length between two directly-connected nodes, give an exact expression for the total time taken by the algorithm.

Course Outcome 4(CO4):

1. Show how the two-dimensional matrix-vector multiplication program needs to be changed so that it will work correctly for a matrix of size $n \times m$ on a $q \times r$ process grid
2. One of the advantages of non-blocking communication operations is that they allow the transmission of the data to be done concurrently with computations. Discuss the type of restructuring that needs to be performed on a program to allow for the maximal overlap of computation with communication. Is the sending process in a better position to benefit from this overlap than the receiving process

Course Outcome 5(CO5):

1. Implement a multi-access threaded queue with multiple threads inserting and multiple threads extracting from the queue. Use mutex-locks to synchronize access to the queue. Document the time for 1000 insertions and 1000 extractions each by 64 insertion threads (producers) and 64 extraction threads (consumers).
2. Implement a producer-consumer framework in OpenMP using sections to create a single producer task and a single consumer task. Ensure appropriate synchronization using locks.

Course Outcome 6 (CO6):

1. Consider a hypothetical block with 8 threads executing a section of code before reaching a barrier. The threads require the following amount of time (in microseconds) to execute the sections: 2.0, 2.3, 3.0, 2.8, 2.4, 1.9, 2.6, and 2.9 and to spend the rest of their time waiting for the barrier. What percentage of the total execution time of the thread is spent waiting for the barrier?
2. Write and explain the CUDA program for vector addition.

Model Question Paper

QP CODE: PAGES :3

Reg No: _____

Name: _____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: CST436

Course Name: PARALLEL COMPUTING

Max.Marks:100 **Duration: 3 Hours**

PART A

Answer All Questions. Each Question Carries 3 Marks

1. Explain partitioning of data with an example
2. Which are the characteristics of tasks influencing the selection of mapping scheme?
3. Describe the scatter - gather communication.
4. Explain the Circular Shift operation.
5. Explain the handshaking sequence of Blocking Non-Buffered Send/Receive operation with a neat diagram.
6. Describe the six fundamental routines of MPI.
7. Explain thread cancellation.
8. Explain how concurrent tasks are specified in openMP
9. Explain the architecture of modern GPU with a diagram.
10. Describe how the data transfer between GPU device and the host memories are managed. **(10x3=30)**

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) Describe recursive decomposition with an example. **(8)**

(b) Compare various parallel algorithm models **(6)**

OR

12. (a) Differentiate between static and dynamic task mapping **(8)**

(b) In the algorithm, assume a decomposition such that each execution of Line 7 is a task. Draw a task-dependency graph and a task-interaction graph. **(6)**

```
1. procedure FFT_like_pattern(A, n)
2. begin
3. m := log2 n;
4. for j := 0 to m - 1 do
5. k := 2j;
6. for i := 0 to n - 1 do
7. A[i] := A[i] + A[i XOR 2j];
8. end  // for
9. end    // FFT_like_pattern
```

13. (a) Illustrate the All-to-All Broadcast and Reduction with an example **(8)**

(b) Explain any three techniques to improve the speed of communication operations **(6)**

OR

14. (a) Explain the One-to-All Broadcast and All-to-One Reduction with an example **(8)**

(b) Explain the Ring and Mesh techniques of All-to-All Personalized communication. **(6)**

15. (a) Explain Collective Communication and Computation Operations in MPI **(9)**

(b) Show the impact of finite buffers in message passing. **(5)**

OR

16. (a) Write algorithm for Collective Communication and Computation Operations **(9)**

using MPI.

(b) How is deadlock avoided in *MPI_Send* and *MPI_Recv* **(5)**

17. (a) Explain how mutual exclusion for shared variables are accomplished in threads. **(6)**

(b) Explain the nesting of parallel directives with a suitable example. **(8)**

OR

18. (a) Explain the compilation operations of an example openMP program along with its *pThread* translations. **(4)**

(b) Explain the parallel matrix multiplication using openMP **(10)**

19. (a) Describe the CUDA Kernel functions. **(6)**

(b) How is synchronization between CUDA threads achieved? **(8)**

OR

20. (a) Explain the two-level hierarchical organization of CUDA threads. **(10)**

(b) Write and explain the CUDA program for vector addition. **(4)**

TEACHING PLAN

No	Contents	No of Lecture Hrs (37)
Module – 1 (Basic Introduction to Parallel Processing) (TB-1, Ch. 3) (7 hrs)		
1.1	Basic Introduction to Parallel Processing platforms. Preliminaries	1
1.2	Decomposition Techniques – Recursive, Data	1
1.3	Decomposition Techniques – Exploratory, Speculative, Hybrid	1
1.4	Characteristics of Tasks and Interactions	1
1.5	Mapping Techniques for Load Balancing -Static	1
1.6	Mapping Techniques for Load Balancing - Dynamic	1
1.7	Methods for Containing Interaction Overheads, Parallel Algorithm Models.	1
Module- 2 (Basic Communication Operations) (TB-1, Ch. 4) (6hrs)		
2.1	One-to-All Broadcast and All-to-One Reduction	1
2.2	All-to-All Broadcast and Reduction	1
2.3	All-Reduce and Prefix-Sum Operations, Scallter Gather	1
2.4	All-to-All Personalized Communication	1
2.5	Circular Shift	1
2.6	Improving the Speed of Some Communication Operation	1
Module- 3 (Programming Using the Message Passing Paradigm) (TB-1, Ch. 6) (7 hrs)		
3.1	Principles of Message-Passing Programming, The Building Blocks: Send and Receive Operations	1
3.2	MPI: The Message Passing Interface	1
3.3	MPI: The Message Passing Interface : Illustration	1

3.4	Overlapping Communication with Computation	1
3.5	Overlapping Communication with Computation : Illustration	1
3.6	Collective Communication and Computation Operations	1
3.7	Collective Communication and Computation Operations : Illustration	1
Module 4 (Programming Shared Address Space Platforms) (TB-1, Ch. 7, 8) (8hrs)		
4.1	Thread Basics, Why Threads? The POSIX Thread API	1
4.2	Synchronization Primitives in POSIX	1
4.3	Controlling Thread and Synchronization Attributes	1
4.4	Thread Cancellation, Composite Synchronization Constructs,	1
4.5	OpenMP: a Standard for Directive Based Parallel Programming	1
4.6	Specifying Concurrent Tasks in OpenMP, Synchronization Constructs in OpenMP	1
4.7	Data Handling in OpenMP, OpenMP Library Functions	1
4.8	OpenMP Applications: Parallel algorithm development for Matrix multiplication	1
Module 5 (GPU Programming) (TB-2, Ch. 1, 2) (9 hrs)		
5.1	Heterogeneous Parallel Computing, Architecture of a Modern GPU, Speeding up Real Applications	1
5.2	Data parallel computing – CUDA C Program Structure	1
5.3	Vector Addition Kernel, Device Global Memory and Data Transfer	1
5.4	Kernel Functions and Threading, Kernel Launch	1

5.5	CUDA Thread Organization, Mapping Threads to Multidimensional Data	1
5.6	Synchronization and Transparent Scalability, Resource Assignment, Querying Device Properties, Thread Scheduling and Latency Tolerance	1
5.7	Importance of Memory Access Efficiency, Cuda Memory Types	1
5.8	Tiling for Reduced Memory Traffic	1
5.9	Tiled Matrix Multiplication Kernel, Boundary Checks	1

CST446	DATA COMPRESSION TECHNIQUES	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PEC	2	1	0	3	2019

Preamble: This course helps the learners to understand compression techniques on text, image, audio and video data. It covers lossy &lossless compression, RLE, JPEG, MPEG and its variants. This course enables the students to develop and implement compression algorithms on different domains.

Prerequisite: Knowledge of probability theory, computation on matrices, basic topics in data structures, storage and efficiency

Course Outcomes: After the completion of the course the student will be able to

CO#	CO
CO1	Describe the fundamental principles of data compression**(Cognitive Knowledge level: Understand)**
CO2	Make use of statistical and dictionary based compression techniques for various applications **(Cognitive Knowledge level: Apply)**
CO3	Illustrate various image compression standards. **(Cognitive Knowledge level: Apply)**
CO4	Summarize video compression mechanisms to reduce the redundancy in video.**(Cognitive Knowledge level: Understand)**
CO5	Use the fundamental properties of digital audio to compress audio data.**(Cognitive Knowledge level: Understand)**

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓											✓
CO2	✓	✓	✓		✓							✓
CO3	✓	✓	✓		✓							✓
CO4	✓											✓
CO5	✓	✓	✓									✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and team work
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Life long learning

Assessment Pattern

Bloom's Category	Continuous Assessment Tests		End Semester Examination Marks (%)
	Test 1 (%)	Test 2 (%)	
Remember	30	30	30
Understand	40	40	40
Apply	30	30	30
Analyze			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	50	100	3

Continuous Internal Evaluation Pattern:

Attendance **10 marks**

Continuous Assessment Tests(Average of SeriesTests1& 2) **25 marks**

Continuous Assessment Assignment **15 marks**

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks. First series test shall be preferably conducted after completing the first half of the syllabus and the second series test shall be preferably conducted after completing remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly completed module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly completed module), each with 7 marks. Out of the 7 questions, a student should answer any5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 full questions from each module of which student should answer any one. Each question can have maximum 2 sub-divisions and carries 14 marks.

Syllabus

Module-1 (Modelling and types of compression)) 1

Introduction to Compression Techniques- Lossy compression & Lossless compression, Measures of Performance, Modeling and coding. Mathematical modelling for Lossless and lossy compression - Physical models and probability models.

Module – 2 (Basic Compression Methods)

Basic Compression Technique- Run length encoding, RLE Text compression. Statistical Methods- Prefix Codes, Binary Huffman coding, non-binary Huffman Algorithms, Arithmetic Coding.

Module - 3 (Text & Image Compression)

Dictionary based Coding- LZ77, LZ78 and LZW compression.Image Compression- Image standards, JPEG image Compression- Baseline JPEG, JPEG-LS.

Module - 4 (Video Compression)

Video Compression- Analog video, Digital Video, Motion Compensation. MPEG standards- MPEG 1, MPEG 4

Module - 5 (Audio Compression)

Audio Compression- Basics of Digital Audio, Basic Audio Compression Techniques, MPEG Audio Compression-Layer 1 coding, Layer 2 coding and Layer 3 coding.

Text Book

1. David Solomon, Data compression: the complete reference, 4/e, Springer, January 2007
2. Khalid Sayood, Introduction to data compression, Morgan Kaufmann Publishers,2003.

References

1) Stephen Welstead, Fractal and wavelet Image Compression techniques, PHI, 1999.
2) Sleinreitz, Multimedia System, Addison Wesley.
3) Mark Nelson and Jean-loup Gailly, The Data Compression Book, M&T Books.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Discuss different types of compression performance metrics
2. Explain mathematical model for lossless compression

Course Outcome 2 (CO2):

1. Explain RLE based text compression and identify a example with compression ratio of 2.
2. Given the eight symbols A, B, C, D, E, F, G, and H with probabilities 1/30, 1/30, 1/30, 2/30, 3/30, 5/30, 5/30, and 12/30, draw three different Huffman trees with heights 5 and 6 for these symbols and calculate the average code size for each tree.

Course Outcome 3 (CO3):

1. Differentiate the LZ77 and LZ78 performance with the input given as ‘sirsideastmaneasilyteasesseasickseals’
2. Explain why the continuous-tone images is required for JPEG and the main steps used in image compression.

Course Outcome 4 (CO4):

1. Briefly explain MPEG-4 video compression standard
2. How H.261 video compression is completed.

Course Outcome 5 (CO5):

1. Explain critical bands, thresholding and masking related to audio compression
2. Explain the working of -law encoder and decoder with an example

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 2**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: CST446

Course Name: Data Compression Techniques

Max.Marks:100 **Duration: 3 Hours**

PART A

Answer All Questions. Each Question Carries 3 Marks

1. Specify different quantities used to measure the performance of a data compression technique
2. Explain mathematical model for lossless compression
3. State and prove Kraft-McMillan inequality
4. Compare Huffman and Arithmetic coding
5. Describe LZ77 approach of encoding a string with the help of an example
6. Compare and contrast JPEG and JPEG-LS differences in working.
7. Discuss different components of video
8. Identify the advantage of MPEG-4 over MPEG
9. Explain critical bands, thresholding and masking related to audio compression
10. Explain the working of -law encoder and decoder with an example

(10x3=30)

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) Explain mathematical model for lossy compression and lossless compression **(10)**

    (b) Define compression ratio with an example **(4)**

OR

12. (a) Discuss any probability model and identify the shortcoming of the solution. **(7)**

    (b) Identify the mathematical preliminaries for Lossless Compression **(7)**

13. (a) With a help of flowchart discuss the RLE text compression for text data given below **(10)**

    '*ABBBBBBBBBCDEEEEF*'

    (b) calculate the compression ratio for the example while taking repetitions = 4 **(4)**

OR

14. (a) Illustrate with a example why Huffman coding is preferred than Shannon Fano Algorithm for compression **(10)**

    (b) How Huffman coding is handling the unpredictability of input data stream **(4)**

15. (a) Explain in detail the working of LZ78 with example and dictionary Tree **(10)**

    (b) Illustrate with example, how the compression factor LZW differ from the LZ78 **(4)**

OR

16. (a) How quantization and coding helps in compression and their role in JPEG. **(6)**

    (b) With the help of the given example illustrate the compression ratio of JPEG and JPEG-LS **(8)**

17. (a) With the help of equations discuss Composite and Components Video **(7)**

    (b) Differentiate the major changes in MPEG - 2 and MPEG-4 Video **(7)**

OR

18. (a) Describe in details about functionalities for MPEG-4 **(8)**

    (b) How Motion Compensation help in video compression **(6)**

19. (a) How The Human Auditory System limitations can be taken in audio **(7)**

compressions

(b) Discuss the complexity of Layer III compared to others in MPEG Audio Coding **(7)**

OR

20. (a) Discuss Format of Compressed Data and encoding in layer I and II **(9)**

(b) Differentiate Spectral and Temporal Masking **(5)**

TEACHING PLAN

No	Contents	No of Lecture Hrs (36 Hours)
	Module – 1 (Modelling and types of compression) (7 hrs)	
1.1	Introduction to Compression Techniques- Lossy compression & Lossless compression, Measures of Performance	2
1.2	Modelling and coding.	1
1.3	Physical model for lossless compression	1
1.4	Physical model for lossy compression	1
1.5	Probability model for lossless compression	1
1.6	Probability model for lossly compression	1
	Module - 2 (Basic Compression Methods) (8 hrs)	
2.1	Run length encoding, RLE Text compression	1
2.2	Statistical methods-Prefix Codes	1
2.3	Binary Huffman coding	1
2.4	Illustration of Binary Huffman coding	1
2.5	Non-binary Huffman Algorithms	1
2.6	Arithmetic Coding algorithm	1
2.7	Illustration of Arithmetic Coding algorithm	2

Module - 3 (Text & Image Compression) (8 hrs)		
3.1	LZ77 compression	2
3.2	LZ78 Compression	1
3.3	LZW Compression	1
3.4	Basics of Image compression and Image standards	1
3.5	Baseline JPEG Image compression	1
3.6	JPEG-LS Image compression	1
Module - 4 (Video Compression) (7 hrs)		
4.1	Basics of Video Compression- Analog video and Digital Video.	2
4.2	Motion Compensation	1
4.3	MPEG-1 standard and Video Syntax	1
4.4	MPEG-1 Pel Reconstruction	1
4.5	MPEG-4 standard	1
4.6	Functionalities for MPEG-4	1
Module - 5 (Audio Compression) (6 hrs)		
5.1	Basics of Audio Compression, Digital Audio	1
5.2	Basic Audio Compression Techniques	1
5.3	MPEG Audio Compression basics- Frequency Domain Coding	1
5.4	Encoding: Layers I and II	1
5.5	Encoding: Layer II -Psychoacoustic Models	1
5.6	Psychoacoustic Models - Encoding: Layer III	1

CST466	DATA MINING	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PEC	2	1	0	3	2019

Preamble: This course helps the learner to understand the concepts of data mining and data warehousing. It covers the key processes of data mining, data preprocessing techniques, fundamentals and advanced concepts of classification, clustering, association rule mining, web mining and text mining. It enables the learners to develop new data mining algorithms and apply the existing algorithms in real-world scenarios.

Prerequisite: NIL

Course Outcomes: After the completion of the course the student will be able to

CO#	CO
CO1	Employ the key process of data mining and data warehousing concepts in application domains. (**Cognitive Knowledge Level: Understand**)
CO2	Make use of appropriate preprocessing techniques to convert raw data into suitable format for practical data mining tasks (**Cognitive Knowledge Level**: **Apply**)
CO3	Illustrate the use of classification and clustering algorithms in various application domains (**Cognitive Knowledge Level: Apply**)
CO4	Comprehend the use of association rule mining techniques. (**Cognitive Knowledge Level: Apply)**
CO5	Explain advanced data mining concepts and their applications in emerging domains (**Cognitive Knowledge Level: Understand**)

Mapping of course outcomes with program outcomes

	PO 1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO1 0	PO11	PO1 2
CO1	✓	✓										✓
CO2	✓	✓	✓	✓	✓							✓
CO3	✓	✓	✓	✓	✓							✓

CO4	✓	✓	✓	✓	✓							✓
CO5	✓	✓										✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and team work
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Lifelong learning

Assessment Pattern

Bloom's Category	**Continuous Assessment Tests**		**End Semester Examination Marks (%)**
	Test 1 (%)	**Test 2 (%)**	
Remember	**20**	**20**	**20**
Understand	**30**	**30**	**30**
Apply	**50**	**50**	**50**
Analyze			
Evaluate			
Create			

Mark Distribution

Total Marks	**CIE Marks**	**ESE Marks**	**ESE Duration**
150	**50**	**100**	**3**

Continuous Internal Evaluation Pattern:

Attendance	**10 marks**
Continuous Assessment Test(Average of Internal Test1&2)	**25 marks**
Continuous Assessment Assignment	**15 marks**

Internal Examination Pattern

Each of the two internal examinations has to be conducted out of 50 marks. First series test shall be preferably conducted after completing the first half of the syllabus and the second series test shall be preferably conducted after completing the remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly completed module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly completed module), each with 7 marks. Out of the seven questions, a student should answer any five.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 full questions from each module of which student should answer any one. Each question can have a maximum 2 subdivisions and carries 14 marks.

Syllabus

Module – 1 (Introduction to Data Mining and Data Warehousing)

Data warehouse-Differences between Operational Database Systems and Data Warehouses, Multidimensional data model- Warehouse schema, OLAP Operations, Data Warehouse Architecture, Data Warehousing to Data Mining, Data Mining Concepts and Applications, Knowledge Discovery in Database Vs Data mining, Architecture of typical data mining system, Data Mining Functionalities, Data Mining Issues.

Module - 2 (Data Preprocessing)

Data Preprocessing-Need of data preprocessing, Data Cleaning- Missing values, Noisy data, Data Integration and Transformation, Data Reduction-Data cube aggregation, Attribute subset selection, Dimensionality reduction, Numerosity reduction, Discretization and concept hierarchy generation.

Module - 3 (Advanced classification and Cluster analysis)

Classification- Introduction, Decision tree construction principle, Splitting indices -Information Gain, Gini indexDecision tree construction algorithms-ID3, Decision tree construction with presorting-SLIQ, Classification Accuracy-Precision, Recall.

Introduction to clustering-Clustering Paradigms, Partitioning Algorithm- PAM, Hierarchical Clustering-DBSCAN, Categorical Clustering-ROCK

Module 4: (Association Rule Analysis)

Association Rules-Introduction, Methods to discover Association rules, Apriori(Level-wise algorithm), Partition Algorithm, Pincer Search Algorithm, Dynamic Itemset Counting Algorithm, FP-tree Growth Algorithm.

Module 5 (Advanced Data Mining Techniques)

Web Mining - Web Content Mining, Web Structure Mining- Page Rank, Clever, Web Usage Mining- Preprocessing, Data structures, Pattern Discovery, Pattern Analysis. Text Mining-Text Data Analysis and information Retrieval, Basic measures for Text retrieval, Text Retrieval methods, Text Indexing Techniques, Query Processing Techniques.

Text Books

1. Dunham M H, “Data Mining: Introductory and Advanced Topics”, Pearson Education, New Delhi, 2003.
2. Arun K Pujari, “Data Mining Techniques”, Universities Press Private Limited,2008.
3. Jaiwei Han and Micheline Kamber, “Data Mining Concepts and Techniques”, Elsevier, 2006

Reference Books

1. M Sudeep Elayidom, “Data Mining and Warehousing”, 1st Edition, 2015, Cengage Learning India Pvt. Ltd.
2. MehmedKantardzic, “Data Mining Concepts, Methods and Algorithms”, John Wiley and Sons, USA, 2003.
3. Pang-Ning Tan and Michael Steinbach, “Introduction to Data Mining”, Addison Wesley, 2006.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. (a) Explain the OLAP operations in a multidimensional model.
   (b) Compare the techniques used in ROLAP, MOLAP and HOLAP
2. Explain the various data mining issues with respect to mining methodology, user interaction and diversity of data types.
3. Suppose that a data warehouse consists of the three dimensions time, doctor, and patient, and the two measures count and charge, where charge is the fee that a doctor charges a patient for a visit.
   a) Draw star and snowflake schema diagrams for the data warehouse.
   b) Starting with the base cuboid [day; doctor; patient], what specific OLAP operations should be performed in order to list the total fee collected by each doctor in 2004?

Course Outcome 2 (CO2):

1. Use the methods below to normalize the following group of data:100, 200, 300, 400,550, 600, 680, 850, 1000
   (a) min-max normalization by setting min = 0 and max = 1
   (b) z-score normalization
   (c) Normalization by decimal scaling
   Comment on which method you would prefer to use for the given data, givingreasons as to why.
2. Identify a suitable dataset from any available resources and apply different preprocessing steps that you have learned. Observe and analyze the output obtained. (Assignment)

Course Outcome 3 (CO3):

1. Illustrate the working of ID3 algorithm with the following example

MOTOR	WHEEELS	DOORS	SIZE	TYPE	CLASS
NO	2	0	small	cycle	bicycle
NO	3	0	small	cycle	tricycle
YES	2	0	small	cycle	motorcycle
YES	4	2	small	automobile	Sports car
YES	4	3	medium	automobile	minivan
YES	4	4	medium	automobile	sedan
YES	4	4	large	automobile	sumo

2. Illustrate the working of K medoid algorithm for the given dataset. A1=(3,9), A2=(2,5), A3=(8,4), A4=(5,8), A5=(7,5), A6=(6,4), A7=(1,2), A8=(4,9).

3. Take a suitable dataset from available resources and apply all the classification and clustering algorithms that you have studied on original and preprocessed datasets. Analyze the performance variation in terms of different quality metrics. Give a detailed report based on the analysis. (Assignment)

Course Outcome 4 (CO4):

1. A database has five transactions. Let min sup = 60% and min con f = 80%.

TID	*items_bought*
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y }
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, O, K, I ,E}

   a) Find all frequent item sets using Apriori and FP-growth, respectively. Compare the efficiency of the two mining processes.
   b) List all of the strong association rules (with support s and confidence c) matching the following metarule, where X is a variable representing customers, and$item_i$denotes variables representing items (e.g., "A", "B", etc.)

$$\forall x \in transaction, buys(X, item_1) \wedge buys(X, item_2) \Rightarrow buys\ (X, item_3) \quad [s, c]$$

2. Identify and list some scenarios in which association rule mining can be used, and then use at least two appropriate association rule mining techniques in one of the two scenarios. (Assignment)

Course Outcome 5 (CO5):

1. Consider an e-mail database that stores a large number of electronic mail (e-mail) messages. It can be viewed as a semi structured database consisting mainly of text data. Discuss the following.
   a. How can such an e-mail database be structured so as to facilitate multidimensional search, such as by sender, by receiver, by subject, and by time?
   b. What can be mined from such an e-mail database?
   c. Suppose you have roughly classified a set of your previous e-mail messages as junk, unimportant, normal, or important. Describe how a data mining system may take this as the training set to automatically classify new e-mail messages or unclassified ones.
2. Precision and recall are two essential quality measures of an information retrieval system.
   (a) Explain why it is the usual practice to trade one measure for the other.
   (b) Explain why the F-score is a good measure for this purpose.

(c) Illustrate the methods that may effectively improve the F-score in an information retrieval system.

3. Explain HITS algorithm with an example.

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: CST466

Course Name: Data Mining

Max.Marks:100 **Duration: 3 Hours**

PART A

Answer All Questions. Each Question Carries 3 Marks

1. Differentiate between OLTP and OLAP.
2. Compare the techniques of ROLAP, MOLAP and HOLAP
3. Explain Concept hierarchy with an example.
4. Explain heuristic methods of attribute subset selection techniques.
5. Consider a two-class classification problem of predicting whether a photograph contains a man or a woman. Suppose we have a test dataset of 10 records with expected outcomes and a set of predictions from our classification algorithm.

	Expected	Predicted
1	man	woman
2	man	man
3	woman	woman
4	man	man
5	woman	man
6	woman	woman
7	woman	woman
8	man	man
9	man	woman
10	woman	woman

Calculate precision, recall of the data.

6. Given two objects represented by the tuples (22,1,42,10) and (20,0, 36,8). Compute the Euclideanand Manhattan distance between the two objects.

7. The pincer search algorithm is a bi-directional search, whereas the level wise algorithm is a unidirectional search. Express your opinion about the statement.

8. Define support, confidence and frequent set in association data mining context.

9. Distinguish between focused crawling and regular crawling.

10. Describe any two-text retrieval indexing techniques. **(10x3=30)**

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) Suppose a data warehouse consists of three measures: customer, account and branch and two measures count (number of customers in the branch) and balance. Draw the schema diagram using snowflake schema and star schema. **(7)**

(b) Explain three- tier data warehouse architecture with a neat diagram. **(7)**

OR

12 (a) Illustrate different OLAP operations in multidimensional data model **(7)**

(b) Describe different issues in data mining **(7)**

13 (a) Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70. **(8)**

(a) Use min-max normalization to transform the value 35 for age onto

the
range [0-1].

(b) Use z-score normalization to transform the value 35 for age, where the standard deviation of age is 12.94 years.

(c) Use normalization by decimal scaling to transform the value 35 for age.

(d) Use smoothing by bin means to smooth the above data, using a bin depth of 3. Illustrate your steps. Comment on the effect of this technique for the given data.

(b) With proper illustration, explain how PCA can be used for dimensionality reduction? Explain **(6)**

OR

14 (a) Suppose a group of 12 sales price records has been sorted as follows: 5, 10, 11, 13, 15, 35, 50, 55, 72, 92, 204, 215. Sketch examples of each of the following sampling techniques: SRSWOR, SRSWR, cluster sampling, stratified sampling. Use samples of size 5 and the strata "youth," "middle-aged," and "senior." **(8)**

(b) Partition the above data into three bins by each of the following methods: **(6)**
(i) equal-frequency (equi-depth) partitioning
(ii) equal-width partitioning

15 (a) Explain the concept of a cluster as used in ROCK. Illustrate with examples **(9)**

(b) Consider the following dataset for a binary classification problem. **(5)**

A	B	Class Label
T	F	+
T	T	+
T	T	+
T	F	-
T	T	+
F	F	-
F	F	-
F	F	-
T	T	-
T	F	-

Calculate the gain in Gini index when splitting on A and B respectively. Which attribute would the decision tree induction algorithm choose?

OR

16 (a) For a sunburn dataset given below, find the first splitting attribute for the decision tree by using the ID3 algorithm. **(10)**

Name	Hair	Height	Weight	Lotion	Class
Sarah	Blonde	Average	Light	No	Sunburn
Dana	Blonde	Tall	Average	Yes	None
Alex	Brown	Tall	Average	Yes	None
Annie	Blonde	Short	Average	No	Sunburn
Emily	Red	Average	Heavy	No	Sunburn
Pete	Brown	Tall	Heavy	No	None
John	Brown	Average	Heavy	No	None
Katie	Blonde	Short	Light	Yes	None

(b) Explain the working of SLIQ algorithm. **(4)**

17 (a) Illustrate the working of Pincer Search Algorithm with an example. **(7)**

(b) Describe the working of dynamic itemset counting technique? Specify when to move an itemset from dashed structures to solid structures? **(7)**

OR

18 (a) A database has six transactions. Let min_sup be 60% and min_conf be 80%. **(9)**

TID	*items_bought*
T1	I1, I2, I3
T2	I2, I3, I4
T3	I4, I5
T4	I1, I2, I4
T5	I1, I2, I3, I5
T6	I1, I2, I3, I4

Find frequent itemsets using FP Growth algorithm and generate strong association rules from a three item dataset.

(b) Write partitioning algorithm for finding large itemset and compare its efficiency with apriori algorithm **(5)**

19 (a) Describe web content mining techniques. **(7)**

(b) Write an algorithm to find maximal frequent forward sequences to mine log traversal patterns. Illustrate the working of this algorithm. **(7)**

OR

20 (a) Explain how web structure mining is different from web usage mining and web content mining? Write a CLEVER algorithm for web structure mining. **(7)**

(b) Describe different Text retrieval methods. Explain the relationship between text mining and information retrieval and information extraction. **(7)**

Teaching Plan

No	Contents	No. of lecture hours (36 Hrs)
Module 1(Introduction to Data Mining and Data Warehousing) (Text3) (6 hours)		
1.1	Data warehouse-Differences between Operational Database Systems and Data Warehouses, Multidimensional data model- Warehouse schema	1
1.2	OLAP Operations	1
1.3	DataWarehouse Architecture, Data Warehousing to Data Mining	1
1.4	Datamining Concepts and Applications, Knowledge Discovery in Database Vs Data mining	1
1.5	Architecture of typical data mining system,Data Mining Functionalities	1
1.6	Data Mining Functionalities, Data Mining Issues	1
Module 2(Data Preprocessing) (6 hours) (Text3)		
2.1	Data Preprocessing: Need of Data Preprocessing, Data Cleaning- Missing values, Noisy data.	1
2.2	Data integration	1
2.3	Data transformation	1
2.4	Data Reduction-Data cube aggregation, Attribute subset selection	1
2.5	Data Reduction-Dimensionality reduction	1

2.6	Numerosity reduction, Discretization and concept hierarchy generation	1
Module 3(Advanced classification and Cluster analysis)(9 hours)(Text2,Text3)		
3.1	Classification- Introduction, Decision tree construction principle, Splitting indices-Information Gain, Gini index	1
3.2	Decision Tree- ID3	1
3.3	Decision Tree- ID3	1
3.4	Decision tree construction with presorting- SLIQ	1
3.5	Accuracy and error measures, evaluation	1
3.6	Introduction to clustering, Clustering Paradigms	1
3.7	Partitioning Algorithm- PAM	1
3.8	Hierarchical Clustering-DBSCAN	1
3.9	Categorical Clustering-ROCK	1
Module 4(Association Rule Analysis) (8 hours) (Text2,Text3,Text1)		
4.1	Association Rules: Introduction, Methods to discover association rules	1
4.2	A priori algorithm (Level-wise algorithm)	1
4.3	A priori algorithm (Level-wise algorithm)	1
4.4	Partition Algorithm	1
4.5	Pincer Search Algorithm	1
4.6	Pincer Search Algorithm	1
4.7	Dynamic Itemset Counting Algorithm	1
4.8	FP-tree Growth Algorithm	1
Module 5(Advanced Data Mining Techniques) (7 hours) (Text1, Text3		
5.1	Web Mining - Web Content Mining	1
5.2	Web Structure Mining- Page Rank	1
5.3	Web Structure Mining –Clever algorithm	1
5.4	Web Usage Mining- Preprocessing, Data structures	1

5.5	Web Usage Mining -Pattern Discovery, Pattern Analysis	1
5.6	Text Mining-Text Data Analysis and information Retrieval, Basic measures for Text retrieval	1
5.7	Text Retrieval methods, Text Indexing Techniques Query Processing Techniques	1

SEMESTER VIII

PROGRAM ELECTIVE V

CST418	HIGH PERFORMANCE COMPUTING	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PEC	2	1	0	3	2019

Preamble: This course helps the learners to understand the different architectural features of high-end processors. This course discusses the Basics of high-end processors Architecture, Instruction-Level Parallelism, Data-Level Parallelism, Thread Level Parallelism, and GPU Architectures. This course enables the students to provide solutions to real-world problems making use of the capabilities of HPC systems.

Prerequisite: Basic knowledge in Computer System architecture, Microprocessors, Operating systems, and System software.

Course Outcomes: After the completion of the course the student will be able to

CO1	Describe different types of modern processing environments and parallel computing hardware (**Cognitive Knowledge Level: Understand**)
CO2	Summarize the concepts of Instruction Level Parallelism (**Cognitive Knowledge Level: Understand**)
CO3	Appreciate the idea of Data Level Parallelism (**Cognitive Knowledge Level: Apply**)
CO4	Demonstrate the concept of Thread Level Parallelism (**Cognitive Knowledge Level: Apply**)
CO5	Describe the basics of GPU architecture. (**Cognitive Knowledge Level: Understand**)

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓										✓
CO2	✓	✓										✓
CO3	✓	✓	✓									✓
CO4	✓	✓	✓									✓
C05	✓	✓										✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and teamwork
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Life long learning

Assessment Pattern

Bloom's Category	Continuous Assessment Tests		End Semester Examination Marks (%)
	Test 1 (%)	**Test 2 (%)**	
Remember	**20**	**20**	**20**
Understand	**50**	**50**	**50**
Apply	**30**	**30**	**30**
Analyze			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks.	ESE Duration
150	**50**	**100**	**3**

Continuous Internal Evaluation Pattern:

Attendance	**10 marks**
Continuous Assessment Tests (Average of Internal Tests 1 & 2)	**25 marks**
Continuous Assessment Assignment	**15 marks**

Internal Examination Pattern

Each of the two internal examinations have to be conducted out of 50 marks. The first series test shall be preferably conducted after completing the first half of the syllabus and the second series test shall be preferably conducted after completing the remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly completed module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly completed module), each with 7 marks. Out of the 7 questions, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 full questions from each module of which students should answer anyone. Each question can have a maximum of 2 sub-divisions and carries 14 marks.

Syllabus

Module-1 (Basics of Architecture)

Classes of Computers - Classes of Parallelism and Parallel Architectures – Defining Computer Architecture – Dependability – Quantitative Principles of Computer Design – Basics of Memory Hierarchies – Virtual Memory and Virtual Machines – Pipelining

Module-2 (Instruction-Level Parallelism)

Instruction-Level Parallelism: Concepts and Challenges – Basic Compiler Techniques for Exposing ILP – Reducing Branch Costs With Advanced Branch Prediction – Hardware-Based Speculation – Multithreading: Exploiting Thread-Level Parallelism to Improve Uniprocessor Throughput

Module-3 (Data-Level Parallelism)

Vector Architecture – SIMD Instruction Set Extensions for Multimedia – Graphics Processing Units – Detecting and Enhancing Loop-Level Parallelism

Module-4 (Thread Level Parallelism)

Multiprocessor Architecture: Issues and Approach – Centralized Shared-Memory Architectures – Performance of Symmetric Shared-Memory Multiprocessors– Distributed Shared-Memory and Directory-Based Coherence – Synchronization: The Basics – Introduction to Memory Consistency

Module-5 (GPU Architectures)

The CPU-GPU system as an accelerated computational platform – The GPU and the thread engine – Characteristics of GPU memory spaces – The PCI bus: CPU to GPU data transfer overhead – Multi-GPU platforms – Potential benefits of GPU – accelerated platforms

Text Books

1. John L. Hennessy, David A. Patterson Computer Architecture, Sixth Edition A Quantitative Approach, Morgan Kaufman, Fifth Edition, 2012.
2. Robert Robey, Yuliana Zamora, Parallel and High-Performance Computing, Manning Publications, First Edition, 2021.

Reference Books

1. Thomas Sterling, Matthew Anderson, and MaciejBrodowicz, High-Performance Computing – Modern Systems and Practices, First Edition, 2017.
2. Charles Severance, Kevin Dowd,High-Performance Computing, O'Reilly Media, Second Edition, 1998.
3. Kai Hwang, Faye Alaye Briggs, Computer Architecture and Parallel Processing, McGraw-Hill, 1984.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Differentiate different classes of computer-based on features like microprocessor cost, system cost, and system design issues.
2. Explain the different methods by which computer hardware exploits application-level parallelism.
3. Explain in detail the instruction set architecture
4. Describe the encoding scheme specified as part of ISA

Course Outcome 2 (CO2):

1. Differentiate data, name, and control dependencies with suitable examples.
2. Explain loop unrolling with suitable coding demonstration
3. Explain in detail about Tournament Predictors.
4. Describe the unique features of very long instruction word processors.

Course Outcome 3 (CO3):

1. What are the three things conveyed through a data dependence? Explain the Data Dependencies of the following code:

```
Loop:  fld     f0,0(x1)    //f0=array element
       fadd.d  f4,f0,f2    //add scalar in f2
       fsd     f4,0(x1)    //store result
       addi    x1,x1,-8    //decrement pointer 8 bytes
       bne     x1,x2,Loop  //branch x1≠x2
```

2. Assume a single-issue pipeline. Unroll the loop as many times as necessary to schedule it without any stalls, collapsing the loop overhead instructions. How many times must the loop be unrolled? Show the instruction schedule. What is the execution time per element of the result?
3. Explain the SIMD Instruction Set Extensions for Multimedia.

Course Outcome 4 (CO4):

1. With the help of a neat diagram illustrate a single-chip multicore with a distributed cache.
2. Demonstrate the Implementation of cache coherence in a distributed-memory multiprocessor by adding a directory to each node with a suitable diagram.
3. Consider the following code segments running on two processors P1 and P2. Assume A, and B, are initially 0. Explain how an optimizing compiler might make it impossible for B to be ever set to 2 in a sequentially consistent execution model.

P1:	P2:
A=1;	B=1;
A=2;	While (A <> 1);
While (B == 0);	B= 2;

Course Outcome 5 (CO5):

1. Explain the benefits of potential GPU.
2. Illustrate GPU system as an accelerated computational platform.
3. Discuss CPU to GPU data transfer overhead.

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: CST418

Course Name: High Performance Computing

Max. Marks : 100 **Duration: 3 Hours**

PART A

Answer All Questions. Each Question Carries 3 Marks

1. Differentiate between Data level parallelism and Task level parallelism
2. Explain the principle of locality
3. Define Instruction Level Parallelism with an example.
4. Devise the importance of loop unrolling with an example.
5. What is the equation of CPI (cycles per instruction) for a pipelined processor? How can we set the ideal pipeline CPI?
6. Explain the two types of name dependencies between an instruction i that precedes instruction j in program order.
7. Differentiate between module reliability and module availability measures with suitable examples.
8. Why SMP architectures are called UMA multiprocessors and DSM multiprocessors as NUMA processors.

9. Explain the need for GPU.

10. List the characteristics of GPU memory spaces.

3x10=30

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) Describe the quantitative principle of computer design with Amdahl's law. (8)

(b) Discuss in detail the importance of considering processor performance for the design of an efficient computer system. (6)

OR

12. (a) Illustrate how processes are protected with the help of virtual memory. (7)

(b) Discuss the role played by virtual machines in providing protection for processes. (7)

13. (a) Explain in detail data dependence and hazards. **(8)**

(b) With neat sketches explain how data-level parallelism is achieved in vector, and SIMD architectures. **(6)**

OR

14. (a) Describe the unique features of very long instruction word processors. **(8)**

(b) Consider a three-way superscalar machine renaming these three instructions concurrently: **(6)**

```
addi x1, x1, x1
addi x1, x1, x1
addi x1, x1, x1
```

If the value of x1 starts at 5, then what will be its value when after this sequence is executed?

15. (a) The following loop has multiple types of dependences. Find all the true dependences, output dependencies, and anti-dependencies, and eliminate the output dependencies and anti-dependencies by renaming. **(8)**

```
for (i=0; i<100; i=i+1) {
    Y[i] = X[i] / c;  /* S1 */
    X[i] = X[i] + c;  /* S2 */
    Z[i] = Y[i] + c;  /* S3 */
    Y[i] = c - Y[i];  /* S4 */
}
```

(b) Describe the limitations of Symmetric Shared-Memory Multiprocessors and Snooping Protocols **(6)**

OR

16. (a) Demonstrate the different types of hardware approaches required for the working of multithreading. **(8)**

(b) Consider the following loop: **(6)**

```
for (i=0;i<100;i++) {
 A[i] = A[i] + B[i]; /* S1 */
 B[i+1] = C[i] + D[i]; /* S2 */
}
```

Are there exist dependencies between S1 and S2? Determine whether the above loop is parallel? If not, show how to make it parallel.

17. (a) Consider an 8-processor multicore where each processor has its own L1 and L2 caches. Here snooping is performed on a shared bus among the L2 caches. Assume that the average L2 request is 15 cycles for a coherence miss or other miss and a clock rate of 3.0 GHz, a CPI of 0.7, and a load/store frequency of 40%. If the goal set is that no more than 50% of the L2 bandwidth is consumed by coherence traffic, then what is the maximum coherence miss rate per processor? **(8)**

(b) Explain the basic structure of a centralized shared-memory multiprocessor **(6)**

based on a multicore chip.

OR

18. (a) Suppose an application running on a 100-processor multiprocessor use 1, 50, or 100 processors. If for 95% of the time all 100 processors are used, illustrate how the remaining 5% of the execution time employs 50 processors for a speedup of 80? **(6)**

(b) With a neat diagram, demonstrate invalidate cache coherence protocol for a private write-back cache, showing the states and state transitions for each block in the cache. **(8)**

19. (a) Explain the multi-GPU platform. **(8)**

(b) Explain some of the benefits of GPU. **(6)**

OR

20. (a) Discuss in detail the characteristics of GPU memory spaces. **(8)**

(b) Explain about GPU thread engine. **(6)**

Teaching Plan

No	Contents	No. of Lecture Hours (36 hrs)
Module 1 - Basics of Architecture (7 hours)		
1.1	Classes of Computers	1 hour
1.2	Classes of Parallelism and Parallel Architectures	1 hour
1.3	Dependability	1 hour
1.4	Quantitative Principles of Computer Design.	1 hour

1.5	Basics of Memory Hierarchies	1 hour
1.6	Virtual Memory and Virtual Machines	1 hour
1.7	Pipelining	1 hour
Module -2 (Introduction to Syntax Analysis) (7 hours)		
2.1	Instruction-Level Parallelism: Concepts and Challenges	1 hour
2.2	Basic Compiler Techniques for Exposing ILP	1 hour
2.3	Reducing Branch Costs With Advanced Branch Prediction	1 hour
2.4	Hardware-Based Speculation	1 hour
2.5	Multithreading	1 hour
2.6	Exploiting Thread-Level Parallelism to Improve Uniprocessor Throughput – Lecture 1.	1 hour
2.7	Exploiting Thread-Level Parallelism to Improve Uniprocessor Throughput – Lecture 2.	1 hour
Module- 3 - Data-Level Parallelism (7 hours)		
3.1	Vector Architecture -Lecture 1	1 hour
3.2	Vector Architecture -Lecture 2	1 hour
3.3	SIMD Instruction Set Extensions for Multimedia – Lecture 1	1 hour
3.4	SIMD Instruction Set Extensions for Multimedia – Lecture 2	1 hour
3.5	Graphics Processing Units	1 hour
3.6	Detecting and Enhancing Loop-Level Parallelism – Lecture 1	1 hour
3.7	Detecting and Enhancing Loop-Level Parallelism – Lecture 2	1 hour
Module 4– Thread Level Parallelism (8 hours)		
4.1	Multiprocessor Architecture: Issues and Approach	1 hour
4.2	Centralized Shared-Memory Architectures – Lecture 1	1hour
4.3	Centralized Shared-Memory Architectures – Lecture 2	1hour
4.4	Performance of Symmetric Shared-Memory Multiprocessors	1hour
4.5	Distributed Shared-Memory	1hour
4.6	Directory-Based Coherence	1hour
4.7	Synchronization	1hour

4.8	Introduction to Memory Consistency	1hour
Module 5 – GPU Architectures (7 hours)		
5.1	The CPU-GPU system as an accelerated computational platform	1 hour
5.2	The GPU and the thread engine – Lecture 1	1 hour
5.3	The GPU and the thread engine – Lecture 2	1 hour
5.4	Characteristics of GPU memory spaces	1hour
5.5	PCI bus: CPU to GPU data transfer overhead	1hour
5.6	Multi-GPU platforms	1hour
5.7	Potential benefits of GPU-accelerated platforms	1hour

CST428	BLOCKCHAIN TECHNOLOGIES	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PEC	2	1	0	3	2019

Preamble: The purpose of this course is to create awareness and understanding among students on the foundation of blockchain technology. The course introduces the cryptographic principles behind blockchain and helps the students understand concepts like consensus, crypto-currency, smart contracts, use cases etc. The course enables students to develop simple decentralized applications using blockchain networks such as Ethereum.

Prerequisite: Basic knowledge in data structures and operating systems.

Course Outcomes: After the completion of the course the student will be able to

CO1	Illustrate the cryptographic building blocks of blockchain technology. **(Cognitive Knowledge Level: Understand)**
CO2	Explain the fundamental concepts of blockchain technology. **(Cognitive Knowledge Level: Understand)**
CO3	Summarize the classification of consensus algorithms. (**Cognitive Knowledge Level: Understand)**
CO4	Explain the concepts of first decentralized cryptocurrency bitcoin. (**Cognitive Knowledge Level: Understand)**
CO5	Explain the use of smart contracts and its use cases. **(Cognitive Knowledge Level: Understand)**
CO6	Develop simple applications using Solidity language on Ethereum platform. **(Cognitive Knowledge Level: Apply)**

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	⊘	⊘										⊘
CO2	⊘	⊘										⊘

CO3	✓	✓										✓
CO4	✓	✓										✓
CO5	✓	✓										✓
CO6	✓	✓	✓	✓	✓							✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and team work
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Life long learning

Assessment Pattern

Bloom's Category	**Continuous Assessment Tests**		**End Semester Examination Marks (%)**
	Test 1 (%)	**Test 2 (%)**	
Remember	**30**	**30**	**30**
Understand	**50**	**50**	**50**
Apply	**20**	**20**	**20**
Analyze			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	50	100	3

Continuous Internal Evaluation Pattern:

Attendance **10 marks**

Continuous Assessment Tests(Average of Internal Tests 1 & 2) **25 marks**

Continuous Assessment Assignment **15 marks**

Internal Examination Pattern

Each of the two internal examinations has to be conducted out of 50 marks. First series test shall be preferably conducted after completing the first half of the syllabus and the second series test shall be preferably conducted after completing remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly completed module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly completed module), each with 7 marks. Out of the 7 questions, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 full questions from each module of which student should answer any one. Each question can have maximum 2 sub-divisions and carries 14 marks.

Syllabus

Module – 1 (Fundamentals of Cryptography)

Introduction to Cryptography, Symmetric cryptography – AES. Asymmetric cryptography – RSA. Elliptic curve cryptography, Digital signatures – RSA digital signature algorithms. Secure Hash Algorithms – SHA-256. Applications of cryptographic hash functions – Merkle trees, Distributed hash tables.

Module – 2 (Fundamentals of Blockchain Technology)

Blockchain – Definition, architecture, elements of blockchain, benefits and limitations, types of blockchain. Consensus – definition, types, consensus in blockchain.

Decentralization – Decentralization using blockchain, Methods of decentralization, Routes to decentralization, Blockchain and full ecosystem decentralization.

Module - 3 (Consensus Algorithms and Bitcoin)

Consensus Algorithms, Crash fault-tolerance (CFT) algorithms – Paxos, Raft. Byzantine fault-tolerance (BFT) algorithms – Practical Byzantine Fault Tolerance (PBFT), Proof of work (PoW), Proof of stake (PoS), Types of PoS.
Bitcoin – Definition, Cryptographic keys – Private keys, public keys, addresses. Transactions – Lifecycle, coinbase transactions, transaction validation. Blockchain – The genesis block.
Mining – Tasks of miners, mining algorithm, hash rate. Wallets – Types of wallets.

Module - 4 (Smart Contracts and Use cases)

Smart Contracts – Definition, Smart contract templates, Oracles, Types of oracles, Deploying smart contracts. Decentralization terminology – Decentralized applications, Decentralized Autonomous Organizations.
Use cases of Blockchain technology – Government, Health care, Finance, Supply chain management.
Blockchain and allied technologies – Blockchain and Cloud Computing, Blockchain and Artificial Intelligence.

Module - 5 (Ethereum and Solidity)

Ethereum – The Ethereum network. Components of the Ethereum ecosystem – Keys and addresses, Accounts, Transactions and messages. The Ethereum Virtual Machine, Blocks and blockchain.
The Solidity language – The layout of a Solidity source code, Structure of a smart contract, variables, data types, control structures, events, inheritance, libraries, functions, error handling.
Smart contracts Case study: Voting, Auction.

Text Book

1. Imran Bashir, Mastering Blockchain: A deep dive into distributed ledgers, consensus protocols, smart contracts, DApps, cryptocurrencies, Ethereum, and more, Packt Publishing, Third edition, 2020.

References

2. Ritesh Modi, Solidity Programming Essentials: A beginner's guide to build smart contracts for Ethereum and blockchain, Packt Publishing, First edition, 2018.
3. Kumar Saurabh, Ashutosh Saxena, Blockchain Technology: Concepts and Applications, First Edition, Wiley Publications, First edition, 2020.
4. Chandramouli Subramanian, Asha A George, et al, Blockchain Technology, Universities Press (India) Pvt. Ltd, First edition, August 2020.

5. Lorne Lantz, Daniel Cawrey, Mastering Blockchain: Unlocking the Power of Cryptocurrencies, Smart Contracts, and Decentralized Applications, O'Reilly Media, First edition, 2020.
6. Andreas M. Antonopoulos, Gavin Wood, Mastering Ethereum: Building Smart Contracts and DApps, O'Reilly Media, First edition, 2018.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Distinguish between Symmetric cryptography and asymmetric cryptography.
2. Explain the working of AES algorithm.

Course Outcome 2 (CO2):

1. Categorize consensus mechanism used in blockchain.
2. Define Blockchain. Explain how decentralization of computing or processing power is achieved by a blockchain.

Course Outcome 3 (CO3):

1. Explain how Proof of Stake can achieve consensus among peers.
2. Explain the working of Raft protocol.

Course Outcome 4 (CO4):

1. Describe the use of genesis block.
2. Explain the mining algorithm used in bitcoin.

Course Outcome 5 (CO5):

1. Illustrate how blockchain technology can be used in supply chain management.
2. What are oracles in a blockchain ecosystem? Explain the generic data flow from a smart contract to an oracle.

Course Outcome 6 (CO6):

1. Develop a smart contract for voting process. In this application, delegated voting is allowed and the counting is automatic and completely transparent at the same time.
2. Develop a smart contract for auction process. The contract should be a blind auction where it is not possible to see the actual bid until the bidding period ends.

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 2**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: CST428

Course Name: BLOCK CHAIN TECHNOLOGIES

Max. Marks : 100 **Duration: 3 Hours**

PART A

Answer All Questions. Each Question Carries 3 Marks

1. Discuss the role of secure hash functions in blockchain.
2. List out the properties of digital signatures.
3. Illustrate the blockchain based decentralized system.
4. Explain how Proof of Stake can achieve consensus among peers.
5. If your blockchain network has 5 Byzantine nodes, what is the minimum number of nodes that are required to ensure Byzantine fault tolerance using PBFT protocol?
6. How are transactions verified in a Bitcoin network?
7. Explain how smart contracts can be used for enforcing agreements between parties in the form of business logic.
8. Explain the concept of blockchain-based digital identity cards.
9. Explain error handling in Solidity language.

10. With the help of a figure show the relationship between the transaction, transaction trie, and block header in Ethereum. **(10x3=30)**

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) Explain the design of SHA-256 and its compression function using a diagram. **(9)**

    (b) Explain how hash functions are used to build Merkle trees in blockchain. **(5)**

OR

12. (a) Explain public and private keys. Perform encryption and decryption using RSA for p=3, q=11, e= 7 and M=5. **(7)**

    (b) Explain elliptic curve digital signature algorithm. **(7)**

13. (a) Illustrate and explain how blockchain works using a neat diagram. **(7)**

    (b) Explain the benefits, features and limitations of blockchain. **(7)**

OR

14. (a) Explain consensus mechanisms used in blockchain. List out any six consensus algorithms used in the context of blockchain. **(7)**

    (b) Define blockchain. Explain how decentralization of computing or processing power is achieved by a blockchain. **(7)**

15. (a) Explain and illustrate how Paxos protocol can be used to achieve consensus. **(7)**

    (b) Show how Practical Byzantine Fault Tolerance can achieve consensus in the presence of Byzantine faults. **(7)**

OR

16. (a) Describe the various fields that make up a transaction in Bitcoin. **(7)**

    (b) What is the role of a Bitcoin miner? Explain the mining algorithm used in **(7)**

Bitcoin with the help of a flowchart.

17. (a) Illustrate how blockchain technology can be implemented in finance sector. **(7)**

(b) Discuss oracles in a blockchain ecosystem. Explain the generic data flow from a smart contract to an oracle. **(7)**

OR

18. (a) Explain the design process of decentralized applications with diagrams. **(7)**

(b) Explain the use of blockchain technology in supply chain management. **(7)**

19. (a) Using Solidity language, create a simple bank contract that allows a user to deposit, withdraw and view balance. **(7)**

(b) Define block difficulty. Explain how block difficulty is adjusted in Ethereum blockchain network. **(7)**

OR

20. (a) Using Solidity language, create a simple voting smart contract where a chairperson will give the right to vote to each address individually. **(7)**

(b) Explain the concept of Gas in Ethereum. Explain how transaction cost can be calculated in an Ethereum blockchain network. **(7)**

Teaching Plan

No	Contents	No. of Lecture Hours (35 hours)
Module-1 (Fundamentals of Cryptography) (7 hours)		
1.1	Introduction to cryptography	1 hour
1.2	Symmetric cryptography, AES	1 hour
1.3	Asymmetric cryptography, RSA	1 hour
1.4	Elliptic curve cryptography	1 hour
1.5	Digital signatures – RSA digital signature algorithm	1 hour
1.6	Secure Hash Algorithms – SHA-256	1 hour
1.7	Applications of cryptographic hash functions – Merkle trees, Distributed hash tables	1 hour
Module-2 (Fundamentals of Blockchain Technology) (6 hours)		
2.1	Blockchain – definition and architecture	1 hour
2.2	Elements of blockchain.	1 hour
2.3	Blockchain – benefits and limitations, types.	1 hour
2.4	Consensus – definition, types, consensus in blockchain	1 hour
2.5	Decentralization using blockchain, Methods of decentralization	1 hour
2.6	Routes to decentralization, Blockchain and full ecosystem decentralization	1 hour
Module-3 (Consensus Algorithms and Bitcoin) (7 hours)		
3.1	Consensus Algorithms – Crash fault-tolerance (CFT) algorithms – Paxos, Raft (working is expected).	1 hour
3.2	Byzantine fault-tolerance (BFT) algorithms – Practical Byzantine Fault Tolerance (PBFT) (working is expected).	1 hour
3.3	Proof of work (PoW), Proof of stake (PoS), Types of PoS	1 hour
3.4	Bitcoin – Definition, Cryptographic keys – Private keys, public keys, addresses.	1 hour
3.5	Transactions – Lifecycle, coinbase transactions, transaction validation	1 hour

3.6	Blockchain – The genesis block. Mining – Tasks of miners	1 hour
3.7	Mining – mining algorithm, hash rate. Wallets – Types of wallets.	1 hour
Module-4 (Smart Contracts and Use cases) (6 hours)		
4.1	Smart Contracts – Definition, Smart contract templates	1 hour
4.2	Oracles, Types of oracles, Deploying smart contracts.	1 hour
4.3	Decentralization terminology –Decentralized applications, Decentralized Autonomous Organizations.	1 hour
4.4	Use cases of Blockchain technology – Government, Health care.	1 hour
4.5	Use cases of Blockchain technology – Finance, Supply chain management.	1 hour
4.6	Blockchain and Allied Technologies – Blockchain and Cloud Computing, Blockchain and Artificial Intelligence.	1 hour
Module-5 (Ethereum and Solidity) (9 hours)		
5.1	Ethereum - The Ethereum network, Components of the Ethereum ecosystem – Keys and addresses, Accounts	1 hour
5.2	Components of the Ethereum ecosystem – Transactions and messages	1 hour
5.3	The Ethereum Virtual Machine	1 hour
5.4	Ethereum Blocks and blockchain	1 hour
5.5	The Solidity language – The layout of a Solidity source code, Structure of a smart contract, variables, data types	1 hour
5.6	The Solidity language – control structures, events, inheritance, libraries	1 hour
5.7	The Solidity language – functions, error handling.	1 hour
5.8	Smart contracts Case study: Voting.	1 hour
5.9	Smart contracts Case study: Auction.	1 hour

CST438	IMAGE PROCESSING TECHNIQUE	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PEC	2	1	0	3	2019

Preamble: This course helps the learners understand the core concepts and applications of Digital Image Processing. It covers Digital Image Fundamentals, Image Transforms, Image Enhancement in Spatial and Frequency Domain, Image Restoration & Image Segmentation and Morphological Operations & Representation and Description. The learners will be able to develop new algorithms, tools, and application software for real-world applications involving image processing.

Prerequisite: A basic knowledge of Computer Graphics and Image representation

Course Outcomes: After the completion of the course, the student will be able to

CO1	Explain the concepts of image formation and the basis of digital image processing. **(Cognitive Knowledge Level: Understand)**
CO2	Demonstrate the role of image transforms in representing, highlighting, and modifying image features. **(Cognitive Knowledge Level: Apply)**
CO3	Solve image enhancement problems using spatial and frequency domain techniques. **(Cognitive Knowledge Level: Apply)**
CO4	Make use of the concept of image restoration and image segmentation techniques in real-world problems. **(Cognitive Knowledge Level: Apply)**
CO5	Interpret morphological operations, image representation, and description techniques. **(Cognitive Knowledge Level: Understand)**

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓										✓
CO2	✓	✓			✓							✓
CO3	✓	✓	✓									✓
CO4	✓	✓	✓	✓	✓	✓						✓
CO5	✓	✓										✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and teamwork
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Life long learning

Assessment Pattern

Bloom's Category	Continuous Assessment Tests		End Semester Examination Marks (%)
	Test 1 (%)	Test 2 (%)	
Remember	**30**	**30**	**30**
Understand	**40**	**40**	**40**
Apply	**30**	**30**	**30**
Analyze			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	**50**	**100**	**3**

Continuous Internal Evaluation Pattern:

Attendance	**10 marks**
Continuous Assessment Tests (Average of Internal Tests 1 & 2)	**25 marks**
Continuous Assessment Assignment	**15 marks**

Internal Examination Pattern

Each of the two internal examinations has to be conducted out of 50 marks. The first series test shall be preferably conducted after completing the first half of the syllabus and the second series test shall be preferably conducted after completing the remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly completed module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly completed module), each with 7 marks. Out of the 7 questions, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 full questions from each module of which students should answer anyone. Each question can have a maximum 2 subdivisions and carries 14 marks.

Syllabus

Module – 1 (Digital Image Fundamentals)

Elements of Visual Perception, A Simple Image Formation Model. Spatial and Intensity Resolution. Image Interpolation. Classification of Digital Images. Image Types. Image Storage Mechanisms. Arithmetic and Logical Operations. Geometric Spatial Transformations and Image Registration. Image File Formats. Colour Fundamentals and Colour Models.

Module - 2 (Image Transforms)

Basic concept of spatial domain and frequency domain, Unitary transform, Discrete Fourier Transform- 2D DFT, 4 order DFT Transform coefficients, Forward and inverse transform, Discrete Cosine Transform- 2D DCT, 4 order DCT Transform Coefficients(No derivation needed), Forward and Inverse DCT, Hadamard Transform.

Module - 3 (Image Enhancement in Spatial and Frequency Domain)

Point operations- Clipping and Thresholding, Digital Negative, Intensity Level Slicing, Bit Extraction, Range Compression. Spatial Operations- Fundamentals of spatial convolution and

correlation, Spatial averaging and spatial Low pass filtering, Directional Smoothing, Median Filtering, Unsharp masking and Crispening.

Basics of Filtering in Frequency Domain, Filters, Smoothing Frequency Domain Filters- Sharpening Frequency Domain Filters

Module - 4 (Image Restoration & Image Segmentation)

Image degradation model, Noise models, Mean Filters, Order Statistic filter, Adaptive filters.
Edge Detection, gradient operators, Laplace operators and zero crossings. Thresholding, Basic Global Thresholding, Optimum global thresholding using Otsu method, Multiple thresholds, Variable thresholding, Multivariable thresholding. Region-Based Approach to Segmentation.

Module - 5 (Morphological Operations & Representation and Description)

Structuring Element, Dilation and Erosion, Opening and Closing, Hit or Miss Transformation.

Boundary Following. Chain Codes. Polygonal Approximation. Boundary Descriptors. Regional Descriptors. Relational Descriptors.

Text Books

1. Rafael C. Gonzalez, Richard E. Woods, Digital Image Processing (English) 3rd Edition, Pearson India, 2013
2. A K. Jain, Fundamentals of digital image processing, Prentice Hall of India, 1989.

Reference Books

1. Al Bovik, The Essential Guide to Image Processing, Academic Press, 2009.
2. Milan Sonka, Vaclav Hlavac and Roger Boyle, Image Processing, Analysis, and Machine Vision, Thomson Learning, 2008.
3. S Jayaraman, S Esakkirajan and T Veerakumar, Digital Image Processing, McGraw Hill Education, 2009.

Course Level Assessment Questions

Course Outcome1 (CO1) :

1. Find the number of bits required to store a 256 X 256 image with 32 gray levels.
2. Explain the reasons for blocking artifacts and false contours in an image.

Course Outcome 2 (CO2) :

1. Compare different image transforms based on their roles, properties and applications.
2. Compute the inverse 2D DFT of the transform coefficients F(k,l) given below.

$$F(k, l) = \begin{pmatrix} 64 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

3. Use Discrete Fourier transform to construct 2D DFT for a 4x4 image given below. Assume that indices start from (0,0)

6	6	6	6
6	6	6	6
6	6	6	6
6	6	6	6

Course Outcome 3 (CO3) :

1. Perform intensity level slicing on the 3 BPP (Bit Per Pixel) image. Let r1=3 and r2=5. Draw the modified image with/without background transformations.

$$\begin{bmatrix} 2 & 1 & 2 & 2 & 1 \\ 2 & 3 & 4 & 5 & 2 \\ 6 & 2 & 7 & 6 & 0 \\ 2 & 6 & 6 & 5 & 1 \\ 0 & 3 & 2 & 2 & 1 \end{bmatrix}$$

2. Let y(m) = {2,3,8,4,2}. Obtain the median filter output for the window W = [-1,0,1,2] and show how salt and pepper noise is reduced.
3. Consider a 3*3 spatial mask that averages the four closest neighbors of a point(x,y), but excludes the point itself from the average.

   (a) Find the equivalent filter H(u,v) in the frequency domain.

   (b) Show that H(u,v) is a lowpass filter (ASSIGNMENT)

Course Outcome 4 (CO4) :

1. Compare Region and Edge-based techniques in segmentation.

2. Consider a noisy image that is restored using arithmetic mean filter of size 3x3 and using the geometric mean filter of the same size. Which image will be less blurred and why?
3. Suppose that you want to help a radiologist to extract the tumor portion from an MRI image for volumetric analysis. This volumetric analysis determines the effect of treatment on the patient, which can be judged from the extracted size and shape of the abnormal portion. Manual tracing of the tumor regions is very difficult since the tumor portion on the MRI image is inhomogeneous, with complex shapes and ambiguous boundaries. Suggest a sequence of steps that you may use to automate this process as an image processing student. (ASSIGNMENT)

Course Outcome 5 (CO5) :

4. Explain the significance of structuring elements in morphological operations with example.
5. Explain how chain codes are used to represent boundaries of a region with examples.

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: CST438

Course Name: IMAGE PROCESSING TECHNIQUE

Max. Marks : 100 **Duration: 3 Hours**

PART A

Answer All Questions. Each Question Carries 3 Marks

1. Give an image representation model and describe how the representation changes in different types of images.
2. Describe any three types of color models.
3. Obtain the HADAMARD basis matrix for N=8.
4. Prove that DFT is a unitary transform.
5. Sketch perspective plot of a 2-D ideal low pass filter transfer function and filter cross-section. List its usefulness in Image enhancement.
6. Explain the significance of directional smoothing technique.
7. Specify the significance of the Zero crossing detector.
8. Describe region growing technique for image segmentation.
9. Define 'Structuring Element' used in morphological operations. Give samples for Structuring Elements.
10. Explain image boundary representation using polygonal approximation.

(10x3=30)

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) Explain a Simple Image Formation Model with the help of a neat diagram. **(7)**

(b) Explain the relationship between image size, spatial resolution, and image quality. Compare gray level and intensity resolution. **(7)**

OR

12. (a) Describe arithmetic, logical and geometrical operations on Image. **(7)**

(b) Explain the significance of image interpolation and describe its various types. **(7)**

13. (a) State the advantages of Discrete Cosine Transform over Discrete Fourier Transform. **(4)**

(b) You are given a 4 X 4 image patch Compute 2D DCT for the image patch. Reconstruct the original image patch by neglecting the last four coefficients in 2D DCT. Comment on the observed result. **(10)**

$$\begin{bmatrix} 12 & 4 & 2 & 6 \\ 5 & 10 & 12 & 24 \\ 6 & 8 & 10 & 12 \\ 14 & 12 & 8 & 10 \end{bmatrix}$$

OR

14. (a) Discuss the concept of sequency in Hadamard transform. **(4)**

(b) Find the 2D forward DFT of the image segment **(10)**

1 1 1 1

1 1 1 1

1 1 1 1

Prove the unitary property of the given image segment.

15. (a) Explain the output and application of the following point processing techniques **(9)**

(i)Range Compression (ii) Bit Extraction (iii) Thresholding

(b) State and explain the features of median filtering. Compute the output of the median filtering for Y(m)={2,4,8,3,2} , w={-1,0,1,2} where Y(m) is an array and w is a window. **(5)**

OR

16. (a) Describe the role of Unsharp masking with its applications **(4)**

(b) Explain and compare the basic frequency domain filters for image sharpening **(10)**

17. (a) A 4×4 image is given by **(8)**

2 4 8 7
12 6 9 8
13 7 4 3
8 12 4 9

Filter the above image using

(a) MIN filter (b) MAX filter using the filter mask

0 1 0
1 1 1
0 1 0

(Assume replicate padding of the input image)

(b) Explain any two types of thresholding techniques. Describe the threshold detection algorithm using Otsu's method. **(6)**

OR

18. (a) Explain Image degradation model with the help of a neat diagram. **(8)**

(b) Illustrate the split and merge algorithm for image segmentation using neat sketches. **(6)**

19. (a) Explain the purpose of morphological operations in digital image? Describe the opening and closing operations with examples. **(7)**

(b) Illustrate Hit or Miss Transformation. **(7)**

OR

20. (a) Explain the concept of the chain coding scheme with its applications. **(6)**

(b) Describe in detail any two boundary representation schemes and illustrate with examples. **(8)**

Teaching Plan

No	Contents	No. of Lecture Hours (36 hrs)
Module-1 (Digital Image Fundamentals) (7 hours)		
1.1	Elements of Visual Perception, A Simple Image Formation Model	1
1.2	Spatial and Intensity Resolution, Image Interpolation, Classification of Digital Image.	1
1.3	Image Types, Image Storage Mechanisms.	1
1.4	Arithmetic and Logical Operations.	1
1.5	Geometric Spatial Transformations and Image Registration.	1
1.6	Image File Formats.	1

1.7	Colour Fundamentals and Colour Models.	1
Module-2 (Image Transforms) (8 hours)		
2.1	Basic concept of spatial domain and frequency domain.	1
2.2	Need of Image Transform, Basic properties of unitary transform.	1
2.3	Discrete Fourier transform, Proof DFT is Unitary.	1
2.4	4 order DFT Transform coefficients (Derivation).	1
2.5	Problems (4 order DFT).	1
2.6	Discrete Cosine Transform- 2D DCT.	1
2.7	4 order DCT Transform Coefficients(No derivation needed).	1
2.8	Hadamard Transform.	1
Module-3 (Image Enhancement in spatial and frequency domain) (8 hours)		
3.1	Point operations- Clipping and Thresholding, Digital Negative. Intensity Level Slicing.	1
3.2	Bit Extraction, Range Compression + (Work out problems).	1
3.3	Spatial Operations-Fundamentals of spatial convolution and correlation.	1
3.4	Spatial averaging and spatial Low pass filtering, Directional Smoothing.	1
3.5	Median Filtering, Unsharp masking and Crispening.	1
3.6	Basics of Filtering in Frequency Domain.	1
3.7	Smoothing Frequency Domain Filters : Ideal Low Pass Filter; Gaussian Low Pass Filter; Butterworth Low Pass Filter;	1
3.8	Sharpening Frequency Domain Filters: Ideal High Pass Filter; Gaussian High Pass Filter; Butterworth High Pass filter.	1
Module-4 (Image Restoration & Image Segmentation) (6 hours)		
4.1	Image degradation model, Noise models.	1
4.2	Mean Filters – Order Statistic filter – Adaptive filters.	1
4.3	Edge Detection, Gradient operators, Laplace operators and zero crossings.	1

4.4	Thresholding- Basic Global Thresholding, Optimum global thresholding using Otsu method.	1
4.5	Multiple thresholds, Variable thresholding, Multivariable thresholding.	1
4.6	Region-Based Approach to Segmentation.	1
Module-5 (Morphological Operations & Representation and Description) (7 hours)		
5.1	Structuring Element. Dilation and Erosion,	1
5.2	Morphological Opening, Closing.	1
5.3	Hit or Miss Transformation.	1
5.4	Boundary Following. Chain Codes, Polygonal Approximation.	1
5.5	Boundary Descriptors.	1
5.6	Regional Descriptors.	1
5.7	Relational Descriptors.	1

CST448	INTERNET OF THINGS	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PEC	2	1	0	3	2019

Preamble: This course equips the learners with fundamental of the Internet of Things(IoT) and the IoT ecosystem. It covers the architecture of IoT, communication mechanisms, protocols, hardware, software, data analytics, and the cloud platforms for IoT. This course enables the students to design smart IoT applications for real world problems using Raspberry Pi.

Prerequisite: Basic knowledge in Data Communication, Computer Networks and Programming in Python

Course Outcomes: After the completion of the course the students will be able to

CO1	Outline the fundamentals of IoT and its underlying physical and logical architecture**(Cognitive Knowledge Level: Understand)**
CO2	Explain the hardware architectures for IoT **(Cognitive Knowledge Level : Understand)**
CO3	Outline the Network architectures for IoT**(Cognitive Knowledge Level : Understand)**
CO4	Implement data analytics on the IoT platforms **(Cognitive Knowledge Level : Apply)**
CO5	Appreciate the security considerations in IoT **(Cognitive Knowledge Level : Understand)**
CO6	Implement IoT applications using the available hardware and software. **(Cognitive Knowledge Level : Apply)**

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓									✓
CO2	✓	✓	✓									✓
CO3	✓	✓	✓									✓
CO4	✓	✓	✓	✓	✓							✓

CO5	✓	✓	✓		✓							✓
CO6	✓	✓	✓	✓	✓	✓						✓

Abstract POs Defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and teamwork
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Lifelong learning

Assessment Pattern

Blooms Category	**Continuous Assessment Tests**		**End Semester Examination Marks**
	Test 1 (Percentage)	**Test 2 (Percentage)**	
Remember	30	20	30
Understand	60	50	40
Apply	10	30	30
Analyze			

Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	**50**	**100**	**3 Hours**

Continuous Internal Evaluation Pattern:

Attendance	**10 marks**
Continuous Assessment Tests	**25 marks**
Continuous Assessment Assignment	**15 marks**

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks. First Internal Examination shall be preferably conducted after completing the first half of the syllabus, and the Second Internal Examination shall be preferably conducted after completing the remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly covered module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly covered module), each with 7 marks. Out of the 7 questions in Part B, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 questions from each module of which a student should answer anyone. Each question can have a maximum 2 subdivisions and carries 14 marks.

Syllabus

Module- 1 (IoT Architecture)

What is IoT, Genesis of IoT, IoT and Digitization, IoT Impact, Convergence of IT and IoT, IoT Challenges, IoT Network Architecture and Design, Drivers Behind New Network Architectures, Comparing IoT Architectures, A Simplified IoT Architecture, The Core IoT Functional Stack, IoT Data Management and Compute Stack.

Module- 2 (Engineering IoT Networks)

Smart Objects: The "Things" in IoT, Sensors, Actuators, and Smart Objects, Sensor Networks, Connecting Smart Objects, Communications Criteria, IoT Access Technologies

Module- 3 (IoT Network Layer)

IP as the IoT Network Layer, The Business Case for IP, The need for Optimization, Optimizing IP for IoT, Profiles and Compliances, Application Protocols for IoT, The Transport Layer, IoT Application Transport Methods

Module 4 (Data Analytics for IoT)

Data and Analytics for IoT, An Introduction to Data Analytics for IoT, Machine Learning, Big Data Analytics Tools and Technology, Edge Streaming Analytics, Network Analytics, Securing IoT, A Brief History of OT Security, Common Challenges in OT Security, Differences between IT and OT Security Practices and Systems, Formal Risk Analysis Structures: OCTAVE and FAIR.

Module 5 (Developing IoT Systems)

IoT Logical Design using Python, IoT Physical Devices and Endpoints - Raspberry Pi interfaces, Programming Raspberry Pi using Python, Other IoT devices, IoT Physical devices and Cloud offerings, Cloud Storage Models, WAMP - Autobahn for IoT, Django, Designing RESTful Web API, Cloud Web Services for IoT.

Textbooks

1. David Hanes, Gonzalo Salgueiro, Patrick Grossetete, Robert Barton, Jerome Henry, "IoT Fundamentals: Networking Technologies, Protocols, and Use Cases for the Internet of Things", 1st Edition, Pearson Education (Cisco Press Indian Reprint)

2. Arshadeep Bahga, Vijay Madisetti, “Internet of Things: A hands-on approach”, University Press, 2015 (First edition)

References

1. Rajkamal, “Internet of Things: Architecture and Design Principles”, McGraw Hill (India) Private Limited
2. Dieter Uckelmann, Mark Harrison, Michahelles Florian (Ed.), Architecting the internet of things, Springer, 2011
3. Dr. Ovidiu Vermesan, Dr. Peter Friess, Internet of Things: Converging Technologies for Smart Environments and Integrated Ecosystems, River Publishers, 2013
4. Simon Monk, “Programming Arduino: Getting Started with Sketches”, McGraw Hill Publications

Sample Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Write a short note on the impact of IoT in the real world
2. Explain the challenges of IoT.
3. Compare OT and IT Technology.
4. Describe the elements of one M2M architecture of IoT

Course Outcome 2 (CO2):

1. Mention any four wireless technologies and its architectural characteristics
2. Comment things in IoT
3. Compare biosensors and biodegradable sensors used in IoT
4. Explain the term NBIoT(Narrow Band IoT)

Course Outcome 3 (CO3):

1. Discuss the need for optimization
2. Compare MQTT and COAP
3. Explain different schedule management and packet forwarding models of 6TiSCH

Course Outcome 4(CO4):

1. Compare Bigdata and edge analytics
2. Compare structured and unstructured data
3. Describe the components of FNF

Course Outcome 5(CO5):

1. What are the major challenges in IoT security?
2. Explain the impact of OT Network Characteristics onIoT Security.

Course Outcome 6(CO6):

1. Implement LDR interfacing with Raspberry Pi
2. Explain the development of a RESTful web API.

Model Question Paper

QP CODE: PAGES :3

Reg No: _____

Name: _____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: CST448

Course Name: Internet of Things

Max.Marks : 100 **Duration: 3 Hours**

PART A

Answer All Questions. Each Question Carries 3 Marks

1. Explain the role of IoT in connected roadways,
2. Describe the functions of the various layers of simplified IoT Architecture Model.
3. Explain the communication protocols employed in Wireless Sensor Networks
4. What are the essential performance considerations of constrained-node networks?
5. Explain the parameters to be considered while choosing between IP adaptation / adoption for last mile communication.
6. With neat diagrams compare the IoT protocol stacks using 6LoWPAN and IP.
7. Differentiate the types of IoT data analytics results.

8. How can the insecure operational protocols be characterized?

9. Write a program to interface an LED and a switch with Raspberry Pi

10. List down the Raspberry Pi interfaces and explain. **(10x3=30)**

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) Illustrate the impact of IoT in at least 2 domains of normal human life. **(9)**

(b) Describe the Application and Analytics sublayer of IoT Architecture **(6)**

OR

12. (a) Describe the Standardized IoT architectures. **(8)**

(b) Explain the functions of Access Network Sublayer of IoT Architecture **(6)**

13. (a) Describe the LoRaWAN technology as an IoT communication paradigm. **(10)**

(b) Describe various types of sensors. **(4)**

OR

14. (a) Define actuators. Describe the roles of actuators in IoT systems. **(6)**

(b) Explain the IEEE 802.15.4 standard for wireless communication. **(8)**

15. (a) Explain Message Queuing Telemetry Transport framework and message format. **(8)**

(b) Explain tunneling of legacy SCADA over IP Networks with a neat diagram. **(6)**

OR

16. (a) Explain SCADA Transport over LLNs with MAP-T. **(7)**

(b) Explain RPL encryption and authentication on constrained nodes. **(7)**

17. (a) Explain the Hadoop ecosystem with a neat diagram. **(7)**

(b) Explain the Flexible NetFlow Architecture. **(7)**

OR

18. (a) Explain the "The Purdue Model for Control Hierarchy" and OT network characteristics. **(8)**

(b) Explain any twp formal risk analysis structures **(6)**

19. (a) Explain the working of WAMP protocol. **(8)**

(b) Describehow AWS supports IoT development **(6)**

OR

20. (a) Demonstrate an example of Raspberry Pi applications for Industrial IoT. **(8)**

(b) Explain the Django Architecture **(6)**

TEACHING PLAN

No	Contents	No of Lecture Hrs (35 Hrs)
Module – 1 (IoT Architecture) (6 hrs) (TB-1, Chapter 1,2)		
1.1	What is IoT, Genesis of IoT, IoT and Digitization,	1
1.2	IoT Impact, Convergence of IT and IoT, IoT Challenges	1
1.3	IoT Network Architecture and Design	1
1.4	Drivers Behind New Network Architectures, Comparing IoT Architectures	1
1.5	A Simplified IoT Architecture,	1

1.6	The Core IoT Functional Stack, IoT Data Management and Compute Stack.	1
Module- 2 (Engineering IoT Networks) (7hrs)(TB-1, Chapter 3,4)		
2.1	Smart Objects: The "Things" in IoT,	1
2.2	Sensors, Actuators, and Smart Objects	1
2.3	Sensor Networks	1
2.4	Connecting Smart Objects	1
2.5	IoT Access Technologies –IEEE 802.15.4 (g/e), 1901.2a	1
2.6	IoT Access Technologies - 802.11ah, LoRaWAN	1
2.7	IoT Access Technologies – LoRaWAN, NBIoT, LTE	1
Module- 3 (IoT Network Layer) (7 hrs)(TB-1, Chapter 5,6)		
3.1	IP as the IoT Network Layer, The Business Case for IP	1
3.2	The need for Optimizing IP for IoT	1
3.3	Optimizing IP for IoT, Profiles, and Compliance	1
3.4	Application Protocols for IoT - CoAP	1
3.5	Application Protocols for IoT - MQTT	1
3.6	The Transport Layer, IoT Application Transport Methods	1
3.7	The Transport Layer, IoT Application Transport Methods	1
Module 4 (Data Analytics for IoT) (6hrs)(TB-1, Chapter 7,8)		
4.1	An Introduction to Data Analytics for IoT, Machine Learning	1
4.2	Big Data Analytics Tools and Technology	1
4.3	Edge Streaming Analytics, Network Analytics	1

4.4	A Brief History of OT Security, Common Challenges in OT Security	1
4.5	Differences between IT and OT Security Practices and Systems	1
4.6	Formal Risk Analysis Structures: OCTAVE and FAIR	1
Module 5 (Developing IoT Systems)(9 hrs) (TB-2, Chapter 6,7,8)		
5.1	IoT Logical Design using Python,	1
5.2	IoT Physical Devices and Endpoints	1
5.3	Raspberry Pi interfaces, Programming Raspberry Pi using Python	1
5.4	Other IoT devices	1
5.5	Cloud Storage Models	1
5.6	WAMP-Autobahn for IoT	1
5.7	Django	1
5.8	Designing RESTful Web API	1
5.9	Cloud Web Services for IoT.	1

CST458	SOFTWARE TESTING	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PEC	2	1	0	3	2019

Preamble: This is a course in theoretical computer science that introduces the concepts and methods in software testing. It covers various techniques for test case design used to test software artifacts, including requirements, design, and code, the different techniques for test case design based on graphs, programming language syntaxes and symbolic execution using PEX tool. It enables the learners to follow a systematic software testing approaches while developing applications.

Prerequisite: Nil

Course Outcomes: After the completion of the course the student will be able to:-

CO1	List a range of different software testing techniques and be able to apply specific unit testing method to the projects using Junit.**(Cognitive Knowledge Level: Understand)**
CO2	Illustrate using appropriate tools the mutation testing method for a given piece of code to identify hidden defects that can't be detected using other testing methods.**(Cognitive Knowledge Level: Apply)**
CO3	Explain graph coverage criteria in terms of control flow graph and data flow graph for a given program.**(Cognitive Knowledge Level: Understand)**
CO4	Demonstrate the importance of black-box approaches in terms of domain and functional testing.**(Cognitive Knowledge Level: Apply)**
CO5	Illustrate the use of PEX tool with symbolic execution.**(Cognitive Knowledge Level: Apply)**

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO 9	PO10	PO11	PO12
CO1	✓	✓	✓									✓
CO2	✓	✓	✓	✓	✓					✓		✓
CO3	✓	✓	✓							✓		✓
CO4	✓	✓	✓	✓								✓

CO5	✓	✓	✓	✓	✓					✓		✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	PO7	Environment and Sustainability
PO2	Problem Analysis	PO8	Ethics
PO3	Design/Development of solutions	PO9	Individual and team work
PO4	Conduct investigations of complex problems	PO10	Communication
PO5	Modern tool usage	PO11	Project Management and Finance
PO6	The Engineer and Society	PO12	Life long learning

Assessment Pattern

Bloom's Category	**Continuous Assessment Tests**		**End Semester Examination Marks**
	Test 1 (Marks)	**Test 2 (Marks)**	
Remember	30	30	30
Understand	40	40	40
Apply	30	30	30
Analyze			
Evaluate			
Create			

Mark Distribution

Total Marks	**CIE Marks**	**ESE Marks**	**ESE Duration**
150	**50**	**100**	**3 hours**

Continuous Internal Evaluation Pattern:

Attendance	**10 marks**
Continuous Assessment Tests(Average of SeriesTests1& 2)	**25 marks**
Continuous Assessment Assignment	**15 marks**

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks. First series test shall be preferably conducted after completing the first half of the syllabus and the second series test shall be preferably conducted after completing remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly completed module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly completed module), each with 7 marks. Out of the 7 questions, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 questions from each module of which a student should answer any one. Each question can have maximum 2 sub-divisions and carries 14 marks.

Syllabus

Module - 1 (Introduction to Software Testing)

Some Popular Errors – Ariane 5, Therac 25, Intel Pentium Bug. What is Software testing? Why should it be tested? Software Quality, Role of Testing. Testing Process - Level 0 thinking, Level 1 thinking, Level 2 thinking, Level 3 thinking, Level 4 thinking. Software Testing Terminologies - Verification, Validation and Testing, Faults, Error and Bug, Test cases, Coverage Criteria. Types of Testing- Unit testing, integration testing, System testing, Acceptance testing, Beta testing, Functional testing, Stress testing, Performance testing, Usability testing and Regression testing. Testing Methods - Black Box testing, White Box testing, Grey Box testing.

Module - 2 (Unit Testing)

Concept of Unit testing. Static Unit testing. Dynamic Unit testing - Control Flow testing, Data Flow testing, Domain testing, Functional Program testing. Mutation testing - Mutation and Mutants, Mutation operators, Mutation score. Junit - Framework for Unit testing. Case Study - Mutation testing using Junit and Muclipse.

Module - 3 (Unit Testing - White Box Approaches)

Overview of Graph Coverage Criteria. Structural Graph Coverage Criteria - Node/vertex coverage, Edge coverage, Edge pair coverage, Path coverage, Complete path coverage, Prime path coverage, Complete round trip coverage, Simple round trip coverage. Data Flow Criteria - du paths, du pairs. Subsumption Relationships among Graph Coverage Criteria. Graph Coverage for Source Code - Control flow graphs for code, CFG: If statement, CFG: If statement with return, CFG: Switch-case, CFG: Loops, CFG: Exceptions (try-catch). Example program – Statistics. Graph Coverage for Design Elements - Call graphs and classes, Class inheritance testing: Coverage criteria, Coverage criteria on inheritance graph, Data flow at the design level, Inter-procedural DU pairs, Coupling du-pairs example. Example - Quadratic Root. Case Study - Graph Based testing using JUnit Framework.

Module - 4 (Unit Testing - Black Box Approaches)

Domain Testing / Input Space Partitioning - Partitions of a set. Input domain modelling - Interface-based approach, Functionality-based approach. Identifying values. Multiple partitions of the input domain - All Combinations Coverage (ACoC), Each Choice Coverage (ECC), Pair-wise Coverage, T-wise Coverage, Base Choice Coverage, Multiple Base Choices Coverage. TriTyp example. Functional Testing - Functional Testing Concepts of Howden. Functional testing - Important Steps. Types of Functional testing - Equivalence Class Partitioning, Boundary Value Analysis, Decision Tables, Random Testing. Case Study - Black Box testing approaches using JUnit.

Module - 5 (Grey Box Testing Approaches)

Introduction to Grey Box testing - Why Grey Box testing, Gray Box Methodology, Advantages and Disadvantages. Techniques of Grey Box Testing - Matrix Testing, Regression Testing, Orthogonal Array Testing or OAT, Pattern Testing. An Introduction to PEX - Parameterized Unit Testing, The Testing Problem. Symbolic Execution – Example, Symbolic execution tree. PEX application Case Study – PEX.

Text Books

1. Paul Ammann and JeffOffutt , Introduction to Software Testing, Cambridge University Press
2. Kshirasagar Naik and Priyadarshi Tripathy, Software Testing And Quality Assurance: Theory And Practice, Wiley.

Reference Materials

1. King, James C, “Symbolic Execution and Program Testing”, Association for Computing Machinery, July 1976.

Sample Course Level Assessment Questions

Course Outcome 1 (CO1):

Explain the following types of testing methods with examples.

(i) Balck-box testing.

(ii) White-box testing.

(iii) Grey-box testing.

Course Outcome 2 (CO2):

Define 12 mutants for the following method *power*() using effective mutation operators. Try to use each mutation operator at least once. Approximately, how many mutants do you think there would be, if all mutants for *power*() were created?

```
public static int power (int left, int right)
{
//*************************************
// Raises Left to the power of Right
// precondition : Right >= 0
// postcondition: Returns Left**Right
//*************************************
```

```
intrslt;
rslt = Left;
if (Right == 0)
{
rslt = 1;
}
else
{
for (int i = 2; i <= Right; i++)
rslt = rslt * Left;
}
return (rslt);
}
```

Course Outcome 3 (CO3):

Draw the control flow graph and data flow graph of given piece of code.

```
public static double ReturnAverage(int value[],int AS, int MIN, int MAX){
/*
Function: ReturnAverageComputes the averageof all those numbers in the input array in the positive range [MIN, MAX]. The maximum size of the array is AS. But, the array size could be smaller than AS in which case the end of input is represented by -999.
*/
int i, ti, tv, sum;
doubleav;
i = 0; ti = 0; tv = 0; sum = 0;
while (ti< AS && value[i] != -999) {
ti++;
if (value[i] >= MIN && value[i] <= MAX) {
tv++;
sum = sum + value[i];
}
i++;
}
```

```
if (tv> 0)
av = (double)sum/tv;
else
av = (double) -999;
return (av);
}
```

Course Outcome 4 (CO4):

Explain the following with examples.

1. Input domain modelling.
2. All Combinations Coverage (ACoC)
3. Each Choice Coverage (ECC)
4. Pair-wise Coverage
5. T-wise Coverage
6. Base Choice Coverage
7. Multiple Base Choices Coverage.

Course Outcome 5 (CO5):

Draw the symbolic execution tree for the following program code and explain the symbolic execution of testme (α1, α2).

```
int twice (int v) {
 return 2 ∗ v;
 }
 void testme (int x, int y ) {
 z = twice ( y);
 if ( z == x ){
 if ( x > y + 10)
ERROR;
 }
 }
int main() {
 x = sym input();
 y = sym input();
testme ( x , y);
```

return(0);

Model Question Paper

QP CODE: **PAGES: 3**

Reg No:_____ Name :_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: CST458

Course Name: Software Testing

Max.Marks:100 **Duration: 3 Hours**

PART A

Answer all Questions. Each question carries 3 Marks

1. Explain the differences between Validation and Verification?
2. Explain the differences between Fault, Error, and Bug?
3. Define Ground string, Mutation score, and Mutants?
4. What are the functions of Test driver and Test stubs in dynamic unit testing?
5. Define Node coverage, Edge coverage and Prime path coverage in a control flow graph?
6. What are du paths and du pairs in a data flow graph?
7. Explain the two approaches in input domain modelling?
8. Explain the difference between Equivalence Class Partitioning and Boundary Value Analysis?
9. Briefly explain three techniques of Grey box testing?
10. Explain the concept of symbolic execution with the help of a toy example?

(10x3=30)

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) Explain the following types of testing

(i) Black Box testing (ii) White Box testing (iii) GreyBox testing (iv) Unit testing (v) Integration testing (vi) System testing (vii) Acceptance testing **(14)**

OR

12. (a) Explain the following coverage criterias based on the code fragment given below? (i) Functional coverage (ii) Statement coverage (iii)Conditional coverage (iv)Branch coverage **(8)**

```
int foo (int x, int y){
    int z = 0;
    if ((x > 0) && (y > 0)){
    z = x;}
    return z;
    }
```

(b) Write positive and negative test cases for an ATM Machine? **(6)**

13. (a) Explain Dynamic unit test environment with a neat figure. **(8)**

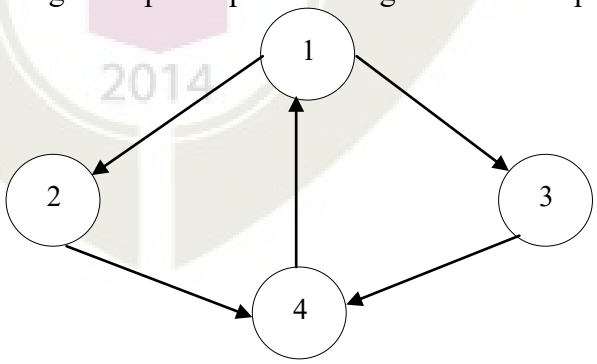
(b) Explain the major difference between control flow testing and data flow testing. **(6)**

OR

14. (a) Explain seven types of mutation operators with neat examples? **(14)**

15. (a) Explain touring, side trips and detours with a neat example **(7)**

(b) Explain simple path coverage and prime path coverage with the help of CFG given below? **(7)**



OR

16. (a) Draw CFG fragment for **(7)**

(i) Simple *if* (ii) Simple *while* loop (iii) Simple *for* loop

(b) Explain the following concepts with examples? **(7)**

(i)Call graph (ii) Inheritance graph (iii) Coupling du-pairs

17. (a) What are the four important steps in functional testing? **(7)**

(b) Briefly explain input domain modelling approaches? **(7)**

OR

18. (a) Consider the triangle classification program with a specification: **(6)**

The program reads floating values from the standard input. The three values *A*, *B*, and *C* are interpreted as representing the lengths of the sides of triangle. The program then prints a message to the standard output that states whether the triangle, if it can be formed, is scalene, isosceles, equilateral, orright angled. Determine the following for the above program:

(i) For the boundary condition $A + B > C$ case (scalene triangle), identify test cases to verify the boundary.

(ii) For the boundary condition $A = C$ case (isosceles triangle), identify testcases to verify the boundary.

(iii) For the boundary condition $A = B = C$ case (equilateral triangle), identify testcases to verify the boundary.

(b) Develop a decision table to generate test cases for this specification. **(8)**

19. (a) Explain the importance of grey box testing, its advantages and disadvantages? **(9)**

(b) Explain the concept of symbolic execution tree? **(5)**

OR

20. (a) Consider the code fragment given below: - **(7)**

1. POWER: PROCEDURE(X, Y);
2. Z ← 1;

3. J ← 1;
4. LAB: IF Y ≥ J THEN
5. DO; Z← Z * X;
6. J ← J + 1;
7. GO TO LAB; END;
8. RETURN (Z) ;
9. END;

a) Explain Symbolic execution of POWER (α1, α2).

(b) Explain Execution tree for POWER (α1, α2). **(7)**

TEACHING PLAN

No	Contents	No of Lecture Hrs (35 hrs)
Module 1 (Introduction to Software Testing) -(7 Hours)		
1.1	Some Popular Errors– Ariane 5, Therac 25, Intel Pentium Bug.	1 Hour
1.2	What is Software testing? Why should it be tested? Software Quality, Role of Testing.	1 Hour
1.3	Testing Process - Level 0 thinking, Level 1 thinking, Level 2 thinking, Level 3 thinking, Level 4 thinking.	1 Hour
1.4	Software Testing Terminologies- Verification, Validation and Testing, Faults, Error and Bug, Test cases, Coverage Criteria.	1 Hour
1.5	Types of Testing- Unit testing, integration testing, System testing, Acceptance testing, Beta testing	1 Hour
1.6	Functional testing, Stress testing, Performance testing, Usability testing and Regression testing.	1 Hour
1.7	Testing Methods - Black Box testing, White Box testing, Grey Box testing.	1 Hour
Module 2 (Unit testing)- (6 Hours)		
2.1	Concept of Unit testing, Static Unit Testing	1 Hour

2.2	Dynamic Unit testing - Control Flow testing, Data Flow testing, Domain testing, Functional Program testing.	1 Hour
2.3	Mutation testing - Mutation and Mutants, Mutation operators, Mutation score.	1 Hour
2.4	Junit - Framework for Unit testing.	1 Hour
2.5	Case Study - Mutation testing using Junit	1 Hour
2.6	Case Study - Mutation testing using Muclipse	1 Hour
Module 3 (Unit Testing:- White Box Approaches)- (8 Hours)		
3.1	Structural Graph Coverage Criteria - Node/vertex coverage, Edge coverage, Edge pair coverage, Path coverage	1 Hour
3.2	Complete path coverage, Prime path coverage, Complete round trip coverage, Simple round trip coverage.	1 Hour
3.3	Data Flow Criteria - du paths, du pairs	1 Hour
3.4	Subsumption Relationships among Graph Coverage Criteria	1 Hour
3.5	Graph Coverage for Source Code – Control Flow Graphs (CFG) for code, CFG: If statement, CFG: If statement with return, CFG: Switch-case, CFG: Loops, CFG: Exceptions (try-catch). Example program - Statistics	1 Hour
3.6	Graph Coverage for Design Elements – Structural graph coverage and data flow graph coverage for design elements	1 Hour
3.7	Case Study - Graph Based testing using JUnit Framework. (Lecture 1)	1 Hour
3.8	Case Study - Graph Based testing using JUnit Framework. (Lecture 2)	1 Hour
Module 4 (Unit Testing:- Black Box Approaches) -(7 Hours)		
4.1	Domain Testing / Input Space Partitioning - Partitions of a set.	1 Hour
4.2	Input domain modelling - Interface-based approach, Functionality-based approach.	1 Hour

4.3	Multiple partitions of the input domain - All Combinations Coverage (ACoC), Each Choice Coverage (ECC), Pair-wise Coverage, T-wise Coverage, Base Choice Coverage, Multiple Base Choices Coverage.	1 Hour
4.4	Functional Testing - Functional Testing Concepts of Howden. Important Steps.	1 Hour
4.5	Types of Functional testing - Equivalence Class Partitioning, Boundary Value Analysis	1 Hour
4.6	Decision Tables, Random Testing.	1 Hour
4.7	Case Study - Black Box testing approaches using JUnit.	1 Hour
Module 5 (Grey Box Testing Approaches)- (7 Hours)		
5.1	Introduction to Grey Box testing - Why Grey Box testing, Gray Box Methodology, Advantages and Disadvantages.	1 Hour
5.2	Techniques of Grey Box Testing - Matrix Testing, Regression Testing, Orthogonal Array Testing or OAT, Pattern Testing.	1 Hour
5.3	An Introduction to Pex - Parameterized Unit Testing, The Testing Problem.	1 Hour
5.4	Symbolic Execution – Example, Symbolic execution tree.	1 Hour
5.5	Case Study – PEX (Lecture 1)	1 Hour
5.6	Case Study – PEX (Lecture 2)	1 Hour
5.7	Case Study – PEX (Lecture 3)	1 Hour

CST468	BIOINFORMATICS	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PEC	2	1	0	3	2019

Preamble: This course helps the learners to understand the fundamental concepts in Molecular Biology, Genomics, Proteomics and Modelling. This course introduces bio macromolecules such as genes and proteins, different biological databases, and tools and algorithms for biological data processing, analysis and interpretation, and the elements of the systems approach to Molecular Biology. This course enables the learners to contribute towards drug discovery and computational analysis and modelling of biological process.

Prerequisite: Basic background in higher secondary biology

Course Outcomes: After the completion of the course, the student will be able to

CO 1	Describe the basic concepts of Bioinformatics with an emphasis on structure, function and synthesis of biomolecules (**Cognitive knowledge level : Understand**)
CO 2	Identify biological data formats and databases, retrieve bio-sequences, and align bio-sequences to identify similarity (**Cognitive knowledge level : Apply)**
CO 3	Employ similarity searching tools and algorithms to align sequences to highlight the similarity, and describe the structure of genes (**Cognitive knowledge level : Apply)**
CO 4	Demonstrate Protein Structure, visualize protein structure using tools, and explain how proteins interact (**Cognitive knowledge level : Apply)**
CO 5	Explain the fundamental aspects of Systems Biology, Computational Modeling and properties of models (**Cognitive knowledge level : Understand**)

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓										✓
CO2	✓	✓	✓	✓	✓							✓
CO3	✓	✓	✓	✓	✓							✓

CO4	✓	✓	✓	✓								✓
CO5	✓	✓			✓							✓

PO#	Broad PO	PO#	Broad PO
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and team work
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Life long learning

Assessment Pattern

Bloom's Category	Continuous Assessment Tests		End Semester Examination
	Test1 (%)	**Test2 (%)**	
Remember	**30**	**30**	**30**
Understand	**50**	**50**	**50**
Apply	**20**	**20**	**20**
Analyse			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	50	100	3

Continuous Internal Evaluation Pattern:

Attendance	**10 marks**
Continuous Assessment Tests (Average of Series Tests 1& 2)	**25 marks**
Continuous Assessment Assignment	**15 marks**

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks. First series test shall be preferably conducted after completing the first half of the syllabus and the second series test shall be preferably conducted after completing remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly completed module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly completed module), each with 7 marks. Out of the 7 questions, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 full questions from each module of which student should answer any one. Each question can have maximum 2 sub-divisions and carries 14 marks.

Syllabus

Module-1 (Introduction to bioinformatics)

Introduction to bioinformatics, Nature & Scope of Bioinformatics, DNA, RNA, and Protein: The Central Dogma, Messenger RNA, tRNA, rRNA, Genetic code, Gene Structure and Control, Transcription, translation

Module-2 (Introduction to bio sequences and analysis)

Introduction to Biological Databases, NCBI, Genbank, Bio sequence formats- FASTA, Sequence alignment- Global Alignment and Local Alignment, Dot Matrix Method, Dynamic Programming Method, Gap Penalties, Amino Acid Scoring Matrices - PAM and BLOSUM

Module-3 (Database Similarity Searching and genomics)

Database Similarity Searching, BLAST – Variants -BLASTN, BLASTP, BLASTX, Statistical Significance, Needleman and Wunsch and Smith–Waterman Method, Multiple Sequence Alignment, scoring function, Clustal, introduction to structure of prokaryotic and eukaryote gene

Module-4 (Proteomics)

Protein Structure, Ramachandran Plot, Hierarchies of Protein Structure, Determination of Protein three-dimensional structure, protein structure database-PDB, Protein structure visualization, introduction to Protein protein interaction, STRING database

Module-5 (Systems Biology)

Introduction to Systems Biology, Models and Modelling, Properties of models, Systems state and steady state, Variables, Parameters, and Constants in modelling, Purpose and Adequateness of Models, Advantages of Computational Modelling, Model Development, Network Versus Elements, Modularity, Robustness and Sensitivity, Data Integration

Text books

1. Zvelebil, Marketa J., and Jeremy O. Baum. *Understanding bioinformatics*. Garland Science, 2007.
2. Xiong, Jin. *Essential bioinformatics*. Cambridge University Press, 2006.
3. Klipp, E., Herwig, R., Kowald, A., Wierling, C., &Lehrach, H. *Systems biology in practice: concepts, implementation and application*. John Wiley & Sons. 2005

References

1. Baxevanis, Andreas D., Gary D. Bader, and David S. Wishart, eds. *Bioinformatics*. John Wiley & Sons, 2020.
2. Shaik, Noor Ahmad, et al. *Essentials of Bioinformatics, Volume I.* Springer, 2019

3. Selzer, Paul M., Richard J. Marhöfer, and Andreas Rohwer, *Applied bioinformatics*. *An introduction–Springer, Verlag,*, 2008.
4. S C Rastogi, N Mendiratta and PRastogi, *Bioinformatics: Methods and Applications* , PHI Learning Private Limited, New Delhi, 2015.
5. D E Krane and M L Raymer, *Fundamental Concepts of Bioinformatics*, Pearson Education, 2006.
6. Andreas D.Baxevanis, B F Francis Ouellette, *Bioinformatics - A Practical Guide to the Analysis of Genes and Proteins*, Third Edition, John Wiley & Sons INC. , U.K. 2006
7. Neil C Jones and Pavel A Pevzner, *An Introduction to Bioinformatics Algorithms*, MIT press, 2004.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Compare and contrast the DNA and RNA on the basis of structure and functions.
2. Demonstrate with the help of a flow diagram the generation of protein using the transcription and translation process.

Course Outcome 2 (CO2):

1. Download DNA sequence of human insulin form NCBI
2. Identify the following qualifiers for GenBank and give their definitions: [ACCN], [ALL], [AUTH], [ECNO], [FKEY], [GENE], [JOUR], [KYWD]
3. Construct a dot plot and find the sequence alignment between the following two sequences:
   Sequence1: GATTCTATCTAACTA, Sequence2: GTTCTATTCTAAC

Course Outcome 3 (CO3):

1. Apply Needleman-Wunsch Algorithm to perform sequence alignment for the following sequences: CGTGAATTCAT (sequence #1), GACTTAC (sequence #2)
2. Construct a BLAST procedure for sequence alignment(HSP) if a sequence and its corresponding database sequence are given. Assume the necessary data and demonstrate the procedure.

Course Outcome 4 (CO4):

1. Differentiate between the different protein molecular structure visualizations. Also mention the advantages and uses of each visualization technique.
2. Make use of an example and demonstrate the steps in protein comparison. Show how root mean square deviationis calculated while comparing two proteins.

Course Outcome 5 (CO5):

1. Explain how systems biology is used in data integration.
2. Explain the process of model development.

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 3**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: CST468

Course Name: Bioinformatics

Max. Marks : 100 **Duration: 3 Hours**

PART A

Answer All Questions. Each Question Carries 3 Marks

1. Differentiate DNA, Gene, genome and chromosome.
2. What are the functions of mRNA, tRNA and rRNA?
3. What do you mean by Gene expression?
4. Write difference between local and global alignment.
5. Write short note on Gap penalties and its usage in comparing Biological sequences.
6. List any three typesof BLAST and make short description on each.
7. What are the principle underlying the formation of Ramachandran plot?.
8. What are the experimental methods for determining protein structure?
9. What do you mean by steady state in a biological system.
10. Justify the statement - systems are modular. **(10x3=30)**

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) What is the central dogma of molecular biology? **(6)**

    (b) Explain the steps involved in the process of transcription. How is the primary transcript produced by a prokaryote different from that produced by a eukaryotic cell? **(8)**

OR

12. (a) Discuss translation process in protein synthesis. **(6)**

    (b) Explain bio-molecules involved in central dogma, its structure and types. **(8)**

13. (a) Explain the importance of Primary and secondary databases in Bioinformatics **(6)**

    (b) Illustrate the methods of pairwise sequence alignment. What is the use of assigning gap penalties in alignment? **(8)**

OR

14. (a) Illustrate sequence alignment. What are the applications of sequence alignment in Bioinformatics? **(7)**

    (b) What is the use of scoring matrices? Differentiate between PAM and BLOSUM matrices and its usage in alignment. **(7)**

15. (a) Using Needleman and Wunsch dynamic programming method, construct the partial alignment score table for the following two sequences, using the scoring parameters: match score: +5, mismatch score: -1, gap penalty: -2. **(9)**
    CCATGCU
    GATTACA
    Also write down the optimal global alignment between these sequences along with the optimal score.

    (b) Interpret the blast result and statistical significance of the alignment by analyzing the results. **(5)**

OR

16. (a) Using Smith Waterman method construct the partial alignment scoring table and obtain the optimal local alignment of the following two sequences: **(9)**
    ACGTATCGCGTATA
    GATGCTCTCGGAJAA

    (b) Illustrate multiple sequence alignment. **(5)**

17. (a) Discuss hierarchies of protein structure. **(6)**

    (b) Explain how the protein structure is determined by using experimental techniques. **(8)**

OR

18. (a) Discuss protein interaction. How it contributes to the complexity of an organism? **(9)**

    (b) Discuss on Protein Structure Database. **(5)**

19. (a) Discuss systems biology approach of understanding complex biological systems. **(6)**

(b) Explain on Variables, Parameters, and Constants in modeling biological systems. **(8)**

OR

20. (a) Explain on advantages of Computational Modeling of biological system. **(7)**

(b) What are the properties of models in biological system? **(7)**

TEACHING PLAN

No	Contents	No of Lecture (36 Hrs)
	Module-1 (Introduction to bioinformatics)(8 hrs) Text 1 (Relevant topics from chapter 1.1, 1.2, 1.3)	
1.1	Introduction to bioinformatics	1
1.2	Nature & Scope of Bioinformatics	1
1.3	DNA, RNA, and Protein	1
1.4	The Central Dogma introduction	1
1.5	Messenger RNA, tRNA, rRNA,	1
1.6	Genetic code,	1
1.7	Gene Structure and Control	1
1.8	Transcription, Translation	1
	Module-2 (Introduction to bio sequences and analysis) (7 hrs) Text 2 (Relevant topics from chapter 2, 3)	
2.1	Introduction to Biological Databases	1
2.2	NCBI Sequence retrieval	1
2.3	Genbank, Bio sequence formats- FASTA	1
2.4	Sequence alignment- Global Alignment and Local Alignment	1
2.5	Dot Matrix Method, Dynamic Programming Method	1

2.6	Gap Penalties	1
2.7	Amino Acid Scoring Matrices – PAM, BLOSUM	1
Module-3 (Database Similarity Searching and genomics) (7 hrs) Text 2 (Relevant topics from chapter 4 5 and 8)		
3.1	Database Similarity Searching, BLAST, Variants of BLAST - BLASTN, BLASTP, BLASTX	1
3.2	BLAST Analysis - Statistical Significance	1
3.3	Needleman and Wunsch Method	1
3.4	Smith–Waterman Method	1
3.5	Multiple Sequence Alignment, scoring function	1
3.6	Clustal tool	1
3.7	Gene Structure of prokaryotic, eukaryote	1

Module-4 (Proteomics) (7 hrs) Text 2 (Relevant topics from chapter 12, 13 and 19)		
4.1	Protein Structure, Ramachandran Plot	1
4.2	Hierarchies of Protein Structure	1
4.3	Determination of Protein three-dimensional structure	1
4.4	protein structure database-PDB	1
4.5	Protein structure visualization	1
4.6	Protein protein interaction	1
4.7	Protein protein interaction networks, STRING database	1
Module-5 (Systems Biology) (7 hrs) Text 3 (Relevant topics from Section 1.1-1.4)		
5.1	Introduction to Systems Biology, Properties of models	1
5.2	Systems state and steady state	1
5.3	Variables, Parameters, and Constants in modelling	1
5.4	Purpose and Adequateness of Models	1
5.5	Advantages of Computational Modelling ,Model Development (introduction only)	1
5.6	Network Versus Elements, Modularity,	1

5.7	Robustness and Sensitivity, Data Integration	1

CST478	COMPUTATIONAL LINGUISTICS	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PEC	2	1	0	3	2019

Preamble: The course aims to teach the basics of Computational Linguistics to the students viewing language phenomena from a computational/statistical standpoint. This involves ideas about statstical and computational models and how these could be linked with various language processing tasks. The course helps the learner to appreciate the complexities involved in language processing tasks using a machine, in contrast with the ease with which human beings handle them. Some practical aspects are also discussed using the Python and NLTK framework to equip the student with the capability to design solutions to linguistic problems.

Prerequisite: Nil

Course Outcomes: After the completion of the course the student will be able to

CO#	CO
CO1	Explain the fundamental concepts of language processing **(Cognitive Knowledge Level: Understand)**
CO2	Demonstrate the concepts of probability, statistical inference and hidden Markov model. **(Cognitive Knowledge Level: Apply)**
CO3	Compare and summarize the various methods of word sense disambiguation, lexical acquisition and selectional preferences. (**Cognitive Knowledge Level: Apply)**
CO4	Make use of different Part-of-Speech Tagging methods for language modelling. **(Cognitive Knowledge Level: Apply)**
CO5	Examine Probabilistic Context Free Grammars and various probabilistic parsing methods **(Cognitive Knowledge Level: Apply)**
CO6	Develop simple systems for linguistic tasks using Python and NLTK. **(Cognitive Knowledge Level: Apply)**

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓											✓
CO2	✓	✓										✓
CO3	✓	✓	✓									✓
CO4	✓		✓		✓							✓
CO5	✓	✓	✓									✓
CO6	✓	✓	✓	✓	✓							✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and team work
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Life long learning

Assessment Pattern

Bloom's Category	Continuous Assessment Tests		End Semester Examination Marks (%)
	Test 1 (%)	Test 2 (%)	
Remember	30	30	30
Understand	30	30	30
Apply	40	40	40
Analyze			
Evaluate			

Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	50	100	3Hrs

Continuous Internal Evaluation Pattern:

Attendance **10 marks**

Continuous Assessment Tests(Average of SeriesTests1& 2) **25 marks**

Continuous Assessment Assignment **15 marks**

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks. First series test shall be preferably conducted after completing the first half of the syllabus and the second series test shall be preferably conducted after completing remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly completed module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly completed module), each with 7 marks. Out of the 7 questions, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 full questions from each module of which student should answer any one. Each question can have maximum 2 sub-divisions and carries 14 marks.

Syllabus

Module- 1 (Preliminaries)

Introduction: Rationalist and Empiricist Approaches to Language-Questions that linguistics should answer-Noncategorical phenomena in language-Language and cognition as probabilistic phenomena

The Ambiguity of Language: Why natural language processing is difficult-Lexical resources-Word counts-Zipf's laws-Collocations-Concordances

Linguistic Essentials:

Parts of Speech and Morphology -Nouns and pronouns-Words that accompany nouns: Determiners and adjectives-Verbs-Other parts of speech-Phrase Structure-Phrase structure grammars -Semantics and Pragmatics-Corpus Based Work

Module -2 (Mathematical Essentials:)

Probability Theory-Probability spaces-Conditional probability and independence-Bayes' theorem-Random variables-Expectation and variance-Notation-Joint and conditional distributions-Standard distributions-Bayesian statistics

Statistical Inference: n-gram Models over Sparse Data-Bins: Forming Equivalence Classes-Reliability vs discrimination-n gram models

Markov Models-Hidden Markov Models-Why use HMMs?-General form of an HMM-Finding the probability of an observation-Finding the best state sequence

Module -3 (Word Sense Disambiguation)

Methodological Preliminaries- Supervised and unsupervised learning-Pseudowords-Upper and lower bounds on performance-Supervised Disambiguation-Bayesian classification-Dictionary based Disambiguation-Disambiguation based on sense definitions-Thesaurus based disambiguation

Lexical Acquisition-Evaluation Measures-Verb Subcategorization -Attachment Ambiguity-PP attachment- Selectional Preferences

Semantic Similarity: Vector space measures-Probabilistic measures

Module -4 (Grammar)

Part-of-Speech Tagging-The Information Sources in Tagging-Markov Model Taggers-Hidden Markov Model Taggers-Applying HMMs to POS tagging-The effect of initialization on HMM training-Transformation Based Learning of Tags

Probabilistic Context Free Grammars-Some Features of PCFGs-Questions for PCFGs -The Probability of a String -Using inside probabilities-Using outside probabilities-Finding the most likely parse for a sentence-parsing for disambiguation-parsing model versus language model

Module -5 (Language Processing with Python)

Introduction to NLTK, Text Wrangling and Text cleansing : Sentence Splitter, Tokenization, Stemming, Lemmatization, Stop word removal , Rare word Removal, Spell Correction. Part of Speech Tagging and NER. Parsing Structure in Text: Shallow versus deep parsing, different types of parsers and dependency parsing.

Text Books :

1. C.D. Manning and H. Schutze. Foundations of Statistical Natural Language Processing. MIT Press.
2. Steven Bird, Ewan Klein, Edward Loper, Natural Language Processing with Python and NLTK. O'reilly Pub.

References:

1. D. Jurafsky and J.H. Martin: Speech and Language Processing: Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition, PHI.
2. James Allen: Natural Language Understanding. Pearson Pub.
3. Nitin Hardeniya, Jacob Perkins, Deepti Chopra, Nisheeth Joshi, ItiMathur: Natural Language Processing: Python and NLTK., **1**st Edition. Packt Publishing

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. What do you understand by the term *collocations?* List their properties.
2. Define the term phrase structure grammar formally.

Course Outcome 2 (CO2):

1. State Bayes' theorem and explain briefly. Comment on it's usefulness in NLP.
2. How can n-grams be used to model natural language statistically?

Course Outcome 3 (CO3):

1. What is meant by attachment ambiguity? Show it using English sentences
2. What is meant by Word Sense Disambiguation (WSD)? Outline any one WSD algorithm

Course Outcome 4 (CO4):

1. How can HMM be used for Parts of speech tagging?
2. Outline an implementation procedure for HMM

Course Outcome 5 (CO5):

1. Show with an example how can probabilistic grammars be used to model human preferences in parsing.
2. Give the technique of Transformation-Based Learning of Tags

Course Outcome 6 (CO6):

1. Implement a python program for stop word removal in a simple paragraph.
2. Write a code to access a weather site and extract the forecast top temperature for your town or city today.

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 3**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: CST478

Course Name: Computational Linguistics

Max. Marks : 100 **Duration: 3 Hours**

PART A

Answer All Questions. Each Question Carries 3 Marks

1. Define Zipf's law.
2. List the uses of a corpus in language processing?
3. What is a Hidden Markov Model?
4. State Bayes' theorem and explain briefly. Comment on its usefulness in NLP.
5. What is meant by supervised disambiguation? What are its prerequisites ?
6. Consider the sentence: "the children ate the cake with a spoon". Construct the parse tree for it and explain the attachment ambiguity.
7. Discuss the properties of Markov chain useful in POS tagging.
8. Explain the features of PCFG.
9. What is NLTK? How is it useful in text processing ?
10. Write a Python program to extract different date formats from a text document.

(10x3=30)

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) Write a note on the following terms with example: **(9)**
(i) Collocations (ii) Concordances (iii) Phrase structure grammars

(b) Differentiate stemming and lemmatization with examples. **(5)**

OR

12. (a) Write a note on all parts of speech tags of English language **(9)**

(b) What are the differences between Rationalist and Empiricist to Language approaches **(5)**

13. (a) What do you mean by a probability distribution? **(5)**
What are the approaches used in SNLP to estimate probability distribution of linguistic events?

(b) Give a formal definition of Hidden Markov Model (HMM) and state the relevant assumption while using HMM for language modeling **(9)**

OR

14. (a) Assume that a particular type of syntactic error detected by a system A occurs once in 1,00,000 sentences on an average. This system detects an error correctly with a probability 0.05. Suppose the system reports an error in a test sentence. What is the probability that this is true? **(5)**

(b) List some of the problems associated with sparse data in SNLP. **(9)**
Write a note on n-gram Models over Sparse Data

15. (a) What do you understand by Disambiguation based on sense definitions. **(9)**
Write and explain any one algorithm for this.

(b) With the help of Bayes' rule, explain the Bayesian disambiguation algorithm. **(5)**

OR

16. (a) Write a note on selectional preferences with an example **(5)**

(b) What is meant by attachment ambiguity? List different attachment issues. **(9)**

17. (a) Write a note on Transformation-Based Learning of tags. Give it's algorithm **(9)**

(b) How can HMM be used for parts of speech tagging **(5)**

OR

18. (a) Write the formal definition of PCFG. **(5)**

Apply probabilistic parsing on the following sentence and find the correct parsing using the given grammar

Sentence: Astronomers saw stars with ears.

Probabilistic grammar:

S → NP VP	1.0	NP → NP PP	0.4
PP → P NP	1.0	NP → *astronomers*	0.1
VP → V NP	0.7	NP → *ears*	0.18
VP → VP PP	0.3	NP → *saw*	0.04
P → *with*	1.0	NP → *stars*	0.18
V → *saw*	1.0	NP → *telescopes*	0.1

(b) How do you find the probability of a string using inside and outside probabilities ? **(9)**

19. (a) Write a Python program for PoS tagging using the necessary Python packages. **(9)**

(b) Explain the process of Named Entity Recognition. **(5)**
List its uses and challenges involved.

OR

20. (a) Write a regular expression for removing punctuations, numbers and white spaces in a piece of text. **(9)**

(b) Write a Python program to count the number of sentences, words and line numbers in a given piece of text. Display each sentence along with that. **(5)**

TEACHING PLAN

No	Contents	No of Lecture Hrs (36 hrs)
Module - 1 (Preliminaries) (9 hrs)		
1.1	Introduction: Rationalist and Empiricist Approaches to Language-Questions that linguistics should answer-	1
1.2	Non-categorical phenomena in language-Language and cognition as probabilistic phenomena	1
1.3	The Ambiguity of Language: Why natural language processing is difficult	1
1.4	Lexical resources-Word counts	1
1.5	Zipf's laws-Collocations-Concordances	1
1.6	Linguistic Essentials: Parts of Speech and Morphology -Nouns and pronouns	1
1.7	Words that accompany nouns: Determiners and adjectives-Verbs-Other parts of speech	1
1.8	Phrase Structure-Phrase structure grammars	1
1.9	Semantics and Pragmatics-Corpus Based Work	1
Module – 2 (Mathematical Essentials) (7 hrs)		
2.1	Probability Theory-Probability spaces	1
2.2	Conditional probability and independence-Bayes' theorem	1
2.3	Random variables-Expectation and variance-Notation	1
2.4	Joint and conditional distributions-Standard distributions-Bayesian statistics	1
2.5	Statistical Inference: n-gram Models over Sparse Data-Bins: Forming Equivalence Classes	1
2.6	Markov Models-Hidden Markov Models: Why use HMMs?	1
2.7	General form of an HMM-Finding the probability of an observation-Finding the best state sequence	1
Module – 3 (Word Sense Disambiguation) (7 hrs)		
3.1	Methodological Preliminaries-Supervised and unsupervised learning	1
3.2	Upper and lower bounds on performance-Supervised Disambiguation	1
3.3	Bayesian classification-Dictionary based Disambiguation-	1
3.4	Disambiguation based on sense definitions-Thesaurus based disambiguation	1
3.5	Lexical Acquisition-Evaluation Measures	1

3.6	Verb Subcategorization-Attachment Ambiguity, PP attachment-Selectional Preferences	1
3.7	Semantic Similarity:Vector space measures-Probabilistic measures	1

Module – 4 (Grammar) (8 hrs)		
4.1	Part-of-Speech Tagging-The Information Sources in Tagging	1
4.2	Markov Model Taggers-Hidden Markov Model Taggers-	1
4.3	Applying HMMs to POS tagging-The effect of initialization on HMM training-	1
4.4	Transformation-Based Learning of Tags	1
4.5	Probabilistic Context Free Grammars-Some Features of PCFGs	1
4.6	Questions for PCFGs	1
4.7	The Probability of a String -Using inside probabilities Using outside probabilities	1
4.8	Finding the most likely parse for a sentence-parsing for disambiguation, parsing model *vs.* language model	1
Module - 5 (Language Processing with Python) (5 hrs)		
5.1	Introduction to NLTK	1
5.2	Text Wrangling and Text cleansing : Sentence Splitter, Tokenization, Stemming,	1
5.3	Lemmatization, Stop word removal , Rare word Removal, Spell Correction.	1
5.4	Part of Speech Tagging and NER.	1
5.5	Parsing Structure in Text: Shallow versus deep parsing, types of parsers	1

SEMESTER VIII

MINOR

CSD482	MINI PROJECT	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PWS	0	0	3	4	2019

Preamble: The objective of this course is to apply the fundamental concepts of different courses learned in respective Minor Streams: Software Engineering, Machine Learning and Networking. This course helps the learners to get an exposure to the development of application software/hardware solutions/ software simulations in the field of Computer Science and Engineering. It enables the learners to understand the different steps to be followed such as literature review and problem identification, preparation of requirement specification &design document, testing, development and deployment. Mini project enables the students to boost their skills, widen the horizon of thinking and their ability to resolve real life problems.

Prerequisite: A sound knowledge in courses studied in respective minor stream.

Course Outcomes: After the completion of the course the student will be able to

CO#	CO
CO1	Identify technically and economically feasible problems **(Cognitive Knowledge Level: Apply)**
CO2	Identify and survey the relevant literature for getting exposed to related solutions. **(Cognitive Knowledge Level: Apply)**
CO3	Perform requirement analysis, identify design methodologies and develop adaptable & reusable solutions of minimal complexity by using modern tools & advanced programming techniques (**Cognitive Knowledge Level: Apply)**
CO4	Prepare technical report and deliver presentation **(Cognitive Knowledge Level: Apply)**
CO5	Apply engineering and management principles to achieve the goal of the project **(Cognitive Knowledge Level: Apply)**

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
CO2	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓
CO3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CO4	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓
CO5	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and team work
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Lifelong learning

Assessment Pattern

Mark Distribution

Total Marks	CIE Marks	ESE Marks
150	**75**	**75**

Continuous Internal Evaluation Pattern:

Attendance **10 marks**

Project Guide **15 marks**

Project Report **10 marks**

Evaluation by the Committee (will be evaluating the level of completion and demonstration of functionality/specifications, presentation, oral examination, work knowledge and involvement) : **40 marks**

Student Groups with 4 or 5 members should identify a topic of interest in consultation with a Faculty Advisor/Project Coordinator/Guide. Review the literature and gather information pertaining to the chosen topic. State the objectives and develop a methodology to achieve the objectives. Carryout the design/fabrication or develop codes/programs to achieve the objectives by strictly following steps specified in the teaching plan. Innovative design concepts, performance, scalability, reliability considerations, aesthetics/ergonomic, user experience and security aspects taken care of in the project shall be given due weight.

The progress of the mini project is evaluated based on a minimum of two reviews. The review committee may be constituted by a senior faculty member, Mini Project coordinator and project guide. The internal evaluation shall be made based on the progress/outcome of the project, reports and a viva-voce examination, conducted internally by a 3-member committee. A project report is required at the end of the semester. The project has to be demonstrated for its full design specifications.

End Semester Examination Pattern:

The marks will be distributed as

Presentation : **30 marks**
Demo : **20 marks**
Viva : **25 marks.**
Total : **75 marks**.

TEACHING PLAN

Students are expected to follow the following steps.

1. Review of Literature and Identification of a problem
2. Create an abstract with a problem statement, solution approach, technology stack, team, etc.
3. Create Requirements Specification
4. Create Design Document . This may include designs like,
    a. System Architecture Design
    b. Application Architecture Design
    c. GUI Design
    d. API Design
    e. Database Design
    f. Technology Stack
5. Deployment, Test Run & Get Results
6. Prepare Project Report

Guidelines for the Report preparation

A bonafide report on the mini project shall be submitted within one week after the final presentation. Minimum number of pages should be 40.

- Use Times New Roman font for the entire report – Chapter/Section Title – Times New Roman18, Bold; Heading 2 – Times New Roman16, Bold; Heading 3 – Times New Roman14, Bold; Body- Times New Roman 12, Normal.
- Line Spacing – Between Heading 2 – 3 lines, between lines in paragraph 1.5 lines.
- Alignments – Chapter/Section Title – Center, Heading 2 & 3 should be Left Aligned. Ensure that all body text is paragraph justified.
- Figures & Tables – Ensure that all Figures and Tables are suitably numbered and given proper names/headings. Write figure title under the figure and table title above the table.

- **Suggestive order of documentation:**

  i. Top Cover
  ii. Title page
  iii. Certification page
  iv. Acknowledgement
  v. Abstract
  vi. Table of Contents
  vii. List of Figures and Tables
  viii. Chapters
  ix. Appendices, if any
  x. References/Bibliography

SEMESTER VIII

HONOURS

CSD496	MINI PROJECT	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PWS	0	0	3	2	2019

Preamble: The objective of this course is to apply the fundamental concepts of courses learned in respective Honors Streams: Security in Computing, Machine Learning and Formal Methods. This course helps the learners to get an exposure to the development of application software/hardware solutions/ software simulations in the field of Computer Science and Engineering. It enables the learners to understand the different steps to be followed such as literature review and problem identification, preparation of requirement specification &design document, testing, development and deployment. Mini project enables the students to boost their skills, widen the horizon of thinking and their ability to resolve real life problems.

Prerequisite: A sound knowledge in courses studied in respective honor stream.

Course Outcomes: After the completion of the course the student will be able to

CO#	CO
CO1	Identify technically and economically feasible problems **(Cognitive Knowledge Level: Apply)**
CO2	Identify and survey the relevant literature for getting exposed to related solutions. **(Cognitive Knowledge Level: Apply)**
CO3	Perform requirement analysis, identify design methodologies and develop adaptable & reusable solutions of minimal complexity by using modern tools & advanced programming techniques (**Cognitive Knowledge Level: Apply)**
CO4	Prepare technical report and deliver presentation **(Cognitive Knowledge Level: Apply)**
CO5	Apply engineering and management principles to achieve the goal of the project **(Cognitive Knowledge Level: Apply)**

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
CO2	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓
CO3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CO4	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓
CO5	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and team work
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Lifelong learning

Assessment Pattern

Mark Distribution

Total Marks	CIE Marks	ESE Marks
150	**75**	**75**

Continuous Internal Evaluation Pattern:

Attendance **10 marks**

Project Guide **15 marks**

Project Report **10 marks**

Evaluation by the Committee (will be evaluating the level of completion and demonstration of functionality/specifications, presentation, oral examination, work knowledge and involvement) : **40 marks**

Student Groups with 4 or 5 members should identify a topic of interest in consultation with a Faculty Advisor/Project Coordinator/Guide. Review the literature and gather information pertaining to the chosen topic. State the objectives and develop a methodology to achieve the objectives. Carryout the design/fabrication or develop codes/programs to achieve the objectives by strictly following steps specified in the teaching plan. Innovative design concepts,

performance, scalability, reliability considerations, aesthetics/ergonomic, user experience and security aspects taken care of in the project shall be given due weight.

The progress of the mini project is evaluated based on a minimum of two reviews. The review committee may be constituted by a senior faculty member, Mini Project coordinator and project guide. The internal evaluation shall be made based on the progress/outcome of the project, reports and a viva-voce examination, conducted internally by a 3-member committee. A project report is required at the end of the semester. The project has to be demonstrated for its full design specifications.

End Semester Examination Pattern:

The marks will be distributed as

Presentation : **30 marks**
Demo : **20 marks**
Viva : **25 marks.**
Total : **75 marks**.

TEACHING PLAN

Students are expected to follow the following steps.

1. Review of Literature and Identification of a problem
2. Create an abstract with a problem statement, solution approach, technology stack, team, etc.
3. Create Requirements Specification
4. Create Design Document . This may include designs like,
    a. System Architecture Design
    b. Application Architecture Design
    c. GUI Design
    d. API Design
    e. Database Design
    f. Technology Stack
5. Deployment, Test Run & Get Results
6. Prepare Project Report

Guidelines for the Report preparation

A bonafide report on the mini project shall be submitted within one week after the final presentation. Minimum number of pages should be 40.

- Use Times New Roman font for the entire report – Chapter/Section Title – Times New Roman18, Bold; Heading 2 – Times New Roman16, Bold; Heading 3 – Times New Roman14, Bold; Body- Times New Roman 12, Normal.
- Line Spacing – Between Heading 2 – 3 lines, between lines in paragraph 1.5 lines.

- Alignments – Chapter/Section Title – Center, Heading 2 & 3 should be Left Aligned. Ensure that all body text is paragraph justified.
- Figures & Tables – Ensure that all Figures and Tables are suitably numbered and given proper names/headings. Write figuretitle under the figure and table title above the table.

- **Suggestive order of documentation:**

  i. Top Cover
  ii. Title page
  iii. Certification page
  iv. Acknowledgment
  v. Abstract
  vi. Table of Contents
  vii. List of Figures and Tables
  viii. Chapters
  ix. Appendices, if any
  x. References/Bibliography